БРАТІЙЧУК М.С., ЧЕЧЕЛЬНИЦЬКИЙ О.А.

ЛЕКЦІЇ зі СТОХАСТИКИ

Ймовірність. Статистика. Випадкові процеси

Київ 2021

ББК22.17я73
УДК 519.2

Рецензенти:
Чабанюк Я.М.–доктор фіз.-мат. наук, професор Львівського національного університету імені Івана Франка;
Самойленко І.В.– доктор фіз.-мат. наук, доцент Київського національного університету імені Тараса Шевченка.

Рекомендовано Вченою радою факультету клмп'ютерних наук і кібернетики. Протокол N 3 від 25 жовтня 2021р.

Братійчук М.С., Чечельницький О.А.– . – Навчальний посібник. Київ

У посібнику представлені головні розділи курсу теорії ймовірностей, математичної статистики та випадкових процесів. Наведені визначення ймовірносного простору та випадкової величини. Розкривається зміст головних характеристик випадкових величин: функції розподілу, математичного сподівання, дисперсії. Для послідовностей випадкових величин вивчаються різні види їх збіжностей, закон великих чисел та центральна гранична теорема. Представлені головні розділи класичної статистичної теорії: оцінювання невідомих параметрів, перевірка гіпотез, асимптотичні властивості головних типів статистик та оцінок, інтервальне оцінювання, побудова найпотужніших тестів. Значна увага приділена опису тестів, які часто використовуються на практиці.

Розглянуті головні класи випадкових процесів з дискретьним та неперервеним часом: ланцюги Маркова з дискретним фазовим простором, процеси з незалежними приростами, гіллясті процеси та процеси відновлення. Наведено приклади застосування теорії випадкових процесів в різних прикладних задачах (теорії обслуговування та теорії ризика)

Призначений для студентів факультетів прикладної математики, кібернетики, а також для студентів політехнічних інститутів.

ISBN 978-966-96123-3-5

Зміст

Передмова

Даний посібник розрахований, перш за все, на студентів університетів, які навчаються на факультетах прикладної математики та кібернетики. Він також може бути корисним для студентів технічних університетів, які в своїх дослідженнях використовують статистичні та ймовірносні методи. Посібник написаний на основі лекцій, які автори читали на факультеті кібернетики та комп'ютерних наук Київського національного університету імені Тараса Шевченка та Технічного університету в Глівіцах (Польща).

У посібнику не використовується стандартний поділ матеріалу на розділи. Все оформлено у вигляді лекцій, і при виборі матеріалу автори орієнтувались на курси теорії ймовірностей, математичної статистики та випадкових процесів для факультетів кібернетики та прикладної математики, а також факультетів технічних університетів, для яких математика є профільним предметом.

Від читача потрібні знання в об'ємі стандартного курсу математичного аналізу та лінійної алгебри для прикладних факультетів. Знання теорії міри та інтеграла Лебега не вимагаються. Деякі поняття, які не завше можна знайти в програмах прикладних факультетів університетів (наприклад, інтеграл Стілтьєса), але без яких подати матеріал на відповіному науковому рівні проблематично, формулюються окремо, в об'ємі, необхідному для даного посібника.

Більшість результатів подано з доведенням, оскільки ми вважаємо, що їх вивчення не лише підвищує загальну математичну культуру, але й дає можливість подивитися на проблему, так би мовити, "зсередини". Тому цей посібник розрахований на читача, який не обмежується формальним застосуванням відповідних процедур, а прагне зрозуміти їх сутність. Результати, обґрунтування яких виходить за рамки даного посібника, лише формулюються, але при цьому робляться посилання на літературу, де зацікавлений читач може знайти відповідну інформацію. Тексти всіх доведень та прикладів набрані дрібнішим шрифтом.

Нумерація теорем, лем, формул, зауважень, наслідків, задач є своєю в кожній лекції. Тому при посиланні на теорему, лему і т.п. з актуальної лекції вказується лише їх номер, а при посиланні на ці самі об'єкти з іншої лекції вказується сторінка, на якій знаходиться даний об'єкт.

Більшість підрозділів закінчується завданнями, які можна використовувати при проведенні практичних занять. До деяких з них, але не до всіх, подаються відповіді. Частина фактів, які використовуються в доведеннях, подаються у вигляді задач для самостійного розв'язування.

Знак ◀ в тексті означає кінець доведення.

Даний посібник розрахований, перш за все, на студентів університетів, які навчаються на факультетах прикладної математики та кібернетики. Він також може бути корисний для студентів технічних університетів, які в своїх дослідженнях використовують статистичні методи.

Автори

ТЕОРІЯ ЙМОВІРНОСТЕЙ

Лекція 1. Стохастичний експеримент, випадкова подія, ймовірність

В кожній теорії є первісні поняття, або ті, сенс яких не визначається. В арифметиці таким поняттям є число, в геометрії – точка, лінія, площина.

Первісними поняттями в теорії ймовірностей є поняття: *стохастичного експерименту*, а також *простору елементарних подій* (*множини елементарних подій*).

Визначення 1. Під *стохастичним* експериментом розуміють експеримент S, результат якого заздалегідь не можна передбачити.

Результати експерименту, які виключають один одного, будемо називати *елементарною подією.* Елементарну подію зазвичай позначають грецькою літерою ω.

Визначення 2. Множину елементарних подій стохастичного експерименту будемо називати *простором елементарних подій* і позначати літерою Ω.

Приклад 1. *(підкидання монети). Цей експеримент має два можливих результати: поява "орла" (О), або "цифри" (Ц), отже* $\Omega = \{О, Ц\}$

Приклад 2. *Припустимо, що ми підкидаємо монету до того часу, поки не випаде перший "орел", після чого експеримент закінчуємо. Результат такого експерименту можна представити у вигляді послідовності (ЦЦ...ЦО). Кількість таких послідовностей є вочевидь злічена величина.*

1.1 Дискретний простір елементарних подій. Поняття випадкової події та ймовірності.

Припустимо, що простір елементарних подій є не більш ніж зліченний (в частковому випадку скінченний)

$$\Omega = \{\omega_1, \omega_2, ...\}.$$

У цьому випадку будемо казати, що простір Ω є *дискретним.*

Визначення 3. *Випадковою подією* будемо називати будь-яку підмножину $A \subset \Omega$ простору елементарних подій.

Будемо казати, що подія A *настає*, якщо настає якась елементарна подія, яка належить до A. Іншими словами, якщо результатом експерименту є елементарна подія ω і $\omega \in A$, то ми кажемо, що настала подія A. В цьому випадку елементарна подія ω називається *елементарною подією сприятливою для події* A.

Позаяк завжди $\Omega \subset \Omega$ і $\emptyset \subset \Omega$, то згідно з наведеним визначенням множини Ω і $\emptyset$ є подіями. Множина Ω називається *достовірною* подією . Порожню множину $\emptyset$ будемо називати *неможливою* подією.

Оскільки випадкові події визначені як підмножини певної універсальної множини Ω, то операції на подіях можна визначити в той самий спосіб. Це означає що, наприклад, сума двох подій A і B може бути визначена як подія C, яка складається тільки з тих елементарних подій, які належать до A або до B. Але ми будемо використовувати мову, притаманну теорії ймовірностей, це означає, що ми будемо казати про реалізацію або не реалізацію відповідних подій .

Так *сумою* двох випадкових подій A і B будемо називати подію, яка відбувається тоді і тільки тоді, коли відбувається принаймні одна з цих подій. Суму подій A і B будемо позначати $A \cup B$. З визначення суми випливає, що подія $A \cup B$ складається тільки з тих елементарних подій, які належать до A або до B. Аналогічно *добутком* двох випадкових подій A і B будемо називати подію, яка відбувається тоді і тільки тоді, коли відбуваються обидві ці події одночасно. Добуток подій A і B будемо позначати наступним чином: $A \cap B$ або AB.

Подібним чином визначимо інші дії над подіями, а саме: *різницю* випадкових подій A і B (позначаємо $A \backslash B$), подію *протилежну* до події A або *доповнення* до події A (позначаємо символом $\bar{A} = \Omega \backslash A$.)

Якщо $A \cap B = \emptyset$, тоді події A і B будемо називати *несумісними*.

Якщо $A \subset B$, тоді будемо казати, що поява події A *тягне за собою* подію B (або подія B є *наслідком* події A, або (менш уживано) подія A *міститься* в події B.)

Позаяк ми визначаємо події як підмножини простору Ω, тоді дії над подія мають такі самі властивості як і дії над множинами. В зв'язку з чим, наведемо відповідні твердження без доведення.

Теорема 1. *Дії над подіями мають наступні властивості:*

1) $AB = BA$ – *комутативність добутку;*

2) $A \cup B = B \cup A$ – *комутативність суми;*

3) $A(BC) = (AB)C$ – *асоціативність добутку;*

4) $A \cup (B \cup C) = (A \cup B) \cup C$ – *асоціативність суми;*

5) $A(B \cup C) = AB \cup AC$ – *дистрибутивність добутку;*

6) $A \cup BC = (A \cup B)(A \cup C)$ – *дистрибутивність суми;*

7) $\overline{A \cup B} = \bar{A} \cap \bar{B}$, $\overline{A \cap B} = \bar{A} \cup \bar{B}$ – *закони д'Моргана.*

Визначення 4. Будемо казати, що задана *ймовірність елементарних подій*, якщо на $\Omega = \{\omega_1, \omega_2, ...\}$ визначена функція $\mathbf{P}(\cdot)$ така, що:
а) $\mathbf{P}(\omega_i) \geq 0$ для всіх $i = 1, 2, ...$,

б) $\sum_{i=1}^{\infty} \mathbf{P}(\omega_i) = 1$.

Визначення 5. *Ймовірністю події A* будемо називати число

$$\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\omega).$$

Можна зауважити, що, взагалі кажучи, ймовірність (функція $\mathbf{P}(\cdot)$) може бути визначена різним чином і тому конкретні числові значення цієї функції не будуть нас цікавити. Це є справа тієї чи іншої вибраної моделі. Так, наприклад, якщо в прикладі 1 монета симетрична, то логічно взяти $\mathbf{P}(O) = \mathbf{P}(\text{Ц}) = 1/2$.

Теорема 2. *Ймовірність має наступні властивості:*

1) $\mathbf{P}(\Omega) = 1$, $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$, $\mathbf{P}(\emptyset) = 0$;

2) *якщо A і B є несумісні, тоді* $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$;

3) *якщо $A \subset B$, тоді* $\mathbf{P}(B \backslash A) = \mathbf{P}(B) - \mathbf{P}(A)$ *і* $\mathbf{P}(A) \leq \mathbf{P}(B)$, *отже,* $\mathbf{P}(A) \leq 1$ *для всіх $A \subset \Omega$;*

4) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$ *для довільних подій A, B;*

5) $\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$ *для довільних подій A, B.*

Доведення цієї теореми легко випливає з властивості абсолютної збіжності рядів. Так рівність $\mathbf{P}(\Omega) = 1$, є власне умовою б) у визначенні ймовірності. Оскільки,

$$\mathbf{P}(\bar{A}) + \mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\omega) + \sum_{\omega \in \bar{A}} \mathbf{P}(\omega) = \sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1,$$

тоді будемо мати $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$. Якщо в цій рівності взяти $A = \emptyset$, тоді отримаємо $1 = \mathbf{P}(\Omega) = 1 - \mathbf{P}(\emptyset)$. Звідки $\mathbf{P}(\emptyset) = 0$, що доводить пункт 1) нашого твердження. Доведення решти властивостей ймовірності залишаємо читачеві.

Властивість 2) може бути легко узагальнена наступним чином: якщо послідовність подій $A_i, i = 1, 2, ..., n$, $n \leq \infty$ є такою, що $A_i A_j = \emptyset$, $i \neq j$, тоді

$$\mathbf{P}(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbf{P}(A_i). \tag{1}$$

Властивість 5) так само може бути легко узагальнена

$$\mathbf{P}(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \mathbf{P}(A_i), \tag{2}$$

Як і раніш, випадок $n = \infty$ не є виключеним (задача 1.)

1.2 Класична схема

Нехай простір Ω складається з $n < \infty$ елементів $\Omega = \{\omega_1, ..., \omega_n\}$. Припустимо також, що всі елементарні події є однаково ймовірні (тобто $\mathbf{P}(\omega_i) = \mathbf{P}(\omega_j)$ для всіх $i, j = 1, ..., n$. В цій ситуації будемо говорити, що ми маємо справу з *класичною ймовірносною схемою.* Зрозуміло, що тоді $\mathbf{P}(\omega) = 1/n$ для кожного $\omega \in \Omega$. При цьому ймовірність довільної події A, яка містить k елементарних сприятливих подій, визначається наступним чином:

$$\mathbf{P}(A) = \frac{k}{n}. \tag{3}$$

Символічно цей вираз можемо записати так:

$$\mathbf{P}(A) = \frac{\text{число елементарних подій, сприятливих для } A}{\text{число всіх елементарних подій}}.$$

Це *класичне визначення ймовірності*, автором якого є Лаплас.

Використання формули (3) зазвичай пов'язано з розділом математики, який називається "Комбінаторика". В зв'язку з чим наведемо деякі поняття і формули з цієї області, які найчастіше використовуються при розв'язанні задач класичної схеми. Доведення цих формул можна знайти в довільному стандарному підручнику з комбінаторики.

Основне правило комбінаторики. *Припустимо, що треба послідовно виконати k дій. Якщо першу дію можна виконати n_1 способами, другу n_2 способами, т.д. Зрештою k-ту дію n_k способами, тоді всі k дій можна виконати $n_1 \times n_2 \times \cdots \times n_k$ способами.*

Розглянемо множину H, яка складається з n різних елементів.

- Кожне впорядкування елементів з H будемо називати *перестановкою.* Число всіх перестановок з n елементів дорівнює

$$P_n = n!.$$

- Впорядковану множину з n елементів множини H, серед яких певні елементи повторюються відповідно $n_1, n_2, ..., n_k$ раз, будемо називати n–елементною *перестановка з повтореннями.* Число всіх таких перестановок дорівнює:

$$P_n(n_1, n_2, ..., n_k) = \frac{n!}{n_1! \cdot n_2! \cdot ... \cdot n_k!}.$$

- Підмножини по k елементів множини H, які розрізняються між собою принаймні одним елементом, називаються k–елементними *комбінаціями* з n–елементної множини. Число всіх таких комбінацій визначається формулою:

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

- Будь-яку впорядковану підмножину з k елементів множини H називають *розміщенням.* Число всіх k–елементних розміщень з n–елементної множини визначається формулою:

$$A_n^k = \frac{n!}{(n-k)!}.$$

- Множини з k елементів множини H, які можуть різнитись або не різнитись між собою, при цьому порядок їх розташування не є важливим, називають k–елементними *комбінаціями з повтореннями* з n–елементної множини. Число всіх таких комбінацій визначається формулою:

$$\bar{C}_n^k = C_{n+k-1}^k = \frac{(n+k-1)!}{k!(n-1)!}.$$

- Нехай $k \geq n$. Число k–елементних комбінацій з повтореннями, в яких кожний елемент з n–елементної множини бере участь принаймні раз, дорівнює C_{k-1}^{n-1}.

Дану лекцію закінчимо корисною формулою, яка є узагальненням властивості 4 теореми 2.

Теорема 3. *Для довільних подій $A_1, ..., A_n$*

$$\mathbf{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{i<j} \mathbf{P}(A_i A_j) + ... + (-1)^{n-1}\mathbf{P}(A_1...A_n). \quad (4)$$

Д о в е д е н н я. Нехай $\omega \in \cup_{i=1}^n A_i$, а k дорівнює числу всіх подій, до яких входить ω. Вочевидь, $1 \leq k \leq n$. Число $\mathbf{P}(\omega)$ є доданком в лівій стороні (4). Підрахуємо, скільки разів доданок $\mathbf{P}(\omega)$ зустрічається в правій стороні рівності.

В першій сумі $\mathbf{P}(\omega)$ зустрічається k разів, в другій сумі $\mathbf{P}(\omega)$ зустрічається стільки разів, скільки добутків $A_i A_j$, $i < j$ містять ω, а таких добутків є C_k^2. В третій сумі $\mathbf{P}(\omega)$ зустрічається C_k^3 разів і т.д. Тепер можна підрахувати, скільки разів доданок $\mathbf{P}(\omega)$ зустрічається в правій частині рівності (4):

$$k - C_k^2 + C_k^3 - ... + (-1)^{k-1}C_k^k = 1 - \left[1 - C_k^1 - C_k^2 + C_k^3 - ... + (-1)^k C_k^k\right] = 1 - (1-1)^k = 1.$$

Отже, в лівій і правій частині співвідношення (4) беруть участь однакові доданки,що й доводить рівність (4). ◀

Як приклад використування формули (4) розглянемо наступну задачу.

Приклад 1. *Нехай задано n елементів, розміщених у певному порядку. Кожне розташування відбувається випадковим чином (кожна з $n!$ перестановок є однаково ймовірна). Якою є ймовірність того, що хоча б один елемент буде займати своє місце?*

Р о з в' я з о к. Нехай подія A_k означає, що k–тий елемент знаходиться на своєму місці. Число всіх перестановок дорівнює $n!$, а перестановок, сприятливих для події A_k, є $(n-1)!$ (дійсно, треба поставити k–тий елемент на своє місце і зробити всі можливі перестановки з інших елементів). Отже,

$$\mathbf{P}(A_k) = \frac{(n-1)!}{n!}.$$

Подія A_iA_j означає, що елементи i–тий і j–тий знаходяться на своїх місцях. Число перестановок, які є сприятливими цій події, становить $(n-2)!$, тоді $\mathbf{P}(A_iA_j) = \frac{(n-2)!}{n!}$, і. д., і врешті

$$\mathbf{P}(A_1...A_n) = \frac{(n-(n-1))!}{n!} = \frac{1}{n!}.$$

Подія $\cup_{k=1}^{n} A_k$ полягає в тому, що принаймні один елемент знаходиться на своєму місці. Беручи до уваги доведене вище, і те, що в $\sum_{i<j}$ є докладно C_n^2 доданків, в $\sum_{i<j<k}$ – C_n^3 доданків і т.д. Понадто, всі вони однаково ймовірні. Тоді

$$\mathbf{P}(\cup_{k=1}^{n} A_k) = n\frac{(n-1)!}{n!} - C_n^2\frac{(n-2)!}{n!} + C_n^3\frac{(n-3)!}{n!} - ... + (-1)^{n-1}\frac{1}{n!} =$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - ... + (-1)^{n-1}\frac{1}{n!} = 1 - \quad 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + ... + (-1)^{n}\frac{1}{n!} \quad .$$

Вираз у дужках є розкладом e^{-1} в ряд, а, отже,

$$\mathbf{P}(\cup_{k=1}^{n} A_k) \underset{n\to\infty}{\longrightarrow} 1 - e^{-1}.$$

ЗАВДАННЯ

1. Довести нерівність (2) для $n = \infty$.
2. В урні знаходяться 4 білих, 3 чорних і 5 зелених куль. Знайти ймовірність витягнути з урни білу або чорну кулю. ($\frac{7}{12}$)
3. З колоди у 32 карти два рази без повернення вибирають дві карти. Яка є ймовірність того, що

   а) лише одна з цих карт буде валетом? ($\frac{7}{31}$)

   б) принаймні одна з цих карт буде валетом? ($\frac{59}{248}$)
4. З чисел 1, 2, ..., n вибираємо два числа. Якою є ймовірність того, що одне з них буде строго меншим, а друге строго більшим від даного числа k, $1 < k < n$? ($\frac{2(k-1)(n-k)}{n(n-1)}$)
5. В урні є n чорних і m білих куль. Вибираємо без повернення k куль ($k < n, k < m$). Знайти ймовірність того, що серед цих k куль буде рівно r чорних. ($\frac{C_n^r C_m^{k-r}}{C_{n+m}^k}$)
6. Двадцять осіб: 10 чоловіків і 10 жінок поділено випадковим чином на пари. Підрахувати ймовірність того, що кожна пара буде складатися з осіб різної статі. ($\frac{10!}{19!!}$)

7. У конверті є 100 світлин і серед них одна, яку шукають. З конверту виймають 10 світлин. Знайти ймовірність того, що серед них є ця потрібна світлина. (0.1)

8. При наборі номеру телефону абонент забув три останні цифри. Він пам'ятає тільки, що ці цифри різні, і набирає їх навмання. Знайти ймовірність того, що він набрав правильні цифри. (1/720)

9. У продаж надійшло n лотерейних білетів, з них m є виграшними. Якою є ймовірність виграшу при купівлі k білетів? $(1-\frac{(n-m)!(n-k)!}{n!(n-m-k)!})$

10. (Задача de Mare) Що є більш ймовірним: при одному підкиданні чотирьох гральних кубиків отримати принаймні одну одиницю, чи при 24 підкиданнях двох кубиків отримати принаймні один раз дві одиниці? $(1-(\frac{5}{6})^4 > 1-(\frac{35}{36})^{24})$

11. Якою є ймовірність того, що 12 осіб мають дні народження в різних місяцях? $(\frac{12!}{12^{12}})$

12. Підрахувати ймовірність того, що серед k осіб, вибраних випадковим чином, кожна особа відзначає день народження в інший день року. (C_{365}^k/C_{364+k}^k)

13. Маємо n конвертів з різними адресами і n листів, які повинні бути розіслані за цими адресами. Випадковим чином вкладають листи до конвертів. Якою є ймовірність того, що m листів потраплять за своїми адресами? $(\frac{1}{m!}[\frac{1}{2!}-\frac{1}{3!}+...+(-1)^{n-m}\frac{1}{(n-m)!}]$

14. Маємо 30 однакових м'ячиків, які кидають до 5 коробок. Якою є ймовірність того, що жодна коробка не є пустою? $(\frac{C_{29}^4}{C_{34}^{30}})$

Лекція 2. Незалежність подій. Умовна ймовірність

Поняття, які будуть розглянуті в цій лекції, є одними з найважливіших понять теорії ймовірностей.

2.1 Незалежність подій

Визначення 1. Події A і B будемо називати *незалежними*, якщо

$$\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B).$$

Так події Ω і A для довільного A є завжди незалежні, позаяк $\mathbf{P}(\Omega A) = \mathbf{P}(A) = 1 \cdot \mathbf{P}(A) = \mathbf{P}(\Omega)\mathbf{P}(A)$. Якщо $\mathbf{P}(A) = 0$, тоді для довільної події B події A і B є незалежні. Це випливає зі співвідношення $0 = \mathbf{P}(AB) = 0 \cdot \mathbf{P}(B) = \mathbf{P}(A)P(B)$.

Теорема 1. *Якщо події A і B є незалежними, тоді незалежними будуть події A і $\overline{B}$.*

Д о в е д е н н я. Позаяк $A\overline{B} = A \backslash AB$, тоді

$$\mathbf{P}(A\overline{B}) = \mathbf{P}(A \backslash AB) = \mathbf{P}(A) - \mathbf{P}(AB) = \mathbf{P}(A) - \mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(A)(1 - \mathbf{P}(B)) = \mathbf{P}(A)\mathbf{P}(\overline{B}).$$

◀

У випадку більшої кількості подій визначення їх незалежності формулюється наступним чином.

Визначення 2. Події $A_1, \ldots, A_n$ будемо називати *незалежними*, якщо

$$\mathbf{P}(A_1 A_2 \cdots A_n) = \mathbf{P}(A_1)\mathbf{P}(A_2) \cdots \mathbf{P}(A_n).$$

Визначення 3. Події $A_2, ..., A_n$ будемо називати *незалежними у сукупності*, якщо для кожної підмножини індексів $i_1, i_2, \ldots, i_k$, $1 \leq k \leq n$, з послідовності $1, 2, ..., n$ події $A_{i_1}, A_{i_2}, ..., A_{i_k}$ є незалежними.

В цьому визначенні без зміни його суті n може бути нескінченним.

З незалежності подій $A_1, A_2, ..., A_n$ не виникає незалежність підмножини цих подій (задача 3).

Крім того, довільна підмножина подій зі сукупності $A_1, \dots, A_n$ може складатися з незалежних подій, але з цього не буде випливати незалежність всієї сукупності. Це демонструє наступний приклад.

Приклад 1. (Приклад Бернштейна). *Проведемо наступний експеримент. На площину будемо кидати тетраедр, три грані якого пофарбовані відповідно в червоний, синій і зелений кольори, а четверта грань містить всі три кольори. Подія C описує ситуацію, коли тетраедр впав на грань, на якій є червоний кольор. Подія N буде означати, що він впав на грань, на якій є синій кольор. Подія Z - на грань з зеленим кольором.*

Позаяк кожний колір присутній на двох гранях, будемо мати

$$\mathbf{P}(C) = \mathbf{P}(N) = \mathbf{P}(Z) = \frac{1}{2}.$$

Легко зрозуміти, що $\mathbf{P}(CN) = 1/4$, $\quad \mathbf{P}(CZ) = \mathbf{P}(ZN) = 1/4$. Отже,

$$\mathbf{P}(CN) = \frac{1}{4} = \mathbf{P}(C)\mathbf{P}(N), \quad \mathbf{P}(CZ) = \frac{1}{4} = \mathbf{P}(C)\mathbf{P}(Z), \quad \mathbf{P}(ZN) = \frac{1}{4} = \mathbf{P}(Z)\mathbf{P}(N),$$

а тому кожна підмножина сукупності подій C, N, Z складається з незалежних подій, але

$$\mathbf{P}(CNZ) = \frac{1}{4} \neq \mathbf{P}(C)\mathbf{P}(N)\mathbf{P}(Z) = \frac{1}{8}.$$

2.2 Умовна ймовірність

Розпочнемо з прикладу. Припустимо, що задана класична схема з n елементарними подіями, і нас цікавить певна подія A. Нехай додатково відомо, що станеться подія B. Якою тоді буде ймовірність того, що відбудеться також подія A. Позначимо цю ймовірність через $\mathbf{P}(A/B)$.

Нехай подія B складається з m елементарних подій, а подія AB з k елементарних подій. Позаяк з m елементарних подій B тільки k є сприятливими для A, тоді згідно з класичною ймовірносною схемою повинно бути:

$$\mathbf{P}(A/B) = \frac{k}{m} = \frac{\frac{k}{n}}{\frac{m}{n}} = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)}$$

Цей вираз є вказівкою для визначення загального поняття умовної ймовірності.

Визначення 4. Нехай A і B є довільними подіями, крім того $\mathbf{P}(B) > 0$. *Умовну ймовірність* події A за умови, що станеться подія B, визначимо як

$$\mathbf{P}(A/B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)}. \tag{1}$$

З (1) маємо:

$$\mathbf{P}(AB) = \mathbf{P}(B)\mathbf{P}(A/B) \tag{2}$$

Вираз (2) має назву *формула ймовірності добутку подій.*

Доведення наступного твердження, яке описує властивості умовної ймовірності, легко випливає з її визначення.

Теорема 2. *Умовна ймовірність має наступні властивості:*

1) $0 \leq \mathbf{P}(A/B) \leq 1, \quad \mathbf{P}(\Omega/B) = 1, \quad \mathbf{P}(B/B) = 1$*:*

2) *якщо події* A, B *є незалежними, тоді* $\mathbf{P}(A/B) = \mathbf{P}(A)$*;*

3) *якщо події* $A_1, ..., A_n$*,* $n \leq \infty$ *є такими, що* $A_iA_j = \emptyset$*,* $i \neq j$*, тоді*

$$\mathbf{P}(\cup_{i=1}^{n} A_i/B) = \sum_{i=1}^{n} \mathbf{P}(A_i/B).$$

2.3 Формула повної ймовірності і формула Байєса

При підрахунку ймовірності складних подій постає природне питання: "А чи не можна представити складну подію через більш прості, ймовірності яких можна легше знайти? ". А потім, використовуючи їх, підрахувати ймовірність складної події. Відповідь на це питання дає наступна теорема, в якій формула (3) має назву *формула повної ймовірності.*

Теорема 3. *Нехай* $A \subset \cup_{j=1}^{n} B_j$*, де* $B_iB_j = \emptyset, i \neq j$ *і* $\mathbf{P}(B_i) > 0$*,* $i = 1, ..., n$*. Тоді*

$$\mathbf{P}(A) = \sum_{j=1}^{n} \mathbf{P}(B_j)\mathbf{P}(A/B_j). \tag{3}$$

Зауваження 1. Якщо $\cup_{j=1}^{n} B_j = \Omega$, тоді умова $A \subset \cup_{j=1}^{n} B_j$ виконується автоматично.

Д о в е д е н н я. Позаяк $A = \cup_{j=1}^{n} B_jA$ (це випливає з нашого припущення $A \subset \cup_{j=1}^{n} B_j$), і події $AB_1, AB_2, ..., AB_n$ є несумісні (оскільки такими є події $B_1, B_2, ..., B_n$), то будемо мати

$$\mathbf{P}(A) = \sum_{j=1}^{n} \mathbf{P}(AB_j) = \ \{\text{в силу (2)}\} = \sum_{j=1}^{n} \mathbf{P}(B_j)\mathbf{P}(A/B_j). \qquad \triangleleft$$

Наступна формула, яка має назву *формула Байєса*, є в певному сенсі протилежною до (3). Вона дозволяє підрахувати ймовірність, що станеться одна з взаємно несумісних подій B_1, B_2, ..., B_n, якщо відомо, що відбудеться подія A. Іншими словами отримаємо формулу для $\mathbf{P}(B_i/A)$.

Теорема 4. (Формула Байєса). *Нехай* $A \subset \cup_{j=1}^{n} B_j$, *де* $B_i B_j = \emptyset, i \neq j$ *і* $\mathbf{P}(B_i) > 0$, $i = 1, ..., n$. *Тоді*

$$\mathbf{P}(B_i/A) = \frac{\mathbf{P}(B_i)\mathbf{P}(A/B_i)}{\mathbf{P}(A)}. \tag{4}$$

Д о в е д е н н я. Легко випливає з визначення умовної ймовірності.

2.4 Моделі експериментів, приклади розрахунку ймовірностей

Схема бінарного експерименту

Розглянемо певний експеримент S з двома можливими результатами A і $\overline{A}$. Нехай

$$\mathbf{P}(A) = p, \qquad \mathbf{P}(\overline{A}) = q = 1 - p, \qquad 0 \leq p \leq 1.$$

Визначення 5. Експеримент з двома можливими результатами A і $\overline{A}$ будемо називати *схемою бінарного експерименту.* В цьому випадку $\mathbf{P}(A) = p$, $\mathbf{P}(\overline{A}) = q = 1 - p$, $0 \leq p \leq 1$.

Зрозуміло, що сукупність чисел $\{p, 1-p\}$, $0 \leq p \leq 1$ є ймовірностями.

Зауваження 2. Назва цієї схеми походить з того, що часто випадок, коли відбувається A, позначають числом 1, а подію $\overline{A}$ – числом 0.

Геометрична схема експерименту

Будемо тепер проводити наш експеримент S багаторазово незалежним чином. Нехай $A_i, \overline{A}_i$ описують можливі результати в i–тому експерименті. Зрозуміло, що події $A_i, i = 1, 2, ...$ - незалежні і $\mathbf{P}(A_i) = p$, $\mathbf{P}(\overline{A}_i) = q = 1 - p$, $0 \leq p \leq 1$. Будемо повторювати експеримент S доти, аж поки не станеться подія A_i. Простір елементарних подій в цьому випадку буде наступним:

$$\Omega = \{A_n : A_n = \underbrace{\overline{A}_1 \overline{A}_2 \cdots \overline{A}_n}_{n} A_{n+1}, n \geq 0\}.$$

Тоді

$$\mathbf{P}(\overline{A}_1\overline{A}_2\cdots\overline{A}_nA_{n+1}) = \underbrace{\mathbf{P}(\overline{A}_1)\mathbf{P}(\overline{A}_2)\cdots\mathbf{P}(\overline{A}_n)}_{n}\mathbf{P}(A_{n+1}) = q^n p.$$

Визначення 6. Описаний вище експеримент будемо називати *геометричною схемою*, а сукупність чисел $\{(1-p)^n p, \quad n \geq 0\}$, яка описує ймовірність результатів цього експерименту, будемо називати *геометричною ймовірністю.*

Схема випробувань Бернуллі

Будемо проводити наш експеримент S незалежним чином n разів. Як і раніше, нехай $A_i, \overline{A}_i$ означають можливі результати в i–тому експерименті. Треба знайти ймовірність події: результат A_i з'явиться k разів, $0 \leq k \leq n$. Позначимо цю подію $A(n,k)$. Кожний результат нашого експерименту можемо представити собі у вигляді вектора з n компонентами. Якщо в i-тому випробуванні сталася подія A_i, тоді i-та компонента дорівнюватиме 1, а в протилежному випадку - 0. Треба знайти ймовірність того, що по проведенню n разів експерименту k компонент вектора дорівнюватимуть одиниці, а решта нулям. Якщо взяти один з таких векторів з n компонентами, то його ймовірність становитеме $p^k q^{n-k}$. Але загальне число таких векторів, в яких k компонент дорівнює одиниці, а решта нулям, визначається як C_n^k. Отже, будемо мати:

$$\mathbf{P}(A(n,k)) = C_n^k p^k q^{n-k}. \tag{5}$$

Визначення 7. Описаний вище експеримент будемо називати *схемою Бернуллі* з параметрами (n,p), а сукупність чисел $\{C_n^k p^k q^{n-k}, \quad 0 \leq k \leq n \geq 0\}$, яка описує ймовірності результатів цього експерименту, називається *біноміальним розподілом.*

Якщо n і k - великі, тоді підрахунок ймовірності по формулі (5) є непростою справою. Наступна теорема подає наближений вираз для ймовірності (5).

Теорема 5. (Пуассона) *Якщо в схемі Бернуллі k є сталим, а n і $p = p(n)$ змінюються так, що $a(n) \equiv np(n) \leq c$, де c- стала, тоді*[1]

$$\mathbf{P}(A(n,k)) = C_n^k p^k(n) q^{n-k}(n) \sim \frac{(a(n))^k}{k!}e^{-a(n)}, \qquad \text{\textit{при}} \quad n \to \infty. \tag{6}$$

[1] Як завжди запис $f(n) \sim g(n), n \to \infty$ означає, що $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

Д о в е д е н н я. Проведемо розрахунки

$$C_n^k p^k(n) q^{n-k}(n) = \frac{n(n-1)...(n-k+1)}{k!} \left(\frac{a(n)}{n}\right)^k \left(1-\frac{a(n)}{n}\right)^{n-k} =$$

$$= \frac{n(n-1)...(n-k+1)}{k!n^k} (a(n))^k \left(1-\frac{a(n)}{n}\right)^{n-k}. \qquad (7)$$

Позаяк для сталого k

$$\lim_{n\to\infty}(1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{k-1}{n}) = 1$$

і, якщо $a(n) \le c$, тоді легко довести, що

$$\left(1-\frac{a(n)}{n}\right)^{n-k} \sim e^{-a(n)}.$$

Отже, з (7) випливає (6). ◄

Наслідок 1. *Нехай* $\lim_{n\to\infty} np(n) = \lambda$. *Тоді для кожного* k *маємо:*

$$\lim_{n\to\infty} \mathbf{P}(A(n,k)) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

Позаяк $\lambda^k e^{-\lambda}/k! > 0$ для кожного $k \ge 0$, а

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1,$$

тоді сукупність чисел $\{\lambda^k e^{-\lambda}/k!, \quad k = 0,1,2...\}$ є розподілом ймовірностей, який називається *розподілом Пуассона з параметром* λ.

ЗАВДАННЯ

1. Двічі підкидають гральний кубик. Подія A означає: за першим разом випаде 5, а подія B: сума очок при двох підкиданнях більша за 8. Дослідити події A і B на незалежність.

2. З колоди 52 карт виймаємо одну карту. Подія A означає, що ця карта виявилась піковою, подія B – тузом. Перевірити події A і B на незалежність.

3. Експеримент полягае в підкиданні симетричного грального кубика. Позначимо через A_1 випадіння числа не більшого 4, через A_2 випадіння числа не більшого 3, а через A_3 появу одного з чисел 3, 4, 5. Показати, що події A_1, A_2, A_3, а також A_1, A_3 є незалежними, але події A_1, A_2 і A_2, A_3 - залежні.

4. Події A і B_1, а також A і B_2 є незалежними, при цьому $B_1 B_2 = \emptyset$. Довести, що події A і $B_1 \cup B_2$ також є незалежними.

5. Дві особи по черзі підкидають симетричну монету. Виграє та особа, у якої вперше з'явиться орел. Знайти ймовірність виграшу для кожного з гравців. ($\frac{2}{3}, \frac{1}{3}$)

6. Скільки разів треба підкидати трьома монетами, аби ймовірність появи принаймні раз трьох орлів одночасно була бульше за 0,8? ($n \geq 12$)

7. Що є більш ймовірним при грі з противником рівного рівня (виключаємо нічиї):

   а) виграти три з чотирьох партій, чи виграти п'ять з восьми партій? ($\frac{1}{4} > \frac{7}{32}$)

   б) виграти не менш ніж три з чотирьох партій, чи виграти не менш ніж п'ять з восьми партій? ($\frac{5}{16} < \frac{93}{256}$)

8. При підкиданні п'яти монет принаймні на одній випав орел. Якою є ймовірність того, що на решті монет також випали орли? ($\frac{1}{31}$)

9. Шість осіб пробують відгадати невідомий їм результат підкидання грального кубика. Якою є ймовірність того, що це зробить принаймні одна особа? ($1 - (\frac{5}{6})^6$)

10. Маємо дві урни з кулями. Урна I містить 6 білих куль і 3 чорних, урна II містить 4 білих кулі і 6 чорних.

    а) Після того, як куля виймається з урни, вона знову до неї повертається. Знайти ймовірність наступних подій:

    1. Дві кулі виявились білими, якщо їх виймали з урни I. ($\frac{4}{9}$)

    2. Три кулі виявились чорними, якщо їх виймали з урни II. ($\frac{27}{125}$)

    3. з двох куль - одна біла і одна чорна, якщо їх виймали по одній з кожної урни. ($\frac{8}{15}$)

    4. З трьох куль - одна біла і дві чорні, якщо з I урни виймали дві кулі, а з II урни – одну. ($\frac{14}{45}$)

    б) Кулі виймаються з урни без повернення. Знайти ймовірність вийняти:

    1. Дві білі кулі, якщо їх виймали з урни I. ($\frac{5}{12}$)

2. Три чорні кулі, якщо їх виймали з урни II. ($\frac{1}{6}$)

3. Три кулі - одну білу і дві чорні, якщо з урни I виймають дві кулі, а з урни II –одну. ($\frac{1}{3}$)

11. Дві урни містять відповідно 4 і 7 білих, а також 6 і 8 чорних куль. З кожної урни взяли навмання по одній кулі, а потім з цих двох випадковим чином вибрали одну.

    а) Якою є ймовірність того, що ця куля є білою? ($\frac{13}{30}$)

    б) Куля виявилась білою. Якою є ймовірність того, що ця куля знаходилась у першій урні? ($\frac{6}{13}$)

12. Одна партія містить 12 виробів, а друга 10 виробів, при цьому в кожній з них є по одному бракованому виробу. Навмання вибраний виріб з першої партії перекладається до другої, після цього:

    а) з другої партії навмання вибирається один виріб. Якою є ймовірність того, що він виявиться бракованим? ($\frac{13}{132}$)

    б) з першої партії навмання вибирається один виріб. Якою є ймовірність того, що він виявиться бракованим? ($\frac{1}{12}$)

    в) з другої партії навмання вибирається один виріб і він виявився бракованим. Якою є ймовірність того, що він належить до першої партії? ($\frac{1}{13}$)

13. Маємо дві монети – нормальну і „фальшиву", яка вдвічі частіше випадає орлом. Підкидаємо навмання вибрану монету. Випав орел. Якою є ймовірність того, що ми підкидали „фальшиву"монету? ($\frac{4}{7}$)

14. Дві фабрики виробляють певний виріб. Перша виробляє в середньому 3% бракованих виробів, друга – 5%. Навмання вибирається одна з фабрик, після чого купують 100 штук виробу. Якою є ймовірність того, що докладно два вироби будуть бракованими?

15. В урні знаходиться одна куля: біла або чорна. До урни кладуть додатково білу кулю, після чого навмання виймається одна куля, яка виявилась білою. Якою є ймовірність того, що з початку в урні була чорна куля. ($\frac{1}{3}$)

Лекція 3. Аксіоматичне визначення ймовірності

До цього часу ми розглядали експерименти, в яких множина результатів складалась не більш ніж зліченної кількості елементів, тобто простір Ω мав скінченне або зліченне число елементарних подій. Тоді ми казали, що кожна підмножина простору Ω є випадковою подією. Але ж може бути й так, що простір Ω є множиною незліченною. Якщо кинути точку на проміжок $[0, 1]$ і вважати результатом експерименту отримане число з цього проміжку, тоді простором елементарних подій буде множина $\Omega = [0, 1]$, яка є незліченною. У цьому випадку, якщо б ми прийняли кожну підмножину Ω за випадкову подію, перед нами повстали б значні труднощі. По-перше, не можемо написати $\sum_{\omega \in A} \mathbf{P}(\omega)$, адже множина A може бути незліченою. По-друге (і то є найважливіше), якщо ми захочемо визначити ймовірність на всіх підмножинах простору Ω таким чином, щоб мала місце наступна природня властивість:

$$\mathbf{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbf{P}(A_n), \qquad A_i A_j = \emptyset,\ i \neq j, \tag{1}$$

то в загальному випадку це можливо єдине тоді, коли $\mathbf{P}(A) = 0$ для всіх $A \subset \Omega$. Це пов'язано з тим, що у випадку незліченного простору Ω послідовностей $A_i,\ i = 1, 2, ...,\ A_i A_j = \emptyset,\ i \neq j$ є забагато, аби можна було в нетривіальний спосіб визначити ймовірність так, щоб для всіх таких послідовностей мала місце властивість (1).

З цією метою поняття випадкової події та ймовірності у випадку незліченного простору Ω будемо визначати іншим чином. Зрозуміло, що цей новий підхід не повинен суперечити ситуації зліченного простору Ω, який розглядався в попередніх лекціях.

Почнемо з введення нових понять, котрі визначають спеціальні класи підмножин з Ω, і які ми будемо вважати випадковими подіями.

3.1 Алгебра і σ-алгебра подій

Визначення 1. Клас U підмножин з Ω будемо називати *алгеброю*, якщо виконані наступні умови:

а) $\Omega \in U$;

б) якщо $A \in U$, тоді $\bar{A} \in U$;

в) якщо $A \in U$, $B \in U$, тоді $A \cup B \in U$.

Неважко довести, якщо U є алгеброю, $A_1, A_2, ..., A_n \in U$ і $n < \infty$, тоді $\cup_{k=1}^{n} A_k \in U$, $\cap_{k=1}^{n} A_k \in U$ (завдання1).

З цього випливає, що наведене вище визначення алгебри еквівалентне наступному.

Визначення 2. Клас множин U з Ω називається алгеброю, якщо він є замкненим стосовно скінченного числа операцій доповнення, сум та добутків (а, отже, і різниць).

Приклад 1. *Множина* $\{\emptyset, \Omega\}$ *є алгеброю подій (вона часто називається тривіальною або мінімальною).*

Приклад 2. *Нехай U буде множиною всіх підмножин Ω. Вочевидь, тоді U є алгеброю, яку часто називають максимальною.*

Цікавим є наступний приклад алгебри.

Приклад 3. *Нехай $\Omega = R^1 = [-\infty, \infty)$. Визначимо клас U підмножин з Ω наступним чином:*

$$U = \{A \subset R^1 : \ A = \cup_{i=1}^{n}[a_i, b_i), -\infty \leq a_i < b_i \leq \infty, \ 1 \leq n < \infty.\}$$

Легко довести (завдання 2), що множина U є алгеброю.

Визначення 3. Клас $\mathcal{U}$ підмножин з Ω будемо називати *σ-алгеброю,* якщо виконані наступні умови:

а) $\Omega \in \mathcal{U}$;

б) якщо $A \in \mathcal{U}$, тоді $\bar{A} \in \mathcal{U}$;

в) якщо $A_n \in \mathcal{U}$, $n = 1, 2, ...$, тоді $\cup_{n=1}^{\infty} A_n \in \mathcal{U}$.

Як і раніше, можна довести (завдання 1), що наведене визначення σ-алгебри еквівалентне наступному.

Визначення 4. Клас підмножин з Ω називається σ-алгеброю, якщо він є замкненим стосовно зліченного числа операцій доповнень, сум, добутків (а, отже, і різниць).

В прикладах 1-2, наведених вище, клас U є одночасно і σ-алгеброю, а в прикладі 3 ні, позаяк, наприклад, згідно з визначенням класу U маємо $\{1\} \notin U$, в той же час:

$$\{1\} = \cap_{n=1}^{\infty}[1; 1 + n^{-1}).$$

Наведемо кілька простих фактів стосовно σ-алгебр.

Теорема 1. *Нехай $\mathcal{U}_t$, $t \in T$ є σ-алгебрами, де T - довільна множина індексів. Тоді*

$$\mathcal{U} = \cap_{t \in T} \mathcal{U}_t$$

є σ-алгеброю.

Визначення 5. Будемо казати, що σ-алгебра $\mathcal{U}_1$ є *меншою* за σ-алгебру $\mathcal{U}_2$, якщо $\mathcal{U}_1 \subset \mathcal{U}_2$.

Визначення 6. Нехай K є певним класом підмножин з Ω. Найменша σ-алгебра, яка містить клас K називається *σ-алгеброю, побудованою над класом* K, і позначається як $\sigma(K)$.

Вочевидь, σ-алгебра $\sigma(K)$ має наступні властивості
1. $K \subset \sigma(K)$.
2. Для кожної σ-алгебри $\mathcal{U}$ такої, що $K \subset \mathcal{U}$, маємо

$$\sigma(K) \subset \mathcal{U}.$$

Те, що для кожного класу така найменша σ-алгебра існує, випливає з наступного твердження.

Теорема 2. *Для кожного класу множин K існує найменша σ-алгебра, яка містить K.*

Д о в е д е н н я. Завжди існує принаймні одна σ-алгебра, яка містить K. Такою σ-алгеброю є, наприклад, σ-алгебра всіх підмножин з Ω, або з K. Нехай $\mathcal{K}_t$, $t \in T$ - всі σ-алгебри, які містять K. Тоді $\cap_{t \in T} \mathcal{K}_t$ є найменшою σ-алгеброю. Те, що $\cap_{t \in T} \mathcal{K}_t$ є σ-алгеброю, випливає з теореми 1, а те, що вона - найменша, випливає з визначення цієї σ-алгебри. ◄

3.1.1 Борелівські σ-алгебри

Визначення 7. Нехай $\Omega = R^1 = [-\infty, +\infty)$, а множина K буде множиною всіх інтервалів вигляду $[a, b)$, $a < b$ (тобто $K = \{[a, b),\ a < b\}$). Тоді найменша σ-алгебра, яка містить K називається σ-алгеброю *борелівських множин в* R^1. Будемо позначати її наступним чином: $\mathcal{B}(R^1)$. Множини $B \in \mathcal{B}(R^1)$ називаються *борелівськими*.

Можна довести, що $\mathcal{B}(R^1) = \sigma(U)$, де U є алгеброю з прикладу 3.

Зауваження 1. Якщо у визначенні 7 замість Ω взяти довільний інтервал з R^1, наприклад $[A;B]$, а замість K взяти всі інтервали $[a,b)$, $A \leq a < b \leq B$, тоді відповідна σ-алгебра називається *борелівською σ-алгеброю множин з інтервалу* $[A;B]$. Вона записується як $\mathcal{B}_{[A;B]}$.

Наступна теорема, доведення якої полишаємо читачеві (завдання 3), описує деякі властивості множин з $\mathcal{B}(R^1)$.

Теорема 3. *Нехай $\mathcal{B}(R^1)$ є σ-алгеброю борелівських множин в R^1. Тоді*
а) $\{a\} \in \mathcal{B}(R^1)$ *для кожного* $a \in R^1$.
б) *Інтервали* $[a,b], (a,b], (a,b)$ *належать до* $\mathcal{B}(R^1)$ *для всіх* $a, b \in R^1$.
в) *Замкнута або відкрита множина в* R^1 *належить до* $\mathcal{B}(R^1)$.

3.2 Аксіоми теорії ймовірностей

Нехай задані простір елементарних подій Ω і якась σ-алгебра $\mathfrak{B}$ підмножин з Ω.

Визначення 8. Пару $(\Omega, \mathfrak{B})$ будемо називати *вимірним простором*, а кожну множину $A \in \mathfrak{B}$ *випадковою подією.*

Нехай заданий вимірний простір $(\Omega, \mathfrak{B})$. Згідно з визначенням всі множини з $\mathfrak{B}$ є подіями. Введемо тепер поняття ймовірності подій з σ-алгебри $\mathfrak{B}$.

Визначення 9. *Ймовірність* це числова функція $\mathbf{P}(\cdot)$, визначена на σ-алгебрі $\mathfrak{B}$, яка задовільняє наступним умовам:
P_1. $\mathbf{P}(A) \geq 0$ для всіх $A \in \mathfrak{B}$.
P_2. $\mathbf{P}(\Omega) = 1$.
P_3. Якщо послідовність $A_n,\ n = 1, 2, ...$ подій є такою, що $A_i A_j = \emptyset$ для $i \neq j$, тоді

$$\mathbf{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbf{P}(A_n). \tag{2}$$

Визначення 10. Трійку $(\Omega, \mathfrak{B}, \mathbf{P})$ будемо називати *ймовірносним простором.*

У такому вигляді аксіоматика теорії ймовірностей була сформульована А.М. Колмогоровим у 1933 році у праці [**?**]. Вона використовується в усіх сферах науки, які застосовують поняття і результати теорії ймовірностей.

У першій лекції, де простір елементарних подій Ω був скінченний або зліченний, ми вже визначали поняття ймовірності (див. визначення 4 на стор. 15). Тепер нам треба показати, що те визначення є частковим випадком аксіоматичного визначення ймовірності. Отже, розглянемо визначення ймовірності зі сторінки 15. Візьмемо за σ-алгебру $\mathfrak{B}$ множину всіх підмножин з Ω. Далі для довільного $A = \{\omega_{i_1}, \omega_{i_2}, ..., \omega_{i_k}, ...\} \in \mathfrak{B}$ визничимо функцію $\mathbf{P}(\cdot)$ наступним чином:

$$\mathbf{P}(A) \overset{def}{=} \sum_{k=1}^{\infty} a_{i_k}.$$

Якщо взяти до уваги пункти а), б) з визначення 4, стор. 15 а також рівність (1) зі сторінки 288 (з $n = \infty$), тоді легко зрозуміти, що для визначених $\mathfrak{B}$ і $\mathbf{P}(\cdot)$ справедливі аксіоми $P_1 - P_3$ з визначення 9.

З іншої сторони, нехай для вимірного простору $(\Omega, \mathfrak{B})$, де $\Omega = \{\omega_1, \omega_2, ...\}$, і σ-алгебра $\mathfrak{B}$– множина всіх підмножин з Ω, визначена ймовірність у сенсі визначення 9.

Тепер кожна підмножина з Ω буде подією, що відповідає визначенню на стор. 14. Позаяк кожна елементарна подія ω_i може бути трактована як одноелементна множина $\{\omega_i\}$, яка є підмножиною Ω, тоді згідно з визначеням 9 задані числа $a_i \overset{def}{=} \mathbf{P}(\omega_i) \geq 0$, отже, умова а) зі стор. 15 є виконаною. Далі, множину Ω можна трактувати як $\cup_{i=1}^{\infty}\{\omega_i\}$, і оскільки множини $\{\omega_i\}$, $\{\omega_j\}$ є різними для $i \neq j$, тоді згідно з властивостями P_2, P_3 будемо мати:

$$1 = \mathbf{P}(\Omega) = \mathbf{P}(\cup_{i=1}^{\infty}\{\omega_i\}) = \sum_{i=1}^{\infty} \mathbf{P}(\{\omega_i\}) = \sum_{i=1}^{\infty} a_i.$$

Отже, умова б) зі стор. 15 також виконана. Якщо тепер A є будь-якою подією, тоді $A = \cup_{i:\omega_i \in A}\{\omega_i\}$ і, застосовуючи знову умову P_3, будемо мати:

$$\mathbf{P}(A) = \mathbf{P}(\cup_{i:\omega_i \in A}\{\omega_i\}) = \sum_{\omega_i \in A} \mathbf{P}(\omega_i).$$

Ми отримали визначення ймовірності, яким користувались у випадку зліченного простору елементарних подій.

ЗАВДАННЯ

1. Довести, що якщо U є алгеброю, $A_1, A_2, ..., A_n \in U$ і $n < \infty$, тоді $\cup_{k=1}^{n} A_k \in U$ і $\cap_{k=1}^{n} A_k \in U$. А якщо U є σ-алгеброю, $A_1, A_2, ... \in U$, тоді $\cup_{k=1}^{\infty} A_k \in U$ і $\cap_{k=1}^{\infty} A_k \in U$.

2. Довести, що множина U з прикладу 3 є алгеброю.

3. Довести теорему 3.

   Вказівка При доведенні пункту в) потрібно використати факт, що кожну відкриту множину з R^1 можна представити у вигляді скінченної, або зліченної суми інтервалів, зазначених у пункті б), а кожна замкнута множина є доповненням відкритої.

4. Нехай $\Omega = [-\infty, \infty)$ і $\mathcal{F}$ є $\sigma-$ алгебра всіх підмножин Ω . Для кожного$A \in \mathcal{F}$ покладемо

   $$\mathbf{P}(A) = \sum_{n \in A} 2^{-n},$$

   де набуває цілих додатніх значень. Чи є $(\Omega, \mathcal{F}, \mathbf{P})$ ймовірнісним простором?

5. Нехай $A_1,\ A_2$ випадкові події, $\mathcal{M}-$ найменша σ–алгебра події, яка містить події $A_1,\ A_2$. Довести, що кожна подія з $\mathcal{M}$ є сумою деякого числа випадкових подій $A_1 \cap A_2,\ A_1 \cap \overline{A}_2,\ \overline{A}_1 \cap A_2,\ \overline{A}_1 \cap \overline{A}_2$. Скільки різних подій містить σ–алгебра $\mathcal{M}$?

Лекція 4. Властивості ймовірності

З формальної точки зору теорему про властивості ймовірності (теорема 2, стор.15) треба тепер довести на підставі аксіом $P_1 - P_3$. Але тепер ця теорема набуває іншого вигляду, позаяк, наприклад, властивість 2) з цієї теореми тепер випливають з аксіоми P_3. Теж саме можна сказати про рівність (1) на стор. 288.

Теорема 1. *Для довільних подій A, B мають місце наступні властивості.*

1) $\mathbf{P}(\emptyset) = 0, \quad \mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A), \quad \mathbf{P}(A) \leq 1$;

2) *якщо* $A \subset B$, *тоді* $\mathbf{P}(B \backslash A) = \mathbf{P}(B) - \mathbf{P}(A)$ *і* $\mathbf{P}(A) \leq \mathbf{P}(B)$;

3) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$;

4) $\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$.

Д о в е д е н н я.

1). Позаяк $\Omega \cap \emptyset = \emptyset$ і $\Omega \cup \emptyset = \Omega$, то на підставі аксіом P_1, P_3 маємо:

$$1 = \mathbf{P}(\Omega) = \mathbf{P}(\Omega \cup \emptyset) = \mathbf{P}(\Omega) + \mathbf{P}(\emptyset) = 1 + \mathbf{P}(\emptyset).$$

Звідки $\mathbf{P}(\emptyset) = 0$. Оскільки множини A , $\overline{A}$ є різними, тоді знову на підставі аксіом P_1, P_3 маємо:

$$1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup \bar{A}) = \mathbf{P}(A) + \mathbf{P}(\bar{A}),$$

звідки $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$, а, отже, $\mathbf{P}(A) \leq 1$.

2). Якщо $A \subset B$, то $B = B \backslash A \cup A$, отже, (позаяк $(B \backslash A) \cap A = \emptyset$)

$$\mathbf{P}(B) = \mathbf{P}(B \backslash A) + \mathbf{P}(A).$$

3). $A \cup B = A \backslash AB \cup B$ і $(A \backslash AB) \cap B = \emptyset$, а тому

$$\mathbf{P}(A \cup B) = \mathbf{P}(A \backslash AB \cup B) = \mathbf{P}(B) + \mathbf{P}(A \backslash AB) = \mathbf{P}(B) + \mathbf{P}(A) - \mathbf{P}(AB).$$

4). Ця властивість легко випливає з попередньої. ◄

Зауваження 1. У випадку зліченного простору елементарних подій ми ввели такі поняття як незалежність подій, умовна ймовірність. Крім того, вивели формулу повної ймовірності, формулу Байєса тощо. Легко зауважити, що при визначенні цих понять і виводі тверджень факт зліченності простору Ω не використовувався (крім доведення формули (4), стор. 18.) Нам була потрібна теорема 2, стор.15, а оскільки ця теорема в загальному випадку випливає з аксіом ймовірності, тоді всі поняття, визначення, теореми з попередніх лекцій залишаються в силі і в загальному випадку. Це стосується також формули 4, стор. 18 (завдання 4).

4.1 Властивості ймовірності як функції множини

Вивчимо властивості функції $\mathbf{P}(\cdot\}$ як функції, визначеної на множинах. Для цього нам будуть потрібні певні нові поняття.

Нехай $A_n, n = 1, 2, ...$ є нескінченною послідовністю подій $A_n \subset \Omega$.

Визначення 1. Послідовність подій A_n будемо називати *незростаючою*, якщо

$$A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq A_{n+1} \supseteq \cdots .$$

Тоді подію $A = \cap_{n=1}^{\infty} A_n$ називають *границею послідовності* A_n і позначають

$$\lim_{n\to\infty} A_n = A = \cap_{n=1}^{\infty} A_n.$$

Визначення 2. Послідовність подій A_n будемо називати *неспадною,* якщо

$$A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq A_{n+1} \subseteq \cdots .$$

Тоді подію $A = \cup_{n=1}^{\infty} A_n$ називають *границею послідовності* A_n і позначають за аналогією

$$\lim_{n\to\infty} A_n = A = \cup_{n=1}^{\infty} A_n.$$

Незростаючі або неспадні послідовності називають *монотонними.*

Теорема 2. *Якщо послідовність подій A_n, $n = 1, 2, ...$ є монотонною, тоді*

$$\mathbf{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbf{P}(A_n). \qquad (1)$$

Зауваження 2. Якщо послідовність A_n, $n = 1, 2, \ldots$ є неспадною, то рівність (1) може бути переписана наступним чином:

$$\mathbf{P}(\cup_{n=1}^{\infty} A_n) = \mathbf{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

Аналогічно для незростаючих послідовностей цю рівність можна переписати так:

$$\mathbf{P}(\cap_{n=1}^{\infty} A_n) = \mathbf{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

Доведення. Нехай послідовність A_n, $n = 1, 2, \ldots$ буде неспадною, тобто $A_n \subseteq A_{n+1}$. Позначимо

$$B_n = A_{n+1} \backslash A_n, \ i = 1, 2, \ldots$$

Легко перевірити, що $B_i B_j = \emptyset, i \neq j$ і

$$A_n = \cup_{i=1}^{n-1} B_i, \quad n < \infty, \qquad \cup_{i=1}^{\infty} A_i = \cup_{i=1}^{\infty} B_i.$$

Отже,

$$\mathbf{P}(\lim_{n\to\infty} A_n) = \mathbf{P}(\cup_{n=1}^{\infty} A_n) = \mathbf{P}(\cup_{n=1}^{\infty} B_n) = \sum_{i=1}^{\infty} \mathbf{P}(B_i) =$$
$$= \lim_{n\to\infty} \sum_{i=1}^{n-1} \mathbf{P}(B_i) = \lim_{n\to\infty} \mathbf{P}(\cup_{i=1}^{n-1} B_i) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

Якщо послідовність A_n, $n \geq 1$ - незростаюча, тоді $\bar{A}_n$, $n = 1, 2, \ldots$ є неспадною і

$$\mathbf{P}(\lim_{n\to\infty} A_n) = \mathbf{P}(\cap_{n=1}^{\infty} A_n) = 1 - \mathbf{P}(\overline{\cap_{n=1}^{\infty} A_n}) = 1 - \mathbf{P}(\cup_{n=1}^{\infty} \bar{A}_n) =$$
$$= 1 - \mathbf{P}(\lim_{n\to\infty} \bar{A}_n) = 1 - \lim_{n\to\infty} \mathbf{P}(\bar{A}_n) = \lim_{n\to\infty} (1 - \mathbf{P}(\bar{A}_n)) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

◀

При побудові конкретного ймовірносного простору $(\Omega, \mathfrak{B}, \mathbf{P})$ найважчим завданням є побудова ймовірності $\mathbf{P}(\cdot)$. Основною причиною цього є те, що повинна бути виконана властивість P_3. Наступне твердження часто допомагає у розв'язанні цієї проблеми.

Теорема 3. *Нехай Ω є простором елементарних подій, а U - алгеброю підмножин з Ω. Нехай функція $Q(\cdot)$, визначена на U, є такою, що $Q(\Omega) = 1$, $Q(A) \leq 1$, $A \in U$, і виконані наступні умови:*

1) *якщо $A_i \in U$, $i = 1, 2, \ldots, n < \infty$, і $A_i A_j = \emptyset$ для $i \neq j$, тоді*

$$Q(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} Q(A_i); \tag{2}$$

2) *для кожної незростаючої послідовності* $A_n \in U$, $n = 1, 2, ...$ *і такої, що* $\lim_{n\to\infty} A_n = \emptyset$ *маємо:*

$$\lim_{n\to\infty} Q(A_n) = 0. \tag{3}$$

Тоді для кожної послідовності подій $A_n \in U$, $n = 1, 2, ...$ *такої, що* $A_i A_j = \emptyset$ *для* $i \neq j$, *і* $\cup_{i=1}^{\infty} A_i \in U$, *будемо мати:*

$$Q(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} Q(A_n). \tag{4}$$

Зауваження 3. Властивість (2) для функції $Q(\cdot)$ називається *властивістю скінченної адитивності*, а властивість (4) має назву *властивість зліченної адитивності* функції $Q(\cdot)$.

Д о в е д е н н я. Нехай $A_n \in U$, $n = 1, 2, ...$ є такою послідовністю подій, що $A_i A_j = \emptyset$ для $i \neq j$, і $\cup_{i=1}^{\infty} A_i \in U$. Позначимо

$$A = \cup_{n=1}^{\infty} A_n, \qquad C_n = A \setminus \cup_{i=1}^{n} A_i, \; n = 1, 2, ...$$

Тоді $C_{n+1} \subseteq C_n$. Крім того, неважко зрозуміти, що $\lim_{n\to\infty} C_n = \emptyset$. Отже, маємо

$$0 = \lim_{n\to\infty} \mathbf{P}(C_n) = \lim_{n\to\infty} \mathbf{P}(A \setminus \cup_{i=1}^{n} A_n) = \mathbf{P}(A) - \lim_{n\to\infty} \mathbf{P}(\cup_{i=1}^{n} A_n) =$$

$$= \mathbf{P}(A) - \lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{P}(A_i) = \mathbf{P}(A) - \sum_{i=1}^{\infty} \mathbf{P}(A_i).$$

Звідки

$$\mathbf{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbf{P}(A_i).$$

◀

Зауваження 4. Схема застосування цієї теореми для побудови ймовірносного простору $(\Omega, \mathfrak{B}, \mathbf{P})$ є наступною. На просторі елементарних подій Ω визначаємо алгебру подій U. Далі нам треба визначити на цій алгебрі певну функцію множин Q, для якої виконуються умови теореми 3. Тоді згідно з цією теоремою функція Q буде злічено адитивною на U. Потім нам треба скористатися теоремою Каратеодорі з теорії міри. Ця теорема стверджує, що функція Q може бути однозначно продовжена[1] на σ- алгебру, побудовану над U (тобто на $\sigma(U)$) до функції $\mathbf{P}(\cdot)$ такої, що $\mathbf{P}(\cdot)$ буде зліченно адитивною на $\sigma(U)$. Отже, $(\Omega, \sigma(U), \mathbf{P})$ буде ймовірносним простором. Приклад застосування цієї схеми наведемо нижче.

[1] Будемо казати, що функція φ, визначена на U, продовжена до функції $\widehat{\varphi}$, визначеної на $\widehat{U}$, якщо $U \subset \widehat{U}$ і $\varphi(x) = \widehat{\varphi}(x)$ для $x \in U$.

4.1.1 Геометричне визначення ймовірності

Розглянемо наступний експеримент. На проміжок $[a, b]$ навмання кидаємо точку. Вираз "навмання" тут означає, що для двох довільних інтервалів $[a_1, b_1] \subset [a, b]$, $[a_2, b_2] \subset [a, b]$ рівної довжини (тобто $b_1 - a_1 = b_2 - a_2$) ймовірність того, що точка потрапить до $[a_1, b_1]$ дорівнює ймовірності того, що точка потрапить до інтервалу $[a_2, b_2]$. Вочевидь, в нашому експерименті $\Omega = [a, b]$. За алгебру візьмемо

$$U(a, b) = \{A \subset [a, b] : \; A = \cup_{i=1}^{n} [\alpha_i, \beta_i), a \le \alpha_i < \beta_i \le b, 1 \le n < \infty\}.$$

На U визначимо функцію $\lambda(\cdot)$ наступним чином:

$$\lambda(A) = \sum_{i=1}^{n} (\beta_i - \alpha_i), \tag{5}$$

де $A = \cup_{i=1}^{n} [\alpha_i, \beta_i)$, й інтервали $[\alpha_i, \beta_i)$, $i = 1, ..., n$ є несумісними. Позаяк кожна множина з $U(a, b)$ може бути записана у вигляді суми несумісних інтервалів, то функція $\lambda(\cdot)$ є визначеною на $U(a, b)$.

Неважко перевірити, що для функції

$$Q(A) = \frac{\lambda(A)}{b - a}$$

виконуються умови теореми 3. Отже, згідно з зауваженням 4 на σ-алгебрі борелівських множин з $[a, b]$, тобто на $\mathcal{B}_{[a,b]} = \sigma(U(a, b))$ існує функція $\mathbf{P}(\cdot)$ така, що $(\Omega, \mathcal{B}_{[a,b]}, \mathbf{P})$ є ймовірносним простором. Зрозуміло, що значення $\lambda(A)$ для $A \in U$ є звичайною довжиною інтервалу або суми інтервалів. В теорії міри Лебегом доведено, що поняття "довжина" може бути узагальнено на множини з $\sigma(U(a, b))$, тобто функція $\lambda(\cdot)$ може бути продовжена з U на $\sigma(U(a, b))$, і для кожної множини $A \in \sigma(U(a, b))$ визначена величина $\lambda(A) \ge 0$, яку будемо називати "довжиною" множини A. Тоді для кожної множини $A \in \mathcal{B}_{[a,b]}$ маємо:

$$\mathbf{P}(A) = \frac{\lambda(A)}{b - a}, \qquad A \in \mathcal{B}_{[a,b]}. \tag{6}$$

А якщо $A \in U$, то в (6) беруть участь значення функції λ з (5).

Наведена схема побудови ймовірносного простору називається *геометричним визначенням ймовірності.*

Аналогічно можна дослідити ситуацію, коли в ролі Ω бере участь певна область евклідового простору R^n. Розглянемо стисло тільки випадок R^2. Нехай область $\Omega \subset R^2$ і, крім того, $0 < S(\Omega) < \infty$, де $S(A)$ означає площу області A. З теорії міри Лебега випливає, що сукупність підмножин з Ω, для яких можемо визначити поняття "площа" є σ-алгеброю, яку будемо позначати $\mathcal{B}_\Omega$. Наш експеримент полягає в тому, що в область Ω випадково кидається точка x. Як і раніше, слово "випадково" означає, що якщо є дві множини $A_1 \subset \mathcal{B}_\Omega$, $A_2 \subset \mathcal{B}_\Omega$ такі що $S(A_1) = S(A_2)$, то ймовірність потрапляння до них точки x є однаковою. Визначимо ймовірність подій в нашому експерименті наступним чином:

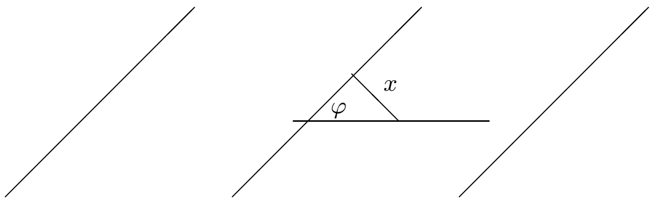
$$\mathbf{P}(A) = \frac{\lambda_n(A)}{\lambda_n(\Omega)}, \qquad A \in \mathcal{B}_\Omega. \tag{7}$$

Легко перевірити, що для введеної таким чином функції $\mathbf{P}(\cdot)$ виконані аксіоми ймовірності, а отже, ми побудували ймовірносний простір $(\Omega, \mathcal{B}_\Omega, \mathbf{P})$. Як і раніш, ймовірність, визначену формулой (7), будемо називати геометричною.

Застосуємо поняття геометричної ймовірності до розв'язання наступної задачі, відомої як *задача Бюффона для голки.*

Приклад 1. *Площина поділена паралельними прямими, відстань між якими дорівнює $2a$. На цю площину випадковим чином кидається голка довжини $2l$, $l < a$. Яка є ймовірність того, що вона перетне одну з паралельних прямих.*

Р о з в' я з о к. Нехай x буде відстанню від середини голки до найблищої прямої, а ϕ - кутом, який утворює голка з цією прямою



Числа x і ϕ точно визначають положення голки по відношенню до найблищої прямої. Таким чином, ймовірносним простором буде прямокутник:

$$\Omega = \{(x, \phi) : 0 \le x \le a, 0 \le \phi \le \pi\}.$$

Голка перетинає найблищу пряму тоді і тільки тоді, коли $x \le l \sin \phi$. Нехай

$$C = \{(x, \phi) : x \le l \sin \phi, (x, \phi) \in \Omega\}$$

Оскільки кожний результат нашого експерименту є однаково можливим, то ми можемо застосувати схему геометричної ймовірності, і тоді ймовірність нашої події дорівнює $S(C)/S(\Omega)$, де $S(A)$ площа множини A. Після підрахунків

$$S(\Omega) = a\pi, \qquad S(C) = \int_0^\pi l \sin \phi d\phi = 2l,$$

Отже, остаточно маємо $\frac{S(C)}{S(\Omega)} = \frac{2l}{a\pi}$.

4.2^*. Незалежність σ- алгебр

Нехай $(\Omega, \mathfrak{B}, \mathbf{P})$ буде ймовірнісним простором а $\mathcal{A}_1,\ \mathcal{A}_2 \subset \mathfrak{B}$ - деякі класи подій з $\mathfrak{B}$.

Визначення 3. Класи $\mathcal{A}_1,\ \mathcal{A}_2$ називаються *незалежними*, якщо довільні події $A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2$ є незалежними, тобто якщо $\mathbf{P}(A_1A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2)$ для довільних подій $A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2$.

Нехай тепер $\mathcal{A}_1,\ \mathcal{A}_2 \subset \mathfrak{B}$ будуть алгебрами подій а $\mathcal{U}_1, \mathcal{U}_2$ позначають σ- алгебри, згенерованими алгебрами $\mathcal{A}_1,\ \mathcal{A}_2$ відповідно, тобто $\mathcal{U}_1 = \sigma(\mathcal{A}_1), \mathcal{U}_2 = \sigma(\mathcal{A}_2)$. Зрозуміло, що $\mathcal{U}_1, \mathcal{U}_2$ є підмножинами σ- алгебри $\mathfrak{B}$, тобто $\mathcal{U}_1, \mathcal{U}_2 \subset \mathfrak{B}$.

Ціллю цього підрозділу є доведення наступного результату.

Теорема 4. *Якщо алгебри $\mathcal{A}_1,\ \mathcal{A}_2$ є незалежними, то σ-алгебри $\mathcal{U}_1, \mathcal{U}_2$ також є незалежними.*

Для доведення цього твердження нам буде потрібен деякий допоміжний результат, але спочатку

Визначення 4. *Різниця симетрична* множин A, B позначається символом Δ і означається наступним чином $A\Delta B = (A\backslash B) \cup (B\backslash A)$.

Легко довести, що якщо $A\Delta B = \emptyset$, то $A = B$ (завдання 11).

Лема 1. *Нехай $\mathcal{A} \subset\ \mathfrak{B}$ буде деякою алгеброю подій і $\mathcal{U} = \sigma(\mathcal{A})$. Тоді для кожної події $A \in \mathcal{A}$ існує послідовність подій $A_n \in \mathcal{A}, n \geq 1$, така що*

$$\lim_{n\to\infty} \mathbf{P}(A_n \Delta A) = 0. \tag{8}$$

Зауваження 5. З означення симетричної різниці випливає, що якщо має місце (8), то $\lim_{n\to\infty} \mathbf{P}(A_n\backslash A) = \lim_{n\to\infty} \mathbf{P}(A\backslash A_n) = 0$, а тому з рівності $\mathbf{P}(A) = \mathbf{P}(A_n) - \mathbf{P}(A_n\backslash A) + \mathbf{P}(A\backslash A_n)$ отримуємо $\mathbf{P}(A) = \lim_{n\to\infty} \mathbf{P}(A_n)$. Тепер з леми 1 випливає, що кожна подія з $\mathcal{U}$ може бути апроксимована подіями з породжуючої її алгебри з докладністю до події нульової ймовірності: тобто для кожної події $A \in \mathcal{U} = \sigma(\mathcal{A})$ існує послідовність подій $A_n \in \mathcal{A}$, така що $\mathbf{P}(A) = \lim_{n\to\infty} \mathbf{P}(A_n)$.

Зауваження 6. Послідовність A_n з леми 1 будемо називати *апроксимуючою* послідовністю для A.

Для апроксимуючою послідовності маємо $\mathbf{P}(A) = \lim_{n\to\infty} \mathbf{P}(A_n)$.

Д о в е д е н н я леми 4. Означимо множину подій наступним чином

$$\mathcal{U}^* = \{A \in \mathcal{U} : \exists\, A_n \in \mathcal{A}, n \geq 1 \text{ така що } \lim_{n\to\infty} \mathbf{P}(A_n \Delta A) = 0.\}$$

Іншими словами, множина $\mathcal{U}^*$ складається з подій з $\mathcal{U}$ для яких існує послідовність подій з $\mathcal{A}$ така, що виконується співвідношення (8). Очевидно $\mathcal{A} \subset \mathcal{U}^*$ (для $A \in \mathcal{A}$ досить взяти $A_n = A$) а отже якщо ми покажемо, що $\mathcal{U}^*$ є σ-алгеброю, то тоді автоматично $\mathcal{U} = \sigma(\mathcal{A}) \subset \mathcal{U}^*$ і наше твердження буде доведено.

Доведемо спочатку, що $\mathcal{U}^*$ є алгеброю. Нехай $A \in \mathcal{U}^*$ і $A_n \in \mathcal{A}, n \geq 1$ послідовність для якої $\lim_{n\to\infty} \mathbf{P}(A_n \Delta A) = 0$. Тоді для послідовності $\overline{A}_n \in \mathcal{A}, n \geq 1$ маємо

$$\lim_{n\to\infty} \mathbf{P}(\overline{A}_n \Delta \overline{A}) = \lim_{n\to\infty} \mathbf{P}(\overline{A}_n \backslash \overline{A}) + \lim_{n\to\infty} \mathbf{P}(\overline{A} \backslash \overline{A}_n) = \lim_{n\to\infty} \mathbf{P}(\overline{A}_n A) +$$
$$+ \lim_{n\to\infty} \mathbf{P}(\overline{A} A_n) = \lim_{n\to\infty} \mathbf{P}(A \backslash A_n) + \lim_{n\to\infty} \mathbf{P}(A_n \backslash A) = \lim_{n\to\infty} \mathbf{P}(A_n \Delta A) = 0$$

а отже $\overline{A} \in \mathcal{U}^*$.

Аналогічно можна довести, що якщо $A, B \in \mathcal{U}^*$ і послідовності $A_n, B_n,\ n \geq 1$ апроксимують відповідно A та B, то послідовність $A_n \cup B_n,\ n \geq 1$ апроксимує $A \cup B$ (завдання 12). Отже $\mathcal{U}^*$ є алгеброю.

Доведемо тепер, що $\mathcal{U}^*$ є σ - алгеброю. Тут нам буде потрібен такий простий результат.

Лема 2. *Якщо* $\lim_{n\to\infty} \mathbf{P}(A_{nk} \Delta A_k) = 0$ *для кожного фіксованого* $k = 1, 2, ...$ *і* $\lim_{k\to\infty} \mathbf{P}(A_k \Delta A) = 0$, *то існує послідовність* $k = k(m),\ n = n(m),\ m = 1, 2, ...$ *така, що* $k = k(m) \to \infty,\ n = n(m) \to \infty$ *якщо* $m \to \infty$ *і* $\lim_{m\to\infty} \mathbf{P}(A_{n(m)k(m)} \Delta A) = 0$.

Д о в е д е н н я. Для довільних n, k маємо (завдання 13)

$$\mathbf{P}(A_{nk} \Delta A) = \mathbf{P}(A_{nk} A_k \overline{A}) + \mathbf{P}(A_{nk} \overline{A}_k \overline{A}) + \mathbf{P}(A \overline{A}_{nk} A_k) + \mathbf{P}(A \overline{A}_{nk} \overline{A}_k) \leq$$
$$\leq \mathbf{P}(A_k \overline{A}) + \mathbf{P}(A_{nk} \overline{A}_k) + \mathbf{P}(\overline{A}_{nk} A_k) + \mathbf{P}(A \overline{A}_k) =$$
$$= \mathbf{P}(A_k \Delta A) + \mathbf{P}(A_{nk} \Delta A_k). \tag{9}$$

Нехай послідовність $\varepsilon_m > 0$ буде такою, що $\lim_{m\to\infty} \varepsilon_m = 0$. З умови леми випливає, що для кожного m ми можемо знайти $k = k(m)$ таке, що $\mathbf{P}(A_{k(m)} \Delta A) \leq \varepsilon_m / 2$. Аналогічно для вибраного $k = k(m)$ ми можемо знайти $n = n(m)$ таке, що $\mathbf{P}(A_{n(m)k(m)} \Delta A_{k(m)}) \leq \varepsilon_m / 2$. А тому з (9) маємо

$$\mathbf{P}(A_{n(m)k(m)} \Delta A) \leq \varepsilon_m / 2 + \varepsilon_m / 2 = \varepsilon_m.$$

Останяя нерівність доводить нашу лему.

Тепер вже легко завершити доведення леми. Нехай $A_i \in \mathcal{U}^*,\ i \geq 1$ і послідовності $A_{ni} \in \mathcal{A}, n \geq 1$ апроксимують $A_i,\ i \geq 1$. Позначимо $C = \cup_{i=1}^{\infty} A_i$, $C_k = \cup_{i=1}^{k} A_i$. Легко зрозуміти, що $\lim_{k\to\infty} \mathbf{P}(C_k \Delta C) = 0$. Оскільки $\mathcal{U}^*$ є алгеброю, то $C_k = \cup_{i=1}^{k} A_i \in \mathcal{U}^*$ для всіх $k < \infty$, а отже існує послідовність подій $D_{nk} \in \mathcal{A}$ таких, що $\lim_{n\to\infty} P(D_{nk}\Delta C_k)$ для кожного $k = 1, 2, ...$ Тепер з леми 2 випливає, що для деякої послідовності $D_m \in \mathcal{A},\ m = 1, 2, ...$ маємо $\lim_{m\to\infty} \mathbf{P}(D_m \Delta C) = 0$, а отже $C = \cup_{i=1}^{\infty} A_i \in \mathcal{U}^*$, що і доводить наше твердження.

Д о в е д е н н я теореми 4. Нехай $A \in \mathcal{U}_1$, $B \in \mathcal{U}_2$. Тоді згідно леми 1 існують послідовності подій $A_n \in \mathcal{A}_1$, $B_n \in \mathcal{A}_2$ такі, що $\mathbf{P}(A_n \Delta A) \to 0$, $\mathbf{P}(B_n \Delta B) \to 0$, якщо $n \to \infty$ і події A_i, B_j є незалежними для всіх i, j. З співвідношення

$$\mathbf{P}(AB\Delta A_nB_n) = \mathbf{P}(AB\overline{A_nB_n}) + \mathbf{P}(\overline{AB}A_nB_n) = \mathbf{P}\big(AB(\overline{A}_n \cup \overline{B}_n)\big)+$$
$$+\,\mathbf{P}\big((\overline{A} \cup \overline{B})A_nB_n\big) \leq \mathbf{P}(AB\overline{A}_n)+\mathbf{P}(AB\overline{B}_n)+\mathbf{P}(\overline{A}A_nB_n)+\mathbf{P}(\overline{B}A_nB_n) \leq$$
$$\leq \mathbf{P}(A\overline{A}_n)+\mathbf{P}(B\overline{B}_n)+\mathbf{P}(\overline{A}A_n)+\mathbf{P}(\overline{B}B_n)=\mathbf{P}(A\Delta A_n)+\mathbf{P}(B\Delta B_n) \underset{n\to\infty}{\longrightarrow} 0$$

випливає, що послідовність A_nB_n апроксимує подію AB. Тепер маємо

$$\mathbf{P}(AB) = \lim_{n\to\infty} \mathbf{P}(A_nB_n) = \lim_{n\to\infty} \mathbf{P}(A_n)\mathbf{P}(B_n) = \mathbf{P}(AB),$$

що і доводить нашу теорему.

ЗАВДАННЯ

1. Нехай $A_n = \left(\frac{1}{2n}, \frac{1}{n}\right)$. Описати множини $A = \cup_{n=1}^{\infty} A_n$, $A = \cap_{n=1}^{\infty} A_n$.

2. Довести, що якщо події A і B є незалежними, і $A \subset B$, тоді або $\mathbf{P}(A) = 0$, або $\mathbf{P}(B) = 1$.

3. Нехай $AB = \emptyset$. Довести, що $\mathbf{P}(A) \le \mathbf{P}(\bar{B})$.

4. Довести, що для довільних подій $A_1, ..., A_n$ має місце рівність:

$$\mathbf{P}(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbf{P}(A_i) - \sum_{i<j} \mathbf{P}(A_i A_j) + ... + (-1)^{n-1} \mathbf{P}(A_1 ... A_n).$$

5. Довести, що множина значень $\mathbf{P}(A)$ у випадку довільного простору $(\Omega, \mathfrak{B}, \mathbf{P})$ є замкненою множиною.

6. Дріт довжиною 7 см загнуто під прямим кутом у випадково вибраному місці. Якою є ймовірність того, що відстань між його кінцями є більшою за 5 см?

7. На площині проведені паралельні прямі на відстані $2a$. На площину випадковим чином кидається квадрат зі стороною b ($\sqrt{2}b < 2a$). Якою є ймовірність того, що квадрат не перетне жодної прямої?

8. У колі з радіусом R випадковим чином проведена хорда. Визначити ймовірність того, що її довжина не перевищує сторони рівностороннього трикутника, який є вписаним в це коло.

9. Визначити ймовірність того, що розв'язки квадратного рівняння $x^2 + 2ax + b = 0$ є дійсними, якщо коефіцієнти можуть з однаковими ймовірностями прийняти значення в прямокутнику:

   а) $-k \le a \le k,\ -k^2 \le b \le k^2,$ б) $-k \le a \le k, -l \le b \le l,\ l < k^2,$

   с) $-k \le a \le k,\ -l \le b \le l,\ l > k^2.$

10. На проміжку довжиною l випадковим чином взято дві точки. Якою є ймовірність того, що відстань між ними не перевищує r, $0 \le r \le l$? $\left(\frac{r(2l-r)}{l^2}\right)$

11. Довести, що якщо $A \Delta B = \emptyset$, то $A = B$.

12. Довести, що якщо $A, B \in \mathcal{U}^*$ і послідовності $A_n \in \mathcal{A},\ B_n \in \mathcal{A},\ n \ge 1$ апроксимують відповідно A та B, то послідовність $A_n \cup B_n,\ n \ge 1$ апроксимує $A \cup B$.

13. Довести співвідношення (9).

Лекція 5. Випадкова величина і функція розподілу

5.1 Випадкова величина

Нехай є заданим довільний ймовірносний простір $(\Omega, \mathfrak{B}, \mathbf{P})$.

Визначення 1. *Випадковою величиною* ξ будемо називати функцію $\xi = \xi(\omega)$, що відображає множину Ω в множину дійсних чисел R^1, для якої виконується наступна умова:

$$\{\omega : \xi(\omega) < x\} \in \mathfrak{B} \tag{1}$$

для кожного $x \in R^1$.

Зауваження 1. Умова (1) є рівнозначною наступній:

$$\{\omega : \xi(\omega) \in B\} \in \mathfrak{B}$$

для кожної борелівської множини $B \subset R^1$ (див. [**?**], стор. 55)).

Нехай $\xi(\omega)$ є випадковою величиною. Позаяк $\{\omega : \xi(\omega) < x\} \in \mathfrak{B}$ для всіх $x \in R^1$, тоді можемо визначити ймовірність $\mathbf{P}(\{\omega : \xi(\omega) < x\})$, яку для зручності будемо записувати як $\mathbf{P}(\xi < x)$. Цю ймовірність можна трактувати як функцію від змінної x і ця функція має спеціальну назву, що є змістом наступного визначення.

Визначення 2. Функцію $F(x) = \mathbf{P}(\{\omega : \xi(\omega) < x\})$ будемо називати *функцією розподілу* випадкової величини ξ.

Наступна теорема описує головні властивості функції розподілу.

Теорема 1. *Функція розподілу $F(x)$ випадкової величини ξ має наступні властивості:*

1) *якщо $x_1 < x_2$, то $F(x_1) \le F(x_2)$;*

2) $\lim_{x\to-\infty} F(x) = 0$ *і* $\lim_{x\to\infty} F(x) = 1$*;*

3) $\lim_{x\uparrow x_0} F(x) = F(x_0)$;

4) $\mathbf{P}(\xi \in [a;b)) = F(b) - F(a)$ *для довільних* $a,\ b \in R^1$, $a < b$.

Д о в е д е н н я. 1) Якщо $x_1 < x_2$, то $\{\omega : \xi(\omega) < x_1\} \subseteq \{\omega : \xi(\omega) < x_2\}$, і згідно властивості 2 теореми 1 (стор. 35) будемо мати:

$$F(x_1) = \mathbf{P}(\xi < x_1) \leq \mathbf{P}(\xi < x_2) = F_2(x).$$

2) Розглянемо числову послідовність x_n, $n = 1, 2, \ldots$, таку що x_n спадає і $x_n \to -\infty$, $n \to \infty$. Позначимо $A_n = \{\xi < x_n\}$. З того, що послідовність $x_n \to -\infty$ монотонно, випливає, що послідовність множин A_n є незростаючою і, вочевидь, $\cap_{n=1}^{\infty} A_n = \emptyset$. На підставі теореми 2 (стор. 36) маємо $\lim_{n\to\infty} \mathbf{P}(A_n) = 0$, або $\lim_{n\to\infty} F(x_n) = 0$. З цього результату та з монотонності функції $F(x)$ випливає, що

$$\lim_{x\to -\infty} F(x) = 0.$$

Доведення властивості $\lim_{x\to\infty} F(x) = 1$ є аналогічним (завдання 3).

3) Нехай $A = \{\xi < x_0\}$, $A_n = \{\xi < x_n\}$, де послідовність $x_n, n = 1, 2, \ldots$ зростає і $x_n \uparrow x_0$. Послідовність множин A_n є, вочевидь, неспадаючою і $\lim_{n\to\infty} A_n = A$. Будемо мати:

$$\lim_{x_n \uparrow x_0} F(x) = \lim_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(A) = F(x_0).$$

4) Позаяк $\{\xi \in [a, b)\} = \{\xi < b\} \backslash \{\xi < a\}$ і $\{\xi < a\} \subset \{\xi < b\}$, то

$$\mathbf{P}(\xi \in [a, b)) = \mathbf{P}(\xi < b) - \mathbf{P}(\xi < a) = F(b) - F(a).$$

◀

Твердження 1 можна трактувати так: умови 1)-3) є необхідними, аби функція $F(x)$ була функцією розподілу. Наступна теорема показує, що ці умови є також достатніми, аби функція $F(x)$ була функцією розподілу.

Теорема 2. *Якщо функція* $F(x)$ *має властивості* $1) - 3)$ *з попередньої теореми, то вона є функцією розподілу певної випадкової величини, тобто існує ймовірносний простір* $(\Omega, \mathfrak{B}, \mathbf{P})$ *і випадкова величина* ξ, *визначена на цьому просторі, така, що* $F(x) = \mathbf{P}(\xi < x)$.

Д о в е д е н н я. Нехай за простір елементарних подій буде $\Omega = R^1 = [-\infty, \infty)$. За алгебру подій виберемо

$$U = \{A \subset R^1 :\ A = \cup_{i=0}^{n} [a_i, b_i),\ -\infty \leq a_i < b_i \leq \infty,\ 0 \leq n < \infty.\}$$

На множинах з U визначимо функцію $Q(\cdot)$ наступним чином: якщо$A = \cup_{i=0}^{n} [a_i, b_i)$, де інтервали $[a_i, b_i)$ є несумісні, то

$$Q(A) = \sum_{i=0}^{n} (F(b_i) - F(a_i)). \tag{2}$$

З монотонності $F(x)$ випливає, що $\mathbf{P}(A) \geq 0$, а з властивості 2), що $\mathbf{P}(\Omega) = 1$. Позаяк кожна множина з U може бути записана у вигляді суми несумісних інтервалів, то визначення (2) є коректним. Також легко зрозуміти, якщо множини $A_1, ..., A_n,\ n < \infty$ належать до U і є несумісні, то

$$Q(\cup_{i=1}^n A_i) = \sum_{i=0}^n Q(A_i).$$

Отже, функція Q є скінченно адитивна. Доведемо, що якщо послідовність $A_n \in U$ $n \geq 1$ є спадною і такою, що $\lim_{n\to\infty} A_n = \emptyset$, то $\lim_{n\to\infty} Q(A_n) = 0$. Достатньо це зробити для множин A_n вигляду $A_n = [a_n, b_n)$ з $a_n \leq a_{n+1}, b_{n+1} \leq b_n,\ n \geq 1$. Отже, нехай $\lim_{n\to\infty}[a_n, b_n) = \emptyset$. Тоді існує таке n_0, що $b_n = b \geq a_n$ для $n \geq n_0$ (завдання 4), отже, $\lim_{n\to\infty} a_n = b$. На підставі властивості 3) будемо мати:

$$\lim_{n\to\infty} Q([a_n, b_n)) = \lim_{n\to\infty}(F(b_n) - F(a_n)) = F(b) - \lim_{n\to\infty} F(a_n) = 0.$$

Тепер з теореми 3 (стор. 37) випливає, що функція множин Q є зліченно адитивною на U, а з зауваження 4 (стор.38) випливає, що ця функція може бути однозначно продовжена на σ-алгебру $\mathcal{B}(R^1)$, побудовану на U, тобто σ- алгебру борелівських множин $\mathcal{B}(R^1) = \sigma(U)$. Таким чином, на $\mathcal{B}(R^1)$ побудована невід'ємна зліченно адитивна функція множин $\mathbf{P}(\cdot)$ така, що $\mathbf{P}(\Omega) = 1$ і $\mathbf{P}(A) = Q(A)$ для $A \in U$. Доведемо, що $(\Omega, \mathcal{B}(R^1), \mathbf{P})$ і є потрібний нам ймовірносний простір. Визначимо випадкову величину наступним чином: $\xi(z) = z,\ z \in \Omega = R^1$. Маємо

$$\mathbf{P}(\{z : \xi(z) < x\}) = \mathbf{P}([-\infty, x)) = Q([-\infty, x)) = F(x) - F(-\infty) = F(x).$$

Отже, випадкова величина ξ, визначена на ймовірносному просторі $(\Omega, \mathcal{B}(R), \mathbf{P})$, має функцію розподілу $F(x)$.

Визначення 3. Будемо казати, що функція розподілу $F(x)$ випадкової величини ξ має в точці x_0 *атом* величини $p > 0$, якщо функція $F(x)$ має в цій точці стрибок величини p, тобто

$$p = F(x_0 + 0) - F(x_0) > 0,$$

де $F(x_0 + 0) = \lim_{x\downarrow x_0} F(x)$.

Часом також кажуть, що випадкова величина ξ має атом в точці x_0. Обгрунтовує це наступне твердження.

Теорема 3. *Функція розподілу $F(x)$ випадкової величини ξ має в точці x_0 атом тоді і тільки тоді, якщо*

$$\mathbf{P}(\xi = x_0) > 0. \tag{3}$$

Д о в е д е н н я. Для довільного $\varepsilon > 0$ маємо (властивість 4 теореми 1)

$$\mathbf{P}(x_0 \leq \xi < x_0 + \varepsilon) = F(x_0 + \varepsilon) - F(x_0)$$

і якщо $\varepsilon \to 0$, то з цього співвідношення випливає

$$\mathbf{P}(\xi = x_0) = F(x_0 + 0) - F(x_0). \tag{4}$$

Тепер, якщо функція $F(x)$ має стрибок в точці x_0, то права частина в (4) є більшою від нуля, а тоді і ліва частина є більшою від нуля і навпаки. ◀

5.2 Типи функцій розподілу

Функції розподілу дискретного типу

Визначення 4. Будемо казати, що випадкова величина ξ є *дискретного* типу, якщо для певної скінченної, або зліченної множини $\{x_1, x_2, \ldots, x_n, ..\}$ має місце:

а) $p_i = \mathbf{P}(\xi = x_i) > 0, \quad i = 1, 2, ..,$

б) $\sum_{i=1}^{\infty} p_i = 1.$

З цього визначення випливає, що випадкова величина ξ приймає значення тільки з множини $\{x_1, x_2, \ldots, x_n, ..\}$.

Зауваження 2. Сукупність чисел $\{p_i, i = 1, 2, ...\}$ з попереднього визначення називається *розподілом* випадкової величини ξ.

Визначення 5. Функцію розподілу випадкової величини ξ дискретного типу будемо називати *функцією розподілу дискретного* (або *сходинкового*) *типу.*

Легко зрозуміти, що функція розподілу випадкової величини ξ може бути записана у вигляді

$$F(x) = \mathbf{P}(\xi < x) = \sum_{x_i < x} \mathbf{P}(\xi = x_i) = \sum_{i: x_i < x} p_i,$$

де $\sum_{x_i < x}$ ($\sum_{i: x_i < x}$) означає, що сумування відбувається по всім x_i меньшим від x (всім i, для яких x_i меньше від x). Якщо $x_k < x \leq x_{k+1}$, тоді будемо мати:

$$F(x_{k+1}) - F(x) = \sum_{x_i < x_{k+1}} \mathbf{P}(\xi = x_i) - \sum_{x_i < x} \mathbf{P}(\xi = x_i) = \sum_{x \leq x_i < x_{k+1}} \mathbf{P}(\xi = x_i) = 0,$$

а, отже, функція розподілу дискретного типу є сталою між атомами.

Функції розподілу неперервного типу.

Визначення 6. Будемо казати, що функція розподілу $F(x)$ випадкової величини ξ є *неперервного типу*, якщо $F(x)$ є неперервною функцією змінної x.

З теореми 3 маємо

Наслідок 1. *Функція розподілу $F(x)$ випадкової величини ξ є неперервного типу тоді і тільки тоді, якщо* $\mathbf{P}(\xi = x) = 0$ *для всіх* $x \in R^1$.

Теорема 4. (Жордана) *Кожна функція розподілу може бути однозначно представлена у вигляді:*

$$F(x) = pF_d(x) + (1-p)F_c(x), \tag{5}$$

де $0 \le p \le 1$, *і* $F_d(x)$, $F_c(x)$ *є відповідно функція розподілу дискретного і неперервного типу.*

Д о в е д е н н я. Доведемо спочатку однозначність цього представлення. Припустимо, що існує ще одне представлення

$$F(x) = qG_d(x) + (1-q)G_c(x). \tag{6}$$

Тоді, для всіх x з (5), (6) маємо

$$pF_d(x) - qG_d(x) = (1-q)G_c(x) - (1-p)F_c(x).$$

З цієї рівності випливає, що функція $(1-q)G_c(x) - (1-p)F_c(x)$ має дискретний тип, що (оскільки є неперервною) можливе лише тоді, коли

$$(1-q)G_c(x) - (1-p)F_c(x) \equiv c = const.$$

Звідки при $x \to -\infty$ випливає $c = 0$, а при $x \to \infty$ маємо $1-p = 1-q$, отже, $p = q$ і $G_c(x) = F_c(x)$. Але ж тоді і $G_d(x) = F_d(x)$.

Нехай функція $F(x)$ буде довільною функцією розподілу. Якщо функція $F(x)$ є неперервною, тоді твердження є правдивим і в цьому випадку $p = 0$. Припустимо тепер, що функція $F(x)$ не є неперервною. Позаяк $F(x)$ є зростаючою, то вона може мати не більш ніж злічене число стрибків. Позначимо всі точки стрибків через $x_1, x_2, \ldots$, а через $p_1, p_2, ..$ відповідно величини цих стрибків, і нехай

$$p = \sum_{n=1}^{\infty} p_n.$$

Можна вважати, що $0 < p < 1$, позаяк в протилежному випадку функція $F(x)$ є або неперервна ($p = 0$), або дискретна ($p = 1$).

Легко зрозуміти, що функція

$$F_d(x) = \frac{1}{p} \sum_{k: x_k < x} p_k$$

є функцією розподілу дискретного типу. Покладемо

$$F_c(x) \overset{def}{=} \frac{F(x) - pF_d(x)}{1-p} = \frac{F(x) - \sum_{x_k < x} p_k}{1-p} \tag{7}$$

і доведемо, що таким чином визначена функція є функцією розподілу неперервного типу. Дійсно,

$$F_c(-\infty) = 0, \qquad F_c(\infty) = \frac{1-p}{1-p} = 1.$$

Нехай тепер $x < y$, а x_k, $k \in B$ будуть стрибками функції $F(x)$, які містяться в інтервалі $[x, y)$. Тоді

$$F_c(y) - F_c(x) = \frac{F(y) - F(x) - p(F_d(y) - F_d(x))}{1-p} = \frac{\mathbf{P}(x \le \xi < y) - \sum_{k \in B} \mathbf{P}(\xi = x_k)}{1-p} =$$
$$= \frac{\mathbf{P}(x \le \xi < y, \xi \ne x_k, k \in B)}{1-p} \ge 0.$$

Тому

$$F_c(x) \le F_c(y).$$

Функція $\sum_{x_k<x} p_k$ представляє стрибки функції $F(x)$, а отже, функція $F(x) - \sum_{x_k<x} p_k$ не має стрибків, тобто, є неперервною і згідно теореми 2 (стор. 47) є функцією розподілу. Позаяк, як це випливає з (7),

$$F(x) = pF_d(x) + (1-p)F_c(x),$$

то теорема доведена. ◄

Функція розподілу абсолютно непервного типу.

Визначення 7. Будемо казати, що функція розподілу $F(x)$ є *абсолютно неперервного типу*, якщо існує функція $f(x) \ge 0$ така, що

$$F(x) = \int_{-\infty}^{x} f(y)dy, \qquad x \in (-\infty, \infty). \tag{8}$$

Визначення 8. Функція $f(x) \ge 0$ в представленні (8) називається *щільністю* функції розподілу $F(x)$.

Позаяк $F(\infty) = 1$, то з (8) випливає, що для щільності завше виконується умова $\int_{-\infty}^{\infty} f(y)dy = 1$. Ця умова, а також невід'ємність функції $f(x)$ гарантують те, що вона є щільністю певної функції розподілу. Це випливає з наступного твердження.

Теорема 5. *Для того, щоб інтегрована на* $(-\infty, \infty)$ *функція* $f(x)$ *була щільністю певної функції розподілу необхідно і достатньо:*

а) $f(x) \ge 0$ *для всіх* $x \in R^1$,

б) $\int\limits_{-\infty}^{\infty} f(x)dx = 1.$

Д о в е д е н н я. Необхідність умов а), б) випливає з визначення, а для достатності вистачить довести, що функція

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

є функцією розподілу. Маємо

$$F(-\infty) = 0, \qquad F(\infty) = \int_{-\infty}^{\infty} f(x)dx = 1.$$

Якщо $x_1 < x_2$, то

$$F(x_1) = \int\limits_{-\infty}^{x_1} f(y)dy \leq \int\limits_{-\infty}^{x_1} f(y)dy + \int\limits_{x_1}^{x_2} f(y)dy = \int\limits_{-\infty}^{x_2} f(y)dy = F(x_2).$$

Вочевидь, функція $F(x)$ є неперервною, отже, згідно теореми 2 (стор. 47) є функцією розподілу. ◄

Якщо $f(x)$ є щільністю функції розподілу $F(x)$ випадкової величини ξ, тоді для $a < b$ маємо

$$\mathbf{P}(a \leq \xi \leq b) = \mathbf{P}(\xi \leq b) - \mathbf{P}(\xi < a) = \int\limits_{-\infty}^{b} f(x)dx - \int\limits_{-\infty}^{a} f(x)dx = \int\limits_{a}^{b} f(x)dx. \quad (9)$$

Якщо щільність є неперервною в точці x, то з аналізу відомо, що права частина в (8) є диференційованою в цій точці і

$$\frac{dF(x)}{dx} = f(x). \quad (10)$$

Зауваження 3. Якщо щільність не є непрерервною в точці x, то, взагалі кажучи, формула (10) не має місця, але можна довести (див., наприклад, [28], розділ 4, §3), що для довільної щільності $f(x)$ рівність (10) виконується майже скрізь, тобто, для всіх x за вийнятком, може, множини точок, яка має міру Лебега нуль.

ЗАВДАННЯ

1. Довести, що умови

   а) $\{\omega : \xi(\omega) < x\} \in \mathfrak{B}$ для кожного $x \in R^1$,

   б) $\{\omega : \xi(\omega) \leq x\} \in \mathfrak{B}$ для кожного $x \in R^1$,

   в) $\{\omega : \xi(\omega) > x\} \in \mathfrak{B}$ для кожного $x \in R^1$,

   г) $\{\omega : \xi(\omega) \geq x\} \in \mathfrak{B}$ для кожного $x \in R^1$

   є рівнозначними.

2. Навести приклад ймовірнісного простору $(\Omega, \mathfrak{B}, \mathbf{P})$ та відображення $\xi : \Omega \to R^1$, яке не буде випадковою величиною в сенсі означення 1 .

3. Довести, що $\lim_{x\to\infty} F(x) = 1$ для кожної функції розподілу $F(x)$.

4. Нехай $A_n = [a_n, b_n)$ є такими, що $a_n \leq a_{n+1}, b_{n+1} \leq b_n,\ n \geq 1$ і $\lim_{n\to\infty}[a_n, b_n) = \emptyset$. Довести, що існує таке n_0, що $b_n = b$ для $n \geq n_0$ і $a_n \uparrow b$.

5. Випадкова величина ξ є такою, що:

x	-4	-3	-2	-1	1	2	3	4
$\mathbf{P}(\xi = x)$	1/16	1/16	1/8	1/4	1/4	1/8	1/16	1/16

   а) Знайти функцію розподілу ξ і намалювати її графік.

   б) Знайти функцію розподілу випадкової величини $\eta = \xi^2 - 1$.

6. Які з поданих нижче функцій є функціями розподілу;

   а) $F(x) = \frac{3}{4} + \frac{1}{2\pi}\operatorname{arctg}\ x,$ б) $F(x) = \begin{cases} 0, & x \leq 0; \\ \frac{[x]}{2}, & 0 < x \leq 2; \\ 1, & x > 2; \end{cases}$

   в) $F(x) = \begin{cases} 0, & x \leq 0; \\ \frac{x}{1+x}, & x > 0; \end{cases}$ г) $F(x) = \begin{cases} 0, & x \leq 0; \\ 1 - \frac{1-e^{-x}}{x}, & x > 0; \end{cases}$

   д) $F(x) = e^{-e^{-x}}$.

   Знайти відповідні щільності (у випадках, коли вони існують).

7. Довести, що якщо $F(x)$ є функцією розподілу, то для довільного $h > 0$ функції

   а) $G(x) = \frac{1}{h}\int\limits_{x}^{x+h} F(v)dv, \quad Q(x) = \frac{1}{2h}\int\limits_{x-h}^{x+h} F(v)dv$

   є функціями розподілу.

   б) Якщо $F(0) = 0$, то функція $G(x) = F(x) - F(x^{-1}),\ x \geq 1$ і $G(x) = 0,\ x < 1$ також є функцією розподілу.

8. Розподіл випадкової величини ξ називається *симетричним*, якщо функції щільностей випадкових величин ξ і $-\xi$ є однаковими. Довести, що розподіл випадкової величини ξ з функцією розподілу $F(x)$ є симетричним тоді і тільки тоді, коли $F(-x) = 1 - F(x+0)$ для довільного x.

9. Функція розподілу $F(x)$ є визначеною наступним чином

$$F(x) = \begin{cases} \frac{1}{8\pi}\operatorname{arctg} x + \frac{1}{16}, & x \leq 0, \\ \frac{1}{8}x + \frac{1}{4}, & 0 < x \leq 1, \\ \frac{2}{\pi}\operatorname{arctg} x, & x > 1. \end{cases}$$

   Знайти атоми і записати розклад Жордана для цієї функції.

10. Довести, що функція

$$F(x) = \begin{cases} \frac{1}{4}e^{x}, & x \leq 0, \\ \frac{1}{2}(1 + \frac{x^2}{2}), & 0 < x \leq 1, \\ 1 - \frac{1}{8}e^{1-x}, & x > 1 \end{cases}$$

   є функцією розподілу. Знайти її атоми і розклад Жордана.

11. Випадкова величина ξ приймає натуральні k з ймовірностями $\mathbf{P}(\xi = k) = \frac{c}{3^k}$, де c є сталою величиною. Визначити c, а також підрахувати $\mathbf{P}(\xi \geq 4)$.

12. Довести, що функція, визначена наступним чином:

$$f(x) = \begin{cases} 0, & x < 0, \\ 3e^{-3x}, & x \geq 0 \end{cases}$$

   є щільністю розподілу. Знайти її функцію розподілу $F(x)$. Підрахувати $\mathbf{P}(1 < \xi < 3)$.

13. Для якого значення C функція

$$f(x) = \frac{C}{e^x + e^{-x}}$$

є щільністю розподілу випадкової величини. Знайти функцію розподілу цієї випадкової величини.

14. Випадковаі величини мають щільність

а) $f(x) = \frac{1}{\pi(1+x^2)}$;

б) $f(x) = \frac{1}{2}|x|e^{-|x|}$.

Знайти функції розподілів цих випадкових величини.

Лекція 6. Приклади розподілів та незалежність випадкових величин

В цій лекції ми розглянемо деякі приклади розподілів випадкових величин, які часто зустрічаються в застосуваннях, а також розглянемо деякі інші поняття пов'язані з випадковими величинами.

6.1 Приклади розподілів випадкових величин

Будемо казати, що випадкова величина ξ має:

- *розподіл нуля-одиниці,* якщо $\mathbf{P}(\xi = 1) = p \geq 0$, $\mathbf{P}(\xi = 0) = q \geq 0$, $p + q = 1$.
- *геометричний розподіл*, якщо

$$\mathbf{P}(\xi = n) = q^n p = (1-p)^n p \qquad n \geq 0,\ 0 \leq p \leq 1. \tag{1}$$

Цей розподіл має цікаву властивість. Підрахуємо $\mathbf{P}(\xi \geq n + k/\xi \geq k)$, $n, k \geq 0$, де випадкова величина ξ має розподіл (1). Згідно з визначенням умовної ймовірності:

$$\mathbf{P}(\xi = n + k/\xi \geq k) = \frac{\mathbf{P}(\xi = n + k)}{\mathbf{P}(\xi \geq k)} = \frac{pq^{n+k}}{p\sum\limits_{i=k}^{\infty} q^i} = pq^n.$$

А, отже,

$$\mathbf{P}(\xi = n + k/\xi \geq k) = \mathbf{P}(\xi = n). \tag{2}$$

Таким чином, розподіл залишається геометричним. Властивість (2) називається *властивістю відсутності пам'яті.* Можна довести, що серед дискретних розподілів тільки геометричний має таку властивість. Нижче ми розглянемо показниковий розподіл, для якого має місце подібна властивість. Там ми подамо інтерпретацію цієї властивості;

- *розподіл Бернуллі* або *біноміальний*, якщо

$$\mathbf{P}(\xi = k) = C_n^k p^k (1-p)^{n-k}, \qquad 0 \le p \le 1,\ 0 \le k \le n.$$

- *розподіл Пуассона з параметром* λ, якщо

$$\mathbf{P}(\xi = n) = \frac{\lambda^n}{n!} e^{-\lambda},\ n = 0, 1, 2...,\ \ 0 < \lambda < \infty.$$

- *рівномірний розподіл*, якщо його щільність має вигляд:

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b, \\ 0, & x \notin [a, b]. \end{cases}$$

  В цих виразах a і b є довільними сталими, при чому $a < b$.

- *показниковий* (*експоненційний*) *розподіл* з параметром $\lambda > 0$, якщо його щільність є функцією вигляду:

$$f(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x \ge 0. \end{cases}$$

  Інтегруючи щільність, отримаємо функцію розподілу: $F(x) = 1 - e^{-\lambda x},\ x \ge 0$ і $F(x) = 0,\ x < 0$. Численні застосування показникового розподілу пов'язані з його властивістю, яка є аналогічна властивості геометричного розподілу. Якщо випадкова величина ξ має показниковий розподіл, тоді для всіх $x, y \ge 0$ виконується рівність:

$$\mathbf{P}(\xi > x + y / \xi > x) = \mathbf{P}(\xi > y). \tag{3}$$

  Доведення цієї властивості можна провести так само, як це ми робили для геометричного розподілу.

  Властивість (3) також має назву *властивість відсутності пам'яті* і може бути інтерпретована наступним чином: якщо випадкова величина ξ з показниковим розподілом є часом безаварійної роботи певного елементу, тоді незалежно від того, скільки він пропрацював до певного моменту часу, подальший інтервал часу роботи не залежить від "минулого" і має такий самий розподіл як увесь час роботи елементу від самого початку. Можна довести, що серед неперервних розподілів таку властивість має лише показниковий розподіл.

- *гамма розподіл з параметрами* $\lambda > 0$, $p > 0$, якщо його щільність є функцією вигляду

$$f(x,p,\lambda) = \begin{cases} 0, & x < 0, \\ \frac{\lambda^p x^{p-1}}{\Gamma(p)} e^{-\lambda x}, & x \geq 0, \end{cases}$$

де через $\Gamma(p)$ позначена гамма функція, тобто функція

$$\Gamma(p) = \int\limits_0^\infty x^{p-1} e^{-x} dx, \qquad p > 0.$$

Розподіли такого типу часто використовуються в таких прикладних дисциплінах як теорія обслуговування, теорія надійності і математична статистика. Наведемо ці розподіли.

Перш за все, можна зауважити, що для $p = 1$, це буде відомий нам вже показниковий розподіл з параметром λ.

  - Нехай $p = n$, де n натуральне число. Тоді

  $$f(x,n,\lambda) = \begin{cases} 0, & x < 0, \\ \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, & x \geq 0. \end{cases}$$

  Розподіл з такою щільністю має спеціальну назву: *розподіл Ерланга з параметрами* (n, λ) (або *розподіл Ерланга порядку* n *з параметром* λ, або просто *розподіл Ерланга*). Цей розподіл часто використовується в теорії ослуговуванні, а також при моделюванні комп'ютерних мереж та мереж звя'зку.

  - Нехай $p = n/2, \lambda = 1/2$. Тоді

  $$f(x,n/2,1/2) = \begin{cases} 0, & x < 0, \\ \frac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)} e^{-\frac{x}{2}}, & x \geq 0. \end{cases}$$

  Розподіл з такою щільністю також має спеціальну назву: *розподіл chi-квадрат з* n *ступенями свободи,* який за звичай позначають як $\chi_n^2(x)$. Цей розподіл часто зустрічається в задачах математичної статистики.

- *нормальний розподіл* або *гауссовський* з параметрами $m \in R^1$, $\sigma^2 > 0$, якщо щільність є функцією вигляду:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \qquad -\infty < x < \infty. \tag{4}$$

Нормальний розподіл з параметрами m, σ^2 зазвичай позначають символом $N(m, \sigma^2)$ і запис $\xi \in N(m, \sigma^2)$ означає, що щільність випадкової величини ξ має вигляд (2). В окремому випадку, коли $m = 0$, $\sigma^2 = 1$ кажуть про стандартний нормальний розподіл (*стандартний нормальний розподіл*), з яким ми будемо часто зустрічатися в подальшому. Щільність цього розподілу зазвичай позначають літерою φ, а функцію розподілу – Φ, тобто,

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{y^2}{2}} dy, \qquad -\infty < x < \infty.$$

- *розподіл Коші* з параметрами $m \in R^1$, $\lambda > 0$, якщо його щільність є функцією вигляду:

$$f(x) = \frac{\lambda}{\pi(\lambda^2 + (x-m)^2)}, \qquad -\infty < x < \infty.$$

6.2 Незалежні випадкові величини. Функції від випадкових величин.

Нехай на ймовірносному просторі $(\Omega, \mathfrak{B}, \mathbf{P})$ задані випадкові величини ξ, η.

Визначення 1. Випадкові величини ξ, η будемо називати *незалежними*, якщо для всіх $x, y \in R^1$ події $\{\omega : \xi(\omega) < x\}$, $\{\omega : \eta(\omega) < y\}$ є незалежними. Іншими словами, якщо має місце рівність:

$$\mathbf{P}(\xi < x, \eta < y) = \mathbf{P}(\xi < x)\mathbf{P}(\eta < y)$$

для довільних $x, y \in R^1$, тоді випадкові величини ξ, η називають незалежними.

Якщо випадкові величини ξ, η є дискретними і $\xi \in \{x_1, x_2, ...\}$, $\eta \in \{y_1, y_2, ...\}$, то ці випадкові величини називають незалежними, якщо

$$\mathbf{P}(\xi = x_i, \eta = y_j) = \mathbf{P}(\xi = x_i)\mathbf{P}(\eta = y_j)$$

для довільних x_i, y_j.

Нехай числова функція $g(x)$ є визначеною на R^1, а ξ - випадкова величина, задана на ймовірносному просторі $(\Omega, \mathfrak{B}, \mathbf{P})$. Чи буде $g(\xi)$ випадковою величиною? Взагалі кажучи, без додаткових припущень стосовно $g(\cdot)$ відповідь на це питання є негативною. Наведене нижче твердження відповідає на сформульоване питання, але спочатку одне необхідне визначення.

Визначення 2. Нехай $\mathcal{B}(R^1)$ є борелівська σ- алгебра множин з R^1. Функція $g : R^1 \to R^1$ називається *борелівською*, якщо $\{x : g(x) \in B\} \in \mathcal{B}(R^1)$ для довільного $B \in \mathcal{B}(R^1)$.

Можна довести, що умова $\{x : g(x) \in B\} \in \mathcal{B}(R^1)$ для довільного $B \in \mathcal{B}(R^1)$ є рівнозначною наступній: $\{x : g(x) < z\} \in \mathcal{B}(R^1)$ для довільного $z \in R^1$.

Теорема 1. *Нехай ξ є випадковою величиною на ймовірносному просторі $(\Omega, \mathfrak{B}, \mathbf{P})$. Тоді для довільної борелівської функції $g : R^1 \to R^1$ значення $g(\xi)$ буде випадковою величиною на тому ж ймовірносному просторі.*

Доведення легко виникає з того, що множина $B_x = \{z : g(z) < x\} \in \mathcal{B}(R^1)$ а отже $\{\omega : g(\xi(\omega)) < x\} = \{\omega : \xi(\omega) \in B_x\} \in \mathfrak{B}$, де останнє включення випливає з того, що $\xi(\omega)$ є випадковою величиною.

Клас борелівських функцій містить функції, які принаймні є одностороннє неперервними, а також границі таких функцій.

Наступна проблема: нехай ξ та η є випадковими величинами, заданими на ймовірносному просторі $(\Omega, \mathfrak{B}, \mathbf{P})$. Чи буде, наприклад, $\xi + \eta$ випадковою величиною?

Наступне твердження описує деякі ситуації, коли дії над випадковими величинами дають в результаті випадкові величини.

Теорема 2. *Нехай ξ, η, ξ_n, $n = 1, 2, ...$ є випадковими величинами, визначеними на ймовірносному просторі $(\Omega, \mathfrak{B}, \mathbf{P})$. Тоді*

1) *$\xi \pm \eta$, $\xi\eta$ і ξ/η за умови, що $\eta \neq 0$, є випадковими величинами;*

2) *$\sup_n \xi_n$, $\inf_n \xi_n$ є випадковими величинами;*

3) *якщо існують наступні границі $\lim_{n\to\infty} \xi_n$, $\overline{\lim}_{n\to\infty}\xi_n$, $\underline{\lim}_{n\to\infty}\xi_n$, то вони є випадковими величинами;*

Д о в е д е н н я. Нехай Q позначає множину раціональних чисел в R^1. Для довільно фіксованого x маємо

$$\{\omega : \xi(\omega) + \eta(\omega) < x\} = \cap_{r\in Q}\{\omega : \xi(\omega) < r < x - \eta(\omega)\} = \cap_{r\in Q}\{\omega : \xi(\omega) < r\}\cap$$
$$\cap \{\omega : r < x - \eta(\omega)\} = \cap_{r\in Q}\{\omega : \xi(\omega) < r\} \cap \{\omega : \eta(\omega) < x - r\} \in \mathfrak{B}$$

Аналогічно розглядаються випадки $\xi - \eta$ та ξ/η.

Для доведення другого та третього пункт потрібно скористуватися з наступних співвідношень (завдання 3)

$$\{\omega : \sup_n \xi_n < x\} = \Omega \backslash \cup_{n=1}^{\infty} \{\omega : \xi_n(\omega) \geq x\} \in \mathfrak{B}; \quad (5)$$

$$\{\omega : \inf_n \xi_n < x\} = \cup_{n=1}^{\infty} \{\omega : \xi_n(\omega) < x\} \in \mathfrak{B}; \quad (6)$$

$$\{\omega : \overline{\lim}_{n\to\infty} \xi_n < x\} = \cup_{k=1}^{\infty} \cup_{n=1}^{\infty} \cap_{j=n}^{\infty} \{\omega : \xi_j(\omega) < x - 1/k\} \in \mathfrak{B}; \quad (7)$$

$$\{\omega : \underline{\lim}_{n\to\infty} \xi_n < x\} = \cup_{k=1}^{\infty} \cup_{n=1}^{\infty} \cap_{j=n}^{\infty} \{\omega : \xi_j(\omega) > x + 1/k\} \in \mathfrak{B}. \quad (8)$$

Дведення того, що $\lim_{n\to\infty} \xi_n$ є випадковою величиною можна знайти в [?].

Легко зрозуміти, що якщо випадкові величини ξ, η є незалежними, а функції $f(x),\ g(x)$ - борелівськими, то випадкові величини $f(\xi),\ g(\eta)$ є також незалежними. Навпаки, взагалі кажучи, так не є (див. завдання 6), але якщо існують обернені $f^{-1}(x),\ g^{-1}(x)$, то з незалежності $f(\xi),\ g(\eta)$ випливає незалежність $\xi,\ \eta$.

6.2*. σ- алгебри, породжені випадковими величинами

Нехай ξ буде випадковою величиною визначеною на деякому ймовірнісному просторі $(\Omega, \mathfrak{B}, \mathbf{P}.)$ Означимо клас подій $\mathfrak{B}_\xi = \{\xi(\omega) \in B\}$, де B- борелівські множини з R^1. Легко довести (завдання 10), що $\mathfrak{B}_\xi$, що є σ-алгеброю і ця алгебра називається σ-алгеброю *продженою випадковою величиною* ξ. Головним результатом цього підрозділу є наступне твердження

Теорема 3. *Випадкові величини $\xi,\ \eta$ є незалежними тоді і лише тоді, коли є незалежними σ -алгебри $\mathfrak{B}_\xi$ та $\mathfrak{B}_\eta$.*

Цю лекцію завершимо означенням інтеграла Стілтьєса (Stieltijes), яке нам буде потрібне в наступній лекції.

6.3 Інтеграл Стілтьєса

Згадаймо визначення інтеграла Рімана (Rieman).

Нехай функція $f(x)$ є обмеженою на скінченному інтервалі $[a, b]$. Для фіксованого натурального n розглянемо розбиття інтервалу $[a, b]$ наступного вигляду $a = a_0(n) < a_1(n) < ... < a_n(n) < a_{n+1}(n) = b$. Будемо позначати таке розбиття як $P_n = \{a_0(n), a_1(n), ..., a_n(n), a_{n+1}(n)\}$. Якщо розглядаємо розбиття P_{n+1}, то завше припускаємо, що $P_n \subset P_{n+1}$, тобто точки розбиття P_n є точками розбиття P_{n+1}. Для $i = 0, ..., n$ позначимо

$$m_i^-(n) = \inf\{f(x) : a_i(n) \leq x < a_{i+1}(n))\}, \quad m_i^+(n) = \sup\{f(x) : a_i(n) \leq x < a_{i+1}(n))\},$$

і нехай $S_n^{\pm}(f) = \sum_{i=0}^n m_i^{\pm} \Delta_i(n)$, $\Delta_i(n) = a_{i+1} - a_i$. Легко довести, що якщо $P_n \subset P_{n+1}$, то $S_{n+1}^+(f) \leq S_n^+(f)$ і $S_n^-(f) \leq S_{n+1}^-(f)$, отже, існують границі $\lim_{\Delta(n)\to 0} S_n^{\pm}(f)$, де $\Delta(n) = \sup_i \Delta_i(n)$. Можна довести, що ці границі не залежать від розбиття $P_i(n)$.

Визначення 3. Якщо

$$\lim_{\Delta(n)\to 0} S_n^+(f) = \lim_{\Delta(n)\to 0} S_n^-(f) \in (-\infty, \infty),$$

тоді ця спільна границя називається *інтегралом Рімана* функції $f(x)$ на інтервалі $[a, b]$ і позначається символом $\int\limits_a^b f(x)dx$.

Інтеграл Рімана на $(-\infty, \infty)$ визначаємо як

$$\int\limits_{-\infty}^{\infty} f(x)dx = \lim_{\substack{b\,\to\,\infty,\\ a\,\to\,-\infty}} \int\limits_a^b f(x)dx,$$

за умови, що остання границя існує.

Можна довести, що якщо функція $f(x)$ є неперервною на скінченному інтервалі $[a, b]$, то

$$\int\limits_a^b f(x)dx = \lim_{\Delta(n)\to 0} \sum_{i=0}^{n} f(x_i(n))\Delta_i(n),$$

де $x_i(n)$ є довільною точкою з інтервалу $[a_i(n), a_{i+1}(n)]$.

Визначення інтегралу Стілтьєса подамо тільки для ситуації, яка розглядається у даному підручнику. Загальний випадок читач може знайти в [28] (розділ 6, § 6).

Нехай на проміжку $[a, b]$, $-\infty < a < b < \infty$ є заданими дві функції: функція розподілу $F(x)$ і функція $g(x)$, яка є неперервною на $[a, b]$. Для розбиття P_n інтервалу $[a, b]$, який ми розглядали вище, утворимо суму

$$S_n(g) = \sum_{i=0}^{n} g(\tilde{x}_i(n))(F(a_{i+1}(n)) - F(a_i(n)),$$

де $\tilde{x}_i(n)$ є довільним числом з $[a_{i+1}(n), a_i(n)]$.

Визначення 4. Нехай $\Delta(n) = \max_{0\le i\le n}(a_{i+1}(n) - a_i(n))$. Якщо існує

$$S(g) = \lim_{\Delta(n)\to 0} S_n(g) = \lim_{\Delta(n)\to 0} \sum_{i=0}^{n} g(\tilde{x}_i(n))(F(a_{i+1}(n)) - F(a_i(n)),$$

то цю границю будемо називати *інтегралом Стілтьєса* і позначати символом

$$S(g) = \int\limits_a^b g(x)dF(x).$$

Нехай тепер на вісі $[-\infty, \infty)$ є заданими дві функції: функція розподілу $F(x)$ і неперервна функція $g(x)$.

Позаяк інтеграл $\int_a^b |g(x)|dF(x)$ монотонним чином залежить від a, b при $a \to -\infty$, $b \to \infty$, то границя $\lim\limits_{\substack{b\to\infty\\ a\to-\infty}} \int_a^b |g(x)|dF(x)$ завше існує, але може дорівнювати

$+\infty$. Якщо ця границя є скінченною, тоді можемо визначити інтеграл Стілтьєса на необмеженому інтервалі $(-\infty, \infty)$ звичним чином:

$$\int_{-\infty}^{\infty} g(x)dF(x) \stackrel{def}{=} \lim_{\substack{b\to\infty \\ a\to-\infty}} \int_a^b g(x)dF(x).$$

Отже, приходимо до наступного визначення:

Визначення 5. Будемо казати, що функція $g(x)$ є *інтегрована в сенсі Стілтьєса* стосовно неспадної функції $F(x)$, якщо

$$\int_a^b |g(x)|dF(x) < \infty. \tag{9}$$

Зауваження 1. Якщо функція $F(x)$ є диференційованою і $F'(x) = f(x)$, тоді $F(a_{i+1}(n)) - F(a_i(n)) = f(\tilde{x}_i(n))(a_{i+1}(n) - a_i(n))$, де $a_i(n) < \tilde{x}_i(n) < a_{i+1}(n)$ і, отже,

$$\int_a^b g(x)dF(x) = \lim_{\Delta(n)\to 0} \sum_{i=0}^{n} g(\tilde{x}_i(n))f(\tilde{x}_i(n))(a_{i+1}(n) - a_i(n)) = \int_a^b g(x)f(x)dx.$$

Таким чином, інтеграл Стілтьєса в цьому випадку співпадає з інтегралом Рімана.

Зауваження 2. Треба зауважити, що наведене визначення інтеграла Стілтьєса враховує можливий стрибок функції $F(x)$ в точці a і не враховує його в точці b. Це є істотним, бо можна було б дати таке визначення (і часом так де-які автори роблять), яке враховує можливі стрибки в обидвох точках. В нашій ситуації будемо вважати, точка a належить області інтегрування, а точка b ні.

Наступне твердження, в якому йде мова про властивості інтеграла Стілтьєса, легко випливає з визначення цього інтеграла (при чому межі a, b можуть приймати значення $a = -\infty, b = \infty$.)

Теорема 4. *Припустимо, що інтеграли Стілтьєса всіх зазначених функцій є визначиними. Тоді:*

1) $\int_a^b dF(x) = F(b) - F(a)$;

2) *для довільних* $a < c < b$ *має місце рівність*

$$\int_a^b g(x)dF(x) = \int_a^c g(x)dF(x) + \int_c^b g(x)dF(x);$$

3) *для довільних сталих* C_1, C_2 *має місце рівність*

$$\int_a^b g(x)d(C_1F(x) + C_2G(x)) = C_1 \int_a^b g(x)dF(x) + C_2 \int_a^b g(x)dG(x);$$

4) *якщо* $f(x) \le g(x)$, *то*

$$\int_a^b f(x)dF(x) \le \int_a^b g(x)dF(x);$$

5) *якщо послідовність* $f_n(x)$ *є такою, що* $\lim_{n\to\infty} f_n(x) = f(x)$, *рівномірно стосовно* x, *то*

$$\lim_{n\to\infty} \int_a^b f_n(x)dF(x) = \int_a^b f(x)dF(x);$$

6) *якщо функція* $G(x)$ *є неперервною на проміжку* $[a,b]$ $-\infty < a < b < \infty$ *і існує* $G'(x) = g(x)$, *то*

$$\int_a^b G(x)dF(x) = F(b)G(b) - F(a)G(a) - \int_a^b F(x)g(x)dx;$$

7) *нехай для функції* $F(x)$ *виконуються умови:* $F'(x) = f_1(x)$ *для* $a < x < c$ *і* $F'(x) = f_2(x)$ *для* $c < x < b$. *Тоді*

$$\int_a^b g(x)dF(x) = g(a)(F(a+0) - F(a)) + \int_a^c g(x)f_1(x)dx + g(c)(F(c+0) - F(c)) +$$

$$+ \int_c^b g(x)f_2(x)dx. \qquad (10)$$

Зауваження 3. З (10) випливає, що якщо функція є сходинковою на інтервалі $[a,b]$, тобто, має вигляд:

$$F(x) = \begin{array}{ll} c_0, & x = a, \\ c_i, & x \in [a_i, a_{i+1}),\ i = 0, ..., n, \end{array}$$

де $a = a_0 < a_1 < ... < a_n < a_{n+1} = b$ і $c_0 < c_1 < ... < c_n$, то

$$\int_a^b g(x)dF(x) = \sum_{i=0}^{n} g(a_i)(c_{i+1} - c_i).$$

Зокрема, якщо $F(x)$ є функцією розподілу дискретного типу з атомами в точках x_i і величинами стрибків p_i, $i = 1, 2, ...$, то

$$\int_{-\infty}^{\infty} g(x)dF(x) = \sum_{i=1}^{\infty} g(x_i)p_i \qquad (11)$$

за умови, що ряд в (11) є абсолютно збіжним.

Як і для інтеграла Рімана має місце наступний результат.

Теорема 5. *Якщо функція $f(x)$ є неперервною на скінченному проміжку $[a,b]$, то*

$$\int_a^b f(x)dF(x) = \lim_{\Delta(n)\to 0} \sum_{i=0}^{n} f(\tilde{x}_i(n))(F(a_{i+1}(n)) - F(a_i(n)),$$

де $\tilde{x}_i(n)$ є довільною точкою з інтервала $[a_i(n), a_{i+1}(n)]$.

ЗАВДАННЯ

1. Знайти щільність розподілу випадкової величини, яка дорівнює площі кола, якщо радіус кола є випадковою величиною зі щільністю

$$f(x) = \begin{cases} \frac{1}{2}, & x \in [0,2], \\ 0, & x \in (-\infty, 0) \cup (2, \infty). \end{cases}$$

2. Знайти щільність розподілу випадкової величини, яка дорівнює обєму куба, якщо його діагональ є випадковою величиною зі щільністю

$$f(x) = \begin{cases} 2e^{-2x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

3. Довести співвідношення (5)-(8).

4. Довести, що якщо ξ є випадковою величиною, то $|\xi|$ теж є випадкова величина.

5. Випадкова величина ξ має щільність $f(x)$. Знайти щільності наступних випадкових величин: а) $1/\xi$, б) e^{ξ}, в) $\sin\xi$, г) $F(\xi)$, де $F(x)$ є функція розподілу для ξ.

6. Нехай випадкова величина (ξ, η) має розподіл $p_{i,j} = \mathbf{P}(\xi = i, \eta = j)$, $i, j = -1, 0, 1$ вигляду $p_{1,1} = p_{-1,1} = 1/32, p_{-1,-1} = p_{1,-1} = p_{1,0} = p_{0,1} = 3/32$, $p_{-1,0} = p_{0,-1} = 5/32, p_{0,0} = 8/32$. Довести, що випадкові величини ξ^2, η^2 є незалежними, а ξ, η ні.

7. Довести, що для довільної функції розподілу $F(x)$ мають місце співвідношення:

$$\lim_{x\to\infty} x\int_x^\infty \frac{1}{y}dF(y)=0, \qquad \lim_{x\to+0} x\int_x^\infty \frac{1}{y}dF(y)=0$$

$$\lim_{x\to-\infty} x\int_{-\infty}^x \frac{1}{y}dF(y)=0 \qquad \lim_{x\to 0} x\int_{-\infty}^x \frac{1}{y}dF(y)=0$$

8. Випадкова величина ξ має рівномірний розподіл на відрізку $[0,1]$. Знайти розподіл випадкової величини $1/\xi$.

9. Випадкова величина ξ має щільність:

   а) $\frac{1}{x^2\sqrt{2\pi}}e^{-1/2x^2}$. Знайти щільність випадкової величини $\eta=\xi^{-1}$. (*нормальна стандардова* $N(0,1)$)

   б) $\frac{1}{\pi}\frac{1}{1+x^2}$. Знайти щільність випадкової величини $\eta=\xi^2$.

   в) рівномірну на $[-\pi/2,\pi/2]$. Знайти щільність випадкової величини $\eta=\sin\xi$. $(\frac{1}{\pi\sqrt{1-x^2}})$

10. Довести, що для довільної випадкової величини клас подій $\{\xi(\omega)\in B\}$, де B- борелівські множини з R^1 є σ-алгеброю.

Лекція 7. Числові характеристики випадкових величин.

З розподілом випадкової величини пов'язані певні числові значення, які ще називають *параметрами* або *характеристиками* розподілу, і які відіграють значну роль у теорії ймовірностей та математичній статистиці. Розкриймо по черзі зміст цих понять.

7.1 Математичне сподівання випадкової величини

7.1.1 Випадок дискретної випадкової величини

Нехай ξ є дискретною випадковою величиною і $\mathbf{P}(\xi = x_k) = p_k > 0, x_k \in R^1, k = 1, 2, ...$

Визначення 1. *Математичне сподівання* (*середнє значення*) випадкової величини ξ будемо позначати символом $\mathbf{M}\xi$ і визначати як:

$$\mathbf{M}\xi = \sum_{k=1}^{\infty} x_k p_k \tag{1}$$

за умови, що ряд в (1) є абсолютно збіжним , тобто збігається ряд

$$\sum_{k=1}^{\infty} |x_k| p_k < \infty.$$

Зауваження 1. З цього означення випливає, що для того щоб знайти математичне сподівання випадкової величини потрібно підрахувати суму добутків всіх можливих значень цієї величини з ймовірностями, з якими вона приймає ці значення.

Зауваження 2. Математичне сподівання випадкової величини ξ є аналогом знаного з фізики такого поняття як центр ваги. Так якщо на числовій вісі в точках з координатами $x_1, x_2, \ldots, x_n$ маємо відповідно маси $m_1, m_2, \ldots, m_n$, то відомо, що центр ваги всіх цих n мас буде в точці з координатою

$$x_0 = \frac{x_1 m_1 + x_2 m_2 + \ldots + x_n m_n}{M}, \qquad M = \sum_{i=1}^{n} m_i.$$

Якщо позначити $p_i = m_i/M$, то тоді $p_i > 0$, $i = 1, 2, ..., n$, а $\sum_{i=1}^{n} p_i = 1$, а отже числа p_i можемо інтерпретувати як ймовірності. Якщо означити випадкову величину ξ, яка прибирає значення $x_i, i = 1, ..., n$ з ймовірностями $\mathbf{P}(\xi = x_i) = p_i$, то тоді

$$x_0 = \sum_{i=1}^{n} x_i p_i = \mathbf{M}\xi.$$

Отже, можна сказати, що математичне сподівання є "середньою" точкою розподілу, довкола якої групуються значення випадкової величини.

Теорема 1. *Якщо математичні сподівання випадкових величин існують, тоді мають місце наступні властивості:*

1) *для довільної* $c = const$ *маємо* : $\mathbf{M}c = c$ *і* $\mathbf{M}c\xi = c\mathbf{M}\xi$;

2) *для довільних* $a, b \in R^1$ *має місце рівність*

$$\mathbf{M}(a\xi + b\eta) = a\mathbf{M}\xi + b\mathbf{M}\eta;$$

3) *якщо випадкові величини* ξ, η *є* <u>*незалежними*</u>, *то* $\mathbf{M}(\xi\eta) = \mathbf{M}\xi\mathbf{M}\eta$;

4) *якщо* $\mathbf{P}(\xi \le \eta) = 1$, *то* $\mathbf{M}\xi \le \mathbf{M}\eta$ *і, зокрема, якщо* $\mathbf{P}(\xi \le C) = 1$, *то* $\mathbf{M}\xi \le C$;

5) $|\mathbf{M}\xi| \le \mathbf{M}|\xi|$.

Доведення.

Справедливість властивостей 1), 4), 5) легко випливає з визначення математичного сподівання і властивостей числових рядів. Доведемо властивості 2), 3).

2). Нехай ξ, η є дискретними випадковими величинами і $\xi \in \{x_1, x_2, ...\}$, $\eta \in \{y_1, y_2, ...\}$. Випадкова величина $a\xi + b\eta$ приймає значення $ax_i + by_j$ з ймовірностями p_{ij}. Якщо тепер скористатися зауваженням 1, то будемо мати:

$$\mathbf{M}(a\xi + b\eta) = \sum_{i,j=1}^{\infty} (ax_i + by_j)\mathbf{P}(\xi = x_i, \eta = y_j) = a\sum_{i=1}^{\infty} x_i \sum_{j=1}^{\infty} \mathbf{P}(\xi = x_i, \eta = y_j) +$$

$$+ b\sum_{j=1}^{\infty} y_j \sum_{i=1}^{\infty} \mathbf{P}(\xi = x_i, \eta = y_j) = a\sum_{i=1}^{\infty} x_i\mathbf{P}(\xi = x_i) + b\sum_{j=1}^{\infty} y_j\mathbf{P}(\eta = y_j) = a\mathbf{M}\xi + b\mathbf{M}\eta.$$

3). Тепер $\mathbf{P}(\xi = x_i, \eta = y_j) = \mathbf{P}(\xi = x_i)\mathbf{P}(\eta = y_j)$ і так само, як і раніше, маємо:

$$\mathbf{M}(\xi\eta) = \sum_{i,j=1}^{\infty} x_i y_j \mathbf{P}(\xi = x_i, \eta = y_j) = \sum_{i,j=1}^{\infty} x_i y_j \mathbf{P}(\xi = x_i)\mathbf{P}(\eta = y_j)$$

$$= \sum_{i}^{\infty} x_i \mathbf{P}(\xi = x_i) \sum_{j=1}^{\infty} y_j \mathbf{P}(\eta = y_j) = \mathbf{M}\xi\mathbf{M}\eta.$$

З доведеної теореми випливає:

Наслідок 1. *Якщо існують метематичні сподівання випадкових величин* $\xi_1, \xi_2, .., \xi_n$, $n < \infty$, *то:*

1) *для довільних* $a_i \in R^1$, $i = 1, 2, ...n$ *маємо*

$$\mathbf{M}\sum_{i=1}^{n} a_i \xi_i = \sum_{i=1}^{n} a_i \mathbf{M}\xi_i;$$

2) *якщо, крім того, випадкові величини є незалежними, то*

$$\mathbf{M}\prod_{i=1}^{n} \xi_i = \prod_{i=1}^{n} \mathbf{M}\xi_i.$$

Приклад 1. *Випадкова величина* ξ *має розподіл нуля-одиниці, тобто* $\mathbf{P}(\xi = 0) = p$, $\mathbf{P}(\xi = 1) = q = 1 - p$, $0 \le p \le 1$. *Знайти* $\mathbf{M}\xi$.

Р о з в' я з о к. Маємо

$$\mathbf{M}\xi = 0 \cdot p + 1 \cdot q = q.$$

Приклад 2. *Знайти* $\mathbf{M}\xi$, *якщо випадкова величина* ξ *має біноміальний розподіл з параметрами* n, p.

Р о з в' я з о к. Маємо

$$\mathbf{M}\xi = \sum_{k=0}^{n} k C_n^k p^k q^{n-k} = \sum_{k=1}^{n} \frac{kn!}{k!(n-k)!} p^k q^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} = np \sum_{l=0}^{n-1} \frac{(n-1)!}{l!(n-1-l)!} p^l q^{n-1-l} = np(p+q)^{n-1} = np.$$

7.1.2 Загальний випадок

Нехай випадкова величина ξ є такою, що $a \le \xi < b$, де $-\infty < a < b < \infty$

Для натурального числа n розглянемо деяке розбиття $a = a_0(n) < a_1(n) < ... < a_n(n) < a_{n+1}(n) = b$ інтервалу $[a, b]$, яке позначимо $P_n = \{a_0(n), ..., a_n(n)\}$. Будемо вважажати, що ці розбиття вибрано так, що $P_n \subset H_{n+1}$ і $max_i(a_{i+1}(n) - a_i(n)) \to 0$ якщо $n \to \infty$. Означимо дискретні випадкові величини $\xi_n^{\pm}$ наступним чином

$$\xi_n^+ = a_{i+1}(n), \quad \xi_n^- = a_i(n) \qquad \text{якщо} \qquad a_i(n) \le \xi < a_{i+1}(n). \qquad (2)$$

Для розподілів випадкових величин $\xi_n^{\pm}$

$$\mathbf{P}(\xi_n^+ = a_{i+1}(n)) = \mathbf{P}(\xi_n^- = a_i(n)) = \mathbf{P}(a_i(n) \le \xi < a_{i+1}(n)) =$$
$$= F(a_{i+1}(n)) - F(a_i(n)) \overset{def}{=} p_i(n), \quad i = 0, ..., n.$$

Зрозуміло також,

$$\xi_n^- \le \xi \le \xi_n^+, \qquad \xi_n^- \uparrow \xi,\ \xi_n^+ \downarrow \xi \text{ якщо } n \to \infty. \tag{3}$$

Маємо

$$\lim_{n\to\infty} \mathbf{M}\xi_n^+ = \lim_{n\to\infty} \sum_{i=0}^{n} a_{i+1}(n)\big(F(a_{i+1}(n)) - F(a_i(n))\big) = \int_a^b x dF(x), \tag{4}$$

$$\lim_{n\to\infty} \mathbf{M}\xi_n^- = \lim_{n\to\infty} \sum_{i=0}^{n} a_i(n)\big(F(a_{i+1}(n)) - F(a_i(n))\big) = \int_a^b x dF(x). \tag{5}$$

Цю спільну границю для $\mathbf{M}\xi_n^{\pm}$ логічно прийняти за означення математичного сподівання випадкової величини ξ і отже маємо

$$\mathbf{M}\xi \overset{def}{=} \lim_{n\to\infty} \mathbf{M}\xi_n^{\pm} = \int_a^b x dF(x). \tag{6}$$

Співвідношення (6) вказує як потрібно означити математичного сподівання в загальному випадку.

Визначення 2. *Нехай випадкова величина ξ має функцію розподілу $F(x)$. Математичне сподівання випадкової величини ξ озачаєтиься наступним чином*

$$\mathbf{M}\xi \overset{def}{=} \int_{-\infty}^{\infty} x dF(x), \qquad \text{якщо} \qquad \int_{-\infty}^{\infty} |x| dF(x) < \infty. \tag{7}$$

Зауваження 3. Формула (1) є частковим випадком формули (1), оскільки якщо випадкова величина ξ є дискретного типу, то $F(x) = \sum_{x_i < x} p_i$, а, отже,

$$\int_{-\infty}^{\infty} x dF(x) = \sum_{k=1}^{\infty} x_k p_k,$$

що випливає з визначення інтеграла Стілтьєса.

Зауваження 4. Якщо функція розподілу $F(x)$ має щільність $f(x)$, тоді з зауваження 1 (стор. 63) випливає:

$$\mathbf{M}\xi = \int_{-\infty}^{\infty} xf(x)dx, \qquad \text{якщо} \qquad \int_{-\infty}^{\infty} |x|f(x)dx < \infty. \tag{8}$$

В теоремі 1 було описано властивості математичного сподівання, але лише для випадку дискретної випадкової величини. Зараз ми покажемо, що це твердження залишається в силі і в загальному випадку. Властивості 1), 4), 5) з цього твердження, очевидно, заново доводити непотрібно, оскільки вони випливають з властивостей інтеграла Стельтєса а тому розглянемо лише властивості 2), 3). Їх доведення буде іншим, і щоб його реалізувати нам будуть потрібні деякі допоміжні факти.

Якщо в конструкції (2) взяти $a = 0, 0 < b < \infty$ (тобто випадкова величина є обмеженою і невід'ємною), то з побудови випадкових величин $\xi_n^{\pm}$ випливає наступний результ

Лема 1. *Для довільної обмеженої, невід'ємною̈ї випадкової величини ξ існуює послідовноість дискретних, обмежених і невід'ємних випадкових величин $\xi_n^{\pm}$ таких, що*

$$\xi_n^- \le \xi \le \xi_n^+, \quad \xi_n^- \uparrow \xi, \ \xi_n^+ \downarrow \xi, \qquad n \to \infty.$$

Зауваження 5. Послідовності $\xi^{\pm}$, про яких говориться в попередній лемі, будемо називати *апроксимуючими послідовностями* для випадкової величини ξ. З їх конструкції та співвідношень (4), (5) випливає, що $\mathbf{M}\xi = \lim_{n\to\infty} \mathbf{M}\xi_n^{\pm}$.

Лема 2. *Довільна випадкова величина ξ може бути записана у вигльді $\xi = \xi_+ - \xi_-$, де $\xi_\pm \ge 0$.*

Для доведення ціє леми потрібно взяти $\xi_+ = \max(0, \xi)$, $\xi_- = \max(0, -\xi)$

Зауваження 6. Легко зрозуміти, що $\mathbf{M}\xi < \infty$ тоді і тільки тоді, коли $\mathbf{M}\xi_\pm < \infty$.

Лема 3. *Нехай для випадкової величини $\xi \ge 0$ маємо $\mathbf{M}\xi < \infty$. Тоді існує послідовність обмежених випадкових величин $\xi_n \ge 0$ таких, що $\mathbf{M}\xi_n \to \mathbf{M}\xi$ при $n \to \infty$.*

Д о в е д е н н я. Нехай

$$\xi_n = \begin{cases} \xi, & 0 \le \xi \le n, \\ n, & \xi > n. \end{cases}$$

Маємо:

$$\mathbf{M}\xi - \mathbf{M}\xi_n = \int_0^\infty x dF(x) - \int_0^n x dF(x) - n(1-F(n)) \le \int_n^\infty x dF(x) +$$

$$+ n(1-F(n)) = \int_n^\infty x dF(x) + n\int_n^\infty dF(x) \le 2\int_n^\infty x dF(x) \underset{n\to\infty}{\longrightarrow} 0. \quad \blacktriangleleft$$

Перейдемо тепер до доведення властивостей 2), 3).

Властивість 2). Якщо взяти до уваги властивість 1), то вистачить розглянути випадок $a = b = 1$. Нехай спочатку випадкові величини ξ, η будуть обмеженими і невід'ємними, а $\xi_n^{\pm}$, $\eta_n^{\pm}$ будуть апроксиммуючими послідовностями для випадкових випадкових величин ξ, η. Легко зрозуміти (завдання 3), що послідовності $\xi_n^{\pm} + \eta_n^{\pm}$ будуть апроксимуючими для випадкової величини $\xi + \eta$. Тоді

$$\mathbf{M}\xi_n^- + \mathbf{M}\eta_n^- = \mathbf{M}(\xi_n^- + \eta_n^-) \le \mathbf{M}(\xi + \eta) \le \mathbf{M}(\xi_n^+ + \eta_n^+) = \mathbf{M}\xi_n^+ + \mathbf{M}\eta_n^+$$

Перехід до границі $n \to \infty$ в цих співвідношеннях дає властивість 2) для обмежених невід'ємних випадкових величин. З леми 3 легко випливає властивість 2) для довільних невід'ємних випадкових величин, а, використовуючи лему 2, завершуємо доведення властивості 2) для довільних випадкових величин.

Тепер доведемо властивість 3). Нехай випадкові величини ξ, η будуть обмеженими і невід'ємними, а $\xi_n^{\pm} \ge 0$, $\eta_n^{\pm} \ge 0$ будуть апроксиммуючими послідовностями для випадкових випадкових величин ξ, η. Неважко зрозуміти, що випадкові величини ξ_n^+, η_n^+ і ξ_n^-, η_n^- також є незалежними. Маємо

$$\xi_n^- \le \xi \le \xi_n^+, \qquad \eta_n^- \le \eta \le \eta_n^+.$$

Звідки

$$\xi_n^- \eta_n^- \le \xi\eta \le \xi_n^+ \eta_n^+ \Longrightarrow \mathbf{M}\xi_n^- \eta_n^- \le \mathbf{M}\xi\eta \le \mathbf{M}\xi_n^+ \eta_n^+.$$

А оскільки випадкові величини е $\xi_n^{\pm}, \eta_n^{\pm}$ є дискретними та незалежними, то з власності 3 теореми 2 маємо

$$\mathbf{M}\xi_n^- \mathbf{M}\eta_n^- \le \mathbf{M}\xi\eta \le \mathbf{M}\xi_n^+ \mathbf{M}\eta_n^+. \tag{9}$$

Переходячи в (9) до границі отримуємо властивість 3) для обмежених, невід'ємних випадкових величин. Як і раніше, з леми 3 випливає властивість 3) для довільних невід'ємних незалежних випадкових величин.

Якщо тепер ξ, η є довільні незалежні випадкові величини, то згідно з лемою 2 маємо $\xi = \xi_+ - \xi_-, \eta = \eta_+ - \eta_-$, де $\xi_\pm, \eta_\pm \geq 0$, і випадкові величини $\xi_\pm, \eta_\pm$ є незалежними. Отже,

$$\begin{aligned}\mathbf{M}(\xi\eta) &= \mathbf{M}(\xi_+ - \xi_-)(\eta_+ - \eta_-) = \mathbf{M}\xi_+\eta_+ - \mathbf{M}\xi_+\eta_- - \mathbf{M}\xi_-\eta_+ + \mathbf{M}\xi_-\eta_- = \\ &= \mathbf{M}\xi_+\mathbf{M}\eta_+ - \mathbf{M}\xi_+\mathbf{M}\eta_- - \mathbf{M}\xi_-\mathbf{M}\eta_+ + \mathbf{M}\xi_-\mathbf{M}\eta_- = \\ &= (\mathbf{M}\xi_+ - \mathbf{M}\xi_-)(\mathbf{M}\eta_+ - \mathbf{M}\eta_-) = \mathbf{M}(\xi_+ - \xi_-)\mathbf{M}(\eta_+ - \eta_-) = \mathbf{M}\xi\mathbf{M}\eta.\end{aligned}$$

Знайдемо математичне сподівання випадкових величин з експоненційним та нормальним розподілами.

Приклад 3. *Випадкова величина ξ має показниковий розподіл з параметром λ. Знайти* $\mathbf{M}\xi$.

Р о з в' я з о к. Інтегруючи частинами, будемо мати

$$\mathbf{M}\xi = \lambda \int\limits_0^\infty x e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_0^\infty + \int\limits_0^\infty e^{-\lambda x} dx = -\frac{e^{-\lambda x}}{\lambda} \Big|_0^\infty = \frac{1}{\lambda}.$$

Приклад 4. *Нехай $\xi \in N(m, \sigma^2)$. Знайти* $\mathbf{M}\xi$.

Р о з в' я з о к. Маємо

$$\mathbf{M}\xi = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^\infty x \exp\{-\frac{(x-m)^2}{2\sigma^2}\} dx = \frac{m}{\sqrt{2\pi}} \int\limits_{-\infty}^\infty e^{-\frac{z^2}{2}} dz + \frac{\sigma}{\sqrt{2\pi}} \int\limits_{-\infty}^\infty z e^{-\frac{z^2}{2}} dz = m.$$

Остання рівність випливає з того, що функція $(2\pi)^{-1/2} e^{-\frac{z^2}{2}}$ є щільністю а функція $z e^{-\frac{z^2}{2}}$ - непарна і інтегрування в останньому інтегралі відбувається на симетричному інтервалі.

7.2 Дисперсія

Нехай ξ є випадковою величиною зі скінченним математичним сподіванням $\mathbf{M}\xi < \infty$.

Визначення 3. *Дисперсію випадкової величини* ξ будемо позначати символом $\mathbf{D^2}\xi$ і визначати як

$$\mathbf{D^2}\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2, \tag{10}$$

за умови, що математичне сподівання в правій частині рівності (10) є скінченним.

Якщо випадкова величина ξ є дискретною і приймає (скінченне) злічення число значень $x_1, x_2, ...$ відповідно з ймовірностями p_1, p_2, ..., тоді

$$\mathbf{D^2}\xi = \sum_{i=1}^{\infty}(x_i - \mathbf{M}\xi)^2 p_i, \tag{11}$$

за умови, що ряд в (11) є збіжним.

Якщо функція $F(x)$ ($f(x)$) є функцією розподілу (щільністю) випадкової величини ξ, то формулу (10) можна записати наступним чином:

$$\mathbf{D^2}\xi = \int_{-\infty}^{\infty}(x - \mathbf{M}\xi)^2 dF(x), \qquad \Big(\mathbf{D^2}\xi = \int_{-\infty}^{\infty}(x - \mathbf{M}\xi)^2 f(x)dx.\Big) \tag{12}$$

У формулах (12) ми припускаємо, що інтеграли є скінченними.

Зауваження 7. Позаяк у визначенні (10) значення $(\xi - \mathbf{M}\xi)^2$ є невід'ємним, то інтеграл

$$\int_{-\infty}^{\infty}(x - \mathbf{M}\xi)^2 dF(x) \tag{13}$$

є завше визначеним, хоча можлива ситуація, коли він дорівнюватиме $+\infty$. Тоді природньо треба прийняти, що $\mathbf{D^2}\xi = +\infty$. Тому в деяких підручниках автори вважають, що дисперсія є визначеною завжди (тобто навіть тоді, коли інтеграл в (13) є нескінченним). В цьому підручнику ми завжди будемо вважати, що дисперсія визначена лише тоді, коли вона скінченна

Наступний простий результат часто використовується для знаходження дисперсії

Теорема 2. *Має місце рівність:*

$$\mathbf{D^2}\xi = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2. \tag{14}$$

Д о в е д е н н я. Маємо

$$\mathbf{D^2}\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2 = \mathbf{M}\ \ \xi^2 - 2\xi\mathbf{M}\xi + (\mathbf{M}\xi)^2\ \ = \mathbf{M}\xi^2 - 2\mathbf{M}\xi\mathbf{M}\xi + (\mathbf{M}\xi)^2 = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2.$$

Наслідок 2. *Якщо* $\mathbf{M}\xi^2 < \infty$, *то*

$$\mathbf{M}\xi^2 \geq (\mathbf{M}\xi)^2. \tag{15}$$

Нерівність (15) випливає з (14), позаяк завше $\mathbf{D^2}\xi \geq 0$. ◀

Теорема 3. *Дисперсія має наступні властивості:*

1) *якщо $c = const$, то $\mathbf{D^2}c = 0$ і $\mathbf{D^2}c\xi = c^2\mathbf{D^2}\xi$;*

2) *якщо випадкові величин ξ, η, є <u>незалежними</u>, тоді для довільних сталих $a, b \in R^1$ має місце рівність*

$$\mathbf{D^2}(a\xi + b\eta) = a^2\mathbf{D^2}\xi + b^2\mathbf{D^2}\eta;$$

3) *для довільної сталої c має місце рівність*

$$\mathbf{M}(\xi - c)^2 \geq \mathbf{M}(\xi - \mathbf{M}\xi)^2,$$

а вираз $\mathbf{M}(\xi - c)^2$ приймає найменше значення, коли $c = \mathbf{M}\xi$, і значення це, вочевидь, дорівнює $\mathbf{D^2}\xi$.

Д о в е д е н н я. Властивостей 1), 2) легко випливає з визначення дисперсії і властивостей математичного сподівання.

Доведемо властивість 3). Маємо

$$\mathbf{M}(\xi - c)^2 = \mathbf{M}((\xi - \mathbf{M}\xi) + (\mathbf{M}\xi - c))^2 = \mathbf{M}(\xi - \mathbf{M}\xi)^2 + 2\mathbf{M}(\xi - \mathbf{M}\xi)(\mathbf{M}\xi - c) + (\mathbf{M}\xi - c)^2$$
$$= \mathbf{M}(\xi - \mathbf{M}\xi)^2 + (\mathbf{M}\xi - c)^2 \geq \mathbf{M}(\xi - \mathbf{M}\xi)^2.$$

Якщо $c = \mathbf{M}\xi$, то в цих співвідношеннях настає рівність. ◀

Наслідок 3. *Якщо випадкові величини $\xi_1, \xi_2, \ldots, \xi_n$ є незалежними, то для довільних $a_i \in R^1$ має місце рівність:*

$$\mathbf{D^2}\sum_{i=1}^{n} a_i\xi_i = \sum_{i=1}^{n} a_i^2\mathbf{D^2}\xi_i.$$

Визначення 4. Корінь квадратний з дисперсії, або $\sqrt{\mathbf{D^2}\xi} = \mathbf{D}\xi$ називають *стандартним відхиленням.*

Приклад 1. *Знайти дисперсію $\mathbf{D^2}\xi$ випадкової величини типу 0-1.*

Р о з в' я з о к. Нехай $\mathbf{P}(\xi = 1) = p \in [0, 1]$ і $\mathbf{P}(\xi = 0) = 1 - p = q$. Позаяк $\mathbf{M}\xi = p$, тоді будемо мати

$$\mathbf{D^2}\xi = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2 = p - p^2 = pq.$$

Приклад 2. *Знайти $\mathbf{D^2}\xi$, якщо випадкова величина ξ має біноміальний розподіл з параметрами n, p.*

Р о з в' я з о к. Випадкову величину, яка має біноміальний розподіл можемо змоделювати схемою Бернуллі з параметрами n, p. Припустимо, що ми маємо n незалежних випробувань типу 0-1 з ймовірністю успіху p. Нехай $\xi_i = 1$, якщо в i-тому випробуванні був успіх і $\xi_i = 0$, якщо неуспіх. Тоді $\xi_i, i = 1, ..., n$ є незалежними однаково розподіленними випадковими величинами типу 0-1. Зрозуміло, що $\xi = \sum_{i=1}^{n} \xi_i$ Позаяк $\mathbf{D^2}\xi_i = pq$, то з властивості дисперсії 2) маємо:

$$\mathbf{D^2}\xi = \mathbf{D^2}\ \sum_{i=1}^{n} \xi_i\ = \sum_{i=1}^{n} \mathbf{D^2}\xi_i = npq.$$

Приклад 3. *Знайти* $\mathbf{D^2}\xi$*, якщо випадкова величина* ξ *має показниковий розподіл з параметром* λ*.*

Р о з в' я з о к. Оскільки $\mathbf{M}\xi = \lambda^{-1}$, то

$$\mathbf{D^2}\xi = \mathbf{M}\xi^2 - \lambda^{-2} = \lambda \int_0^{\infty} x^2 e^{-\lambda x} dx - \lambda^{-2} = \lambda^{-2}.$$

Приклад 4. $\xi \in N(m, \sigma^2)$*. Знайти* $\mathbf{D^2}\xi$*.*

Р о з в' я з о к. Відомо, що $\mathbf{M}\xi = m$. Підрахуємо

$$\mathbf{M}(\xi - m)^2 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - m)^2 \exp\{-\frac{(x - m)^2}{2\sigma^2}\} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} dz =$$

$$= (\text{інтегруючи частинами}) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \sigma^2.$$

Дисперсія, а також стандартне відхилення, є характеристиками, які описують "розпорошення" значень випадкової величини стосовно математичного сподівання. За малої дисперсії випадкова величина приймає свої значення "недалеко" від математичного сподівання. Докладний сенс слова "недалеко" розкриває наступне твердження, яке дає можливість оцінити ймовірність того, що випадкова величина ξ належить до певної області.

Теорема 4. (Нерівність Чебишова). *Якщо випадкова величина* ξ *є такою, що* $\mathbf{D^2}\xi < \infty$*, тоді для довільного* $\varepsilon > 0$ *має місце нерівність*

$$\mathbf{P}(|\xi - \mathbf{M}\xi| \geq \varepsilon) \leq \frac{\mathbf{D^2}\xi}{\varepsilon^2}. \tag{16}$$

Д о в е д е н н я. Позначимо $\mathbf{M}\xi = m$. Для $\varepsilon > 0$ будемо мати:

$$\mathbf{D^2}\xi = \int_{-\infty}^{\infty} (x - m)^2 dF(x) = \int_{|x-m|<\varepsilon} (x - m)^2 dF(x) + \int_{|x-m|\geq\varepsilon} (x - m)^2 dF(x) \geq$$

$$\geq \int_{|x-m|\geq\varepsilon} (x - m)^2 dF(x) \geq \varepsilon^2 \int_{|x-m|\geq\varepsilon} dF(x) = \varepsilon^2 \mathbf{P}(|\xi - m| \geq \varepsilon).$$

Отже, $\mathbf{D^2}\xi \geq \varepsilon^2 \mathbf{P}(|\xi - m| \geq \varepsilon)$, що є рівнозначним (16).

Наслідок 4. *Якщо* $\mathbf{D^2}\xi = 0$, *то* $\mathbf{P}(\xi = \mathbf{M}\xi) = 1$.

Справді, якщо $\mathbf{D^2}\xi = 0$, то з нерівності (16) випливає, що для кожного $\varepsilon > 0$ будемо мати: $\mathbf{P}(|\xi - \mathbf{M}\xi| \geq \varepsilon) = 0$, і при $\varepsilon \downarrow 0$ отримаємо $\mathbf{P}(|\xi - \mathbf{M}\xi| > 0) = 0$. Це є, вочевидь, можливим тільки тоді, коли $\mathbf{P}(\xi = \mathbf{M}\xi) = 1$. ◀

ЗАВДАННЯ

1. Довести, що для кожної події A має місце рівність

$$\mathbf{M}\chi_A(\xi) = \mathbf{P}(\xi \in A),$$

де $\chi_A(\xi)$ – характеристична функція множини A.

2. Знайти математичне сподівання випадкової величини ξ з наступним дискретним розподілом:

x_i	-1	2	3	7
$\mathbf{P}(\xi = x_i)$	1/8	1/4	1/8	1/2

3. Довести, що якщо послідовності а $\xi_n^\pm$, $\eta_n^\pm$ є апроксиммуючими для випадкових величини ξ, η, то послідовності $\xi_n^\pm + \eta_n^\pm$ будуть апроксимуючими для випадкової величини $\xi + \eta$.

4. Знати математичні сподівання і дисперсії випадкових величин, які мають наступні розподіли: геометричний, рівномірний, гамма і розподіл Пуассона.

5. Навести приклад дискретної випадкової величини, яка не має математичного сподівання.

6. Довести, що якщо $\mathbf{M}\xi$ існує, то $\mathbf{M}\xi = \int_0^\infty (1-F(x))dx - \int_0^{-\infty} F(x)dx$ і $\lim_{x\to\infty} x(1 - F(x)) = \lim_{x\to-\infty} xF(x) = 0$.

7. Знайти математичне сподівання випадкової величини, яка має щільність:

$$f(x) = \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}}, \qquad f(x) = \frac{1}{2} e^{-|x|}.$$

8. Випадкова величина ξ має розподіл:

   a) $\mathbf{P}(\xi = k) = \frac{a^k}{(1+a)^{k+1}}, \quad a > 0, k = 0, 1, 2, \ldots (i \ );$

   b)

$$\mathbf{P}(\xi = k) = \Big(\frac{\alpha\lambda}{1 + \alpha\lambda}\Big)^k \frac{(1+\alpha)\ldots(1+(k-1)\alpha)}{k!} p_0,$$
$$k = 1, 2, \ldots, \quad \alpha > 0, \lambda > 0$$

   і

$$\mathbf{P}(\xi = 0) = (1 + \alpha\lambda)^{-1/\alpha}$$

   Цей розподіл називають *розподілом Поля (Polya).*

   Знайти $\mathbf{M}\xi$, $\mathbf{D^2}\xi$.

9. Навести приклад випадкової величини, яка має скінченне математичне сподівання, але нескінченну дисперсію.

10. Знайти дисперсії випадкових величин, які мають наступні щільності:

    а) $f(x) = \begin{cases} \frac{1}{2}\cos x, & |x| \le \frac{\pi}{2}, \\ 0, & |x| > \frac{\pi}{2}, \end{cases}$ б) $f(x) = \frac{1}{2}e^{-|x|}$,

    в)$f(x) = \begin{cases} \frac{3}{4}(2x - x^2), & 0 \le x \le 2, \\ 0, & x < 0, \end{cases}$ г) $f(x) = \begin{cases} \frac{1}{\pi\sqrt{1-x^2}}, & |x| < 1, \\ 0, & |x| \ge 1. \end{cases}$

11. Нехай $\xi_1, ..., \xi_n$– незалежні рівномірно розподілені на $[0, 1]$ випадкові величини. Позначимо $\eta = \xi_1 \cdot \xi_2 \cdot .... \cdot \xi_n$. Обчислити $\mathbf{M}\eta$.

Лекція 8. Інші числові характеристики випадкових величин

Як ми вже говорили, що такі характеристики як $\mathbf{M}\xi$, $\mathbf{D^2}\xi$ вже дають певну інформацію про випадкову величину ξ. Так, наприклад, за допомогою нерівності Чебишева можемо оцінити відхилення випадкової величини від її середньої вартості. Існують і інші числові характеристики, які теж дозволяють зрозуміти деякі особливості випадкових величини або її розподілу. Головні з них це міри середніх значень (до них належать $\mathbf{M}\xi$, $\mathbf{D^2}\xi$), мінливості, асиметрії та концентрації.

8.1 Міри положення, асиметрії та розпорошеності

Визначення 1. *Звичайним моментом порядку* k і *абсолютним моментом порядку* k випадкової величини ξ будемо називати відповідно:

$$m_k = \mathbf{M}\xi^k = \int\limits_{-\infty}^{\infty} x^k dF(x), \qquad \nu_k = \mathbf{M}|\xi|^k = \int\limits_{-\infty}^{\infty} |x|^k dF(x), \tag{1}$$

де $k = 1, 2, ...$

Визначення 2. *Центральним моментом порядку* k і *абсолютним центральним моментом порядку* k випадкової величини ξ будемо називати відповідно:

$$\mu_k = \mathbf{M}(\xi - \mathbf{M}\xi)^k, \qquad \varkappa_k = \mathbf{M}|\xi - \mathbf{M}\xi|^k. \tag{2}$$

Зауваження 1. Зауважимо, що для абсолютних моментів значення $k = 1, 2, ...$ можна замінити на $k > 0$.

Визначення 3. *Квантиллю порядку* $0 < p < 1$ випадкової величини ξ з функцією розподілу $F(x)$ будемо називати число

$$k(p) = \inf\{x : F(x) > p\}. \tag{3}$$

Якщо функція розподілу $F(x)$ є неперервною в точці $k(p)$, тоді $k(p)$ є найменшим розв'язком рівняння $F(x) = p$, а якщо на додаток функція розподілу $F(x)$ зростає в точці $k(p)$, то $k(p)$ є єдиним розв'язком рівняння $F(x) = p$.

Квантиль $k(1/2)$ називають *медіаною* випадкової величини ξ і часто позначають символом m_e.

Визначення 4. *Модою* або *домінантою* випадкової величини ξ, позначають символом m_0, будемо називати:

а) у випадку дискретної випадкової величини – значення випадкової величини, яке є найбільш ймовірним,

б) у випадку абсолютно неперервної випадкової величини – значення аргументу, в якому щільність досягає локального максимуму.

Значення моди у випадку а) є тим значенням, яке випадкова величина приймає найчастіше. У випадку б) випадкова величина найчастіше приймає значення, які лежать в околі моди (таких величин може бути кілька, отже, мода може приймати не єдине значення).

Приклад 1. *Знайти модальне значення випадкової величини ξ зі щільністю:*

$$f(x) = \begin{array}{rl} xe^{-2x/\sqrt{3}}, & x > 0, \\ -xe^{2x}, & x \le 0. \end{array}$$

Р о з в' я з о к. Треба знайти локальний максимум щільності. Для $x > 0$ маємо:

$$f'(x) = e^{-2x/\sqrt{3}} \;\; 1 - \frac{2x}{\sqrt{3}} \;\; ,$$

отже, $f'(x) = 0$ для $x_0 = \sqrt{3}/2$. Легко довести, що це є локальний максимум щільності на піввісі $x > 0$.

Для $x < 0$ аналогічно можна показати, що $x_2 = -1/2$ є локальним максимумом щільності на піввісі $x \le 0$. Випадкова величина ξ має два модальні значення.

Визначення 5. *Показником мінливості* випадкової величини ξ називають число

$$\gamma_1 = \frac{\mathbf{D^2}\xi}{|\mathbf{M}\xi|}, \qquad \mathbf{M}\xi \ne 0. \tag{4}$$

Як ми вже говорили, дисперсія характеризує розпорошення випадкової величини, але коли нам потрібно порівняти розпорошенність двох випадкових величин, то зробити це за допомогою дисперсії неможливо, оскільки якщо середні значення цих величин сильно різняться, наприклад, 1 та 100, то відхилення на ± 1 від середнього значення для першої випадкової величини означає зміну середнього значення на 100% а для другої-на 1%. В цьому випадку використовуємо показник мінливості.

Визначення 6. *Показником асиметрії* випадкової величини ξ називають число

$$\gamma_2 = \frac{\mu_3}{\mathbf{D}^3\xi}. \tag{5}$$

Якщо $\gamma_2 > 0$, то кажуть, що асиметрія розподілу є *додатньою*, в цьому випадку графік щільність випадкової величини ξ має асиметрію правостронню, а якщо $\gamma_2 < 0$, то кажуть, що асиметрія є *від'ємною* і в цьому випадку графік щільність випадкової величини ξ має асиметрію лівостронню. У випадку $\gamma_2 = 0$, графік щільність є, більш-менш, симетричним.

Для нормального розподілу цей показник дорівнює нулеві.

Визначення 7. *Ексцесом* випадкової величини ξ називають значення

$$\gamma_3 = \frac{\mu_4}{\mathbf{D}^4\xi} - 3. \tag{6}$$

Цей показник служить для порівняння скупчення даного розподілу зі скупченням нормального розподілу. Додатній ексцес вказує на те, що графік даного розподілу є вищим і більш сконцентрованим довкола вісі $x = \mathbf{M}\xi$, ніж графік кривої нормального розподілу. Від'ємний ексцес має протилежне значення.

Наступне поняття характеризує зв'язок між випадковими змініими ξ, η .

Визначення 8. Нехай випадкові величини ξ, η є такими, що $\mathbf{M}\xi^2 < \infty, \mathbf{M}\eta^2 < \infty$. *Коефіцієнтом кореляції* випадкових величин ξ, η називається число

$$\rho_{\xi\eta} = \frac{\mathbf{M}[(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta)]}{\mathbf{D}\xi\mathbf{D}\eta}. \tag{7}$$

Теорема 1. *Коефіцієнт кореляції має такі властивості:*

1) $|\rho_{\xi\eta}| \le 1$;

2) *якщо величини ξ, η є незалежними, то $\rho_{\xi\eta} = 0$;*

3) $\rho_{a\xi+b,c\eta+d}{=}\rho_{\xi\eta}$ *для довільних чисел $a, b, c, d \in R^1$ таких, що $ac > 0$;*

4) *якщо $|\rho_{\xi\eta}| = 1$, то $\alpha\xi + \beta\eta = \gamma$ з ймовірністю 1 для деяких $\alpha, \beta, \gamma \in R^1$.*

Д о в е д е н н я. Пункт перший випливає віразу з нерівності Коші-Буняковського (нерівність (9) нижче.)

Доведення пунктів 2), 3) є легким а тому зоставимо його читачеві. Доведемо тепер четвертий пункт для випадку $\rho_{\xi\eta} = 1$. Згідно з (7) маємо

$$\mathbf{M}[(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta)] = \mathbf{D}\xi\mathbf{D}\eta. \tag{8}$$

Позначимо $\xi_1 = (\xi - \mathbf{M}\xi)/\mathbf{D}\xi$, $\eta_1 = (\eta - \mathbf{M}\eta)/\mathbf{D}\eta$. Тоді $\mathbf{M}\xi_1 = 0$, $\mathbf{D}^2\xi_1 = 1$, $\mathbf{M}\eta_1 = 0$, $\mathbf{D}^2\eta_1 = 1$, а тому з (8) дістаємо $\mathbf{M}(\xi_1\eta_1) = 1$. Маємо

$$\mathbf{D}^2(\xi_1 - \eta_1) = \mathbf{M}(\xi_1 - \eta_1)^2 = \mathbf{M}\xi_1^2 - 2\mathbf{M}(\xi_1\eta_2) + \mathbf{M}\eta_1^2 = 0.$$

Отже $\mathbf{D}^2(\xi_1 - \eta_1) = 0$ і це означає, що з ймовірністю 1 $\xi_1 - \eta_1 = c = const$, або

$$\frac{\xi - \mathbf{M}\xi}{\mathbf{D}\xi} - \frac{\eta - \mathbf{M}\eta}{\mathbf{D}\eta} = c,$$

що, очевидно, доводить пункт 4).

Властивість 3) означає, що лінійне перетворення випадкових величин ξ, η не впливає на силу їх зв'язку а властивість 4) означає, що чим ближче $\rho_{\xi\eta}$ до ± 1 тим міцніший лінійна залежність між величинами ξ, η, і та залежність є на 100% лінійною якщо $\rho_{\xi\eta} = \pm 1$. Тому можна сказати, що коефіцієнт кореляції характеризує лінійний зв'язок між випадковими величинами.

Якщо $\rho_{\xi\eta} = 0$, то говоримо, що випадкові величини ξ, η є *нескорельованими*, якщо $\rho_{\xi\eta} > 0$, то говоримо, що випадкові величини ξ, η є *скорельовані додатньо*, а у випадку $\rho_{\xi\eta} < 0$ говоримо, що випадкові величини ξ, η є *скорельовані від'ємно.*

8.2 Нерівності для моментів

Теорема 2. *Якщо випадкові величини ξ, η є такими, що $\mathbf{M}\xi^2 < \infty$ і $\mathbf{M}\eta^2 < \infty$, то*

1) (Коші-Буняковського (Cauchy-Buniakowski))

$$|\mathbf{M}\xi\eta| \leq \sqrt{\mathbf{M}\xi^2\mathbf{M}\eta^2}. \tag{9}$$

2) (Шварца (Schwarz))

$$\sqrt{\mathbf{M}(\xi+\eta)^2} \leq \sqrt{\mathbf{M}\xi^2} + \sqrt{\mathbf{M}\eta^2}. \tag{10}$$

Д о в е д е н н я. Позаяк

$$\mathbf{M}|\xi\eta| \leq \frac{1}{2}(\mathbf{M}\xi^2 + \mathbf{M}\eta^2) < \infty,$$

то значення $\mathbf{M}\xi\eta$ є визначеним.

Нехай $x \in R^1$. Тоді

$$0 \le \mathbf{M}(\xi + x\eta)^2 = x^2\mathbf{M}\eta^2 + 2x\mathbf{M}\xi\eta + \mathbf{M}\xi^2. \tag{11}$$

Якщо розглядати праву частину в (11) як квадратичну функцію стосовно x, то з (11) випливає, що ця функція є невід'ємною, що є можливим лише тоді, коли дискримінат цієї квадратичної функції не більший нуля, тобто

$$4(\mathbf{M}\xi\eta)^2 - 4\mathbf{M}\xi^2\mathbf{M}\eta \le 0,$$

що є рівнозначним (9).

Для доведення (10) використаємо нерівність (9). Маємо:

$$\mathbf{M}(\xi+\eta)^2 \le \mathbf{M}\xi^2 + 2\mathbf{M}\xi\eta + \mathbf{M}\eta^2 \le \mathbf{M}\xi^2 + 2\sqrt{\mathbf{M}\xi^2\mathbf{M}\eta^2} + \mathbf{M}\eta^2 = (\sqrt{\mathbf{M}\xi^2} + \sqrt{\mathbf{M}\eta^2})^2.$$

◄

Наступні 2 твердження, які включають ряд важливих нерівностей, подамо без доведення. Їх можна знайти в [1] або [**?**]. Зазначимо лише, що нерівність (12) є узагальненням нерівності Чебишова і має подібне доведення.

Теорема 3.

1) (Маркова) *Якщо* $\mathbf{M}|\xi|^r < \infty$, $r > 0$, *то для довільного* $\varepsilon > 0$ *має місце нерівність*

$$\mathbf{P}(|\xi| \ge \varepsilon) \le \frac{\mathbf{M}|\xi|^r}{\varepsilon^r}. \tag{12}$$

2) (Гельдера (Galder)) *Нехай* $p > 0$, $q > 0$, $1/p + 1/q = 1$, *а випадкові величини* ξ, η *є такими, що* $\mathbf{M}|\xi|^p < \infty$ *і* $\mathbf{M}|\eta|^q < \infty$, *тоді*

$$\mathbf{M}|\xi\eta| \le \big(\mathbf{M}|\xi|^p\big)^{1/p}\big(\mathbf{M}|\eta|^p\big)^{1/p}. \tag{13}$$

3) (Ляпунова) *Якщо* $0 < s < t$, *то*

$$\big(\mathbf{M}|\xi|^s\big)^{1/s} \le \big(\mathbf{M}|\xi|^t\big)^{1/t}. \tag{14}$$

З останньої нерівності отримаємо:

Наслідок 1. *Якщо* $\mathbf{M}|\xi|^t < \infty,\ t > 0,$ *to* $\mathbf{M}|\xi|^s < \infty,\ 0 < s < t.$

Якщо в (12) замість ξ взяти $\xi - \mathbf{M}\xi$, то ми отримаємо нерівність для центральних моментів:

Наслідок 2. *Якщо* $\mathbf{M}|\xi|^r < \infty$, $r > 0$, *то для довільного* $\varepsilon > 0$ *має місце*

$$\mathbf{P}(|\xi - \mathbf{M}\xi| \geq \varepsilon) \leq \frac{\mu_r}{\varepsilon^r}. \tag{15}$$

Теорема 4. (Колмогорова) *Нехай* $\xi_1, \xi_2, ..., \xi_n$ *послідовність незалежних випадкових величин таких, що* $\mathbf{M}\xi_i^2 < \infty$, $i = 1, ...n$. *Позначимо* $S_n = \sum_{k=1}^n (\xi_k - \mathbf{M}\xi_k)$. *Для довільного* $\varepsilon > 0$ *має місце нерівність*

$$\mathbf{P}(\max\{|S_1|, ..., |S_n|\} \geq \varepsilon)) \leq \frac{\mathbf{D^2} S_n}{\varepsilon^2}. \tag{16}$$

Наступна таблиця містить числові характеристики розподілів, які найчастіше використовуються.

Назва розподілу	Щільність або розподіл	Мат. спод.	Дисперсія	Асиметрія	Ексцес
Нуля-одиниці	$p, q = 1 - p$	p	pq	$-\frac{2p-1}{\sqrt{pq}}$	$\frac{1-6p+6p^2}{pq}$
Біноміальний	$C_n^k p^k q^{n-k}$	np	npq	$-\frac{2p-1}{\sqrt{npq}}$	$\frac{1-6p+6p^2}{npq}$
Геометричний	$(1-p)^{n-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{p+1}{\sqrt{p}}$	$\frac{1+4p+p^2}{p}$
Пуассона	$e^{-\lambda}\frac{\lambda^k}{k!}$	λ	λ	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\lambda}$
Показниковий	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	2	6
Ерланга	$\frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$	$\frac{2}{\sqrt{n}}$	$\frac{6}{n}$
χ_n^2	$\frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)}$	n	$2n$	$\sqrt{\frac{8}{n}}$	$\frac{12}{n}$
Рівномірний	$\frac{1}{b-a}, x \in (a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	0	$-\frac{6}{5}$
Нормальний	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2	0	0
Гама	$\frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}$	$\frac{p}{\lambda}$	$\frac{p}{\lambda^2}$	$\frac{2}{\sqrt{p}}$	$\frac{6}{p}$

ЗАВДАННЯ

1. Для розподілу зі щільністю

$$f(x) = \begin{cases} \frac{3}{32}(4x - x^2), & \text{для} \quad 0 \le x \le 4, \\ 0, & \text{для} \quad \text{інших } x \end{cases}$$

   знайти m_e, m_o.

2. Випадкова величина ξ приймає значення -3, -1, 2, 4 з відповідними ймовірностями: 1/8, 1/8, 1/2, 1/4.

   а) Знайти медіану і моду цього розподілу,

   б) знайти розподіл випадкової величини $\eta = \xi^2$, а також її медіану і моду.

3. Для розподілу зі щільністю

$$f(x) = \begin{cases} \frac{1}{2}x, & 0 \le x \le 2, \\ 0, & \text{для інших } \ x \end{cases}$$

   знайти квантилі $k(1/4)$, $k(3/4)$.

4. Підрахувати дисперсію, третій звичайний і центральний моменти для випадкової величини ξ з наступним дискретним розподілом:

x_i	-1	2	3	7
$\mathbf{P}(\xi = x_i)$	1/8	1/4	1/8	1/2

5. Нехай випадкова величина $\xi \in N(m, \sigma^2)$ довести, що

$$\mathbf{M}(\xi - m)^n = \begin{cases} 0, & \text{якщо} \quad n = 2k+1, \quad k = 1, 2, \dots \\ \\ (2k-1)!!\sigma^{2k}, & \text{якщо} \quad n = 2k, \qquad k = 1, 2, \dots \end{cases}$$

6. Для випадкової величини неперервного типу ξ маємо: $\mathbf{M}\xi$=2, $\mathbf{D^2}\xi$=$0,001$. Оцінити зверху наступну ймовірність $\mathbf{P}(\xi \notin [1,9; 2,1])$.

7. Нехай $\mathbf{M}e^{c\xi} < \infty, \ c > 0$. Показати, що для довільного $\varepsilon > 0$ будемо мати $\mathbf{P}(\xi > \varepsilon) \le e^{-c\varepsilon}\mathbf{M}e^{c\xi}$.

8. Нехай $\xi \in N(m, \sigma^2)$. Довести, що для довільного $\varepsilon > 0$

$$\mathbf{P}(|\xi - a| > \varepsilon\sigma) \le \frac{2}{\varepsilon\sqrt{2\pi}} e^{-\varepsilon^2/2}.$$

Лекція 9. Багатовимірні випадкові величини

9.1 Багатовимірна функція розподілу

Нехай $\xi_1, \xi_2, \ldots, \xi_n$ є випадковими величинами, визначеними на ймовірносному просторі $(\Omega, \mathfrak{B}, \mathbf{P})$. Кожному $\omega \in \Omega$ ці випадкові величини ставлять у відповідність n-вимірний вектор $\bar{\xi}(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$. Точку з координатами $(\xi_1(\omega), \xi_2(\omega), \ \ldots \ , \xi_n(\omega))$ можна також трактувати як точку n-вимірного евклідового простору R^n. Отже, можна казати, що випадкові величини $\xi_1, \xi_2, \ldots, \xi_n$ визначають відображення простору Ω в простір R^n. Підсумуємо це все у вигляді наступного визначення.

Визначення 1. *Випадковим вектором* або *багатовимірною випадковою величиною* будемо називати сукупність випадкових величин $\bar{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$.

Визначення 2. Функцію

$$F_{\xi_1 \xi_2 \ldots \xi_n}(x_1, \ldots, x_n) = \mathbf{P}(\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n)$$

будемо називати *функцією розподілу випадкового вектора* $(\xi_1, \xi_2, \ldots, \xi_n)$ або *спільною функцією розподілу випадкових величин* $\xi_1, \xi_2, \ldots, \xi_n$.

Визначення 3. Випадкові величини $\xi_1, \xi_2, , .., \xi_n$, які визначені на одному ймовірносному просторі, будемо називати *незалежними*, якщо для всіх $x_i \in R^1, i = 1, .., n$ виконана умова:

$$\mathbf{P}(\xi_1 < x_1, \xi_2 < x_2, , .., \xi_n < x_n) = \prod_{i=1}^{n} \mathbf{P}(\xi_i < x_i)$$

або, що є еквівалентним,

$$F_{\xi_1, \xi_2, , .., \xi_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{\xi_i}(x_i),$$

де $F_{\xi_i}(x_i)$ є функціями розподілу випадкових величин $\xi_i, \ i = 1, \ldots, n$.

Визначення 4. Будемо казати, що функція розподілу випадкового вектора $(\xi_1, \xi_2, .., \xi_n)$ є *абсолютно неперервного* типу, якщо її можна записати у вигляді:

$$F(x_1, \dots, x_n) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \cdots dt_n, \qquad (1)$$

де $f(t_1, \dots, t_n) \geq 0$. В цьому випадку функція $f(t_1, \dots, t_n)$ називається *спільною щільністю* випадкових величин $(\xi_1, \xi_2, .., \xi_n)$.

З визначення щільності випливає, що

$$\int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f(t_1, \dots, t_n) dt_1 \cdots dt_n = F(\infty, \infty, \dots, \infty) = 1.$$

Має місце наступне твердження:

Теорема 1. *Для того, щоб функція $f(x_1, x_2, \dots, x_n)$ була щільністю певної функції розподілу n-вимірного випадкового вектора необхідно і достатньо, щоб були виконані наступні умови:*

а) $f(x_1, x_2, \dots, x_n) \geq 0$,

б) $\int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f(t_1, \dots, t_n) dt_1 \cdots dt_n = 1.$

Якщо зазначені умови виконані, то відповідна функція розподілу може бути представлена у вигляді (1).

Якщо функція $f(t_1, \dots, t_n)$ є неперервною в точці $(x_1, x_2, \dots, x_n)$, тоді, так само як і в одновимірному випадку, маємо:

$$\frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n} = f(x_1, x_2, \dots, x_n). \qquad (2)$$

Зокрема, якщо функція $f(t_1, \dots, t_n)$ неперервна, то рівність (2) має місце для всіх x_1, x_2, $\dots x_n$.

У загальному випадку можна стверджувати, що існує така борелівська множина $B \subset R^n$, $\lambda_n(B) = 0$, де $\lambda_n(\cdot)$ є мірою Лебега, що рівність (2) має місце для всіх $x = (x_1, x_2, \dots, x_n) \in R^n \backslash B$.

9.2 Двовимірна неперервна випадкова величина

Нехай заданий двовимірний випадковий вектор (ξ, η) і $F(x, y) = \mathbf{P}(\xi < x, \eta < y)$ є його функцею розподілу. Позначимо через $F_\xi(x)$, $F_\eta(x)$

функцію розподілу ξ, η, тобто $F_\xi(x) = \mathbf{P}(\xi < x)$, $F_\eta(x) = \mathbf{P}(\eta < x)$. Вочевидь,

$$F_\xi(x) = F(x, \infty), \qquad F_\eta(x) = F(\infty, x).$$

Визначення 5. Функції $F_\xi(x)$, $F_\eta(x)$ називають *маргінальними функціями розподілу* випадкового вектора (ξ, η).

Якщо функція $F(x, y)$ є абсолютно неперервного типу, тобто

$$F(x, y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f(u, v) dv du, \tag{3}$$

то

$$F_\xi(x) = F(x, \infty) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{\infty} f(u, v) dv du = \int\limits_{-\infty}^{x} f_\xi(u) du,$$
$$F_\eta(y) = F(\infty, y) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{y} f(u, v) dv du = \int\limits_{-\infty}^{y} f_\eta(v) dv,$$

де

$$f_\xi(u) = \int_{-\infty}^{\infty} f(u, v) dv, \quad f_\eta(v) = \int_{-\infty}^{\infty} f(u, v) du. \tag{4}$$

Ми фактично довели наступне твердження:

Теорема 2. *Якщо функція розподілу випадкового вектора (ξ, η) є абсолютно неперервною зі щільністю $f(x, y)$, то маргінальні функції розподілу $F_\xi(x)$, $F_\eta(y)$ також абсолютно неперервні зі щільностями $f_\xi(x)$, $f_\eta(y)$ з (4).*

Якщо ξ, η є незалежним, то згідно з визначенням 3 маємо:

$$F(x, y) = F_\xi(x) F_\eta(y). \tag{5}$$

Крім того, якщо функція розподілу $F(x, y)$ є абсолютно неперервного типу з щільністю $f(x, y)$, то

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y), \quad \frac{dF_\xi(x)}{dx} = f_\xi(x), \quad \frac{dF_\eta(y)}{dy} = f_\eta(y),$$

і згідно з (5) маємо:

$$f(x,y) = f_\xi(x) f_\eta(y). \tag{6}$$

Нехай функція розподілу випадкового вектора (ξ, η) є абсолютно неперервного типу і має місце рівність (6). Тоді

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dvdu = \int_{-\infty}^{x} f_\xi(u)du \int_{-\infty}^{y} f_\eta(v)dv = F_\xi(x)F_\eta(y).$$

Отже, ми довели наступне твердження:

Теорема 3. *Нехай функція розподілу випадкового вектора (ξ, η) є абсолютно неперервного типу зі щільністю $f(x,y)$, а $f_\xi(x)$, $f_\eta(y)$ - маргінальні щільності. Тоді для незалежності випадкових величин ξ, η необхідно і достатньо, щоб була виконана рівність:*

$$f(x,y) = f_\xi(x) f_\eta(y).$$

Приклад 1. *Спільна щільність випадкових величин ξ, η має вигляд*

$$f(x,y) = \begin{cases} \frac{1}{9}xy, & 1 \le x \le 2,\ 2 \le y \le 4, \\ 0, & \textit{для решти } x, y. \end{cases}$$

Знайти $f_\xi(x)$, $f_\eta(y)$ та перевірити на незалежність ці випадкові величини.

Р о з в' я з о к. Маємо

$$f_\eta(y) = \frac{1}{9}y\int_1^2 xdx = \frac{1}{6}y \quad \text{для} \quad 2 \le y \le 4,$$

$$f_\xi(x) = \frac{1}{9}x\int_2^4 ydy = \frac{2}{3}x \quad \text{для} \quad 1 \le x \le 2,$$

а для решти значень x, y відповідні щільності рівні нулеві. Ослільки, $f(x,y) = f_\xi(x)f_\eta(y)$, то випадкові величини ξ, η є незалежними.

Приклад 2. *Спільна щільність випадкових величин ξ, η має вигляд*

$$f(x,y) = \begin{matrix} \frac{1}{2\sqrt{xy}}, & 0 < x \le y \le 1, \\ 0, & \textit{для решти } x, y. \end{matrix}$$

Знайти $f_\xi(x)$, $f_\eta(y)$ та перевірити на незалежність ці випадкові величини.

Р о з в' я з о к. Маємо

$$f_\xi(x) = \int\limits_{-\infty}^{\infty} f(x,v)dv = \int\limits_{x}^{1} \frac{1}{2\sqrt{xv}}dv = \sqrt{\frac{v}{x}}\,\Big|_x^1 = \sqrt{\frac{1}{x}} - 1, \quad 0 < x \leq 1.$$

Аналогічно

$$f_\eta(y) = \int\limits_{-\infty}^{\infty} f(u,y)du = \int\limits_{0}^{y} \frac{1}{2\sqrt{uy}}du = \sqrt{\frac{u}{y}}\,\Big|_0^y = 1, \quad 0 < y \leq 1.$$

Оскільки, $f(x,y) \neq f_\xi(x) f_\eta(y)$, то випадкові величини ξ, η є залежними.

З виразу (3) легко випливає, що

$$\mathbf{P}(a \leq \xi_1 \leq b, c \leq \eta \leq d) = \int\limits_a^b \int\limits_c^d f(u,v)dvdu. \tag{7}$$

Узагальненяям формули (7) є наступна: для довільної борелівської множини $B \in R^2$ має місце рівність:

$$\mathbf{P}((\xi,\eta) \in B) = \int\int_B f(u,v)dvdu. \tag{8}$$

9.2.1 Двовимірний нормальний розподіл

Розгдянемо тепер один дуже важливий приклад двовимірного розподілу, який часто зустрічається в різних застосуваннях. Почнемо з наступного визначення.

Визначення 6. Будемо казати, що випадковий вектор (ξ, η) має двовимірний *нормальний розподіл*, якщо його щільність визначається формулою:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big\{-\frac{1}{2(1-\rho^2)}\Big[\frac{(x-m_1)^2}{\sigma_1^2} - \\ - 2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\Big]\Big\}, \tag{9}$$

де $-\infty < x, y < \infty,\ \sigma_1, \sigma_2 \geq 0,\ |\rho| \leq 1, -\infty < m_1, m_2 < \infty.$

Знайдемо маргінальні розподіли випадкових величин ξ і η.
Нехай

$$\frac{x-m_1}{\sigma_1} = u, \quad \frac{y-m_2}{\sigma_2} = v. \tag{10}$$

Тоді

$$\frac{(x-m_1)^2}{\sigma_1^2} - 2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2} = (v-\rho u)^2 + (1-\rho^2)u^2, \qquad (11)$$

і

$$f_\xi(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} e^{-\frac{u^2}{2}} \int_{-\infty}^{\infty} \exp \quad -\frac{1}{2(1-\rho^2)}(v-\rho u)^2 \quad dv$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{(x-m_1)^2}{\sigma_1^2}}. \qquad (12)$$

Остання рівність випливає з того, що

$$\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp \quad -\frac{1}{2(1-\rho^2)}(v-\rho u)^2 \quad dv = 1$$

оскільни підінтегральна функція є щільністтю розподілу $N(\rho u, 1-\rho^2)$. Аналогічно

$$f_\eta(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{\frac{(y-m_2)^2}{\sigma_2^2}}. \qquad (13)$$

Отже, маємо цікавий наслідок.

Наслідок 1. *Маргінальними розподілами двовимірного нормального розподілу є нормальні розподіли $N(m_1, \sigma_1), N(m_2, \sigma_2)$.*

Знайдемо тепер $\mathbf{M}[(\xi - m_1)(\eta - m_2)]$. Використовуючи заміну (10) і рівність (11) маємо:

$$\mathbf{M}[(\xi - m_1)(\eta - m_2)] = \int_{-\infty}^{\infty} (x-m_1)(y-m_2) f(x,y)dy =$$

$$= \frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} u e^{-\frac{u^2}{2}} \int_{-\infty}^{\infty} v \exp\left\{-\frac{1}{2(1-\rho^2)}(v-\rho u)^2\right\} dvdu =$$

$$= \frac{\sigma_1\sigma_2\rho}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2}} du = \sigma_1\sigma_2\rho.$$

Отже,

$$\rho = \frac{\mathbf{M}[(\xi - m_1)(\eta - m_2)]}{\sigma_1\sigma_2} = \frac{\mathbf{M}[(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta)]}{\mathbf{D}\xi\mathbf{D}\eta}.$$

Це означає, що ρ є коефіцієнтом кореляції для випадкових величин ξ, η.

Якщо ці випадкові величини незалежні, то $\rho = 0$, а з виразів (9), (12), (13) випливає, що якщо $\rho = 0$, то

$$f(x, y) = f_\xi(x) f_\eta(y).$$

Це означає, що випадкові величини ξ, η є незалежними. Отже, маємо такий результат:

Теорема 4. *Якщо випадкові величини ξ, η мають спільний нормальний розподіл, то вони є незалежними тоді і тільки тоді, коли є некорельованими.*

9.2.2 Дискретна двовимірна випадкова величина

Якщо існує сукупність чисел (скінченна або злічена) $\{x_i, x_j \in R^1, i, j = 1, 2, ...\}$ така, що

$$\sum_{ij=1}^{\infty} \mathbf{P}(\xi = x_i, \eta = y_j) = 1,$$

то будемо говорити, що двовимірна випадкова величина (ξ, η) є *дискретного* типу.

Сукупність чисел $\{p_{ij} = \mathbf{P}(\xi = x_i, \eta = y_i), i, j = 1, 2, ...\}$ називається *розподілом* двовимірної випадкової величини.

Для маргінальних розподілів тепер будемо мати:

$$\mathbf{P}(\xi = x_i) = \sum_{j=1}^{\infty} \mathbf{P}(\xi = x_i, \eta = y_j) = \sum_{j=1}^{\infty} p_{ij} \overset{def}{=} p_i,$$
$$\mathbf{P}(\eta = y_j) = \sum_{i=1}^{\infty} \mathbf{P}(\xi = x_i, \eta = y_j) = \sum_{i=1}^{\infty} p_{ij} \overset{def}{=} q_i.$$

Значення, які приймає дискретна випадкова величина, і відповідні цим значенням ймовірності зручно подавати у вигляді *кореляційної таблиці*:

$\xi \backslash \eta$	y_1	y_2	$\dots$	y_n	$\dots$	Σ
x_1	p_{11}	p_{12}	$\dots$	p_{1n}	$\dots$	p_1
x_2	p_{21}	p_{22}	$\dots$	p_{2n}	$\dots$	p_2
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_n	p_{n1}	p_{n2}	$\dots$	p_{nn}	$\dots$	p_n
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
Σ	q_1	q_2	$\dots$	p_n	$\dots$	1

Якщо випадкові величини ξ, η - незалежні, то, вочевидь,

$$p_{ij} = p_i q_j$$

для всіх $i, j = 1, 2, \ldots$.

Функцію розподілу двовимірної випадкової величини ξ, η можна записати:

$$F(x, y) = \sum_{x_i < x} \sum_{y_j < y} \mathbf{P}(\xi = x_i, \eta = y_j).$$

Як і у одновимірному випадку, для дискретної випадкової величини ξ, η більш зручним використовувати поняття розподілу, а не функції розподілу.

ЗАВДАННЯ

1. Задана функція

$$F(x, y) = \begin{cases} 1 - (1+x)^{-p} - (1+y)^{-p} + (1+x+y)^{-p}, & 0 \le x,\ 0 \le y, \\ 0, & \text{для інших } x,\ y, \end{cases}$$

   де $p > 0$. Показати, що $F(x, y)$ є абсолютно неперервного типу і знайти її щільність.

2. Випадкові величини ξ, η мають спільну щільність

$$f(x, y) = \frac{1}{\pi} \exp\{-(x^2 - 2xy + 2y^2)\}.$$

   Знайти $f_\xi(x)$, $f_\eta(x)$.

3. Двовимірна випадкова величина (ξ, η) має розподіл зі щільністю:

$$f(x, y) = \begin{cases} \frac{x+y}{2} e^{-(x+y)}, & 0 \le x,\ 0 \le y, \\ 0, & \text{для інших } x, y. \end{cases}$$

   Знайти:

   а) функцію розподілу цієї випадкової величини,

   б) щільність $f_\xi(x), f_\eta(x)$,

   в) дослідити незалежність випадкових величин ξ, η,

   г) знайти $\mathbf{P}(0 \le \xi < 1, \eta > 1)$.

4. Нехай випадковий вектор (ξ, η) має функцію розподілу:

$$F(x,y) = \begin{cases} (1-e^{-\lambda x})(1-e^{-\mu y}), & x \geq 0, y \geq 0, \\ 0, & \text{для інших } x, y. \end{cases}$$

а) Знайти $F_\xi(x)$, $F_\eta(y)$, $f_\xi(x)$, $f_\eta(y)$ і дослідити незалежність випадкових величин ξ, η.

б) Знайти $\mathbf{P}(\xi + \eta < 1/2)$.

5. Випадкова величина (ξ, η) має щільність, визначену виразом:

$$f(x,y) = \begin{cases} 4x(1-y), & 0 \leq x \leq 1,\ 0 \leq y \leq 1, \\ 0, & \text{для інших } x, y. \end{cases}$$

а) Знайти одновимірні щільності.

б) Дослідити, чи випадкові величини ξ, η є незалежними.

в) Підрахувати $\mathbf{P}(\xi + \eta < 1/2)$.

6. Визначити сталу c так, щоб функція

$$f(x,y) = \begin{cases} cxy(2-x-y), & 0 \leq x \leq 1,\ 0 \leq y \leq 1, \\ 0, & \text{для інших } x, y. \end{cases}$$

була щільністю двовимірної випадкової величини (ξ, η).

а) Знайти $\mathbf{P}(\xi + \eta \leq 1)$.

б) Дослідити, чи випадкові величини ξ, η є незалежними.

7. Визначити сталу c так, щоб задана функція була щільністю випадкової величини (ξ, η).

$$f(x,y) = \begin{cases} x(cy+1), & 0 \leq x \leq 2,\ 0 \leq y \leq 1, \\ 0, & \text{для інших } x, y. \end{cases}$$

а) Знайти функцію розподілу $F(x,y)$.

б) Підрахувати ймовірність $\mathbf{P}(1 \leq \xi, \eta \leq 1/2)$.

8. Двовимірна випадкова величина має розподіл зі щільністю:

$$f(x,y) = \frac{1}{2\pi(1+x^2+y^2)^{\frac{3}{2}}}.$$

Знайти маргінальні та умовні щільності.

9. Нехай F, G є одновимірними функціями розподілу з щільностями f, g. Довести, що для $|\alpha| \le 1$

$$U(x, y) = F(x)G(y)[1 + \alpha(1 - F(x))(1 - G(y))]$$

є двовимірною функцією розподілу з маргінальними щільностями f і g.

10. Виймаємо одну карту з колоди 24 карт.Визначимо через ξ випадкову величину, яка приймає значення 0, якщо карта є трефою, значення 1, якщо є бубною, значення 2 та 3 якщо є відповідно червою та пікою. Позначимо через η випадкову величину яка приймає значення 5, 4, 3 у випадках відповаідно туза, короля, дами і значення 0 для решти випадків.

    а) Знайти розподіл випадкової величини (ξ, η), а також маргінальні розподіли.

    б) Дослідити ці випадкові величини на незалежність.

    в) Підрахувати ймовірність $\mathbf{P}(\xi \ge \eta)$.

Лекція 10. Умовні розподіли

10.1 Умовна функція розподілу двох неперервних випадкових величин

При дослідженні двох випадкових величин ξ, η, які є, взагалі кажучи, залежними часто виникає потреба знаходження ймовірностей подій, які стосуються однієї випадкової величини при фіксованому значенні іншої. Нехай, наприклад, ми хочемо знайти розподіл випадкової величини ξ за умови, що $\eta = y$, тобто $\mathbf{P}(\xi < x/\eta = y)$. Якщо $\mathbf{P}(\eta = y) > 0$, то згідно з визначенням умовної ймовірності (формула (1), стор. 22) будемо мати:

$$\mathbf{P}(\xi < x/\eta = y) = \frac{\mathbf{P}(\xi < x, \eta = y)}{\mathbf{P}(\eta = y)}. \tag{1}$$

Але, якщо $\mathbf{P}(\eta = y) = 0$, а так буде, коли функція розподілу випадкової величини η є неперервною в точці y, то ми не можемо використати вираз (1). Як тоді підрахувати ймовірність у лівій частині (1)? Можемо зробити наступним чином. Нехай $f(x, y)$ є щільністю випадкових величин ξ, η. Припустимо додатково, що функція $f(x, y)$ є неперервною стосовно своїх аргументів, і нехай y буде таким, що $f_\eta(y) > 0$. Тоді для всіх $\varepsilon > 0$ маємо:

$$\mathbf{P}(y - \varepsilon \le \eta \le y + \varepsilon) = \int\limits_{y-\varepsilon}^{y+\varepsilon} f_\eta(v)dv > 0.$$

Логічно прийняти, що $\lim_{\varepsilon\downarrow 0} \mathbf{P}(\xi \in A/y-\varepsilon \le \eta \le y+\varepsilon) = \mathbf{P}(\xi \in A/\eta = y)$.

Тому

$$\mathbf{P}(\xi < x/\eta = y) = \lim_{\varepsilon\downarrow 0} \mathbf{P}(\xi < x/y - \varepsilon \le \eta \le y + \varepsilon) =$$

$$= \lim_{\varepsilon\downarrow 0} \frac{\mathbf{P}(\xi < x, y - \varepsilon \le \eta \le y + \varepsilon)}{\mathbf{P}(y - \varepsilon \le \eta \le y + \varepsilon)} = \lim_{\varepsilon\downarrow 0} \frac{\int_{-\infty}^{x} \int_{y-\varepsilon}^{y+\varepsilon} f(u, v)dvdu}{\int_{y-\varepsilon}^{y+\varepsilon} f_\eta(v)dv} =$$

$$= \lim_{\varepsilon\downarrow 0} \frac{\int_{-\infty}^{x} \varepsilon^{-1} \int_{y-\varepsilon}^{y+\varepsilon} f(u, v)dvdu}{\varepsilon^{-1} \int_{y-\varepsilon}^{y+\varepsilon} f_\eta(v)dv} = \frac{\int_{-\infty}^{x} f(u, y)du}{f_\eta(y)}.$$

Отже, за наших припущень стосовно $f(x, y)$ можемо записати:

$$\mathbf{P}(\xi < x/\eta = y) = \frac{\int_{-\infty}^{x} f(u, y)du}{f_\eta(y)}.$$

Якщо тепер $N_\eta = \{y : f_\eta(y) = 0\}$, тоді

$$\mathbf{P}(\eta \in N_\eta) = \int\limits_{N_\eta} f_\eta(y)dy = 0.$$

Звідки випливає, що подія $\{\eta = y\}$ має ймовірність нуль для всіх y, для яких $f_\eta(y) = 0$. Тобто на практиці подія $\{\eta = y\}$ не відбуваається, а тому значення $\mathbf{P}(\xi < x/\eta = y)$ для $y \in N_\eta$ не є для нас істотним і ми можемо визначити їх довільним чином, наприклад, як нуль.

Наведені міркування є підставою для наступного визначення.

Визначення 1. Нехай функція розподілу випадкового вектора (ξ, η) має щільність $f(x, y)$. Тоді *умовну функцію розподілу* випадкової величини ξ за умови $\eta = y$ будемо позначати $F_{\xi/\eta}(x/y)$ (або просто $F(x/y)$) і визначати наступним чином:

$$F_{\xi/\eta}(x/y) = \mathbf{P}(\xi < x/\eta = y) \overset{def}{=} \begin{cases} \int\limits_{-\infty}^{x} \frac{f(u,y)}{f_\eta(y)}du, & \text{якщо} \quad f_\eta(y) > 0, \\ \\ 0, & \text{якщо} \quad f_\eta(y) = 0. \end{cases} \tag{2}$$

З (2) випливає, що

$$F_{\xi/\eta}(x/y) = \int\limits_{-\infty}^{x} f_{\xi/\eta}(u/y)du,$$

де функція

$$f_{\xi/\eta}(x/y) \overset{def}{=} \begin{cases} \frac{f(x,y)}{f_\eta(y)}, & \text{якщо} \quad f_\eta(y) > 0, \\ \\ 0, & \text{якщо} \quad f_\eta(y) = 0. \end{cases} \tag{3}$$

називається *умовною щільністю* випадкової величини ξ, за умови $\eta = y$.

Зауваження 1. Як і для умовної функції розподілу, для щільності замість $f_{\xi/\eta}(x/y)$ зазвичай пишемо $f(x/y)$.

З (3) легко випливає, що

$$f_{\xi/\eta}(x/y) = \frac{f_{\eta/\xi}(y/x) f_\xi(x)}{f_\eta(y)},$$

і це є аналогом формули Байєса (вираз (4), стор. 24)

Можемо тепер вивести „неперервний" аналог формули повної ймовірності.

Теорема 1. *Нехай $F_\xi(x)$, $F_\eta(x)$ будуть функціями розподілу випадкових величин ξ, η. Має місце наступна формула*

$$F_\xi(x) = \mathbf{P}(\xi < x) = \int\limits_{-\infty}^{\infty} \mathbf{P}(\xi < x/\eta = y) dF_\eta(y). \tag{4}$$

Формула (4) називається *формулою повної ймовірності.*

Якщо випадкова величина η має щільність $f_\eta(y)$, то

$$F_\xi(x) = \int\limits_{-\infty}^{\infty} \mathbf{P}(\xi < x/\eta = y) f_\eta(y) dy. \tag{5}$$

Д о в е д е н н я. Доведемо тільки (5). Нехай $f_\eta(y) > 0$, тоді з (2) випливає

$$f_\eta(y)\mathbf{P}(\xi < x/\eta = y) = \int\limits_{-\infty}^{x} f(u, y) du. \tag{6}$$

А якщо $f_\eta(y) = 0$, то

$$\int\limits_{-\infty}^{x} f(u, y) du \le \int\limits_{-\infty}^{\infty} f(u, y) du = f_\eta(y) = 0.$$

і, отже, $\int_{-\infty}^{x} f(u, y) du = 0$ для всіх x, а це означає, що рівність (6) виконується, і якщо $f_\eta(y) = 0$. Тому можемо вважати, що ця рівність виконується для всіх y. Інтегруючи її по y від $-\infty$ до $+\infty$, отримаємо:

$$\int\limits_{-\infty}^{\infty} f_\eta(y)\mathbf{P}(\xi < x/\eta = y) dy = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{\infty} f(u, y) dy du = \int\limits_{-\infty}^{x} f_\xi(u) du = F_\xi(x).$$

◄

Зауваження 2. Вираз (4) можна записати в рівнозначному вигляді:

$$\mathbf{P}(\xi \in B) = \int\limits_{-\infty}^{\infty} \mathbf{P}(\xi \in B/\eta = y) f_\eta(y) dy \tag{7}$$

для кожної борелівської множини B.

Визначення 2. Умовне математичне сподівання для ξ при умові $\eta = y$ визначається формулою

$$\mathbf{M}\{\xi/\eta = y\} = \int\limits_{-\infty}^{\infty} x dF(x/y) = \int\limits_{-\infty}^{\infty} x f(x/y) dx$$

при умові, що інтеграли в цій формулі є абсолютно збіжними.

Має місце наступна формула, яку можна розглядати як аналог формули повної ймовірності для математичного сподівання.

$$\mathbf{M}\xi = \int\limits_{-\infty}^{\infty} \mathbf{M}\{\xi/\eta = y\} dF_\eta(y) = \int\limits_{-\infty}^{\infty} \mathbf{M}\{\xi/\eta = y\} f_\eta(y) dy.$$

10.2 Сума, частка, добуток випадкових величин

Розглянемо тепер застосування виразу (4) для знаходження функції розподілу випадкових величин $\zeta = \xi + \eta$, $\theta = \frac{\xi}{\eta}$, $\tau = \xi\eta$ у випадку, коли випадкові величини ξ, η є незалежними.

Теорема 2. *Нехай функції $F_\xi(x), F_\eta(x)$ є відповідно функціями розподілу незалежних випадкових величин ξ, η. Тоді функція розподілу випадкової величини $\zeta = \xi + \eta$ може бути представлена у вигляді:*

$$F_{\xi+\eta}(x) = \int\limits_{-\infty}^{\infty} F_\xi(x-y) dF_\eta(y) = \int\limits_{-\infty}^{\infty} F_\eta(x-y) dF_\xi(y), \tag{8}$$

а якщо випадкові величини ξ, η мають щільності $f_\xi(x)$, $f_\eta(x)$, то

$$F_{\xi+\eta}(x) = \int\limits_{-\infty}^{\infty} F_\xi(x-y) f_\eta(y) dy = \int\limits_{-\infty}^{\infty} F_\eta(x-y) f_\xi(y) dy, \tag{9}$$

$$f_{\xi+\eta}(x) = \int\limits_{-\infty}^{\infty} f_\xi(x-y) f_\eta(y) dy = \int\limits_{-\infty}^{\infty} f_\eta(x-y) f_\xi(y) dy. \tag{10}$$

Д о в е д е н н я. На підставі (4) можемо записати:

$$F_{\xi+\eta}(x) = \mathbf{P}(\xi+\eta < x) = \int\limits_{-\infty}^{\infty} \mathbf{P}(\xi+\eta < x/\eta = y) dF_\eta(y) = \int\limits_{-\infty}^{\infty} \mathbf{P}(\xi < x - y/\eta = y) dF_\eta(y)$$

$$= (\text{позаяк } \xi, \eta \text{ є незалежними}) \int\limits_{-\infty}^{\infty} \mathbf{P}(\xi < x - y) dF_\eta(y) = \int\limits_{-\infty}^{\infty} F_\xi(x-y) dF_\eta(y).$$

Формула (9) випливає з (8), а після диференціонування в (9) отримаємо (10). ◄

Інтеграли у формулі (8) мають спеціальну назву.

Визначення 3. Функція

$$G(x) = \int\limits_{-\infty}^{\infty} G_2(x-y)dG_1(y) = \int\limits_{-\infty}^{\infty} G_1(x-y)dG_2(y)$$

називається *згорткою* функцій $G_1(x)$, $G_2(x)$.

Отже, функція розподілу суми двох незалежних випадкових величин є згорткою їх функцій розподілу.

Зауваження 3. Якщо випадкові величини ξ, η приймають тільки цілі значення з ймовірностями p_n, q_n, $n = 0, \pm 1, \pm 2, \ldots$. Це означає, що $\mathbf{P}(\xi = n) = p_n$, $\mathbf{P}(\eta = n) = q_n$, $n = 0, \pm 1, \pm 2, \ldots$, тоді аналогом (8) є формула

$$\mathbf{P}(\xi + \eta = n) = \sum_{i=-\infty}^{\infty} q_i p_{n-i}.$$

Сума у правій частині останньої рівності називається також згорткою, а повне визначення виглядає наступним чином:

Визначення 4. Нехай послідовності чисел a_i, b_i, $i = 0, \pm 1, \pm 2, \ldots$ є такими, що

$$\sum_{i=-\infty}^{\infty} |a_i| < \infty, \quad \sum_{i=-\infty}^{\infty} |b_i| < \infty,$$

тоді послідовність c_i, $i = 0, \pm 1, \pm 2, \ldots$, яка визначена наступним чином:

$$c_n = \sum_{i=-\infty}^{\infty} a_{n-i} b_i = \sum_{i=-\infty}^{\infty} a_i b_{n-i}$$

називається *згорткою* послідовностей a_i, b_i.

З теореми 2 випливає дуже важливий висновок. Припустимо, що одна з випадкових величин ξ, η має щільність. Нехай, наприклад, це буде ξ. Через $f_\xi(x)$ позначимо щільність її функції розподілу, і нехай $\zeta = \xi + \eta$. Тоді на підставі виразу (8) маємо:

$$F_{\xi+\eta}(x) = \int\limits_{-\infty}^{\infty} F_\xi(x-y)dF_\eta(y) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{x-y} f_\xi(u)du dF_\eta(y) =$$
$$= \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{\infty} f_\xi(u-y)dF_\eta(y)du.$$

Звідки випливає, що функція розподілу $F_{\xi+\eta}(x)$ має щільність, яку можна представити виразом:

$$f_{\xi+\eta}(x) = \int\limits_{-\infty}^{\infty} f_\xi(x-y)dF_\eta(y) \tag{11}$$

Отже, маємо наступний результат.

Наслідок 1. *Якщо принаймні одна з незалежних випадкових величин є абсолютно неперервною, то сума таких випадкових величин є також абсолютно неперервною.*

Розглянемо тепер частку і добуток випадкових величин ξ, η, але тільки в ситуації, коли ці випадкові величини мають щільність.

Теорема 3. *Нехай функції $f_\xi(x)$, $f_\eta(x)$ є щільностями незалежних випадкових величин ξ, η. Тоді для щільності випадкової величини $\theta = \xi/\eta$ має місце формула:*

$$f_\theta(x) = \int\limits_{-\infty}^{\infty} |y| f_\xi(xy) f_\eta(y)dy.$$

Теорема 4. *Нехай функції $f_\xi(x)$, $f_\eta(x)$ є щільностями незалежних випадкових величин ξ, η. Тоді для щільності випадкової величини $\theta = \xi\eta$ має місце формула:*

$$f_\theta(x) = \int\limits_{-\infty}^{\infty} |y|^{-1} f_\xi(xy^{-1}) f_\eta(y)dy.$$

Доведення цих тверджень залишаємо читачеві.

ЗАВДАННЯ

1. Спільна щільність випадкових величин ξ, η задана наступним чином:
$$f(x,y) = \begin{cases} \frac{1}{9}xy, & 1 \le x \le 2,\ 2 \le y \le 4, \\ 0, & \text{для інших } x, y. \end{cases}$$
Знайти $f(x/y)$ і $f(y/x)$, а також $F(x/y)$, $F(y/x)$.

2. Для щільності
$$f(x,y) = \begin{cases} \frac{1}{2\sqrt{xy}}, & 0 < x \le y \le 1, \\ 0, & \text{для інших } x, y \end{cases}$$
Знайти $f(x/y)$, а також $f(y/x)$.

3. Спільна щільність випадкових величин ξ, η є визначеною виразом:
$$f(x,y) = \begin{cases} \frac{x+y}{2}e^{-(x+y)}, & x > 0,\ y > 0, \\ 0, & \text{для інших } x, y. \end{cases}$$
а) Знайти умовні розподіли.

б) Використовуючи формулу повної ймовірності підрахувати $\mathbf{P}(\xi - \eta < 1)$.

4. Двовимірна випадкова величина має щільність:
$$f(x,y) = \frac{1}{\pi}e^{-\frac{1}{2}(x^2+2xy+5y^2)}.$$
Знайти маргінальні та умовні щільності.

5. Нехай випадкові величини ξ, η мають спільний нормальний розподіл вигляду (9), стор.90 . Довести, що:
$$f(x/y) = \frac{1}{\sigma_1\sqrt{2\pi(1-\rho^2)}}\exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x - m_1 - \rho\frac{\sigma_1}{\sigma_2}(y - m_2)\right]^2\right\}.$$

6. Випадкові величини ξ, η є незалежними і мають показниковий розподіл з параметром $\lambda = 1$. Нехай $\eta = \xi + \eta, \zeta = \xi/(\xi + \eta)$. Знайти:

   а) щільність випадкової величини η,
   ($f_\eta(x) = xe^{-x}$, для $x > 0$ і $f_\eta(x) = 0$, для $x < 0$.)

б) щільність випадкової величини ζ,
($f_\zeta(x) = 1$, для $0 < x < 1$ і $f_\zeta(x) = 0$, для інших.)

в) чи випадкові величини η, ζ є незалежними?
(випадкові величини η, ζ є незалежними.)

7. Випадкові величини ξ, η є незалежними і $\xi, \eta \in N(0,1)$. Знайти розподіл квадрату відстані від точки $(\xi,\ \eta)$ до початку координат.

8. Випадкові величини α, β є незалежними і мають рівномірний розподіл на інтервалі (-1,1). Знайти щільність і функцію розподілу випадкової величини $\alpha + \beta$.

9. Випадкові величини ξ, η мають розподіл зі спільною щільністю:
$$f(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}.$$
Знайти розподіл випадкової величини $\xi - \eta$.

10. Знайти розподіл випадкової величини $\frac{\xi}{\eta}$, якщо ξ, η є незалежними випадковими величинами з показниковим розподілом з параметром 1.

11. Довести, що випадкові величини $\xi + \eta$, ξ/η є незалежними, якщо ξ, η є незалежними випадковими величинами з показниковим розподілом з параметром 1.

12. Довести, що у випадку, коли випадкові величини $\xi \in N(0, \sigma^2)$, $\eta \in N(0, \sigma^2)$ є незалежними, то випадкові величини $\xi^2+\eta^2$, ξ/η є також незалежними.

13. Знайти розподіл випадкової величини $\xi+\eta$, якщо ξ, η є незалежними випадковими величинами з наступними розподілами: ξ - має показниковий розподіл з параметром 1, η - рівномірний на інтервалі (0,1).

14. Випадкові величини $x_1, \ldots, \xi_n$ є незалежними і мають показниковий розподіл з параметром λ. Довести, що щільність розподілу суми $\xi + \ldots + \xi_n$ має вигляд: $\frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}$ (*розподіл Ерланга*).

15. Знайти розподіл випадкової величини $\xi + \eta$, якщо ξ та η є незалежними випадковими величинами і мають: ξ - розподіл Коші з параметрами $\lambda = 1$, $m = 0$, а η - дискретний розподіл: $\mathbf{P}(\eta = 0) = 1/3$, $\mathbf{P}(\eta = 1) = 2/3$.

16. (Задача М. Ядренко) Випадкові величини ξ, η - незалежні і мають розподіл зі щільністю:

$$f_\xi(x) = f_\eta(x) = \frac{C}{1+x^4}.$$

Знайти C і довести, що випадкова величина ξ/η має розподіл Коші.

17. (Задача М. Ядренко) Випадкові величини ξ, η є незалежними і мають розподіл зі щільністю:

$$f_\xi(x) = \frac{1}{\pi\sqrt{1-x^2}}, \; |x| < 1, \quad f_\eta(x) = \begin{cases} 0, & x \leq 0 \\ xe^{-x^2/2}, & x > 0. \end{cases}$$

Довести, що випадкова величина $\xi\eta$ має нормальний розподіл.

18. Випадкові величини ξ, η є незалежними і мають показниковий розподіл з параметром λ . Знайти функцію розподілу випадкової величини $\frac{\xi+\eta}{\xi}$.

19. Нехай ξ, η –незалежні випадкові величини і $\tau = \xi + \eta$. Довести, що:

$$F_\tau(x) = \int_{-\infty}^{\infty} F_\xi(x+v)dF_\eta(v) = \int_{-\infty}^{\infty} \big(1 - F_\eta(v-x)\big)dF_\xi(v)$$

20. Випадкові величини ξ, η мають спільний нормальний розподіл і $\mathbf{M}\xi = \mathbf{M}\eta = 0$, $\mathbf{D^2}\xi = \mathbf{D^2}\eta = 1$, $\mathbf{M}\{\xi\eta\} = 1/2$.

Знайти: а) $\mathbf{D^2}\{\eta\xi\}$, б) $\mathbf{D^2}\{\eta^2\xi^2\}$.

Лекція 11. Збіжність випадкових величин

Нехай на ймовірносному просторі $(\Omega, \mathfrak{B}, \mathbf{P})$ задана послідовність випадкових величин ξ_n, $n = 1, 2 \ldots$.

11.1 Збіжність випадкових величин

11.1.1 Збіжність з ймовірністю 1

Визначення 1. Послідовність випадкових величин $\xi_n, n = 1, 2 \ldots$ називають *збіжною з ймовірністю* 1 (або *майже напевно*) до випадкової величини ξ, якщо

$$\lim_{n\to\infty} \xi_n(\omega) = \xi(\omega) \tag{1}$$

для всіх $\omega \in \Omega$ за вийнятком, може бути, множини $B \subset \Omega$ з нульовою ймовірністю: $\mathbf{P}(B) = 0$.

Збіжність з ймовірністю 1 будемо позначати наступним чином:

$$\lim_{n\to\infty} \xi_n = \xi \text{ м.н}, \qquad \xi_n \xrightarrow[n\to\infty]{\text{м.н}} \xi \quad \text{або ще так: } \mathbf{P}(\lim_{n\to\infty} \xi_n = \xi) = 1.$$

Нехай $A = \{\omega : \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\}$. Іншими словами, для всіх $\omega \in A$ границя $\lim_{n\to\infty} \xi_n(\omega)$ існує і дорівнює $\xi(\omega)$. Зрозуміло, що якщо $\mathbf{P}(A) = 1$, то $\xi_n \xrightarrow[n\to\infty]{\text{м.н}} \xi$. Доведемо, що

$$A = \cap_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} \{\omega : |\xi_n(\omega) - \xi(\omega)| \leq \frac{1}{k}\}. \tag{2}$$

Для цього скористаюмось з наступного факту. Числова послідовність c_n збігається до c при $n \to \infty$ тоді і тільки тоді, коли для довільного натурального k існує таке $N(k) \geq 1$, що $|c_n - c| \leq 1/k$, для всіх $n \geq N(k)$. Тепер будемо мати:

$$\begin{aligned}
&\omega \in \cap_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} \{\omega : |\xi_n(\omega) - \xi(\omega)| \leq 1/k\} \Longleftrightarrow \\
&\Longleftrightarrow \forall\ k \geq 1, \omega \in \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} \{\omega : |\xi_n(\omega) - \xi(\omega)| \leq 1/k\} \Longleftrightarrow \\
&\Longleftrightarrow \forall\ k \geq 1, \exists\ N(k) \geq 1, \text{ що } \omega \in \cap_{n=N(k)}^{\infty} \{\omega : |\xi_n(\omega) - \xi(\omega)| \leq 1/k\} \Longleftrightarrow \\
&\Longleftrightarrow \forall\ k \geq 1, \exists\ N(k) \geq 1, \text{ що } |\xi_n(\omega) - \xi(\omega)| \leq 1/k, \forall\ n \geq N(k) \Longleftrightarrow \\
&\Longleftrightarrow \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega) \Longleftrightarrow \omega \in A.
\end{aligned}$$

Позаяк, як це неважко довести, рівності $\mathbf{P}(A_k) = 1, k \geq 1$ і $\mathbf{P}(\cap_{k=1}^{\infty} A_k) = 1$ є рівнозначними, то тепер з (2) отримаємо, що послідовність випадкових величин ξ_n, $n \geq 1$ є збіжною з ймовірністю 1 до випадкової величини ξ тоді і тільки тоді, коли для довільного $k \geq 1$ маємо:

$$\mathbf{P}(\cup_{m=1}^{\infty} \cap_{n=m}^{\infty} \{\omega : |\xi_n(\omega) - \xi(\omega)| \leq 1/k\} = 1 \qquad (3)$$

В останній рівності $1/k$ можемо замінити довільним $\varepsilon > 0$. А, позаяк, послідовність подій $\cap_{n=m}^{\infty}\{\omega : |\xi_n(\omega) - \xi(\omega)| \leq \varepsilon\}$, $m \geq 1$ є зростаючою, то, беручи це до уваги, а також теорему 2 (стор. 36) і рівність (3) маємо такий наслідок.

Наслідок 1. *Послідовність випадкових величин ξ_n, $n \geq 1$ збігається з ймовірністю 1 до випадкової величини ξ тоді і тільки тоді, коли для довільного $\varepsilon > 0$ маємо:*

$$\lim_{m\to\infty} \mathbf{P}(\cap_{n=m}^{\infty}\{\omega : |\xi_n(\omega) - \xi(\omega)| \leq \varepsilon\} = 1. \qquad (4)$$

11.1.2 Збіжність за ймовірністю.

Визначення 2. Послідовність випадкових величин ξ_n, $n = 1, 2\ldots$ називають *збіжною за ймовірністю* до випадкової величини ξ, якщо для кожного $\varepsilon > 0$ виконується умова:

$$\lim_{n\to\infty} \mathbf{P}(|\xi_n - \xi| > \varepsilon) = 0. \qquad (5)$$

Збіжність за ймовірністю будемо позначати наступним чином: $\xi_n \xrightarrow[n\to\infty]{\text{p}} \xi$.

Якщо послідовність випадкових величин ξ_n, $n \geq 1$ є збіжною з ймовірністю 1 до випадкової величини ξ, тоді для довільного $\varepsilon > 0$ з (4) маємо:

$$\lim_{n\to\infty} \mathbf{P}(|\xi_n - \xi| > \varepsilon) = 1 - \lim_{n\to\infty} \mathbf{P}(|\xi_n - \xi| \leq \varepsilon) \leq$$
$$\leq 1 - \lim_{n\to\infty} \mathbf{P}(\cap_{k=n}^{\infty}\{\omega : |\xi_k(\omega) - \xi(\omega)| \leq \varepsilon\} = 0,$$

а, отже, зі збіжності $\xi_n \xrightarrow[n\to\infty]{\text{м.н}} \xi$ випливає збіжність $\xi_n \xrightarrow[n\to\infty]{\text{p}} \xi$. Зворотній факт не має місця (завдання 2), але є справедливе наступне твердження.

Теорема 1. *Якщо $\xi_n \xrightarrow[n\to\infty]{\text{p}} \xi$, то існує послідовність $n_k \to \infty$, така що $\xi_{n_k} \xrightarrow[n_k\to\infty]{\text{м.н}} \xi$.*

Для доведення цього твердження нам буде потрібним результат, який часто використовують в теорії ймовірностей і теорії випадкових процесів.

Лема 1. (Бореля-Кантеллі) *Нехай A_n, $n = 1, 2, \dots$ є послідовністю подій, а B - подія, яка означає, що відбулась нескінченна кількість подій з A_n.*

1) *Для того, щоб серед подій A_n відбулась їх скінченна кількість (тобто $\mathbf{P}(B) = 0$) достатньо, аби*

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty.$$

2) *Якщо події A_n є незалежними, то серед A_n відбудеться скінченна кількість подій тоді і тільки тоді, коли*

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty.$$

Д о в е д е н н я. Насамперед доведемо, що

$$B = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k. \tag{6}$$

Маємо:

$$\omega \in B \Longleftrightarrow \exists\ n_k, k \geq 1, \text{ що } \omega \in A_{n_k} \Longleftrightarrow \omega \in \cup_{k=n}^{\infty} A_k, \forall\ n \geq 1 \ \Longleftrightarrow \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k$$

Доведемо тепер пункт 1) тези. Для кожного $n \geq 1$ з (6) маємо $B \subset \cup_{k=n}^{\infty} A_k$, і, крім того,

$$\mathbf{P}(B) \leq \mathbf{P}(\cup_{k=n}^{\infty} A_k) \leq \sum_{k=n}^{\infty} \mathbf{P}(A_k) \underset{n \to \infty}{\longrightarrow} 0.$$

Отже, $\mathbf{P}(B) = 0$, що й треба було довести.

Д о в е д е н н я пункту 2). Достатність випливає з пункту 1), і треба довести тільки необхідність.

Завдяки припущенням можемо записати:

$$0 = \mathbf{P}(B) = \mathbf{P}(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k) = \lim_{n \to \infty} \mathbf{P}(\cup_{k=n}^{\infty} A_k),$$

а отже, для деякого N маємо: $\mathbf{P}(\cup_{k=N}^{\infty} A_k) < 1$ або $\mathbf{P}(\cap_{k=N}^{\infty} \bar{A}_k) > 0$. В наступних співвідношеннях використовується нерівність $\ln(1-x) \leq -x$, $|x| < 1$. Маємо:

$$0 < \mathbf{P}(\cap_{k=N}^{\infty} \bar{A}_k) = \lim_{n \to \infty} \mathbf{P}(\cap_{k=N}^{n} \bar{A}_k) = \lim_{n \to \infty} \Pi_{k=N}^{n} \mathbf{P}(\bar{A}_k) = \lim_{n \to \infty} e^{\ln\left(\Pi_{k=N}^{n} \mathbf{P}(\bar{A}_k)\right)} =$$
$$= \lim_{n \to \infty} e^{\sum_{k=N}^{n} \ln(1-\mathbf{P}(A_k))} \leq \lim_{n \to \infty} e^{-\sum_{k=N}^{n} \mathbf{P}(A_k)} = e^{-\sum_{k=N}^{\infty} \mathbf{P}(A_k)}.$$

Звідки випливає, що $\sum_{k=N}^{\infty} \mathbf{P}(A_k) < \infty$. Отже, послідовність $\sum_{k=1}^{\infty} \mathbf{P}(A_k)$ є збіжною.

Тепер легко довести теорему 1. Позаяк $\xi_n \xrightarrow[n\to\infty]{\mathrm{P}} \xi$, то для кожного k існує таке натуральне n_k, що

$$\mathbf{P}(|\xi_{n_k} - \xi| > \frac{1}{k}) \le \frac{1}{k^2}.$$

Тоді

$$\sum_{k=1}^{\infty} \mathbf{P}(|\xi_{n_k} - \xi| > \frac{1}{k}) \le \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty,$$

і на підставі леми 1 відбувається тільки скінченне число подій $|\xi_{n_k} - \xi| > \frac{1}{k}$, а це, власне, й означає, що існує таке N (взагалі кажучи, випадкове), що

$$\mathbf{P}(|\xi_{n_k} - \xi| \le \frac{1}{k}) = 1$$

для всіх $k \ge N$, що і означає, що $\lim_{k\to\infty} \xi_{n_k} = \xi$ м.н. ◄

Наступне тведження наводить головні властивості збіжності за ймовірністю.

Теорема 2. *Нехай* $\xi_n \xrightarrow[n\to\infty]{\mathrm{P}} \xi$ *тоді:*

1) *якщо функція* $f(x)$ *є неперервною, то*

$$f(\xi_n) \xrightarrow[n\to\infty]{\mathrm{P}} f(\xi);$$

2) *якщо для фіксованого* $C > 0$ $\mathbf{P}(|\xi_n| \le C) = 1$, $n = 1, 2, \ldots,$ *то*

$$\lim_{n\to\infty} \mathbf{M}\xi_n = \mathbf{M}\xi; \tag{7}$$

3) *якщо для деякого* $\alpha > 0$ $\sup_n \mathbf{M}|\xi_n|^{1+\alpha} < \infty$, *то* $\lim_{n\to\infty} \mathbf{M}\xi_n = \mathbf{M}\xi$.

Доведення наведемо тільки для перших двох пунктів. Доведення третього можна знайти в [**?**] в підрозділі 2, стор.132.

1). Нехай $\varepsilon > 0$ буде усталеним малим числом, і $0 < N < \infty$ є вибраним так, що $\mathbf{P}(\xi \notin [-N, N]) = \mathbf{P}(|\xi| > N) \le \varepsilon$. Позаяк функція $f(x)$ є неперервною, то вона є рівномірно неперервною на інтервалі $[-2N, 2N]$, а, отже, існує $\delta \in (0, N)$ таке, що

$$|f(x) - f(y)| \le \varepsilon \qquad \text{якщо} \quad |x - y| \le \delta, \qquad x, y \in [-2N, 2N]. \tag{8}$$

Тепер будемо мати:

$$\begin{aligned} &\mathbf{P}(|f(\xi_n) - f(\xi)| > \varepsilon) = \mathbf{P}(|f(\xi_n) - f(\xi)| > \varepsilon, |\xi| \le N) + \mathbf{P}(|f(\xi_n) - f(\xi)| > \varepsilon, |\xi| > N) \le \\ &\le \mathbf{P}(|f(\xi_n) - f(\xi)| > \varepsilon, |\xi_n - \xi| \le \delta, |\xi| \le N) + \mathbf{P}(|f(\xi_n) - f(\xi)| > \varepsilon, |\xi_n - \xi| > \delta, |\xi| \le N) + \varepsilon \le \\ &\le \mathbf{P}(|f(\xi_n) - f(\xi)| > \varepsilon, |\xi_n - \xi| \le \delta, |\xi| \le N) + \mathbf{P}(|\xi_n - \xi| > \delta) + \varepsilon. \end{aligned} \tag{9}$$

З того, що $|\xi_n - \xi| \le \delta, |\xi| \le N)| \le N$ і $0 < \delta < N$ випливає $\xi, \xi_n \in [-2N, 2N]$, а, отже, згідно з (8) перший доданок в (9) дорівнює нулеві, і маємо наступну нерівність:

$$\mathbf{P}(|f(\xi_n) - f(\xi)| > \varepsilon) \le \mathbf{P}(|\xi_n - \xi| > \delta) + \varepsilon. \tag{10}$$

Позаяк $\lim_{n\to\infty} \mathbf{P}(|\xi_n - \xi| > \delta) = 0$, то існує натуральне число $n_0(\varepsilon)$ таке, що $\mathbf{P}(|\xi_n - \xi| > \delta) \le \varepsilon$ для всіх $n \ge n_0(\varepsilon)$, і тепер з (10) маємо:

$$\mathbf{P}(|f(\xi_n) - f(\xi)| > \varepsilon) \le 2\varepsilon \qquad \text{якщо} \quad n \ge n_0(\varepsilon),$$

що доводить перший пункт.

2). Якщо $\mathbf{P}(|\xi_n| \le C) = 1$ і $\xi_n \xrightarrow[n\to\infty]{\text{P}} \xi$, to $\mathbf{P}(|\xi| < 2C) = 1$. Дійсно, маємо:

$$\mathbf{P}(|\xi| \ge 2C) = \mathbf{P}(|\xi - \xi_n + \xi_n| \ge 2C) \le \mathbf{P}(|\xi - \xi_n| + |\xi_n| \ge 2C) \le \mathbf{P}(|\xi - \xi_n| \ge C) \xrightarrow[n\to\infty]{} 0.$$

Звідки $\mathbf{P}(|\xi| \ge 2C) = 0$, або $\mathbf{P}(|\xi| < 2C) = 1$.

Тепер з ймовірністю 1 маємо: $|\xi_n - \xi| \le |\xi_n| + |\xi| \le 3C$ і, отже, для довільного $\varepsilon > 0$ отримаємо:

$$\begin{aligned}|\mathbf{M}\xi_n - \mathbf{M}\xi| &\le \mathbf{M}|\xi_n - \xi| = \mathbf{M}|\xi_n - \xi|\chi\{|\xi_n - \xi| > \varepsilon\} + \mathbf{M}|\xi_n - \xi|\chi\{|\xi_n - \xi| \le \varepsilon\} \\ &\le 3C\mathbf{M}\chi\{|\xi_n - \xi| > \varepsilon\} + \varepsilon\mathbf{M}\chi\{|\xi_n - \xi| \le \varepsilon\} \le 3C\mathbf{P}(|\xi_n - \xi| > \varepsilon) + \varepsilon.\end{aligned}$$

Таким чином,

$$\overline{\lim}_{n\to\infty} |\mathbf{M}\xi_n - \mathbf{M}\xi| \le \varepsilon. \tag{11}$$

Позаяк в (11) ε є довільним, то $\lim_{n\to\infty} \mathbf{M}\xi_n = \mathbf{M}\xi$, і (7) є доведеним.

11.1.3 Збіжність в середньому і в середньому квадратичному.

Визначення 3. Послідовність випадкових величин ξ_n, $n = 1, 2 \ldots$ називають *збіжною в середньому* до випадкової величини ξ, якщо $\mathbf{M}|\xi_n| < \infty$, $n = 1, 2, \ldots$, $\mathbf{M}|\xi| < \infty$, і

$$\lim_{n\to\infty} \mathbf{M}|\xi_n - \xi| = 0. \tag{12}$$

Збіжність в середньому будемо позначати наступним чином: $\xi_n \xrightarrow[n\to\infty]{\text{с}} \xi$.

Визначення 4. Послідовність випадкових величин ξ_n, $n = 1, 2 \ldots$ називають *збіжною в середньому квадратичному* до випадкової величини ξ, якщо $\mathbf{M}\xi_n^2 < \infty$, $n = 1, 2, \ldots$,$\mathbf{M}\xi^2 < \infty$, і $\lim_{n\to\infty} \mathbf{M}(\xi_n - \xi)^2 = 0$.

Збіжність у середньому квадратичному позначають так: $\xi_n \xrightarrow[n\to\infty]{\text{с.к}} \xi$.

Зауваження 1. Якщо $\xi_n \xrightarrow[n\to\infty]{} \xi$ з ймовірністю 1 або за ймовірністю, то з цього ще не випливає рівність $\lim_{n\to\infty} \mathbf{M}\xi_n = \mathbf{M}\xi$, але, якщо $\lim_{n\to\infty} \xi_n = \xi$ в середньому, то $\lim_{n\to\infty} \mathbf{M}\xi_n = \mathbf{M}\xi$, що випливає з

$$|\mathbf{M}\xi_n - \mathbf{M}\xi| = |\mathbf{M}(\xi_n - \xi)| \le \mathbf{M}|\xi_n - \xi| \xrightarrow[n\to\infty]{} 0.$$

Наступне твердження наводить залежність між різними типами збіжностей.

Теорема 3.

1. *Якщо* $\xi_n \xrightarrow[n\to\infty]{\text{м.н}} \xi$, *то* $\xi_n \xrightarrow[n\to\infty]{\text{р}} \xi$.
2. *Якщо* $\xi_n \xrightarrow[n\to\infty]{\text{с.к}} \xi$, *то* $\xi_n \xrightarrow[n\to\infty]{\text{с}} \xi$.
3. *Якщо* $\xi_n \xrightarrow[n\to\infty]{\text{с}} \xi$, *то* $\xi_n \xrightarrow[n\to\infty]{\text{р}} \xi$.

Д о в е д е н н я. Пункт 1 був вже доведений в підрозділі 11.1.2 вище. Пункт 2 випливає очевидним чином з нерівності

$$(\mathbf{M}|\xi_n - \xi|)^2 \le \mathbf{M}|\xi_n - \xi|^2,$$

а, користуючись нерівністю Маркова для кожного $\varepsilon > 0$, маємо:

$$\mathbf{P}(|\xi_n - \xi| > \varepsilon) \le \frac{\mathbf{M}|\xi_n - \xi|}{\varepsilon} \underset{n\to\infty}{\longrightarrow} 0.$$

Звідки випливає доведення пункту 3. ◀

Для числових послідовностей має місце наступний результат: якщо $c_n \underset{n\to\infty}{\longrightarrow} \alpha$ і $c_n \underset{n\to\infty}{\longrightarrow} \beta$, то $\alpha = \beta$. Це так зване твердження про однозначність границі. Подібне твердження має місце і для послідовностей випадкових величин.

Теорема 4. *Якщо* $\xi_n \underset{n\to\infty}{\longrightarrow} \xi$ *і* $\xi_n \underset{n\to\infty}{\longrightarrow} \eta$ *в якомусь з вищезазначених сенсах, то*

$$\mathbf{P}(\xi = \eta) = 1.$$

Д о в е д е н н я. Тут буде потрібна нерівність, яку ми іноді будемо використовувати далі. А саме, для довільних випадкових величин ξ, η і довільного ε має місце нерівність:

$$\mathbf{P}(|\xi + \eta| > \varepsilon) \le \mathbf{P}(|\xi| > \varepsilon/2) + \mathbf{P}(|\eta| > \varepsilon/2). \tag{13}$$

Нерівність (6) в очевидний спосіб виникає зі співвідношення:

$$\{|\xi + \eta| > \varepsilon\} \subset \{|\xi| > \varepsilon/2\} \cup \{|\eta| > \varepsilon/2\}, \tag{14}$$

тобто, якщо відбувається подія $\{|\xi + \eta| > \varepsilon\}$, то відбувається хоча б одна з подій $\{|\xi| > \varepsilon/2\}$, $\{|\eta| > \varepsilon/2\}$.

Якщо припустити, що рівність (14) не є правдивою, тобто подія $\{|\xi + \eta| > \varepsilon\}$ відбувається, а обидві події $\{|\xi| > \varepsilon/2\}$, $\{|\eta| > \varepsilon/2\}$ ні. Тоді відбуваються обидві події $\{|\xi| \le \varepsilon/2\}$, $\{|\eta| \le \varepsilon/2\}$ і маємо

$$|\xi + \eta| \le |\xi| + |\eta| \le \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

отже, подія $\{|\xi+\eta|>\varepsilon\}$ не відбувається, що протирічить припущенню.

Доведемо тепер теорему 4. З попереднього твердження випливає, що завжди $\xi_n \xrightarrow[n\to\infty]{\mathrm{P}} \xi$, $\xi_n \xrightarrow[n\to\infty]{\mathrm{P}} \eta$. Тоді для довільного $\varepsilon>0$ маємо:

$$\mathbf{P}(|\xi-\eta|>\varepsilon)=\mathbf{P}(|\xi-\xi_n+\xi_n-\eta|>\varepsilon)\le\mathbf{P}(|\xi-\xi_n|>\frac{\varepsilon}{2})+\mathbf{P}(|\xi_n-\eta|>\frac{\varepsilon}{2}) \underset{n\to\infty}{\longrightarrow} 0.$$

Отже, для кожного $\varepsilon>0$

$$\mathbf{P}(|\xi-\eta|>\varepsilon)=0,$$

що є рінозначним тому, що $\mathbf{P}(\xi=\eta)=1$. ◄

11.2 Критерії збіжності випадкових величин

Для числових послідовностей має місце наступний критерій збіжності. А саме, послідовність c_n є збіжною тоді і тільки тоді, коли

$$c_n-c_m\longrightarrow 0 \qquad \text{якщо} \qquad n\to\infty,\ m\to\infty. \tag{15}$$

Якщо для послідовності c_n, $n\ge 0$ виконується властивість (15), то кажуть, що послідовність c_n є *фундаментальною.* Отже, можемо говорити, що числова послідовність є збіжною тоді і тільки тоді, коли вона є фундаментальною. Виявляється, що для послідовностей випадкових величин ситуація є подібною. Але перед тим розглянемо необхідне визначення.

Визначення 5. Будемо казати, що послідовність випадкових величин ξ_n, $n=1,2,..$ є *фундаментальною послідовністю в сенсі збіжності майже напевно* (*за ймовірністю, в середньому, в середньому квадратичному*), якщо випадкові величини $\xi_n-\xi_m$ збігаються до нуля в тому самому сенсі при $n\to\infty$, $m\to\infty$. При чому, у випадку фундаментальної послідовності в сенсі середньої або середньої квадратичної збіжності вважаємо відповідно, що $\mathbf{M}|\xi_n|<\infty$ або $\mathbf{M}\xi_n^2<\infty$.

Зауваження 2. Для зручності замість „фундаментальна послідовність в сенсі збіжності майже напевно ” будемо писати „послідовність м.н.-фундаментальна” і аналогічне скорочення застосуємо для інших видів фундаментальних послідовностей.

Теорема 5. *Для того, щоб послідовність* $\xi_n, n=1,2,\dots$ *була збіжною в сенсі* м.н. (р, с, ск) *необхідно і достатньо, аби вона була фундаментальною в тому самому сенсі.*

Д о в е д е н н я. Для збіжності майже напевно це твердження випливає з аналогічної властивості для числових рядів.

Розглянемо лише випадок збіжності за ймовірністю. Доведення для інших видів збіжності можна знайти в [**?**], підрозділ 4, стор. 141.

Н е о б х і д н і с т ь.

Якщо $\xi_n \xrightarrow[n\to\infty]{p} \xi$, то необхідність випливає з наступного співвідношення:

$$\mathbf{P}(|\xi_n - \xi_m| > \varepsilon) \le \mathbf{P}(|\xi_n - \xi| > \varepsilon/2) + \mathbf{P}(|\xi_m - \xi| > \varepsilon/2) \underset{n,m\to\infty}{\longrightarrow} 0.$$

Д о с т а т н і с т ь Насамперед доведемо, що можна знайти підпослідовність ξ_{n_k} таку, що $\xi_{n_k} \xrightarrow[k\to\infty]{\mathrm{P}} \xi$, де ξ є певною випадковою величиною.

Маємо

$$\lim_{\substack{n\to\infty\\ m\to\infty}} \mathbf{P}(|\xi_n - \xi_m| > \varepsilon) = 0$$

Оскільки для кожного фіксованого $k = 2, 3, ...$

$$\lim_{\substack{n\to\infty\\ m\to\infty}} \mathbf{P}(|\xi_n - \xi_m| > k^{-2}) = 0,$$

то можна вибрати послідовність n_k, $k = 2, 3, ...$, $n_1 = 1$ таку, що $n_k > n_{k-1}$, і

$$\sup_{m > n_k} \mathbf{P}(|\xi_{n_k} - \xi_m| > k^{-2}) \le k^{-2}.$$

Зрозуміло, що тоді

$$\mathbf{P}(|\xi_{n_k} - \xi_{n_{k+1}}| > k^{-2}) \le k^{-2}.$$

для всіх k. Отже,

$$\sum_{k=1}^{\infty} \mathbf{P}(|\xi_{n_k} - \xi_{n_{k+1}}| > k^{-2}) \le \sum_{k=1}^{\infty} k^{-2} < \infty.$$

Згідно леми Борелі-Кантеллі, існує таке натуральне N (взагалі кажучи випадкове), що

$$\mathbf{P}(|\xi_{n_k} - \xi_{n_{k+1}}| \le \frac{1}{k^2}) = 1, \quad \text{для всіх} \quad k \ge N. \tag{16}$$

В свою чергу з (16) випливає, що існує така множина B з $\mathbf{P}(B) = 0$, що

$$|\xi_{n_k}(\omega) - \xi_{n_{k+1}}(\omega)| \le \frac{1}{k^2}, \quad \text{для} \quad k \ge N(\omega),\ \omega \in \Omega \backslash B. \tag{17}$$

Нерівність (17) показує, що ряд

$$\xi_{n_1} + \sum_{k=1}^{\infty} (\xi_{n_{k+1}} - \xi_{n_k})$$

збігається для всіх $\omega \in \Omega \backslash B$, позаяк його члени від певного моменту є обмеженими членами збіжного ряду

$$\sum_{k=1}^{\infty} \frac{1}{k^2}.$$

Визначимо випадкову величину ξ наступним чином:

$$\xi = \begin{cases} \xi_{n_1} + \sum_{k=1}^{\infty} (\xi_{n_{k+1}} - \xi_{n_k}), & \omega \in \Omega \backslash B, \\ 0, & \omega \in B. \end{cases}$$

Для кожного $\omega \in \Omega \backslash B$ маємо:

$$\lim_{m\to\infty} \xi_{n_m} = \lim_{m\to\infty} \left(\xi_{n_1} + \sum_{k=1}^{m-1} (\xi_{n_{k+1}} - \xi_{n_k}) \right) = \xi_{n_1} + \sum_{k=1}^{\infty} (\xi_{n_{k+1}} - \xi_{n_k}) = \xi, \qquad (18)$$

отже, $\xi_{n_k} \xrightarrow[n\to\infty]{\text{м.н}} \xi$.

Тепер вже легко закінчити доведення. Якщо послідовність ξ_n, $n = 1, 2\ldots$ є р-фундаментальною, то можемо знайти підпослідовність ξ_{n_k}таку, що $\xi_{n_k} \xrightarrow[k\to\infty]{\text{p}} \xi$. Тепер для кожного $\varepsilon > 0$ можемо записати:

$$\mathbf{P}(|\xi_n - \xi| > \varepsilon) \le \mathbf{P}(|\xi_n - \xi_{n_k}| > \varepsilon/2) + \mathbf{P}(|\xi_{n_k} - \xi| > \varepsilon/2) \underset{\substack{n\to\infty \\ k\to\infty}}{\longrightarrow} 0.$$

◄

ЗАВДАННЯ

1. Довести, що послідовність ξ_n має границю в середньому квадратичному сенсі тоді і тільки тоді, коли існує границя $\lim \mathbf{M}(\xi_n \xi_m)$ при $n \to \infty$, $m \to \infty$.

2. Нехай $(\Omega, \mathfrak{B}, \mathbf{P})$, де $\Omega = [0, 1]$, $\mathfrak{B}$–σ-алгебра борелівських множин в $[0, 1]$ а $\mathbf{P}$ означена в спосіб геометричний. Для кожного $n = 1, 2, ...$ нехай l та k будуть числами такими, що $n = l + 2^k$, $0 \le l < 2^k$. Легко зауважити, що числа k, l визначаються однозначно і $k \to \infty$, якщо $n \to \infty$. Покладемо

$$\xi_n(\omega) = \begin{cases} 1, & \omega \in [l2^{-k}, (l+1)2^{-k}), \\ 0, & \text{для інших } \omega. \end{cases}$$

   Довести, що послідовність $\xi_n(\omega)$ є збіжною до нуля за ймовірністю але не з ймовірністю 1.

3. Навести приклад послідовності випадкових величин збіжної за ймовірністю, але не збіжної в середньому.

4. Навести приклад послідовності випадкових величин збіжної в середньому, але не збіжної в середньому квадратичному сенсі.

5. Довести, якщо $\xi_n \xrightarrow[n\to\infty]{\text{p}} \xi$, $\eta_n \xrightarrow[n\to\infty]{\text{p}} \eta$, то $\xi_n \eta_n \xrightarrow[n\to\infty]{\text{p}} \xi\eta$.

6. Випадкові величини ξ_n та η_n збігаються за ймовірністю до ξ та η відповідно. Довести, що якщо випадкові величини ξ_n та η_n є незалежними, то ξ, η теж є незалежними.

7. Задана послідовність незалежних випадкових величин $\{\xi_n\}$ з нульовим математичним сподіванням $\mathbf{M}\xi_n = 0$ Показати, що ряд

$$\sum_{n=1}^{\infty} \xi_n$$

збігається в середньому квадратичному сенсі тоді і тільки тоді, коли $\sum_{n=1}^{\infty} \mathbf{D^2}\xi_n < \infty$.

8. Випадкова величина ξ_n має показникову щільність з параметром n. Довести, що $\xi_n \xrightarrow[n\to\infty]{\mathrm{P}} 0$. Чи буде ця послідовність збіжною з ймовірністю 1.

У наведених нище прикладах операції з випадковмих величинами розуміємо в спосіб натуральний і будемо припускати, що $\xi \equiv \eta$, коли $\mathbf{P}(\xi \neq \eta) = 0$.

9. Нехай L_1 позначає простір випадкових величин зі скінченним математичним сподіванням, заданих на одному ймовірносному просторі, тобто

$$L_1 = \{\xi : \mathbf{M}|\xi| < \infty\}.$$

Визначимо відстань в L_1 для довільних $\xi, \eta \in L_1$ як $\rho(\xi, \eta) = \mathbf{M}|\xi - \eta|$. Довести, що простір L_1 з визначеною в такий спосіб відстанню є простором Банаха.

10. Нехай L_2 позначає простір випадкових величин зі скінченним другим моментом, заданих на спільному ймовірносному просторі, тобто

$$L_2 = \{\xi : \mathbf{M}\xi^2 < \infty\}.$$

Визначимо відстань в L_2 для довільних $\xi, \eta \in L_2$ як $\rho(\xi, \eta) = \mathbf{M}(\xi - \eta)^2$. Довести, що простір L_2 є простором Банаха.

Лекція 12. Збіжність функцій розподілу

Як відомо, найважливішою характеристикою випадкової величини є функція розподілу. Всі види збіжностей, наведені вище, не використовують цього поняття в явний спосіб. Введемо тепер ще один тип збіжності для послідовностей випадкових величин (останній в даному курсі), який називається *слабка збіжність випадкових величин* або *слабка збіжність розподілів.*

Нехай задана послідовність випадкових величин ξ_n, $n = 1, 2, ..$ з функціями розподілу $F_n(x) = \mathbf{P}(\xi_n < x)$, а також випадкова величина ξ з функцією розподілу $F(x) = \mathbf{P}(\xi < x)$.

Визначення 1. Будемо говорити, що послідовність ξ_n, $n = 1, 2, \ldots$ збігається до ξ *за розподілом* при $n \to \infty$, якщо $F_n(x) \underset{n\to\infty}{\longrightarrow} F(x)$ в кожній точці неперервності функції $F(x)$.

Збіжність за розподілом (слабку збіжність) будемо позначати наступним чином:

$$\xi_n \underset{n\to\infty}{\overset{\mathrm{d}}{\longrightarrow}} \xi \quad \text{або} \quad F_n \underset{n\to\infty}{\overset{\mathrm{d}}{\longrightarrow}} F.$$

З цього визначення випливає, що у випадку, коли гранична функція розподілу є неперервною, то $\xi_n \underset{n\to\infty}{\overset{\mathrm{d}}{\longrightarrow}} \xi$ означає, що $F_n(x) \underset{n\to\infty}{\longrightarrow} F(x)$ для кожного x.

Приклад 1. *Нехай* $\Omega = \{\omega_1, \omega_2\}, \mathbf{P}(\omega_1) = \mathbf{P}(\omega_2) = 1/2$. *Визначимо випадкові величини* ξ_n, ξ *наступним чином:*

$$\xi_n(\omega) = \begin{cases} 1/n, & \omega = \omega_1, \\ 1 - 1/n, & \omega = \omega_2, \end{cases} \qquad \xi(\omega) = \begin{cases} 1, & \omega = \omega_1, \\ 0, & \omega = \omega_2. \end{cases}$$

Функції розподілу введених випадкових величин мають вигляд:

$$F_n(x) = \begin{cases} 0, & x \le \frac{1}{n}, \\ \frac{1}{2}, & \frac{1}{n} < x \le 1 - \frac{1}{n}, \\ 1, & x > 1 - \frac{1}{n}, \end{cases} \qquad F(x) = \begin{cases} 0, & x \le 0, \\ \frac{1}{2}, & 0 < x \le 1, \\ 1, & x > 1. \end{cases}$$

Вочевидь, $F_n \xrightarrow[n\to\infty]{\mathrm{d}} F$.

Зауважимо, що в цьому прикладі гранична функція розподілу має дві точки розриву: 0 і 1. Для $x = 0$, вочевидь, $0 = F_n(0) \underset{n\to\infty}{\longrightarrow} F(0) = 0$, але для $x = 1$ маємо $1 = F_n(1) \not\to F(1) = 1/2, \quad n \to \infty$.

Теорема 1. *Якщо* $\xi_n \xrightarrow[n\to\infty]{\mathrm{p}} \xi$*, то* $\xi_n \xrightarrow[n\to\infty]{\mathrm{d}} \xi$.

Д о в е д е н н я. Нехай x є точкою неперервності для функції розподілу $F(x)$, і ε є довільним фіксованим малим числом. Маємо:

$$\begin{aligned}|\mathbf{P}(\xi_n < x) - \mathbf{P}(\xi < x)| &= |\mathbf{P}(\xi_n < x, \xi < x) + \mathbf{P}(\xi_n < x, \xi \geq x)) - \mathbf{P}(\xi < x)| = \\ &= |\mathbf{P}(\xi_n < x, \xi \geq x) - \mathbf{P}(\xi_n \geq x, \xi < x)| \leq \mathbf{P}(\xi_n < x, \xi \geq x) + \mathbf{P}(\xi_n \geq x, \xi < x) \qquad (1)\end{aligned}$$

Існує $\delta > 0$ таке, що $\mathbf{P}(x \leq \xi < x + \delta) = F(x + \delta) - F(x) \leq \varepsilon$. Маємо

$$\begin{aligned}\mathbf{P}(\xi_n < x, \xi \geq x) &= \mathbf{P}(\xi_n < x, \xi \geq x + \delta) + \mathbf{P}(\xi_n < x, x \leq \xi < x + \delta) \leq \\ &\leq \mathbf{P}(|\xi_n - \xi| \geq \delta) + \mathbf{P}(x \leq \xi < x + \delta) \leq \mathbf{P}(|\xi_n - \xi| \geq \delta) + \varepsilon. \qquad (2)\end{aligned}$$

Позаяк $\xi_n \xrightarrow[n\to\infty]{\mathrm{p}} \xi$, то існує n_0 таке, що $\mathbf{P}(|\xi_n - \xi| \geq \delta) \leq \varepsilon$ для $n \geq n_0$, і з (2) випливає, що $\mathbf{P}(\xi_n < x, \xi \geq x) \leq 2\varepsilon$ для $n \geq n_0$. Оскільки ε довільне мале число, то будемо мати, що $\lim_{n\to\infty} \mathbf{P}(\xi_n < x, \xi \geq x) = 0$. З завдання 3 випливає, що також $\lim_{n\to\infty} \mathbf{P}(\xi_n \geq x, \xi < x) = 0$, і тепер з (1) випливає наша теза.

З доведеного твердження маємо, що збіжність $\xi_n \xrightarrow[n\to\infty]{\mathrm{p}} \xi$ тягне за собою збіжність $\xi_n \overset{\mathrm{d}}{\underset{n\to\infty}{\Longrightarrow}} \xi$. Навпаки це не є правдою. Але в одному конкретному випадку, дуже важливому для статистики, збіжність за ймовірністю і слабка збіжність є рівнозначними.

Теорема 2. *Нехай* $F_n(x)$ *є функціями розподілу випадкових величин* $\xi_n, n = 1, 2 \ldots$ *Послідовність випадкових величин* ξ_n *є збіжною за ймовірністю до константи c тоді і тільки тоді, коли*

$$\lim_{n\to\infty} F_n(x) = F(x) = \begin{cases} 0, & x \leq c, \\ 1, & x > c. \end{cases} \qquad (3)$$

Д о в е д е н н я. Зрозуміло, що досить розглянути випадок $c = 0$. Необхідність випливає з попереднього твердження. Нехай тепер має місце (3). Для довільного $\varepsilon > 0$ маємо

$$\begin{aligned}\mathbf{P}(|\xi_n| > \varepsilon) &= \mathbf{P}(\xi_n > \varepsilon) + \mathbf{P}(\xi_n < -\varepsilon) = 1 - \mathbf{P}(\xi_n \leq \varepsilon) + \mathbf{P}(\xi_n < -\varepsilon) \\ &= 1 - F_n(\varepsilon + 0) + F_n(-\varepsilon) \underset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

◄

Розгляд слабкої збіжності закінчимо двома твердженнями (без доведення), які є дуже суттєвими для теорії збіжностей. Доведення цих тверджень читач може знайти в [1], [?].

Теорема 3. (Перша теорема Хеллі) *Кожна послідовність*

$$F_1(x), F_2(x), \dots, F_n(x), \dots$$

функцій розподілу містить підпослідовність

$$F_{n_1}(x), F_{n_2}(x), \dots, F_{n_k}(x), \dots$$

слабко збіжну до певної неспадної, неперервної зліва функції $G(x)$ *такої, що* $0 \leq G(x) \leq 1$.

Зауваження 1. Хоча фукція $G(x)$ є неперервною зліва, неспадною і $0 \leq G(x) \leq 1$, але вона може не бути функцією розподілу, оскільки можлива ситуація, коли $G(\infty) - G(-\infty) < 1$, про що свідчить наступний приклад.

Приклад 2. *Нехай*

$$F_n(x) = \begin{cases} 0, & x < -n, \\ \frac{x+n}{2n}, & |x| \leq n, \\ 1, & x > n. \end{cases}$$

Тоді для кожного x

$$\lim_{n\to\infty} F_n(x) = 1/2.$$

Теорема 4. (Друга теорема Хеллі) *Якщо послідовність функцій розподілу* $F_n(x)$, $n = 1, 2, \dots$ *є слабко збіжною до функції розподілу* $F(x)$, *тоді для кожної неперервної і обмеженої функції* $g(x)$, $-\infty < x < \infty$ *має місце рівность*

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g(x) dF_n(x) = \int_{-\infty}^{\infty} g(x) dF(x) \tag{4}$$

Теорема 5. *Нехай послідовність функцій розподілу* $F_n(x)$, $n = 1, 2, \dots$ *є такою, що для кожної неперервної і обмеженої функції* $g(x)$, $-\infty < x < \infty$ *має місце рівність:*

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g(x) dF_n(x) = \int_{-\infty}^{\infty} g(x) dF(x), \tag{5}$$

де $F(x)$ *є певною функцією розподілу. Тоді послідовність функцій розподілу* $F_n(x)$ *є слабко збіжною до* $F(x)$.

Д о в е д е н н я. Достатньо довести, що кожна слабко збіжна підпослідовність послідовності $F_n(x)$ є збіжною до $F(x)$. Але, насамперед, треба переконатись, що такі послідовності взагалі існують. Насправді, це випливає з першої теореми Хеллі, яка гарантує існування такої послідовності. Нехай тепер послідовність $F_{n_k}(x)$, $k = 1, 2, \ldots$ є такою, що $F_{n_k}(x) \xrightarrow[k\to\infty]{\mathrm{d}} F_0(x)$, де $F_0(x)$ є певною неспадною обмеженою функцією. Доведемо, що $F_0(x) \equiv F(x)$. З другої теореми Хеллі отримаємо:

$$\lim_{k\to\infty} \int_{-\infty}^{\infty} g(x) dF_{n_k}(x) = \int_{-\infty}^{\infty} g(x) dF_0(x) \tag{6}$$

для кожної неперервної обмеженої функції $g(x)$. З (2), (6) випливає

$$\int_{-\infty}^{\infty} g(x) dF(x) = \int_{-\infty}^{\infty} g(x) dF_0(x) \tag{7}$$

для кожної неперервної обмеженої функції $g(x)$. Звідки для $g(x) \equiv 1$ маємо $F_0(-\infty) - F_0(\infty) = F(-\infty) - F(\infty) = 1$ і, отже, $F_0(-\infty) = 0$, $F_0(\infty) = 1$. Тепер з рівності (7) неважко отримати $F(x) = F_0(x)$ (див. завдання 4). ◄

З доведеної теореми і другої теореми Хеллі випливає рівнозначне визначення слабкої збіжності функцій розподілу.

Визначення 2. Будемо говорити, що послідовність $F_n(x)$, $n = 1, 2, \ldots$ функцій розподілу *слабко збігаєтьса* до функції розподілу $F(x)$, якщо для кожної неперервної і обмеженої функції $g(x)$, $-\infty < x < \infty$ має місце рівність:

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g(x) dF_n(x) = \int_{-\infty}^{\infty} g(x) dF(x).$$

ЗАВДАННЯ

1. Кажуть, що послідовність випадкових величин ξ_n є *обмеженою за ймовірністю*, якщо для кожного $\varepsilon > 0$ існує M, а також n_0 такі, що $\mathbf{P}(|\xi_n| > M) < \varepsilon$ для всіх $n > n_0$. Показати, що у випадку, коли послідовність ξ_n є збіжною за розподілом, то вона є обмеженою за ймовірністю.

2. Навести приклад послідовності випадкових величин, яка є збіжною за розподілом, але не збігається за ймовірністю.

3. Довести, що $\lim_{n\to\infty} \mathbf{P}(\xi_n \geq x, \xi < x) = 0)$, якщо x є точкою неперервності для розподілу випадкової величини ξ а послідовність випадкових величин ξ_n збігається за ймовірністю до ξ.

4. Довести, що рівність (7) спричиняє те, що $F(x) = F_0(x)$ для кожного x.

   Вказівка. Для $\varepsilon > 0$ і фіксованого x_0 записати рівність (7) для неперервної функції $g(x)$ такої, що $|g(x)| \leq 1$, $g(x) = 1$, $x \geq x_0$ і $g(x) = 0$, $x \leq x_0 - \varepsilon$, а потім перейти до границі при $\varepsilon \to 0$.

5. Довести, що $F_{S_n} \xrightarrow[n\to\infty]{\mathrm{d}} 0$, де $S_n = \xi_1 + \xi_2 + ... + \xi_n$ а ξ_n, $n = 1, 2, ..$ є послідовністю незалежних однаково розподілених випадкових величин таких, що $\mathbf{P}(\xi_n = 0) < 1$.

6. Чи справедливе твердження: Якщо $\xi_n \xrightarrow[n\to\infty]{\mathrm{d}} \xi$ і $\eta_n \xrightarrow[n\to\infty]{\mathrm{d}} \eta$, то $\xi_n + \eta_n \xrightarrow[n\to\infty]{\mathrm{d}} \xi + \eta$?

7. Довести твердження: Якщо $F_n \xrightarrow[n\to\infty]{\mathrm{d}} F$ і F_n є послідовністю функцій розподілу, то

$$\lim_{n\to\infty} \sup_t |F_n(t) - F(t)| = 0.$$

8. Випадкова величина ξ має розподіл Пуассона з параметром λ. Довести, що розподіл випадкової величини $(\xi - \lambda)/\sqrt{\lambda}$ слабко збігається при $\lambda \to \infty$ до стандартного нормального розподілу.

9. Випадкова величина ξ має розподіл гамма з параметрами λ, p. Довести, що розподіл випадкової величини $(\lambda\xi - p)/\sqrt{p}$ слабко збігається при $p \to 0$ до стандартного нормального розподілу.

10. Величини ξ_n, $n = 1, 2, ...$ мають розподіл

$$\mathbf{P}(\xi_n = k) = \frac{c_n}{k^{2+1/n}}, \qquad c_n^{-1} = \sum_{k=1}^{\infty} k^{-2-1/n}, \quad k = 1, 2, ...$$

   Довести, що послідовність ξ_n, $n = 1, 2, ...$ збігається слабо до величини ξ з розподілом

$$\mathbf{P}(\xi = k) = \frac{c}{k^2}, \qquad c = \sum_{k=1}^{\infty} k^{-2}, \quad k = 1, 2, ...$$

   Довести також, що $\mathbf{M}\xi_n < \infty$ але $\mathbf{M}\xi = \infty$

Лекція 13. Характеристичні функції

В цій лекції ми розглянемо поняття характеристичної функції випадкової величини, яке є одним з головних інструметів вивчення властивостей розподілів цих величин.

13.1 Означення та властивості характеристичних функції

Нехай задана випадкова величина ξ з функцією розподілу $F(x)$.

Визначення 1. *Характеристичною функцією* випадкової величини ξ називають комплекснозначну функцію

$$\varphi(t) = \mathbf{M}e^{it\xi} = \int\limits_{-\infty}^{\infty} e^{itx} dF(x), \tag{1}$$

де t є дійсною змінною, а i – уявною одиницею ($i^2 = -1$).

Якщо функція розподілу має щільність $f(x)$, то (1) можна записати наступним чином:

$$\varphi(t) = \int\limits_{-\infty}^{\infty} e^{itx} f(x) dx. \tag{2}$$

Оскільки

$$e^{itx} = \cos(tx) + i\sin(tx),$$

то вираз (1) можна подати ще так

$$\varphi(t) = \mathbf{M}(\cos(t\xi) + i\sin(t\xi)) = \int\limits_{-\infty}^{\infty} \cos(tx) dF(x) + i \int\limits_{-\infty}^{\infty} \sin(tx) dF(x).$$

Зауваження 1. В загально прийнятій terminології аналізу Фур'є $\varphi(t)$ з (2) є перетворенням Фур'є функції $f(x)$. Це перетворення є визначеним для кожної функції f, оскільки

$$\int_{-\infty}^{\infty} |f(x)|dx < \infty.$$

Термін „характеристична функція" зазвичай застосовують тоді, коли у виразі (2) функція $f(x)$ є щільністю.

Характеристична функція існує для кожної випадкової величини ξ. Це випливає з того, що

$$|\varphi(t)| = |\int_{-\infty}^{\infty} e^{itx} dF(x)| \leq \int_{-\infty}^{\infty} |e^{itx}| dF(x) = \int_{-\infty}^{\infty} dF(x) = 1.$$

Приклад 1. *Випадкова величина має показниковий розподіл з параметром λ. Знадемо її характеристичну функцію.*

$$\varphi(t) = \lambda \int_{0}^{\infty} e^{itx} e^{-\lambda x} dx = \frac{\lambda}{it - \lambda},\ t \in R^1.$$

Теорема 1. *Характеристична функція має наступні властивості:*

1) $\varphi(0) = 1,\ |\varphi(t)| \leq 1$ *для всіх* $t \in R^1$;

2) *для довільних сталих* a, b *має місце рівність:*

$$\varphi_{a\xi+b}(t) = e^{ibt}\varphi_\xi(at);$$

3) $\overline{\varphi}(t) = \varphi(-t)$, *де* $\overline{\varphi}(t)$ *означає спряжене комплексне число до* $\varphi(t)$;

4) *функція* $\varphi(t)$ *є рівномірно неперервною на дійсній осі;*

5) *якщо випадкові величини* ξ, η *є незалежними з відповідними характеристичними функціями* $\varphi_\xi(t)$, $\varphi_\eta(t)$, *то характеристична функція суми* $\xi + \eta$ *має вигляд:*

$$\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\varphi_\eta(t).$$

Д о в е д е н н я пунктів 1-3 випливає з визначення очевидним чином.

Доведемо пункти 4,5. Нехай задане фіксоване $\varepsilon > 0$. Позаяк $F(N) \underset{N\to\infty}{\longrightarrow} 1$, а $F(-N) \underset{N\to\infty}{\longrightarrow} 0$, то можна вказати таке $N_0 > 0$, що

$$1 - F(N_0) + F(-N_0) = \int\limits_{|x|>N_0} dF(x) \le \varepsilon.$$

Для довільного $h \in R^1$ маємо

$$|\varphi(t+h) - \varphi(t)| = \int\limits_{-\infty}^{\infty} e^{itx}(e^{ihx} - 1)dF(x) \le \int\limits_{-\infty}^{\infty} |e^{ihx} - 1|dF(x) =$$

$$\le \int\limits_{|x|\le N_0} |e^{ihx} - 1|dF(x) + \int\limits_{|x|>N_0} |e^{ihx} - 1|dF(x) \le \int\limits_{|x|\le N_0} |e^{ihx} - 1|dF(x) +$$

$$+ 2\int\limits_{|x|>N_0} dF(x) \le \int\limits_{|x|\le N_0} |e^{ihx} - 1|dF(x) + \varepsilon. \qquad (3)$$

Перехід в (3) до границі $\overline{\lim}_{h\to 0}$ дає

$$\overline{\lim}_{h\to 0}|\varphi(t+h) - \varphi(t)| \le 2\varepsilon$$

Позаяк в цій нерівності ε довільне, то пункт 4 є доведеним.

Для доведення останнього пункту зауважимо, що оскільки випадкові величини ξ, η - незалежні, то випадкові величини $e^{it\xi}$, $e^{it\eta}$ також є незалежними. Тоді

$$\varphi_{\xi+\eta}(t) = \mathbf{M}e^{it(\xi+\eta)} = \mathbf{M}e^{it\xi}e^{it\eta} = \mathbf{M}e^{it\xi}\mathbf{M}e^{it\eta} = \varphi_\xi(t)\varphi_\eta(t).$$

◀

Отже, якщо випадкові величини ξ, η - незалежні, то характеристична функція $\xi + \eta$ дорівнює добуткові характеристичних функцій доданків. Але результат навпаки, тобто, якщо характеристична функція суми $\xi + \eta$ дорівнює добуткові характеристичних функцій доданків, то випадкові величини ξ, η є незалежними, не має місця. Натомість справедливе наступне твердження, яке наведемо без доведення (завдання 12).

Теорема 2. *Якщо для довільних $t, s \in R^1$*

$$\mathbf{M}e^{i(t\xi+s\eta)} = \mathbf{M}e^{it\xi}\mathbf{M}e^{is\eta},$$

то випадкові величини ξ, η є незалежними

Як було доведено раніше, якщо $F_\xi(x)$, $F_\eta(x)$ є функціями розподілу незалежних випадкових величин ξ, η, то функція

$$F_{\xi+\eta}(x) = F_\xi \star F_\eta(x) = \int\limits_{-\infty}^{\infty} F_\xi(x-y)dF_\eta(y)$$

є функцією розподілу випадкової величини $\xi+\eta$, а тоді характеристична функція цієї випадкової величини може бути представлена у вигляді:

$$\varphi_{\xi+\eta}(t) = \int\limits_{-\infty}^{\infty} e^{itx} dF_{\xi+\eta}(x) = \varphi_\xi(t)\varphi_\eta(t). \tag{4}$$

У зв'язку з чим можемо сказати, що згортці функцій розподілу відповідає добуток їх характеристичних функцій.

Приклад 2. *Для випадкової величини $\xi \in N(m, \sigma^2)$ знайти характеристичну функцію.*

Р о з в' я з о к. Нехай спочатку $\xi \in N(0,1)$ Тоді

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx.$$

Диференціюючи останній вираз по t під знаком інтегралу, маємо:

$$\frac{d\varphi(t)}{dt} = \frac{i}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} x e^{itx - x^2/2} dx = -\frac{i}{\sqrt{2\pi}} \; e^{itx} e^{-x^2/2} \Big|_{-\infty}^{\infty} - it \int\limits_{-\infty}^{\infty} e^{itx - x^2/2} dx$$

$$= -\frac{t}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{itx - x^2/2} dx = -t\varphi(t).$$

Отримали рівняння $\frac{d\varphi(t)}{dt} = -\varphi(t)$, звідки $\varphi(t) = Ce^{-\frac{t^2}{2}}$. Позаяк $\varphi(0) = 1$, то $C = 1$, і $\varphi(t) = e^{-\frac{t^2}{2}}$.

Якщо $\xi \in N(m, \sigma^2)$, то випадкова величина $\eta = \frac{\xi - m}{\sigma} \in N(0,1)$, і тоді на підставі властивості 2 характеристичних функцій отримаємо:

$$\varphi_\xi(t) = \varphi_{\sigma\eta+m}(t) = e^{imt}\varphi_\eta(\sigma t) = e^{imt} e^{-\frac{\sigma^2 t}{2}} = e^{t(im - \frac{\sigma^2}{2})}.$$

◀

Теорема 3. *Якщо існує k-тий момент випадкової величини ξ (тобто $\mathbf{M}|\xi|^k < \infty$, $k \geq 1$), то існує неперервна l-та $(1 \leq l \leq k)$ похідна функції $\varphi(t)$, і*

$$\varphi^{(l)}(0) = i^l \mathbf{M}\xi^l, \quad \text{а} \quad \mathbf{M}\xi^l = (-i)^l \varphi^{(l)}(0). \tag{5}$$

Д о в е д е н н я. Якщо $1 \leq l \leq k$, то $\mathbf{M}|\xi|^l < \infty$ і можемо записати:

$$\Big| \int\limits_{-\infty}^{\infty} (ix)^l e^{itx} dF(x) \Big| \leq \int\limits_{-\infty}^{\infty} |x|^l dF(x) = \mathbf{M}|\xi|^l < \infty. \tag{6}$$

З (6) випливає, що інтеграл $\int_{-\infty}^{\infty}(ix)^l e^{itx}dF(x)$ є рівномірно збіжним по t для всіх $1 \le l \le k$. Таким чином, можна диференціювати в $\int_{-\infty}^{\infty} e^{itx}dF(x)$ по t під знаком інтегралу l разів. Звідки отримаємо:

$$\varphi^{(l)}(t) = i^l \int_{-\infty}^{\infty} x^l e^{itx} dF(x),$$

отже,

$$\varphi^{(l)}(0) = i^l \int_{-\infty}^{\infty} x^l dF(x) = i^l \mathbf{M}\xi^l.$$

◄

13.2 Генератриси

Якщо випадкова величина ξ є цілочисловою, тобто $\mathbf{P}(\xi = k) = p_k$, $k = 0, \pm1, \pm2, ...$ і $\sum_{k=-\infty}^{\infty} p_k = 1$, то згідно з визначенням характеристичної функції будемо мати:

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x) = \sum_{k=-\infty}^{\infty} e^{itk} p_k = \sum_{k=-\infty}^{\infty} (e^{it})^k p_k,$$

отже, характеристична функція є в кінцевому рахунку функцією комплексної змінної $z = e^{it}$. Позаяк $|z| = 1$, то комплексне число z лежить на колі з радіусом 1 і центром в точці початку координат комплексної площини. У цьому випадку зручніше застосовувати не характеристичну функцію, а генератрису, яку зараз визначимо.

Визначення 2. Якщо випадкова величина ξ є такою, що $\sum_{k=-\infty}^{\infty} \mathbf{P}(\xi = k) = 1$, то *генератрисою* випадкової величини ξ будемо називати функцію

$$P(z) = \mathbf{M}z^{\xi} = \sum_{k=-\infty}^{\infty} z^k \mathbf{P}(\xi = k), \qquad |z| = 1.$$

Якщо $P(z)$ - генератриса випадкової величини ξ, то її характеристична функція, вочевидь, дорівнює $P(e^{it}),\ t \in R^1$.

Приклад 1. *Випадкова величина ξ має біноміальний розподіл*

$$\mathbf{P}(\xi = k) = C_n^k p^k q^{n-k}, \qquad 0 < k \le n.$$

Знайти її генератрису.

Р о з в' я з о к.

$$P(z) = \sum_{k=0}^{n} C_n^k z^k p^k q^{n-k} = (zp+q)^n.$$

Приклад 2. *Знайти генератрису розподілу Пуассона з параметром λ.*

Р о з в' я з о к.

$$P(z) = \sum_{k=0}^{\infty} z^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(z\lambda)^k}{k!} = e^{-\lambda} e^{\lambda z} = e^{-\lambda(1-z)}.$$

Позаяк дійсній вісі $(-\infty, \infty)$ при відображенні $t \to e^{it}$ відповідає коло з радіусом 1 і центом в точці початку координат комплексної вісі, то можна легко написати відповідники теорем 1, 2. А саме:

Теорема 4. (Властивості генератрис).

1. $P(1) = 1,\ |P(z)| \le 1$ *для всіх* $|z| = 1$.

1) *для довільних цілих* k, n *має місце рівність*

$$P_{k\xi+n}(z) = z^n P_\xi(z^k);$$

2) $\overline{P}(z) = P(\overline{z})$;

3) *функція* $P(z)$ *є неперервною на колі* $|z| = 1$*;*

4) *якщо випадкові величини* ξ, η *є незалежними с генератрисами* $P_\xi(z)$ *і* $P_\eta(z)$*, то для генератриси* $\xi + \eta$ *маємо:*

$$P_{\xi+\eta}(z) = P_\xi(z) P_\eta(z), \qquad |z| = 1.$$

Аналог теореми 3 не буде таким простим. Це пов'язано з тим фактом, що у випадку характеристичної функції має місце рівність

$$\frac{d^l e^{itx}}{dt^l} = (ix)^l e^{itx}.$$

А тепер для $l \le k$ маємо

$$\frac{d^l z^k}{dz^l} = k(k-1)\cdots(k-l+1)z^{k-l},$$

отже, степені $(ix)^l$ відповідає добуток $k(k-1)\cdots(k-l+1)z^{-l}$ і в наслідок чого значення похідної $\frac{d^l z^k}{dt^l}|_{z=1}$ вже не буде дорівнювати $\mathbf{M}\xi^l$. Наведемо відповідник теореми 3 лише для $l = 1, 2$.

Теорема 5. *Якщо* $\mathbf{M}|\xi|^2 < \infty$, *то існують перша і друга похідні функції* $P(z)$ *в точці* $z = 1$ *і, крім того,*

$$\mathbf{M}\xi = \frac{dP(z)}{dz}|_{z=1}, \quad \mathbf{M}\xi^2 = \frac{d^2P(z)}{dz^2}|_{z=1} + \mathbf{M}\xi. \tag{7}$$

◀

Приклад 3. (продовження прикладу 1) *Знайти* $\mathbf{M}\xi$, $\mathbf{M}\xi^2$.

Р о з в' я з о к. З формул (7) отримуємо:

$$\mathbf{M}\xi = \frac{d(zp+q)^n}{dz}|_{z=1} = np(zp+q)^{n-1}|_{z=1} = np,$$

$$\mathbf{M}\xi^2 = \frac{d^2(zp+q)^n}{dz^2}|_{z=1} + np = n(n-1)p^2(zp+q)^{n-2}|_{z=1} + np = np(np+q).$$

Наступна таблиця містить приклади розподілів та їх характеристичних функцій або генератрис. Звернімо увагу на характеристичну функцію розподілу під номером 8. Ця функція дорівнює нулеві поза межами скінченного інтервалу, хоча щільність є ненульовою на всій дійсній вісі. Ця особливість дуже часто використовується в аналізі Фур'є.

Lp.	Назва розподілу	Щільність або розподіл	Область визначення	Функ. характер. або генератриса
1	Показниковий	$\lambda e^{-\lambda x}$	$x \geq 0,\ \lambda > 0$	$\frac{\lambda}{\lambda-it}$
2	Ерланга	$\frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}$	$x \geq 0,\ \lambda > 0$	$\frac{\lambda}{\lambda-it}\ ^{n}$
3	χ_n^2	$\frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)}$	$x \geq 0$	$\frac{\lambda}{\lambda-it}\ ^{n/2}$
4	Рівномірний	$1/a$	$0 < x < a$	$\frac{e^{iat}-1}{at}$
5	Рівномірний	$1/2a$	$-a < x < a$	$\frac{\sin(at)}{at}$
6	Трикутний	$\frac{1}{a}\ 1 - \frac{\|x\|}{a}$	$\|x\| \leq a$	$2\frac{1-\cos at}{a^2t^2}$
7	Нормальний	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$	$-\infty < x < \infty$	$e^{-\sigma^2t^2/2+imt}$

Lp.	Назва розподілу	Щільність або розподіл	Область визначення	Функ. характер. або генератриса
8	-	$\frac{1-\cos ax}{a\pi x^2}$	$-\infty < x < \infty$	$1-\frac{\lvert t\rvert}{a},\ \lvert x\rvert \leq a$ $0,\ \lvert x\rvert > a.$
9	Гамма	$\frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}$	$0 < x < \infty,$ $0 < \lambda < \infty,$ $0 < p < \infty$	$\frac{\lambda}{\lambda - it}^{\ p}$
10	Коші	$\frac{\lambda}{\pi(\lambda^2+(x-m)^2)}$	$-\infty < x < \infty,$ $-\infty < m < \infty,$ $\lambda > 0$	$e^{-\lambda\lvert t\rvert + imt}$
11	Нуля-одиниці	$p, q = 1-p$	$0 \leq p \leq 1$	$q + pz$
12	Біноміальний	$C_n^k p^k q^{n-k}$	$0 \leq p \leq 1,$ $0 \leq k \leq n$	$(q+pz)^n$
13	Геометричний	$q^{n-1} p$	$0 \leq p \leq 1,$ $n \geq 1$	$zp(1-qz)^{-1}$
14	Пуассона	$e^{-\lambda} \frac{\lambda^k}{k!}$	$\lambda > 0,\ k \geq 0$	$e^{\lambda(z-1)}$

13.2.1 Гратчасті розподіли

Розглянемо тепер випадкові величини, які є дуже подібні до цілочислових.

Визначення 3. Розподіл випадкової величини ξ будемо називати *гратчастим* (*арифметичним*) *з кроком* $h > 0$, якщо існують такі a і h, що

$$\sum_{k=-\infty}^{\infty} \mathbf{P}(\xi = a + kh) = 1. \tag{8}$$

В цьому випадку будемо казати, що випадкова величина ξ є гратчастою з кроком h.

Приклад 4. *Випадкова величина* ξ *є такою, що* $\mathbf{P}(\xi = \sqrt{2} + \frac{1}{3}k) = \frac{1}{2^k}, k = 1, 2, ...$ Позаяк

$$\sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

то ця ця випадкова величина є гратчастою з кроком $\frac{1}{3}$.

Цілочислові випадкові величини є, вочевидь, гратчастими випадковими величинами з натуральним кроком $h \geq 1$. В той же час, навпаки, якщо для випадкової величини ξ справедливе співвідношення (8), то випадкова величина $\eta = \frac{\xi - a}{h}$ є цілочисловою випадковою величиною (завдання 1).

Теорема 6. *Випадкова величина ξ має гратчастий розподіл з кроком h тоді і тільки тоді, коли*

$$|\varphi(\tfrac{2\pi}{h})| = 1. \tag{9}$$

Д о в е д е н н я. Якщо випадкова величина ξ має гратчастий розподіл і має місце рівність (9), то

$$|\varphi(\tfrac{2\pi}{h})| = |\mathbf{M}e^{i\frac{2\pi\xi}{h}}| = |\sum_{k=-\infty}^{\infty} e^{i\frac{2\pi(a+kh)}{h}}\mathbf{P}(\xi = a + kh)| =$$

$$= |e^{i\frac{2\pi a}{h}} \sum_{k=-\infty}^{\infty} e^{2i\pi k}\mathbf{P}(\xi = a + kh)| = |e^{i\frac{2\pi a}{h}} \sum_{k=-\infty}^{\infty} \mathbf{P}(\xi = a + kh)| = |e^{i\frac{2\pi a}{h}}| = 1.$$

Якщо навпаки, має місце рівність (9), то

$$1 = |\mathbf{M}e^{i\frac{2\pi\xi}{h}}| = |\mathbf{M}\cos\tfrac{2\pi\xi}{h} + i\mathbf{M}\sin\tfrac{2\pi\xi}{h}| = \sqrt{\left(\mathbf{M}\cos\tfrac{2\pi\xi}{h}\right)^2 + \left(\mathbf{M}\sin\tfrac{2\pi\xi}{h}\right)^2}.$$

Отже,

$$\left(\mathbf{M}\cos\tfrac{2\pi\xi}{h}\right)^2 + \left(\mathbf{M}\sin\tfrac{2\pi\xi}{h}\right)^2 = 1. \tag{10}$$

Позаяк завжди $\mathbf{M}\eta^2 \geq (\mathbf{M}\eta)^2$, то

$$\mathbf{M}\left(\cos\tfrac{2\pi\xi}{h}\right)^2 \geq \left(\mathbf{M}\cos\tfrac{2\pi\xi}{h}\right)^2, \quad \mathbf{M}\left(\sin\tfrac{2\pi\xi}{h}\right)^2 \geq \left(\mathbf{M}\sin\tfrac{2\pi\xi}{h}\right)^2. \tag{11}$$

Якщо хоча б одна нерівність (11) була строгою, то, додаючи ці нерівності одна до одної, з урахуванням (10) ми би отримали:

$$1 = \mathbf{M}\left(\cos\tfrac{2\pi\xi}{h}\right)^2 + \mathbf{M}\left(\sin\tfrac{2\pi\xi}{h}\right)^2 > 1.$$

Як наслідок в (11) має місце рівність, це означає, що

$$\mathbf{M}\left(\cos\tfrac{2\pi\xi}{h}\right)^2 = \left(\mathbf{M}\cos\tfrac{2\pi\xi}{h}\right)^2, \quad \mathbf{M}\left(\sin\tfrac{2\pi\xi}{h}\right)^2 = \left(\mathbf{M}\sin\tfrac{2\pi\xi}{h}\right)^2.$$

Це є рівнозначним наступним співвідношенням:

$$\mathbf{D^2}\cos\tfrac{2\pi\xi}{h} = 0, \quad \mathbf{D^2}\sin\tfrac{2\pi\xi}{h} = 0.$$

Отже, отримаємо з ймовірністю 1:

$$\cos\tfrac{2\pi\xi}{h} = c_1, \quad \sin\tfrac{2\pi\xi}{h} = c_2, \tag{12}$$

де c_1, c_2 є таким константами, що $c_1^2 + c_2^2 = 1$. Якщо через α позначити аргумент комплексного числа $c_1 + ic_2$, тобто, якщо $c_1 + ic_2 = \cos\alpha + i\sin\alpha$, то з (12) випливає $\cos\frac{2\pi\xi}{h} + i\sin\frac{2\pi\xi}{h} = \cos\alpha + i\sin\alpha$ або $e^{\frac{2\pi\xi}{h}} = e^{\alpha}$. Звідки

$$\tfrac{2\pi\xi}{h} - \alpha = 2\pi k, \quad k = 0, \pm 1, \pm 2, \ldots$$

або

$$\xi = \frac{\alpha h}{2\pi} + hk, \quad k = 0, \pm 1, \pm 2, \ldots$$

◀

ЗАВДАННЯ

1. Показати, що якщо випадкова величина ξ має гратчастий розподіл, то випадкова величина $(\xi - b)/h$ є цілочисельною.

2. Нехай ξ_i, $i = 1, 2, 3...$ є незалежними однаково розподіленими випадковими величинами з генератрисою $Q(z)$. Випадкова величина η приймає значення $0, 1, 2, ...$ і є незалежною від ξ_i. Функція $g(z)$ є генератрисою η. Нехай

$$\zeta = \xi_1 + \xi_2 + ... + \xi_\eta.$$

   Довести, що генератрисою випадкової величини ζ є $g(Q(z))$.

3. Знайти характеристичну функцію розподілу Лапласа, який має щільність:

$$f(x) = \frac{1}{2}e^{-|x|}.$$

4. Знайти характеристичну функцію розподілу зі щільністю:

$$f(x) = \begin{cases} |x|, & |x| \le 1, \\ 0, & |x| > 1. \end{cases}$$

5. Знайти характеристичну функцію випадкової величини $\xi - \eta$, якщо ξ, η є незалежними і мають однакову характеристичну функцію $\varphi(t)$.

6. Випадкова величина ξ приймає значення 2 і 5 з ймовірностями відповідно $\frac{1}{3}$, $\frac{2}{3}$. Знайти її генератрису і характеристичну функцію.

7. За допомогою характеристичних функцій знайти моменти стандартного нормального розподілу N(0,1) і розподілу гамма.

8. Випадкова величина ξ має генератрису $P(z)$. Знайти генератриси випадкових величин $\xi+2$, 2ξ.

9. Довести теорему 2.

10. Використовуючи таблицю характеристичних функцій, визначити розподіл випадкової величини, характеристична функція якої має вигляд:

    а) $\varphi(t) = e^{e^{it}-1}$, б) $\varphi(t) = \frac{1+e^{it}}{2}$, в) $\varphi(t) = e^{-|t|}$, г) $\varphi(t) = \frac{\sin t}{t}$,

    д)
$$\varphi(t) = \begin{cases} 1-|t| & \text{dla} \quad |t| \le 1, \\ 0 & \text{dla} \quad |t| > 1. \end{cases}$$

11. Довести, що характеристична функція приймає дійсні значення тоді і тільки тоді, коли розподіл випадкової величини є симетричним. (розподіл випадкової величини ξ називається симетричним, якщо функції розподілу випадкових величин ξ та $-\xi$ є ідентичними, тобто, $F(x) = \mathbf{P}(\xi < x) = \mathbf{P}(-\xi < x) = 1 - \mathbf{P}(\xi \ge -x) = 1 - F(-x+0)$)

12. Довести теорему 2.

Лекція 14. Подальші властивості характеристичних функцій

14.1 Формула обернення

Як ми вже з'ясували, кожній функції розподілу відповідає своя характеристична функція. Головною метою цього параграфу є з'ясування таких питань: "Як для заданої характеристичної функції знайти функцію розподілу, яка їй відповідає, і скільки таких функцій розподілу існує? ". Формули, за якими можна для заданої характеристичної функції знайти відповідну функцію розподілу, називаються „ формулами обернення ”.

Теорема 1. *Нехай $\varphi(t)$ є характеристичною функцією для функції розподілу $F(x)$, тоді для довільних $x < y$, в яких функція F є неперервною, будемо мати:*

$$F(y) - F(x) = \frac{1}{2\pi} \lim_{c\to\infty} \int\limits_{-c}^{c} \frac{e^{-itx} - e^{-ity}}{it} \varphi(t) dt.$$

Д о в е д е н н я. З визначення характеристичної функції і після простих перетворень отримаємо:

$$J_c = \frac{1}{2\pi} \int\limits_{-c}^{c} \frac{e^{-itx} - e^{-ity}}{it} \varphi(t) dt = \frac{1}{2\pi} \int\limits_{-c}^{c} \int\limits_{-\infty}^{\infty} \frac{e^{it(z-x)} - e^{it(z-y)}}{it} dF(z) dt =$$

$$= \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \int\limits_{-c}^{c} \frac{e^{it(z-x)} - e^{it(z-y)}}{it} dt dF(z) = \int\limits_{-\infty}^{\infty} \Psi(c,z,x,y) dF(z), \qquad (1)$$

де

$$\Psi(c,z,x,y) = \frac{1}{\pi} \int\limits_{0}^{c} \frac{\sin t(z-x)}{t} - \frac{\sin t(z-y)}{t} \; dt.$$

З аналізу відомо, що

$$\frac{1}{\pi} \lim_{c\to\infty} \int\limits_{0}^{c} \frac{\sin(\alpha t)}{t} dt = \begin{cases} \frac{1}{2}, & \alpha > 0, \\ -\frac{1}{2}, & \alpha < 0 \end{cases} \qquad (2)$$

і ця збіжність є рівномірною стосовно α для всіх $|\alpha| \geq \delta > 0$ з фіксованим δ. Крім того, для всіх $|\alpha| \leq \delta$ і всіх достатньо великих $c > 0$ будемо мати:

$$\frac{1}{\pi}\int_0^c \frac{\sin(\alpha t)}{t}dt \ \leq 1. \qquad (3)$$

Нехай $\delta > 0$ є фіксованим і таким, що $x + \delta < y - \delta$. Маємо

$$J_c = \int_{-\infty}^{x-\delta} + \int_{x-\delta}^{x+\delta} + \int_{x+\delta}^{y-\delta} + \int_{y-\delta}^{y+\delta} + \int_{y+\delta}^{\infty} \Psi(c,z,x,y)dF(z).$$

Звідки

$$|J_c - F(y) + F(x)| \leq \int_{-\infty}^{x-\delta} + \int_{y+\delta}^{\infty} \Psi(c,z,x,y)dF(z) \ + \int_{x-\delta}^{x+\delta}\Psi(c,z,x,y)dF(z) \ +$$

$$+ \int_{y-\delta}^{y+\delta}\Psi(c,z,x,y)dF(z) \ + \int_{x+\delta}^{y-\delta}\Psi(c,z,x,y)dF(z) - F(y) + F(x) \ . \qquad (4)$$

З визначення функції $\Psi(c,z,x,y)$ і зі співвідношень (2), (3) випливають місце наступні властивості цієї функції:

1. $|\Psi(c,z,x,y)| \leq 2$ для всіх z, x, y і достатньо великих c;
2. $\lim_{c\to\infty} \Psi(c,z,x,y) = 0$, якщо $-\infty < z < x - \delta$, або $y + \delta < z < \infty$;
3. $\lim_{c\to\infty} \Psi(c,z,x,y) = 1$, якщо $x + \delta < z < y - \delta$.

З властивостей функції $\Psi(c,z,x,y)$ і (4) випливає:

$$\overline{\lim}_{c\to\infty}|J_c - F(y) + F(x)| \leq 2 \ F(x+\delta) - F(x-\delta) \ +$$

$$+ 2 \ F(y+\delta) - F(y-\delta) \ + \ F(y-\delta) - F(y) + F(x) - F(x+\delta) \ . \qquad (5)$$

Позаяк x, y є точками неперервності функції $F(x)$, то права частина в (5) прямує до нуля, якщо $\delta \to 0$. Цей факт завершує доведення нашого твердження.

Зауваження 1. Границю в теоремі 1 називають „ інтегралом у сенсі Коші ” функції $\varphi(t)(it)^{-1}(\exp(-ixt) - \exp(-iyt))$.

14.1.1 Застосування формул обернення

З теореми про обернення легко випливає дуже важливий результат, який часом називают *теоремою про однозначність.*

Теорема 2. *Характеристична функція випадкової величини однозначно визначає її функцією розподілу.*

Наступний результат часто використовується в математичній статистиці.

Теорема 3. *Нехай випадкові величини ξ_k, $k = 1, \dots, n$ є незалежними і $\xi_k \in N(m_k, \sigma_k)$. Тоді для довільних a_k, $k = 1, \dots, n$ маємо:*

$$\sum_{k=1}^{n} a_k \xi_k \in N(M, D^2), \quad M = \sum_{k=1}^{n} a_k m_k \qquad D^2 = \sum_{k=1}^{n} a_k^2 \sigma_k^2.$$

Д о в е д е н н я наведемо тільки для $n = 2$. Випадкова величина ξ_1 має характеристичну функцію виду $\exp t(-\sigma_1^2/2 + im_1)$, і згідно з пунктом 2 теореми 1 стор. 121 характеристична функція випадкової величини $a_1\xi_1$ є наступною: $\exp t(-a_1^2\sigma_1^2/2 + ia_1m_1)$. Аналогічно для випадкової величини $a_2\xi_2$ характеристична функція має вигляд: $\exp t(-a_2^2\sigma_2^2/2 + im_2a_2)$. Згідно з пунктом 5 цієї самої теореми будемо мати, що характеристична функція суми $a_1\xi_1 + a_2\xi_2$ має вигляд:

$$\varphi_{a_1\xi_1 + a_2\xi_2}(t) = e^{t(-a_1^2\sigma_1^2/2 + ia_1m_1)} e^{t(-a_2^2\sigma_2^2/2 + ia_2m_2} = e^{t \quad -(a_1^2\sigma_1^2 + a_2^2\sigma_2^2)/2 + i(a_1m_1 + a_2m_2)} \quad .$$

Останні вираз в наведенних рівностях є характеристичною функцією нормального розподілу з параметрами $a_1m_1 + a_2m_2$, $a_1^2\sigma_1^2 + a_2^2\sigma_2^2$, отже, на підставі теореми про однозначність маємо: $a_1\xi_1 + a_2\xi_2 \in N(a_1m_1 + a_2m_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2)$, що й завершує доведення. ◄

14.2 Теорема про неперервність

Метою цього пункту є доведення одного з найважливіших тверджень теорії ймовірностей.

Теорема 4. *Нехай задана послідовність функцій розподілу $F_n(x)$, $n \geq 1$ і відповідна їм послідовність характеристичних функцій $\varphi_n(t)$. Якщо $\varphi_n(t) \to \varphi(t)$ при $n \to \infty$ для кожного t і функція $\varphi(t)$ є неперервною в точці $t = 0$, то*

1. *φ є характеристичною функцією певної функції розподілу F,*
2. *$F_n \xrightarrow[n\to\infty]{\mathrm{d}} F$.*

І навпаки, якщо $F_n \xrightarrow[n\to\infty]{\mathrm{d}} F$ і F є функцією розподілу, то $\varphi_n(t) \to \varphi(t)$, при чому $\varphi(t)$ є характеристичною функцією функції розподілу F.

Сенс цього твердження можна прокоментувати наступним чином. Нехай задана послідовність випадкових величин ξ_n, $n \geq 1$. Стоїть задача вивчити збіжність функцій розподілу $F_n(x) = \mathbf{P}(\xi_n < x)$ при $n \to \infty$. Отже, треба з'ясувати, чи існує така функція розподілу $F(x)$, до якої б прямувала послідовність $F_n(x)$ для кожного x (або для тих x, які є точками неперервності $F(x)$ – слабка збіжність). Знаходження явних виразів для функцій розподілу $F_n(x)$ є часто справою дуже важкою. Теорема 4 дає нам іншу можливість розв'язання цієї задачі. А саме, треба знайти характеристичні функції $\varphi_n(t)$ випадкових величин ξ_n (а це часто набагато легше зробити), а потім перевірити, чи існує границя $\lim_{n\to\infty}\varphi_n(t)$. Якщо ця границя існує і дорівнює $\varphi(t)$, то перевіряємо її додатково ще на неперервність у точці $t=0$. Якщо $\varphi(t)$ є в цій точці неперервною, то автоматично $\varphi(t)$ є характеристичною функцією, і послідовність $F_n(x)$ збігається до функції розподілу, якій відповідає $\varphi(t)$. Якщо ж $\lim_{n\to\infty}\varphi_n(t)$ не існує, то F_n не збігається навіть слабко до жодної функції розподілу. Більша частина граничних тверджень теорії ймовірностей і математичної статистики спирається на Твердження 4.

Д о в е д е н н я теореми 4. Друга частина цієї теореми є фактично другою теоремою Хеллі,отже, не потребує доведення. У зв'язки з чим, доведемо тільки пункти 1, 2.

Перша теорема Хеллі стверджує, що завжди існує підпослідовність $F_{n_k}(x)$, $k\geq 1$ і неспадна функція $F(x)$, $0\leq F(-\infty)$, $F(\infty)\leq 1$ такі, що $\lim_{k\to\infty}F_{n_k}(x)=F(x)$ у кожній точці неперервності функції $F(x)$. Доведемо, що функція $F(x)$ є функцією розподілу. Дійсно, треба показати, що $F(\infty)-F(-\infty)=1$. (Позаяк $0\leq F_{n_k}(x)|\leq 1$, то $0\leq F(x)\leq 1$ і, якщо $F(\infty)-F(-\infty)=1$, то з монотонності функції $F(x)$ випливає, що це є можливим лише тоді, коли $F(-\infty)=0$, $F(\infty)=1$). Нехай $F(\infty)-F(-\infty)=\delta<1$. Оскільки $1=\lim_{k\to\infty}\varphi_{n_k}(0)=\varphi(0)$ і функція $\varphi(t)$ є неперервною в нулі, то існують $\tau,\varepsilon>0$ такі, що

$$\frac{1}{2\tau}\int\limits_{-\tau}^{\tau}\varphi(t)dt>\delta+\frac{\varepsilon}{2}. \tag{6}$$

Нехай $K>2/\tau\varepsilon$ таке, що K і $-K$ є точками неперервності функції $F(x)$. Тоді $\lim_{k\to\infty}(F_{n_k}(K)-F_{n_k}(-K))=F(K)-F(-K)\leq\delta$.

Позаяк

$$\int\limits_{-\tau}^{\tau}e^{itx}dt\ \leq 2\tau,\quad \text{i}\quad \int\limits_{-\tau}^{\tau}e^{itx}dt\ =\ \frac{2\sin\tau x}{x}\ \leq\frac{2}{K},\quad \text{dla } x>K,$$

то

$$\int\limits_{-\tau}^{\tau}\varphi(t)dt\ =\lim_{k\to\infty}\int\limits_{-\tau}^{\tau}\varphi_{n_k}(t)dt\ =\lim_{k\to\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\tau}^{\tau}e^{itx}dtdF_{n_k}(x)\ \leq$$

$$\leq\lim_{k\to\infty}\int\limits_{|x|\leq K}\int\limits_{-\tau}^{\tau}e^{itx}dt\ dF_{n_k}(x)+\lim_{k\to\infty}\int\limits_{|x|>K}\int\limits_{-\tau}^{\tau}e^{itx}dt\ dF_{n_k}(x)\leq 2\tau\delta+\frac{2}{K}\leq 2\tau\delta+\tau\varepsilon.$$

Отже,

$$\frac{1}{2\tau}\int\limits_{-\tau}^{\tau}\varphi(t)dt\ \leq\delta+\frac{\varepsilon}{2},$$

що суперечить (6).

ЗАВДАННЯ

1. Знайти розподіл суми двох незалежних випадкових величин з розподілом Коші.

2. Знайти розподіл суми двох незалежних випадкових величин з геометричним розподілом.

3. Довести за допомогою характеристичних функцій аналог теореми 3 для розподілу Пуассона.

4. Для розподілів: показникового, Ерланга, рівномірного, трикутного, нуля-одиниці, біноміального, геометричного дослідити: чи сума двох незалежних випадкових величин з таким розподілом буде мати розподіл цього ж типу?

5. Випадкова величина має характеристичну функцію $\varphi(t) = e^{-\frac{1}{2}t^2}$. Знайти її розподіл.

6. Випадкова величина має характеристичну функцію $\varphi(t) = \cos(t)$. Знайти її розподіл.

7. Нехай послідовність $\varphi_n(t)$, $n = 1, 2, \ldots$ характеристичних функцій є такою, що $\varphi_n(t) \underset{n\to\infty}{\longrightarrow} 1$ для $-\delta < t < \delta$, $\delta > 0$. Довести, що тоді $\varphi_n(t) \underset{n\to\infty}{\longrightarrow} 1$ для всіх t.

8. Нехай ξ є числом успіхів в схемі Бернуллі з параметрами p, N, де ймовірність успіху p є сталою, а параметр N має розподіл Пуассона з параметром a. Знайти розподіл випадкової величини ξ.

Лекція 15. Закон великих чисел

15.1 Слабкий закон великих чисел.

Визначення 1. Нехай $\xi_n, n = 1, 2, ...$ є послідовністю випадкових величин, для яких $\mathbf{M}\xi_n < \infty,\ n = 1, 2, ...$ Будемо казати, що для послідовності $\xi_n, n = 1, 2, ...$ має місце *слабкий закон великих чисел* , якщо

$$\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mathbf{M}\xi_k) \xrightarrow[n\to\infty]{\mathrm{p}} 0.$$

Зауваження 1. З цього визначення випливає, що для великих n послідовність випадкових величин $\frac{1}{n}\sum_{k=1}^{n}\xi_k$ поводить себе як числова послідовність $c_n = \frac{1}{n}\sum_{k=1}^{n}\mathbf{M}\xi_k$.

Якщо випадкові величини є однаково розподіленими, то слабкий закон великих чисел можна записати у вигляді:

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow[n\to\infty]{\mathrm{p}} \mathbf{M}\xi_1.$$

Якщо ж додатково випадкові величини є незалежними, то має місце наступне твердження.

Теорема 1. (Хінчина). *Якщо випадкові величини ξ_k, $k = 1, 2, \ldots$ є незалежними, однаково розподіленими і $\mathbf{M}\xi_k = a$, то*

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow[n\to\infty]{\mathrm{p}} \mathbf{M}\xi_1 = a. \tag{1}$$

Д о в е д е н н я. Якщо $\varphi(t)$ є характеристичною функцією ξ_i, то на підставі формули (5) (стор. 123) маємо: $\lim t^{-1}(1 - \varphi(t)) = \varphi'(0) = i\mathbf{M}\xi_1 = ia$, звідки

$$\varphi(t) = 1 + iat + o(t) \qquad\qquad t \to o.$$

Нехай $S_n = \frac{1}{n}\sum_{k=1}^{n}\xi_k$. Тоді

$$\mathbf{M}e^{iS_n} = \ \mathbf{M}e^{i\xi_1/n}\ ^{2} = \ \varphi(t/n)\ ^{n} = \ 1 + \frac{iat}{n} + o(\frac{t}{n})\ ^{n} \xrightarrow[n\to\infty]{} e^{iat} \tag{2}$$

Позаяк e^{iat} є характеристичною функцією випадкової величини $\xi = a = const$, то на підставі теореми 2 (стор. 116) будемо мати: $S_n \xrightarrow[n\to\infty]{\mathrm{p}} a$. ◀

Наслідок 1. *(* (Бернуллі) *Нехай в схемі випробувань Бернуллі з ймовірністю успіху p випадкова величина μ_n описує число успіхів в n випробуваннях. Тоді*

$$\frac{\mu_n}{n} \xrightarrow[n\to\infty]{\mathrm{p}} p. \tag{3}$$

Д о в е д е н н я. Схему випробувань Бернуллі можна представити як n незалежних випробувань типу "успіх-неуспіх". Визначимо випадкові величини η_k, $k = 1, 2, \ldots, n$ наступним чином:

$$\eta_k = \begin{matrix} 1, & \text{якщо в } k\text{-тому випробуванні був успіх,} \\ 0, & \text{у протилежному випадку.} \end{matrix}$$

Оскільки, згідно з визначенням схеми Бернуллі, всі n випробувань є незалежними і відбуваються в однакових умовах, то випадкові величини η_k, $k = 1, 2, \ldots, n$ є також незалежними і мають однаковий розподіл: $\mathbf{P}(\eta_k = 1) = p$, $\mathbf{P}(\eta_k = 0) = 1-p$. Вочевидь, $\mu_n = \sum_{k=1}^{n} \eta_k$. Позаяк $\mathbf{M}\eta_k = p$, $\mathbf{D^2}\eta_k = (1-p)p$, то є виконаними припущення попереднього наслідку для послідовності η_k, $k = 1, 2, \ldots, n$. Отже,

$$\frac{\mu_n}{n} = \frac{1}{n}\sum_{k=1}^{n} \eta_k \xrightarrow[n\to\infty]{\mathrm{p}} \mathbf{M}\eta_1 = p.$$

◀

Якщо випадкові величини $\xi_n, n = 1, 2, \ldots$ є незалежними, але з різними розподілами, то має місце наступний результат.

Теорема 2. (Чебишова) *Нехай $\xi_n, n = 1, 2, \ldots$ є послідовністю незалежних випадкових величин таких, що $\mathbf{D^2}\xi_n \le C, n = 1, 2, \ldots$ Тоді*

$$\frac{1}{n}\sum_{k=1}^{n} (\xi_k - \mathbf{M}\xi_k) \xrightarrow[n\to\infty]{\mathrm{p}} 0. \tag{4}$$

Д о в е д е н н я. Позначимо

$$S_n = \frac{1}{n}\sum_{k=1}^{n} (\xi_k - \mathbf{M}\xi_k).$$

Тоді

$$\mathbf{M}S_n = 0, \quad \text{і} \quad \mathbf{D^2}S_n = \frac{1}{n^2}\sum_{k=1}^{n} \mathbf{D^2}\xi_k < \frac{C}{n}.$$

На підставі нерівності Чебишова для довільного $\varepsilon > 0$ маємо:

$$\mathbf{P}(|S_n| \ge \varepsilon) \le \frac{\mathbf{D^2}S_n}{\varepsilon^2} \le \frac{C}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0,$$

що й доводить збіжність (4). ◀

Зауваження 2. Доведене твердження залишається в силі, якщо замість $\mathbf{D^2}\xi_n \le C,\ n = 1, 2, \ldots$ вимагати виконання умови: $n^{-2}\sum_{k=1}^{n} \mathbf{D^2}\xi_k \xrightarrow[n\to\infty]{} 0$ (завдання 1).

15.2 Посилений закон великих чисел

Визначення 2. *Посиленим законом великих чисел* називають твердження, що різниця

$$\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mathbf{M}\xi_k) \xrightarrow[n\to\infty]{\text{м.н}} 0.$$

Наступне твердження дає достатні умови для виконання посиленого закону велихих чичел. Відмітимо, що в цьому твердженні не вимагається щоб випадкові величини були одинаклво розподілені.

Теорема 3. *Нехай задана послідовність незалежних випадкових величин $\xi_1, \xi_2, \ldots, \mathbf{M}\xi_k = m_k$ і понадто:*

$$\sum_{k=1}^{\infty}\frac{\mathbf{D^2}\xi_k}{k^2} < \infty.$$

Тоді

$$\frac{1}{n}\sum_{k=1}^{n}(\xi_k - m_k) \xrightarrow[n\to\infty]{\text{м.н}} 0, \tag{5}$$

отже, для послідовності $\xi_1, \xi_2, \ldots$ має місце посилений закон великих чисел.

Д о в е д е н н я. Позначимо $S_n = \sum_{k=1}^{n}(\xi_k - \mathbf{M}\xi_k)$. Достатньо показати, що для довільного $\varepsilon > 0$

$$\lim_{n\to\infty}\mathbf{P}(\sup_{k\geq n}\frac{S_k}{k} > \varepsilon) = 0. \tag{6}$$

Нехай $A_n = \{\max_{2^n\leq k<2^{n+1}}\frac{S_k}{k} > \varepsilon\}$. Зрозуміло, що

$$\{\omega : \sup_{k\geq n}\frac{S_k}{k} > \varepsilon\} = \cup_{k=n}^{\infty}A_k \tag{7}$$

для довільного n. Використовуючи нерівність Колмогорова (стор. 84) маємо

$$\mathbf{P}(A_n) \leq \mathbf{P}(\max_{2^n\leq k<2^{n+1}} S_k > 2^n\varepsilon) \leq \mathbf{P}(\max\{|S_1|, ..., |S_{2^{n+1}}|\} > 2^n\varepsilon) \leq$$

$$\leq \frac{\mathbf{D^2}S_{2^{n+1}}}{\varepsilon^2 2^{2n}} = \frac{\sum_{j=1}^{2^{n+1}}\mathbf{D^2}\xi_j}{\varepsilon^2 2^{2n}}.$$

Отже

$$\sum_{n=1}^{\infty}\mathbf{P}(A_n) \leq \sum_{n=1}^{\infty}\frac{\sum_{j=1}^{2^{n+1}}\mathbf{D^2}\xi_j}{\varepsilon^2 2^{2n}} = \varepsilon^{-2}\sum_{j=1}^{\infty}\mathbf{D^2}\xi_j \sum_{n:2^{n+1}\geq j} 2^{-2n}.$$

Остання сума в цій нерівності менша подвоєного першого доданку а тому $\sum_{n:2^{n+1}\geq j} 2^{-2n} \leq 8j^{-2}$ що дає

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) \leq 8\varepsilon^{-2} \sum_{j=1}^{\infty} \frac{\mathbf{D^2}\xi_j}{j^2} < \infty$$

З леми Бореля-Кантелі випливє, що відбувається лише скінченна кількість подій A_n. Іншими словами, для деякого n_0(взагалі кажучи, випадкового) події $A_n,\ n \geq n_0$ не відбуваються з ймовірністю 1, а тому

$$\mathbf{P}(\max_{2^n \leq k < 2^{n+1}} \frac{S_k}{k} \leq \varepsilon) = \mathbf{P}(\overline{A}_n) = 1, \qquad n \geq n_0$$

З (7) тепер дістаємо

$$\mathbf{P}(\sup_{k \geq n_0} \frac{S_k}{k} \leq \varepsilon) = \mathbf{P}(\cap_{k=n_0}^{\infty} \overline{A}_k) = 1$$

Оскільки ε тут довільне мале число, то це і завершує доведення нашої теореми. ◀

Для однаково розподілених випадкових величин посилений закон великих чисел має наступний вигляд:

$$\frac{1}{n}\sum_{k=1}^{n} \xi_k \underset{n\to\infty}{\overset{\text{м.н}}{\longrightarrow}} \mathbf{M}\xi_1.$$

А, якщо крім того випадкові величини $\xi_i,\ i = 1, ..., n$ є незалежними, то наступне твердження визначає необхідні і достатні умови, коли для цих випадкових величин має місце посилений закон великих чисел.

Теорема 4. (Посилений закон великих чисел Колмогорова) *Нехай $\xi_n,\ n = 1, 2, ...$ є незалежними, однаково розподіленими випадковими величинами. Для того, щоб*

$$\frac{1}{n}\sum_{k=1}^{n} \xi_k \underset{n\to\infty}{\overset{\text{м.н}}{\longrightarrow}} a \tag{8}$$

необхідно і достатньо існування математичного сподівання $\mathbf{M}\xi_k = a$.

Д о в е д е н н я. Д о с т а т н і с т ь. Для довільного натурального k означимо $\xi_k^* = \xi_k\chi(|\xi_k| \leq k)$ і нехай $S_n = \sum_{k=1}^{n} \xi_k$, $S_n^* = \sum_{k=1}^{n} \xi_k^*$. Покажемо, що зі збіжностіь $n^{-1}(S_n^* - \mathbf{M}S_n^*) \underset{n\to\infty}{\overset{\text{м.н}}{\longrightarrow}} 0$ випливає збіжність $n^{-1}(S_n - na) \underset{n\to\infty}{\overset{\text{м.н}}{\longrightarrow}} 0$.

Якщо $A_k = \{\xi_k \neq \xi_k^*\}$, то маємо

$$\sum_{k=1}^{\infty} A_k = \sum_{k=1}^{\infty} \mathbf{P}(|\xi_k| > k) = \sum_{k=1}^{\infty}\sum_{j=k}^{\infty} \mathbf{P}(j < |\xi_1| \leq j+1) = \sum_{j=1}^{\infty} j\mathbf{P}(j < |\xi_1| \leq j+1) \leq \mathbf{M}|\xi_1| \leq \infty.$$

Звідки, згідно леми Бореля-Кантеллі, випливає, що відбувається лише скнченна кількість подїї $\{\xi_k \neq \xi_k^*\}$. А це, в свою чергу, означає, що для кожного $\omega \in \Omega \backslash B$, де $\mathbf{P}(B) = 0$ маємо $\xi_n(\omega) = \xi_n^*(\omega)$, $n \geq n_0$, для деякого скінченного числа n_0 (різного для різних ω), і тому для даного ω маємо

$$\frac{S_n}{n} = \frac{\sum_{k=1}^{n_0} \xi_k - \sum_{k=1}^{n_0} \xi_k^*}{n} + \frac{\sum_{k=1}^{\infty} \xi_k^*}{n} = \frac{\sum_{k=1}^{n_0} \xi_k - \sum_{k=1}^{n_0} \xi_k^*}{n} + \frac{S_n^*}{n},$$

а тому з ймовірністю 1

$$\lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} \frac{S_n^*}{n}. \tag{9}$$

Очевидно $\lim_{k\to\infty} \mathbf{M}\xi_k^* = \mathbf{M}\xi_k = a$, а тому $\lim_{n\to\infty} \mathbf{M}S_n^*/n = a$, що разом з (9) дає

$$\lim_{n\to\infty} \frac{S_n - na}{n} = \lim_{n\to\infty} \frac{S_n^* - \mathbf{M}S_n^*}{n}.$$

Оскільки

$$\mathbf{D^2}\xi_n^* \leq \mathbf{M}(\xi_n^*)^2 \leq \sum_{k=1}^{n} k^2 \mathbf{P}(k-1 \leq |\xi_1| < k),$$

то

$$\sum_{n=1}^{\infty} \frac{\mathbf{D^2}\xi_n^*}{n} \leq \sum_{n=1}^{\infty}\sum_{k=1}^{n} \frac{k^2}{n^2}\mathbf{P}(k-1 \leq \xi_1| < k) = \sum_{k=1}^{\infty} k^2\mathbf{P}(k-1 \leq |\xi_1| \leq k) \sum_{n=k}^{\infty} \frac{1}{n^2} =$$

$$= \mathbf{P}(|\xi_1| \leq 1) \sum_{n=1}^{\infty} \frac{1}{n^2} + \sum_{k=2}^{\infty} k^2 \mathbf{P}(k-1 \leq |\xi_1| < k) \sum_{n=k}^{\infty} \frac{1}{n^2} \leq$$

$$\leq (\text{ в цьому місці використовуємо нерівність} \sum_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{2}{k},\ k \geq 2)$$

$$\leq \sum_{n=1}^{\infty} \frac{1}{n^2} + \sum_{k=2}^{\infty} k\mathbf{P}(k-1 \leq |\xi_1| < k) \leq 1 + \sum_{n=1}^{\infty} \frac{1}{n^2} + \sum_{k=2}^{\infty} (k-1)\mathbf{P}(k-1 \leq |\xi_1| < k) \leq$$

$$\leq 1 + \sum_{n=1}^{\infty} \frac{1}{n^2} + \mathbf{M}|\xi_1| < \infty.$$

Отже для послідовності випадкових величин $\xi_n^*,\ n = 1, 2, ...$ виконуються умови попередньої теореми а тому для неї, а отже і для $\xi_n,\ n = 1, 2, ...$, має місце посилений закон великих чисел.

Н е о б х і д н і с т ь. Нехай тепер виконується (8). Оскільки

$$\frac{\xi_n}{n} = \frac{S_n}{n} - \frac{n-1}{n}\frac{S_{n-1}}{n-1} \xrightarrow[n\to\infty]{\text{м.н}} 0,$$

то подія $\{|\xi_n|/n > 1\}$ відбувається скінченну кількість раз, що згідно леми Бореля-Кантеллі, означає збіжність ряду $\sum_{n=1}^{\infty} \mathbf{P}(|\xi_n| > n)$, а тому

$$\mathbf{M}|\xi_1| \leq \sum_{n=1}^{\infty} (n+1)\mathbf{P}(n \leq |\xi_n| < n+1) = 1 + \sum_{n=1}^{\infty} \mathbf{P}(|\xi_n| > n) < \infty.$$

◄

ЗАВДАННЯ

1. Довести зауваження 2.

2. Послідовність $\xi_n,\ n \geq 1$ випадкових величин (взагалі кажучи, залежних) така, що

$$\lim_{n\to\infty} n^{-2}\mathbf{D^2}\Big(\sum_{k=1}^{n}\xi_k\Big) = 0.$$

Довести, що для цієї послідовності має місце слабкий закон великих чисел.

3. Маємо послідовність $\xi_n,\ n \geq 1$ незалежних випадкових величин, рівномірно розподілених на $[0;1]$ Довести, що $e^n \prod_{k=1}^{n} \xi_k \xrightarrow[n\to\infty]{\text{р}} 1$.

4. Маємо послідовність $\xi_n,\ n \geq 1$ незалежних випадкових величин з показниковим розподілом з параметром λ. Довести, що $\sqrt[n]{\prod_{k=1}^{n}\xi_k} \xrightarrow[n\to\infty]{\text{м.н}} e^{-c}$, $c = \int\limits_0^\infty e^{-x}\ln x dx$ –стала Ейлера.

5. Нехай $\xi_n,\ n \geq 1$ послідовність незалежних випадкових величин таких, що

$$\mathbf{P}(\xi_n = -n) = \mathbf{P}(\xi_n = n) = 1/n, \qquad \mathbf{P}(\xi_n = 0) = 1 - 2/n.$$

а) Довести, що $\sum_{n=2}^{\infty} n^{-2}\mathbf{D^2}\xi_n = \infty$.

б) Посилений закон великих чисел для послідовності $\xi_n,\ n \geq 1$ не виконується.

6. Довести, що необхідною умовою виконання для послідовності незалежних випадкових величин $\xi_n,\ n \geq 1$ з $\mathbf{M}|\xi_n| < \infty,\ \ n \geq 1$ посиленого закону великих чисел є умова $\frac{\xi_n - \mathbf{M}\xi_n}{n} \xrightarrow[n\to\infty]{\text{м.н}} 0$.

Лекція 16. Центральна гранична теорема

16.1 Центральна гранична теорема

У попередній лекції ми довели, що за певних умов стосовно випадкових величин $\xi_n, n = 1, 2, ...$ сума $S_n = \xi_1 + \xi_2 + \ldots + \xi_n$ помножена на n^{-1} при великих n поводить себе як певна числова послідовність (або навіть певне число). Хоча константу теж можна розглядати як певну випадкову величину, яка з ймовірністю 1 приймає тільки одне значення, отже, є, як кажуть, "виродженою". Природньо постає питання, а чи можна знайти числову функцію $f(n)$ таку, що випадкова величина $f(n)S_n$ при великих n поводить себе як певна невироджена випадкова величина. Наступне твердження, яке є одним з найважливіших у теорії ймовірностей, описує таку ситуацію.

Теорема 1. *Якщо ξ_n є послідовністю незалежних, однаково розподілених випадкових величин і $\mathbf{M}\xi_n = a$, $\mathbf{D^2}\xi_n = \sigma^2$, то*

$$\mathbf{P}\left(\frac{\sum_{i=i}^{n}\xi_i - an}{\sigma\sqrt{n}} < x\right) \underset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}}du \tag{1}$$

рівномірно стосовно x, $-\infty < x < \infty$.

Д о в е д е н н я. Позаяк випадкові величини $\eta_k = (\xi_k - a)/\sigma$ є незалежними, однаково розподіленими і $\mathbf{M}\eta_k = 0$, $\mathbf{D^2}\eta_k = 1$, $k = 1, 2, \ldots n$, а

$$\frac{\sum_{k=1}^{n}\xi_k - an}{\sigma\sqrt{n}} = \frac{\sum_{k=1}^{n}\eta_k}{\sqrt{n}},$$

то доведення достатньо провести для $a = 0$ і $\sigma^2 = 1$. Нехай

$$S_n = \frac{\xi_1 + \xi_2 + ... + \xi_n}{\sqrt{n}} \qquad \text{і} \qquad F_n(x) = \mathbf{P}(S_n < x).$$

Оскільки гранична функція в (1) є неперервною для $x \in [-\infty, \infty]$, то рівномірна збіжність в (1) буде випливати зі звичайної збіжності (завдання 2.1), доведення якої ми наведемо.

Якщо $\varphi(t)$ є характеристичною функцією випадкової величини ξ_i, то згідно з пунктом 5 теореми 1 (стор. 121) отримаємо:

$$\varphi_{S_n}(t) = \mathbf{M}e^{itS_n} = \prod_{k=1}^{n}\mathbf{M}^{it\xi_k/\sqrt{n}} = \varphi^n(\tfrac{t}{\sqrt{n}}).$$

Позаяк має місце розклад (див. [?], стор. 577)

$$\psi(t) = 1 - t^2/2 + o(t^2),$$

то

$$\varphi_{S_n}(t) = \varphi^n(\tfrac{t}{\sqrt{n}}) = (1 - t^2/2n + o(t^2/n))^n \underset{n\to\infty}{\longrightarrow} e^{-t^2/2}.$$

Функція $\exp(-t^2/2)$ є характеристичною функцією стандартного нормального розподілу, отже, на підставі теореми про неперервність (стор. 133) збіжність (1) є доведеною. ◄

Зауваження 1. В умовах щойно розглянутого твердження для випадкової величини $\eta_{k,n} \overset{def}{=} (\xi_k - \mathbf{M}\xi_k)/\sigma\sqrt{n}$ маємо:

$$\max_{k\le n} \mathbf{D^2}\eta_{k,n} = \frac{1}{\sqrt{n}} \underset{n\to\infty}{\longrightarrow} 0.$$

Це означає, що сума

$$S_n = \frac{\xi_1 + \xi_2 + ... + \xi_n - na}{\sigma\sqrt{n}} = \eta_{1,n} + \eta_{2,n} + ... + \eta_{n,n}$$

є сумую n „малих” незалежних випадкових величин, тобто таких, що їх дисперсія рівномірно прямує до нуля при $n \to \infty$. Доведене твердження показує, що така сума для великих n має бути нормальною випадковою величиною. Як виявилось, така ситуація є типовою. Отже, якщо є сума великої кількості випадкових величин, дисперсія яких рівномірно мала, то можемо очікувати, що розподіл цієї суми буде близьким до нормального.

Зауваження 2. В умовах попередньої теореми для довільних $a < b$ має місце:

$$\mathbf{P}(a \le \frac{\sum_{k=1}^n \xi_k - an}{\sigma\sqrt{n}} \le b) \underset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-\frac{u^2}{2}} du \qquad (2)$$

рівномірно стосовно x, $-\infty < x < \infty$.

Зауваження 3. Позаяк стандартний нормальний розподіл є неперервним, то збіжність (2) залишається в силі, якщо замість " $\le$ " написати " $<$ ".

Розглянемо схему випробувань Бернуллі, пов’язану з спостереженнями за подією A. Це означає, що в кожному окремому випробуванні нас цікавить лише подія A, яка відбувається з ймовірністю p. Нехай μ_n є числом реалізацій події A при n незалежних випробуваннях, $q = 1 - p$.

Теорема 2. (Інтегральна Муавра-Лапласа (Moivre-Laplace)) *Для довільних $a < b$:*

$$\lim_{n\to\infty} \mathbf{P}\Big(a \le \frac{\mu_n - np}{\sqrt{npq}} \le b\Big) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{u^2}{2}} du = \Phi(b) - \Phi(a), \qquad (3)$$

де $\Phi(x)$, як завжди, означає функцію розподілу стандартного нормального розподілу.

Д о в е д е н н я цієї теореми легко випливає з того, що (див. доведення наслідку 1, стор. 137)

$$\mu_n = \sum_{k=1}^{n} \xi_k,$$

де ξ_k є незалежними випадковими величинами з однаковим розподілом: $\mathbf{P}(\xi_k = 1) = p$, $\mathbf{P}(\xi_k = 0) = 1 - p$. Позаяк $\mathbf{M}\xi_k = p$, $\mathbf{D^2}\xi_k = pq$, то результат випливає з (2).

◄

Нерівність $a \le \mu_n \le b$ є рівнозначною

$$\frac{a - np}{\sqrt{npq}} \le \frac{\mu_n - np}{\sqrt{npq}} \le \frac{b - np}{\sqrt{npq}},$$

отже, з інтегральної теореми Муавра-Лапласа випливає, що для великих n маємо наближену рівність:

$$\mathbf{P}(a \le \mu_n \le b) \approx \Phi(\frac{b - np}{\sqrt{npq}}) - \Phi(\frac{a - np}{\sqrt{npq}}). \qquad (4)$$

Помилка, яка з'являється при застосуванні цієї формули (4), є зворотньо пропорційна до $\sqrt{npq}$ (див. [1], стор. 101), звідки випливає, що цю формулу належить застосовувати при великих n і, якщо p не є достатньо близьким до 0 чи 1. Іншими словами, число $\sqrt{npq}$ не повинно бути малим.

Подамо тепер умови збіжності до нормального розподілу послідовностей сум незалежних випадкових величин ξ_n, $n = 1, 2, ...$, якщо ξ_n не є однаково розподіленими. У цьому випадку припускаємо, що середнє значення і дисперсія кожної випадкової величини існують, але залежать від n, тобто, $\mathbf{M}\xi_n = a_n$, $\mathbf{D^2}\xi_n = \sigma_n^2$. Введемо наступні позначення: $F_n(x)$ – функція розподілу ξ_n, крім того,

$$B_n^2 = \sum_{k=1}^{n} \sigma_k^2, \quad S_n = \frac{\sum_{k=1}^{n} \xi_k - \sum_{k=1}^{n} a_k}{B_n}.$$

припустимо, що є виконаною наступна умова:

$\mathbf{L_0}$ (умова Ліндеберга (Lindeberg)). Для кожного $\tau > 0$

$$L_n(\tau) \stackrel{def}{=} \frac{1}{B_n^2}\sum_{k=1}^{n} \int\limits_{|x-a_k|\geq \tau B_n} (x-a_k)^2 dF_k(x) \underset{n\to\infty}{\longrightarrow} 0.$$

Зауваження 4. Якщо в умові $\mathbf{L_0}$ формально взяти $\tau = 0$, то ми б отримали:

$$L_n(0) \stackrel{def}{=} \frac{1}{B_n^2}\sum_{k=1}^{n} \int\limits_{|x-a_k|\geq 0} (x-a_k)^2 dF_k(x) = \frac{1}{B_n^2}\sum_{k=1}^{n}\sigma_k^2 = 1.$$

Отже, умова Ліндеберга вимагає, аби сумарна частина дисперсії випадкових величин ξ_k, $k = 1, 2, \ldots, n$ по області, яка виходить поза межі сумарного стандартного відхилення B_n, була малою в порівнянні з сумарною дисперсією B_n^2. У свою чергу це спричиняє рівномірну малість дисперсії випадкових величин $\xi_{k,n} = \xi_k - a_k.B_n$. Дійсно. Нехай $A_k(\tau, n) = \{x : |x - a_k| \geq \tau B_n\}$. Тоді

$$\max_{k\leq n} \mathbf{D^2}\xi_{k,n} = \max_{k\leq n}\Big[\frac{1}{B_n^2}\int\limits_{A_k(\tau,n)}(x-a_k)^2 dF_k(x) + \frac{1}{B_n^2}\sum_{k=1}^{n}\int\limits_{\overline{A}_k(\tau,n)}(x-a_k)^2 dF_k(x)\Big]$$
$$\leq L_n(\tau) + \tau^2.$$

Оскільки τ є довільним і виконана умова $\mathbf{L_0}$, це буде означати, що

$$\max_{k\leq n} \mathbf{D^2}\xi_{k,n} \underset{n\to\infty}{\longrightarrow} 0. \tag{5}$$

Позаяк легко зауважити, що

$$\frac{\sum_{k=1}^{n}\xi_k - \sum_{k=1}^{n} a_k}{B_n} = \sum_{k=1}^{n}\xi_{k,n},$$

то згідно з (5) можемо стверджувати, що умова Ліндеберга гарантує нам, що ми маємо суму рівномірно малих доданків, і для великих n така сума ймовірно теж мусить мати нормальний розподіл (див. зауваження 1).

Наступне твердження, яке ми подаємо без доведення, свідчить про те, що саме так є насправді.

Теорема 3. (Центральна гранична теорема). *Якщо ξ_n, $n = 1, 2, \dots$ є послідовністю незалежних випадкових величин з $\mathbf{M}\xi_n = a_n$, $\mathbf{D^2}\xi_n = \sigma_n^2 < \infty$, і виконана умова $\mathbf{L_0}$, то*

$$\mathbf{P}\Big(\frac{\sum_{k=1}^n (\xi_k - a_k)}{B_n} < x\Big) \underset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du$$

рівномірно стосовно x, $-\infty < x < \infty$.

Можна навести також інші достатні умови, щоб центральна гранична теорема для незалежних випадкових величин мала місце. Однією з таких умов є умова Ляпунова. Ця умова є більш обмежувальна ніж умова Ліндеберга, але у багатьох випадках є більш легкою для перевірки.

Теорема 4. (Ляпунова). *Якщо послідовність незалежних випадкових величин ξ_n, $\mathbf{M}\xi_n = a_n$, $n = 1, 2, \dots$ є такою, що для де-якого $\delta > 0$ $\mathbf{M}|\xi_k|^{2+\delta} < \infty$, і виконана умова:*

$$\frac{\sum_{k=1}^n \mathbf{M}|\xi_k - a_k|^{2+\delta}}{\sqrt{\sum_{k=1}^n \mathbf{M}(\xi_k - a_k)^2}} = \frac{\sum_{k=1}^n \mathbf{M}|\xi_k - a_k|^{2+\delta}}{B_n} \underset{n\to\infty}{\longrightarrow} 0, \tag{6}$$

то

$$\lim_{n\to\infty} \mathbf{P}\Big(\frac{\sum_{k=1}^n (\xi_k - a_k)}{B_n} < x\Big) \underset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du$$

рівномірно стосовно x, $-\infty < x < \infty$.

Д о в е д е н н я. Доведемо, що з умови (6) випливає умова Ліндеберга. Нехай $C_n = \sum_{k=1}^n \mathbf{M}|\xi_k - a_k|^{2+\delta}$ і $A_k(\tau, n) = \{x : |x - a_k| \geq \tau B_n\}$. Позаяк з $x \in A_k(\tau, n)$, тобто, $|x - a_k| \geq \tau B_n$, випливає

$$\Big[\frac{|x - a_k|}{\tau B_n}\Big]^{\delta} \geq 1,$$

то будемо мати:

$$L_n(\tau) = \frac{1}{B_n^2} \sum_{k=1}^{n} \int\limits_{A_k(\tau,n)} (x - a_k)^2 dF_k(x) \leq \frac{1}{B_n^2} \sum_{k=1}^{n} \frac{1}{(\tau B_n)^{\delta}} \int\limits_{A_k(\tau,n)} |x - a_k|^{2+\delta} dF_k(x) \leq$$

$$\leq \frac{C_n^{2+\delta}}{\tau^{\delta} B_n^{2+\delta}} = \tau^{-\delta} \ \frac{C_n}{B_n}^{\ 2+\delta} \underset{n\to\infty}{\longrightarrow} 0.$$

◄

16.2 Локальна гранична теорема

У твердженнях попереднього параграфу нас цікавила поведінка ймовірності $\mathbf{P}(\eta_n < x)$ при $n \to \infty$, де η_n є нормованою сумою незалежних випадкових величин. Якщо випадкові величини є дискретного типу, тоді є можливою ситуація, що $\mathbf{P}(\eta_n = a) \neq 0$ для всіх n. Виникає питання: "Як поводить себе ця ймовірність при зростаючому n?"Твердження, які відповідають на такі питання, називають *локальними*, на відміну від тверджень попереднього параграфу, які називають *інтегральними* (наприклад, теорема 2). Зауважимо, якщо серед доданків суми незалежних випадкових величин $\xi_1 + \xi_2 + \ldots + \xi_n$ хоча б одна має неперервний розподіл, то й сама сума має неперервний розподіл, отже, $\mathbf{P}(\xi_1 + \ldots + \xi_n = a) = 0$ для всіх a, і локальна гранична теорема втрачає сенс. Ось чому, у цьому параграфі припускаємо, що випадкові величини $\xi_1, \xi_2, \ldots$ є незалежними і мають дискретний розподіл.

Теорема 5. (Пуассона) *Якщо μ_n дорівнює числу випадків, коли сталась подія A при проведенні n незадежних випробувань, $p(n)$ - ймовірність цієї події в кожному окремому випробуванні, $q(n) = 1 - p(n)$, і виконана умова:* $\lim_{n\to\infty} np(n) = \lambda > 0$. *Тоді для довільного k має місце:*

$$\lim_{n\to\infty} \mathbf{P}(\mu_n = k) = \frac{e^{-\lambda} k^{\lambda}}{k!}.$$

◀

Це твердження вже було доведене на стор. 25. Його можна застосувати до наближеного підрахунку ймовірності $\mathbf{P}(\mu_n = k)$ для великих n і малих p, якщо $0 < np < 20$. Значення k не повинно бути великим. Інакше помилка може бути значною. Якщо умови на λ і k не виконуються, треба застосовувати наближену формулу:

$$\mathbf{P}(\mu_n = k) \simeq \frac{1}{\sqrt{2\pi}\sigma} \exp\Big[-\frac{1}{2}\Big(\frac{k-m}{\sigma}\Big)^2\Big], \tag{7}$$

де $m = np$, $\sigma = \sqrt{npq}$, $q = 1 - p$. Використання цієї формули обумовлене наступним результатом

Теорема 6. (локальне Муавра-Лапласа) *Якщо $k \to \infty$, $n - k \to \infty$ таким чином, що*

$$\lim_{\substack{n\to\infty \\ k\to\infty}} \frac{k - np}{\sqrt{npq}} = y, \tag{8}$$

то

$$\lim_{\substack{n\to\infty\\ k\to\infty}} \sqrt{npq}\,\mathbf{P}(\mu_n = k) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}. \tag{9}$$

Позаяк доведення цього твердження є чисто технічним, наведемо лище схему доведення. З формули Стірлінга (Stirling) маємо:

$$n! = \sqrt{2\pi n} n^n e^{-n}(1+\alpha_n),$$

де $\alpha_n \underset{n\to\infty}{\longrightarrow} 0$. Застосовуючи цю формулу, після простих перетворень отримаємо:

$$\sqrt{npq}\ \mathbf{P}(\mu_n = k) = \frac{1}{\sqrt{2\pi}} R(n,k)S(n,k)T(n,k), \tag{10}$$

де

$$R(n,k) = \sqrt{npq}\sqrt{\frac{n}{k(n-k)}}, \quad S(n,k) = \quad \frac{np}{k}^{\ \ k} \ \ \frac{nq}{n-k}^{\ \ n-k},$$
$$T(n,k) = \frac{1+\alpha_n}{(1+\alpha_k)(1+\alpha_{n-k})}.$$

Вочевидь, $T(n,k) \to 1$, якщо $n, k, n-k \to \infty$, а з (8) випливає

$$k = np + \sqrt{npq}\ y + \beta(n) \tag{11}$$

і $\beta(n) = o(\sqrt{n})$. Тепер маємо:

$$\sqrt{npq}R(n,k) = \frac{\sqrt{pq}}{\sqrt{(p+y\sqrt{pq/n}+\beta(n)/n)((1-p)-y\sqrt{pq/n}-\beta(n)/n)}} \underset{n\to\infty}{\longrightarrow} \frac{\sqrt{pq}}{\sqrt{p(1-p)}} = 1. \tag{12}$$

З відомих результатів математичного аналізу:

$$1+\frac{a}{x}+o(x^{-1})^{\quad xb} \underset{x\to\infty}{\longrightarrow} e^{ab}, \qquad 1+\frac{a}{x}+o(x^{-1})^{\quad o(x)} \underset{x\to\infty}{\longrightarrow} 1,$$

легко довести, що

$$\frac{k}{np}^{\ \ -k} \sim \ \ 1+\frac{y\sqrt{q/p}}{\sqrt{n}}+\frac{\beta(n)}{np}^{\ \ -np} e^{-y^2 q}$$

і

$$\frac{nq}{n-k}^{\ \ n-k} \sim \ \ 1-\frac{y\sqrt{p/q}}{\sqrt{n}}-\frac{\beta(n)}{nq}^{\ \ -nq} e^{-y^2 p},$$

і, отже,

$$S(n,k) \sim \ \ 1+\frac{y\sqrt{q/p}}{\sqrt{n}}+\frac{\beta(n)}{np}^{\ \ -np} \ \ 1-\frac{y\sqrt{p/q}}{\sqrt{n}}-\frac{\beta(n)}{nq}^{\ \ -nq} e^{-y^2}, \tag{13}$$

де $f(n) \sim g(n)$ означає, що $f(n)/g(n) \underset{n\to\infty}{\longrightarrow} 1$. З (13), використовуючи розклад: $\ln(1+z+o(z)) = z - \frac{z^2}{2} + o(z^2),\ z \to 0$, маємо:

$$\lim_{\substack{n\to\infty\\ k\to\infty}} S(n,k) = e^{-y^2} e^{y^2/2} = e^{-y^2/2}. \tag{14}$$

Використовуючи (10),(12) і (14), отримаємо:

$$\lim_{\substack{n\to\infty\\k\to\infty}}\sqrt{npq}\,\mathbf{P}(\mu_n = k) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}.$$

Теорема доведена. ◄

Зауваження 5. Рівність (1) можна записати у такому вигляді:

$$\lim_{\substack{n\to\infty\\k\to\infty}}\sqrt{npq}\,\mathbf{P}\Big(\frac{\mu_n - np}{\sqrt{npq}} = \frac{k-np}{\sqrt{npq}}\Big) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}. \qquad (15)$$

Стосовно (8) маємо $k - np \sim y\sqrt{npq}$ при $k, n, n-k \to \infty$. З цього і з (1) випливає наступний варіант локальної граничної теореми Муавра-Лапласа.

Теорема 7. (локальна Муавра-Лапласа). *Нехай μ_n є числом випадків, коли відбулась подія A при n незалежних випробуваннях, p - її ймовірність у кожному окремому випробуванні, $q = 1-p$. Тоді для довільного y має місце:*

$$\lim_{n\to\infty}\sqrt{npq}\,\mathbf{P}\Big(\frac{\mu_n - np}{\sqrt{npq}} = y\Big) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}.$$

◄

З (1) маємо також наближену формулу (див. формулу (7))

$$\mathbf{P}(\mu_n = k) \approx \frac{1}{\sqrt{2\pi npq}}e^{-\frac{(k-np)^2}{2npq}}.$$

Яку рекомендується застосовувати якщо $n \geq 100$ і $npq > 20$.

ЗАВДАННЯ

1. Довести, що якщо $f_n(x) \underset{n\to\infty}{\longrightarrow} f(x)$ для кожного $x \in (-\infty, \infty)$ і функція $f(x)$ є неперервною на вісі $[-\infty, \infty]$, то ця збіжність є рівномірною стосовно x.

2. Випадкові величини ξ_n, $n = 1, 2, ...$ є незалежними і мають один і той же рівномірний розподіл на інтервалі $(0, 1)$. Підрахувати $\mathbf{P}(\xi_1 + \xi_2 + ... + \xi_{1000} < 490)$.

3. Випадкові величини ξ_n, $n = 1, 2, ...$ є незалежними і мають розподіл $\mathbf{P}(\xi = k) = \frac{e^{-2}2^k}{k!}$, k=0, 1, 2,... . Підрахувати $\mathbf{P}(\xi_1 + \xi_2 + ... + \xi_{100} > 190)$.

4. Випадкові величини ξ_n, $n = 1, 2, ...$ є незалежними і мають розподіл зі щільністю $f(x) = \frac{3}{32}(-x^2+4x)$, $0 < x < 2$. Підрахувати $\mathbf{P}(1900 < \xi_1 + \xi_2 + ... + \xi_{1000} < 2000)$.

5. Підрахувати ймовірність того, що число одиниць, які з'явились при 6000 підкиданнях грального кубика лежить в інтервалі $(950, 990)$.

6. Якою є ймовірність отримати щонайменше 6 орлів при 10 підкиданнях, а також ймовірність отримати щонайменше 600 орлів при 1000 підкиданнях монети?

7. З великої групи осіб вибрали 5000 осіб, серед яких було 2800 жінок. Чи обгрунтованим є припущення, що в цій групі жінок більше ніж чоловіків?

8. Стрілець влучає в ціль з ймовірністю $p = 0.1$. Скільки разів він мусить стріляти, аби з ймовірністю 0.99 мати щонайменше 500 влучень?

9. Знайти значення b таке, щоб ймовірність появи k одиниць при 6000 підкиданнях становила 0.5, де $k \in (900, b)$.

10. Скільки разів треба підкидати монету, аби ймовірність того, що число орлів лежить в інтервалі (0.499,0.501), становила 0.99?

11. Ймовірність влучити в ціль при одному пострілі дорівнює 0,01.

    a) Якою є ймовірність того, що в 500 пострілах буде докладно 6 влучень?

    b) Скільки треба зробити пострілів, аби ймовірність влучити хоча б один раз була більшою за 0,9?

12. Гральний кубик підкидають 144 рази. Підрахувати ймовірність того, що 25 раз з'явиться п'ятірка. (0,0871)

13. У певному виші навчається 10.000 студентів. Припускаючи, що дні народження студентів рівномірно розташовані протягом року, знайти ймовірність того, що докладно 25 студентів відзначають день народження 3 січня.

14. В умовах попереднього завдання знайти ймовірність того, що кількість студентів, які не закінчать навчання, буде знаходитись в інтервалі [300, 400], якщо відомо, що в середньому 5% студентів відраховуються з учбового закладу до закінчення навчання.

МАТЕМАТИЧНА СТАТИСТИКА

На сьогоднішній день можна вказати багато непоганих підручників з математичної статистики в яких знайшли відображення основні результати та тенденцій розвитку сучасної статистичної теорії. Виділимо лише ті з них, які мали найбільший вплив на написання даного посібника. Це, перш за все, класична монографія Крамера [10], а також монографія Боровков О.О. [1] та посібник Лінькова Ю.М. [13, 14]. Останні два підручники розраховані на студентів математичних спеціальностей університетів і вимагають від читача досить ґрунтовних знань не лише з теорії ймовірностей а також з багатьох інших розділів математики. Для студентів технічних університетів а також студентів факультетів кібернетики, інформатики і навіть прикладної математики рівень цих підручників зависокий. Для них більш придатним є підручник Івченка Г.И. та Медведєв Ю.И. [12]. Посібник, який пропонується, в ідейному плані близький до [12], але ми приділяємо більше уваги доведенням а тому цей посібник розрахований на читача, який не обмежується формальним застосуванням відповідних процедур, а прагне зрозуміти їх суть. Результати, обґрунтування яких виходить за рамки даного посібника, лише формулюються, але при цьому подаємо посилання на літературу де зацікавлений читач можне знайти відповідну інформацію.

При виборі матеріалу автори орієнтувались на односеместровий курс математичної статистики для технічних університетів для яких математика є профільним предметом. Як правило такий курс включає такі розділи класичної статистичної теорії як "Оцінка параметрів" та "Теорія гіпотез". Ці два розділи є головними і в даному посібнику. Ми також обмежилися, в основному, випадком скалярного параметра.

Лекція 1. Головні поняття та задачі математичної статистики

Кожна математична дисципліна починається з формулювання певних початкових понять (часом дуже абстрактних). Вони зазвичай є головними елементами алфавіту мови цієї математичної дисципліни. Так в евклідовій геометрії до таких понять відносяться "точка", "площина", "пряма" та деякі інші. Алфавіт математичної статистики включає багато "літер" алфавіту теорії ймовірностей і це природно, оскільки математична статистика базується на поняттях та методах цієї теорії але досліджує свої специфічні задачі використовуючи як методи теорії ймовірностей так і свої власні. Тому в цьому розділі ми опишемо головні задачі та поняття математичної статистики використовуючи при цьому термінологію теорії ймовірностей. Іншими словами, ми опишемо ймовірносно-статистичну модель, яка буде розглядатися в даному посібнику. Такий опис нам буде потрібний ще тому, що в невеликому посібнику практично неможливо описати (навіть коротко) різні розгалуження математичної статистики і тому нам видається сенсовним з самого початку чітко окреслити "територію" в границях якої будуть відбуватися головні події. Крім того, в цій частині ми наведемо ряд результатів які з формальної точки зору можна віднести до теорії ймовірностей (наприклад теорема Фішера) але, як побачимо потім, вони грають важливу роль в математичній статистиці і тому є логічним включати їх до курсу цієї дисципліни.

1.1 Предмет та завдання математичної статистики

Що є математична статистика? Навіщо вона потрібна та які завдання розв'язує? В повсякденному житті ми часто вживаємо слово "статистика", але зазвичай воно асоціюється з таким виразом як "набір даних". Наприклад при вимірюванні показників деякої характеристики зазвичай кажуть, що в результаті цих вимірювань маємо "статистику цієї характеристики". Річний статистичний звіт Державного комітету статистики містить багато різних даних, які стосуються нашого життя, а особливо розвитку та стану економіки країни і в тих звітах також часто вживається це слово. Але значення слова "статистика" в подібного типу довідниках відрізняється від того значення, яке буде розумітися в цьому посібнику. У подальшому також буде вживатися термін "математична статистика". Різницю між словами "статистика" в широкому сенсі і "математична статистика" з'ясуємо на дуже простому прикладі, який також дозволить нам окреслити коло завдань, які стосуються математичної статистики. Нехай ми маємо автомат, який виробляє плитки шоколаду вагою 100 грам. Вочевидь, що в наслідок зміни умов праці цього автомату (коливання електричної напруги, випадкова зміна густини шоколаду і т.і.) ми не завжди будемо отримувати плитки, вага яких точно дорівнює 100 грамам. Крім того може статися, що автомат почне працювати зі збоями і тоді потрібне буде втручання механіка. Виникає питання як маючи значення ваг реально зроблених плиток шоколаду прийняти рішення про якість роботи автомату? Нехай, наприклад, маємо такі результати вимірювань

$$\begin{gathered} 101,\ 100,\ 98.5,\ 99.6,\ 100.3,\ 101.4,\ 98.7,\ 99.3,\ 100.6,\ 99.6, \\ 98.7,\ 99.5,\ 101,\ 101.5,\ 98.8,\ 99.6,\ 100.6,\ 100.7,\ 98.8,\ 99.7. \end{gathered} \tag{1}$$

Неважко знайти, що середнє значення ваги плиток шоколаду з (1) дорівнює 99.89, і тепер виникає таке питання: Чи це відхилення ваги плитки шоколаду від "ідеальної" ваги в 100 гр. є істотним? Якщо відповідь "так", тоді автомат працює не нормально. А якщо відповідь "ні" тоді автомат працює нормально і не треба витрачати зусиль (і коштів) на його перевірку. Власне кажучи відповіді на подібного типу запитань, і є одним з багатьох завдань математичної статистики. Дані з (1) в математичній статистиці називають "вибірка" і фактично кожна така вибірка є результатом вимірювання деякої характеристики, яка нас цікавить. Такі дані є відправною інформацією для математичної статистики, головним завданням якої є запропонувати методи і алгоритми, які на підставі цих даних дозволяють отримати відповіді на питання, що стосуються цієї характеристики. Більш детально будемо говорити про це нижче а тут лише зауважимо, що наведений вище приклад можна формалізувати наступним чином. Нехай H_0 означає, що відхилення середньої величини вибірки (1) від 100 не є істотним, а H_1, що це відхилення є істотним. В математичній статистиці говоримо, що ми маємо справу з двома гіпотезами: головною (H_0) і альтернативною (H_1) і треба знайти алгоритм, який дозволить на основі даних з (1) або прийняти гіпотезу H_0, або відхилити її, і тоді прийняти альтернативну гіпотезу H_1. Це власне і є завдання одного з напрямків математичної статистики, який має назву "Перевірка гіпотез". Можна навести ще один приклад, який пов'язаний з попереднім.

Нехай ξ є реальною вагою шоколадної плитки. З описаної вище ситуації природно вважати, що для нормально працюючого автомату ξ є випадковою величиною з математичним сподіванням 100, тобто $\mathbf{M}\xi = 100$. Докладний вигляд функції розподілу величини ξ нам зазвичай невідомий але в практичних застосуваннях часто вважаємо, що цей розподіл є наближено нормальний (це є власне наслідок центральної граничної теореми) і тому можна вважати, що $\frac{d\mathbf{P}(\xi<x)}{dx} \approx \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-100)^2}{2\sigma^2}\right)$, де σ є невідома нам дисперсія. З теорії ймовірностей відомо, що дисперсія характеризує відхилення випадкової величини від математичного сподівання і може служити характеристикою правильності роботи автомату. Виникає питання: Як на підставі даних (1) оцінити невідоме значення σ? (Бо тоді ми могли б порівнювати різні автомати і використовувати кращі з них). Це є типовою проблемою розділу математичної статистики під назвою "Оцінка параметрів". Можна було б навести і інші приклади, але на підставі наведеного можемо сказати, що головне завдання математичної статистики полягає в розробці і обґрунтуванню методів аналізи даних. Отже можна сказати що математична статистика це прикладна математична дисципліна і, як це вже було зауважено вище, її основою теорія ймовірностей.

Процес розв'язання статистичної проблеми можна (умовно) поділити на такі етапи: I. Формулювання проблеми. II. Збір даних. III. Обробка даних статистичними методами. IV. Прийняття рішень.

В цьому підручнику головна увага буде приділена III та IV етапам.

Перейдемо тепер до, так би мовити, формально-математичного опису задач математичної статистики. Цей опис ми розпочинаємо з впровадження головних понять математичної статистики таких як *генеральна популяція* та *випадковою вибіркою.*

1.2 Генеральна популяція та випадкова вибірка

Вихідними поняттями математичної статистики є "генеральна популяція", "випадкова вибірка", "оцінка" та "статистика". Перші два поняття пов'язані з наступною схемою випадкового експерименту.

Розглянемо наступний випадковий експеримент S: маємо певний набір елементів A однієї природи і експеримент полягає в тому що випадково вибираємо елемент з цього набору, записуємо значення певної характеристики ξ цього елементу і повертаємо елемент до A. Припускається при цьому, що експеримент організовано так, що можливість вибору є однаковою для всіх елементів, тобто жоден елемент немає переваги над іншим.

Визначення 1. Будемо говорити, що даний набір A є *генеральною сукупністю* (*або генеральною популяцією*). *Популяція величин, які спостерігалися під час n реалізацій експерименту S, будемо називати випадковою вибіркою з генеральної популяції*, а вище описаний процес генерації вибірки– *процесом незалежної генерації випадкової вибірки.* Число n називається *об'ємом вибірки.*

Важливою особливістю описаного експерименту є та обставина, що всі виміри характеристики ξ виконуються незалежно і в тих самих умовах, що, очевидно, не завжe може бути реалізовано на практиці.

В даному посібнику розглядається лише процес незалежної генерації випадкової вибірки.

У прикладі, який був описаний вище, в якості ξ виступає вага шоколадної плитки.

Стосовно процесу генерації вибірки зробимо два зауваження, які приведуть нас до математичної моделі, яка буде базовою в наших подальших дослідженнях.

По-перше, до проведення експерименту значення характеристики ξ нам невідомі і лише після вибору елемента з генеральної популяції ми отримаємо значення цієї характеристики. Отже, характеристика ξ може вважатися випадковою величиною $\xi(\omega)$, яка визначена на деякому ймовірносному просторі $(\Omega, \mathcal{B}, \mathbf{P})$, а експеримент S фактично означає те, що ми вибираємо $\omega = \omega_1$, а це, в свою чергу, автоматично фіксує значення нашої випадкової величини $x_1 = \xi(\omega_1)$. Таким чином, значення цієї характеристики є результатом нашого стохастичного експерименту і ми можемо трактувати її як випадкову величину.

По-друге, вочевидь для нас не є істотною природа предметів з A, оскільки нас цікавлять лише значення характеристики ξ. А отже, на процес утворення вибірки можемо дивитись ще наступним чином: робимо n послідовних вимірів випадкової величини ξ, яка є визначеною на деякому ймовірносному просторі $(\Omega, \mathcal{B}, \mathbf{P})$. Істотним є те, що ці виміри ми робимо незалежним чином і в однакових умовах. Все це дає підставу казати, що фактично ми маємо справу з генеральною популяцією, яка представлена випадковою величиною ξ.

Згідно з такою інтерпретацією процесу генерації вибірки ми можемо в інший спосіб трактувати саму вибірку

$$x_1,\ x_2,\ x_3, \ldots,\ x_n. \tag{2}$$

Характеристика ξ має певну функцію розподілу $F_\xi(x)$, яка зазвичай невідома. Якщо тепер запитати яка є ймовірність того, що після проведення експерименту ми отримаємо значення ξ менше деякого усталеного y, то вочевидь маємо взяти ті ω, для яких $\xi(\omega) < y$ тобто $\{x_1 < y\} = \{\omega : \xi(\omega) < y\}$, а це означає, що

$$\mathbf{P}(x_1 < y) = \mathbf{P}(\omega : \xi(\omega) < y) = \mathbf{P}(\xi < y) = F_\xi(y).$$

Такі самі міркування можемо повторити і для x_i, $i = 2, \ldots, n$, а оскільки ми маємо справу з незалежною генерацією вибірки, то можемо зробити наступний важливий висновок,

Компоненти випадкової вибірки (2) *з генеральної популяції* ξ *з функцією розподілу* $F_\xi(x)$ *можемо вважати як послідовність незалежних однаково розподілених випадкових величин з функцією розподілу* $F_\xi(x)$.

Маючи на увазі таку інтерпретацію вибірки часто говорять (і ми теж так будемо говорити) "нехай задана вибірка (2) з деякого розподілу $F_\xi(x)$". Відносно цього розподілу зробимо кілька зауважень. Постійно будемо вважати,що:

а) розподіл $F_\xi(x)$, тобто генеральної популяції, є дискретним або абсолютно неперервним;

б) Якщо розподіл генеральної популяції є абсолютно неперервним і $f(x)$ його щільність, то на кожному скінченому інтервалі функція $f(x)$ є неперервною справа і має границі зліва (за винятком може скінченої кількості точок з цього інтервалу).

Ми також будемо ввсажати, що всі функції $g(u_1, ..., u_n)$, які зустрічаються далі, мають такі самі властивості як і в пункті б) відносно кожного аргументу.

Прокоментуємо ці умови. Моделі для яких виконується умова а) найбільш поширені в застосуваннях і тому є сенсовним обмежитись лише ними. Пункт б) по суті не є окремою умовою, оскільки, якщо функція розподілу є абсолютно неперервною, тоді щільність завжди може бути вибрана так щоб ця умова виконувалася. Як ми побачимо далі, для конструкції статистик використовуються функції типу $g(u_1, ..., u_n)$ (скалярні чи векторові). Але всі сенсовні статистики будуються за допомогою функцій, для яких умова в) є виконана (у більшості випадків функція g є навіть неперервною).

Всі ці умови були б зайві, якщо використовувати поняття і результати теорії міри. Але оскільки ми не будемо використовувати результати цієї теорії, то ці умови нам потрібні, щоб сформулювати ряд результатів (і звичайно подати їх доведення) в математично коректний спосіб.

Міркування, які були наведені вище, можна трохи формалізувати.

В теорії ймовірностей вихідним пунктом є поняття ймовірносного простору $(\Omega, \mathcal{B}, \mathbf{P})$ з простором елементарних подій $\Omega = \{\omega\}$, алгеброю подій $\mathcal{B}$ (тобто множина $\mathcal{B}$ містить Ω та замкнута відносно доповнень та об'єднань не більш ніж зліченої кількості множин) і ймовірністю $\mathbf{P}$(тобто функція множин $\mathbf{P}$ означена на $\mathcal{B}$, зпевними властивостями). Випадкові величини $\xi(\omega)$ за означенням є відображення $\xi(\omega) : \Omega \to \mathbf{R}^1 = (-\infty, \infty)$, такі що $\{\omega : \xi(\omega) < x\} \in \mathcal{B}$ для кожного $x \in \mathbf{R}^1$) і головно задачею теорії ймовірностей є розробка методів дослідження випадкових величин та ймовірностей, пов'язаних з ними. При цьому ми завжди вважаємо, що функція $\mathbf{P}$ є визначеною, а отже визначена і функція розподілу кожної випадкової величини.

В задачах математичної статистики ситуація трохи інша. Як і вище, ми вважаємо, що генеральна популяція ξ визначена на деякому ймовірносному просторі $(\Omega, \mathcal{B}, \mathbf{P})$, але тепер функція $F_\xi(x) = \mathbf{P}(\xi < x)$ нам, взагалі кажучи, невідома або відома лише частково (наприклад, з докладністью до параметра). Як було зазначено вище, відносно випадкової велчини ξ (або популяції генеральної) в математичній статистиці розглядається дві групи завдань: визначення (або оцінювання) характеристик популяції генерально ξ та перевірка гіпотез відносно цієї популяції. Прикладом завдань з першої групи є, наприклад, оцінка наступних параметрів $\mathbf{M}\xi, \mathbf{D^2}\xi$ або параметру λ, якщо відомо, що популяція генеральна має розподіл експоненційний вигляду $1 - e^{-\lambda x}, x \geq 0$, де $\lambda > 0$ є невідомим параметром. Ця група завдань належить до розділу "Оцінка параметрів". Завданям з другої групи є, наприклад, наступне: ввжаємо, що популяція генеральна має розподіл $F_0(x)$ і потрібно перевірити, чи це наше припущення (або гіпотеза) справедливе. Ця група завдань належить до розділу "Перевірка гіпотез". Ці завдання ми мусисо розва'язати маючи поміри популяції генеральної (тобто вибірку (2), користуючись теоретичними знаннаями про випадкові величини та їх власності з розділу "Теорія ймовірностей".

Як ми побачимо пізніше, маючи скінченому вибірку ми практично ніколи не зможемо знайти невідомий нам параметр з точністю до 100%. І це є типовою ситуацією в математичній статистиці. Причиною цього є той факт, що популяція генеральна приймає, взагалі кажучи, нескінченну кількість значень а тому переходячи до скінченна вибірки ми частину інформації про популяцію втрачаємо.

На закінчення цього підрозділу скажемо ще про один важливий факт. Іноді буває так, що окрім вибірки ми ще маємо певну апріорну інформацю про популяцію генеральну. Задачі такого роду виникають як при оцінка параметрів, так і при перевірці гіпотез. Ця обставина часто є дуже корисною. Так, наприклад, маємо вибірку для популяції генеральної ξ і нашим завданням є знайти розподіл цієї популяції. Нехай додатково відомо, що цей розподіл є нормальним $N(m, \sigma^2)$, де параметри m, σ^2 є невідомими. В цьому випадку наше завдання зводиться до оцінки параметрів m σ^2. Неважко зрозуміти, що якість нашої відповіді буде значно кращою в порівняння з ситуацією якби ми нічого не знали про розподіл популяці ξ. А отже, чим менша ступінь невизначеності відносно розподілу генеральної популяції тим глибше і повніше можна дослідити відповідну задачу.

Лекція 2. Елементи описової статистики

2.1 Статистичні характеристики розподілів

З курсу теорії ймовірностей відомо, що для функції розподілу довільної випадкової величини можна ввести цілу низку числових характеристик, які хоча й не визначають цю функцію розподілу однозначно, але у багатьох випадках дають суттєву інформацію про саму випадкову величину. Найчастіше використовують наступні характеристики: математичне сподівання, дисперсію, коефіцієнт асиметрії, ексцес, медіану, моду а також деякі інші. В цьому розділі ми подамо статистичні аналоги цих характеристик (які часто називають описовими характеристиками) і які вираховуються на підставі випадкової вибірки. Їхня головна цінність полягає в тому, що вони мають зрозумілу практичну інтерпретацію і часто вже на основі них можна отримати досить глибоку інформацію про саму випадкову величину (генеральну популяція). Для того щоб ефективно використовувати ці характеристики в практичних дослідженнях від користувача не вимагається фундаментальних математичних знань. Фактично вже випускник середньої школи без жодних проблем може успішно ними користуватися (а отже і зрозуміти матеріал цього розділу). Цим і пояснюється той факт, що ці характеристики часто вживаються при аналізі даних в економічних та соціологічних дослідженнях. В наступному розділі ми більш глибоко вивчимо властивості цих характеристик а також їх зв'язок з відповідними теоретичними аналогами.

2.1.1 Характеристики вибірки

Нехай

$$x_1, \ldots, x_n \tag{1}$$

є n-елементна вибірка, яка представляє собою реалізації деякої випадкової величини ξ. Спочатку зробимо впорядкування елементів вибірки за зростанням і отримаємо нову популяція

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}. \tag{2}$$

Визначення 1. Сукупність (2) будемо називати *варіаційним рядом.*

Визначення 2. *Розмахом* величини ξ у вибірці(1) будемо називати різницю $R = x_{(n)} - x_{(1)}$.

Визначення 3. *Середнім арифметичним* вибірки (1) будемо називати число $x(n)$, яке визначається за формулою $x(n) = n^{-1} \sum_{k=1}^{n} x_k$.

Умовно можна сказати, що середнє арифметичне визначає пункт в околі якого групуються значення випадкової величини ξ.

Визначення 4. *Вибірковою дисперсією* вибірки (1) будемо називати число $S^2(n)$, визначене виразом $S^2(n) = \frac{1}{n} \sum_{k=1}^{n} (x_k - x(n))^2$.

Визначення 5. Величина $S(n) = \sqrt{S^2(n)}$ називається *середньоквадратичним відхиленням* вибірки (1).

Якщо середнє арифметичне $x(n)$ є тим числом, в околі якого групуються значення змінної ξ, то вибіркова дисперсія (а значить і середньоквадратичне відхилення) характеризує ширину цього околу: чим менше $S(n)$, тим більше значень змінної ξ є згрупованими довкола точки $x(n)$. Отже, можемо сказати, що вибіркова дисперсія описує розпорошення значень характеристики ξ.

Якщо підрахувати значення $x(n)$, $S(n)$, то вже на основі цих даних можна зробити певні висновки стосовно змінної ξ. Так відрізок

$$[x(n) - S(n); x(n) + S(n)] \tag{3}$$

називається *типовим інтервалом значень* випадкової величини ξ і має ту властивість, що біля 70% всіх значень вибірки, яка досліджується, лежить в цьому відрізку. На практиці цей факт використовують наступним чином: появу чергового $n+1$-го виміру випадкової величини ξ на 70% очікують у відрізку (3). Аналогічно, відрізок $[x(n) - 3S(n); x(n) + 3S(n)]$ містить майже всі компоненти вибірки. Цей факт ще називають "правилом трьох сигм"[1]

Визначення 6. *Квантилем* $k_n(p)$ *рівня* $0 \le p \le 1$ будемо називати таке перше значення $x_{(k)}$ у варіаційному ряді (2), для якого $k/n > p$. Іншими словами $k_n(p) = \min\{x_{(k)} : k/n > p\}$.

Квантиль рівня 1/2, тобто $k_n(1/2)$, називається *медіаною* і часто позначається символами $m_e(n)$, $M(n)$.

---

[1] Назва походить з того, що якщо $\xi \in \mathbb{N}(m, \sigma^2)$, то $\mathbf{P}(\xi \in [m - 3\sigma; m + 3\sigma]) \approx 0,999$.

Зауваження 1. В деяких підручниках ми можемо зустріти інші визначення квантилів. Так, наприклад, медіана іноді визначається як

$$m_e(n) = \begin{cases} x_{((n+1)/2)}, & \text{якщо n непарне,} \\ \frac{1}{2}\big(x_{(n/2)} + x_{((n+1)/2)}\big), & \text{якщо n парне.} \end{cases} \tag{4}$$

Але, як побачимо згодом, наведене визначення квантилі буде узгоджуватися з визначенням квантилі для функції розподілу (в нашім випадку для емпіричної функції розподілу, яка відповідає вибірці (1)).

В застосуваннях часто використовують ще й інші характеристики, так звані *квартилі*. Найчастіше вживають перші, другі та треті квартилі. Це є відповідно квантилі рівнів 0,25, 0,5 і 0,75. Так перший квартиль $k_n(0,25)$ позначають Q_1 і відділяє він 25% елементів вибірки (1) з меншими значеннями від 75% елементів з більшими значеннями.

Квартиль другого рівня $k_n(0,5)$ позначають Q_2, або $m_e(n)$, є звичайною медіаною, про яку вели мову раніше. Вочевидь, він ділить сукупність (1) навпіл.

Квартиль третього рівня $k_n(0,75)$ позначається Q_3 і відділяє він 75% елементів вибірки з меншими значеннями (1) від 25% з більшими значеннями.

Визначення 7. *Відхиленням четвертинним* будемо називати число $Q(n) = \big(Q_3(n) - Q_1(n)\big)/2$.

Відхиленням четвертинне є аналогом середньоквадратичного відхилення і в принципі не надає жодної додаткової інформації стосовно величини, яка спостерігається, але, позаяк цей показник потребує значно менших розрахунків, його часто використовують для швидкого аналізу даних. Так, наприклад, після підрахунку всіх квартилів можемо записати типовий інтервал значень змінної ξ у вигляді $[M(n) - Q(n); M(n) + Q(n)]$, який не співпадає з інтервалом (3), який зазвичай більш точно описує типові значення характеристики, але оскільки для знаходження інтервалу $[M(n)-Q(n); M(n)+Q(n)]$ потрібно виконати меншу кількість обчислень у порівнянні з інтервалом (3), то якраз його часто використовують на практиці.

Визначення 8. *Модальним значенням* (*модою, домінантою*) вибірки (1) будемо називати такий елемент вибірки m_0, який виступає в цій вибірці найбільшу кількість разів і якщо він не дорівнює $x_{(1)}$ або $x_{(n)}$.

Мода може не існувати а може бути неоднозначна.

Ведене вище середньоквадратичне відхилення $S^2(n)$ є абсолютною мірою варіації певної характеристики і вимірюється в тих самих одиницях, що і сама характеристика. Отже, загалом кажучи, цей показник не можна застосовувати до порівняння ступеня розпорошення двох різних вибірок. З цією метою розглядають відносний показник, який вводиться наступним чином.

Визначення 9. *Показником змінності* характеристики ξ будемо називати число $V(n) = S(n)/x(n)$.

Цей показник найчастіше подають у відсотках.

Для означення наступних показників введемо поняття моментів вищих порядків.

Визначення 10. *Моментом порядку* l з вибірки (1) будемо називати число $m_l(n)$, яке задається виразом $m_l(n) = n^{-1} \sum_{k=1}^{n} x_k^l$.

Визначення 11. *Центральним моментом порядку* l з вибірки (1) будемо називати число $\mu_l(n)$, яке задається виразом $\mu_l(n) = n^{-1} \sum_{k=1}^{n} (x_k - x(n))^l$.

Визначення 12. *Показником асиметрії* вибірки (1) будемо називати число $\gamma_1(n) = \mu_3(n)/S^3(n)$.

Цей показник приймає значення з відрізку $[-1; 1]$ і його сенс можемо подати в наступний спосіб. Нехай випадкова величини ξ має щільність $f(x)$, а число елементів вибірки (1) є великим. Нехай графік функції $f(x)$ є більш витягнутим до правої сторони. Тоді теоретичний показник асиметрії є додатнім $\gamma_1 > 0$ (див. рис. 2.1).

В цьому випадку кажуть, що розподіл випадкової величини ξ має правосторонню асиметрію. Для таких розподілів $\gamma_1(n) > 0$ (загалом кажучи, для великих n).

Якщо графік щільності є більш витягнутим до лівої сторони, тоді теоретичний показник асиметрії $\gamma_1 < 0$. В цій ситуації кажуть, що випадкова величина ξ має лівосторонню асиметрію. Для таких розподілів $\gamma_1(n) < 0$ для великих n.

Якщо ж щільність $f(x)$ є симетрична відносно деякої прямої $x = a$, (див. рис. 2.1), тоді теоретичний показник асиметрії $\gamma_1 = 0$. Для таких розподілів $\gamma_1(n) \approx 0$ (знову для великих n).

Чим ближче показник до $+1$ чи -1, тим сильнішою є відповідна асиметрія. Якщо $\gamma_1(n)$ є близькою до нуля, тоді це означає, що реалізації випадкової величини більш менш рівномірно лягають по обидві сторони від пункту симетрії, яким є значення середнього арифметичного.

Рис. 2.1:

Визначення 13. *Показником концентрації* (*коефіцієнтом сплющення* або *куртозою*) вибірки (1) будемо називати число $\beta(n) = \mu_4(n)/S^4(n)$.

Якщо випадкова величина ξ має нормальний розподіл, тоді $\beta(n) \approx 3$ для великих n. А позаяк в застосуваннях одним з головних розподілів є нормальний розподіл, то часто замість $\beta(n)$ використовують інший показник, який називають *ексцесом* .

Визначення 14. *Ексцесом* вибірки (1) будемо називати число $\gamma_2(n) = \beta(n) - 3$.

Якщо випадкова величина ξ має щільність $f(x)$, а вибірка (1) є великою, тоді наступний графік ілюструє як виглядає щільність $f(x)$, для різних значень $\beta(n)$.

Отже, можемо сказати, що показник концентрації характеризує ступінь сплющення щільності випадкової величини. І, наприклад, зростання цього показника сигналізує, що значення випадкової величини ξ мають тенденцію до концентрації довкола $\mathbf{M}\xi$.

У підручниках [25], [39] можна знайти ряд інших числових характеристик, які підраховують для вибірки (1).

2.2 Розподільчий ряд

При значному розмірі вибірки розрахунок показників, про які ми вели мову в попередньому підрозділі, може бути важким навіть при застосуванні комп'ютерів. Для цього з метою спрощення аналізу, елементи

Рис. 2.2:

вибірки часто групують в *класи* або *групи* і приймають, що всі елементи, які містяться в даному класі, є ідентичними і отже можуть бути представлені одним значенням (як побачимо згодом, серединою цього класу). Вся ця процедура називається групуванням даних і застосовується не лише при розрахунку числових характеристик, а і при побудові статистичних тестів. Опишемо головні етапи цієї процедури.

Крок I: Визначення кількості класів. Існує декілька способів визначення числа k класів в залежності від об'єму n вибірки. Ці способи потрібно сприймати як рекомендації, які виникли радше з практики, а не з теоретичних міркувань, і тому в кожному конкретному випадку число класів визначаємо з огляду на сенс завдання, враховуючи також конкретні дані, якими диспонуємо. Найчастіше застосовуються наступні правила: $k = [1 + 3,222 \ln n]$, $k = [\sqrt{n}]$.

Крок II. Довжина, межі та середини класів

Якщо R є розмахом вибірки, k є числом класів, то за довжину класу приймається значення $b \approx R/k$, але так, аби $bk \geq R$. Іншими словами, якщо береться наближене значення b, то це наближення має бути з надлишком. Точки, які є межами окремих класів, визначаються зазвичай з точністю до $\alpha/2$, де α є докладність, з якою визначалися елементи вибірки.

За ліву межу першого класу належить прийняти величину $l_1 = x_{(1)} - \alpha/2$, а потім за початок $m-$того класу будемо мати $l_m = l_{m-1} + b = x_{(1)} - \alpha/2 + (m-1)b$, а $[l_m, l_{m+1})$ є m-тим класом.

Зауваження 2. На практиці за значення лівої межі першого класу часто беруть $x_{(1)}$, а решту меж визначають так, як описано вище.

Визначення 15. Число значень вибірки, які містяться в i -тому класі будемо називати *розміром* (або *об'ємом*) i -того класу і позначати n_i.

Отже, для кожного класу можна розрахувати такі параметри:
1. Межі класу $[l_i, l_{i+1})$. 2. Розмір класу n_i. 3. Середину класу $\bar{x}_i = l_i + b/2$.

Визначення 16. Сукупність пар $(n_i, \bar{x}_i)$, $i = 1, ..., k$ будемо називати *розподільчим рядом.*

Визначення 17. Число $\omega_i = n_i/n$ будемо називати *частотою* i-того класу.

Приклад 1. *З генеральної популяції вибрано* $n = 50$ *-елементну вибірку і отримано наступні результати:*

$$3,6; 5,0; 4,0; 4,7; 5,2; 5,9; 4,5; 5,3; 5,5; 3,9; 5,6; 3,5; 5,4; 5,2; 4,1; 5,0; 3,1; 5,8; 4,8; 4,4;$$
$$4,6; 5,1; 4,7; 3,0; 5,5; 6,1; 3,8; 4,9; 5,6; 6,1; 5,9; 4,2; 6,4; 5,3; 4,5; 4,9; 4,0; 5,2; 3,3; 5,4;$$
$$4,7; 6,4; 5,1; 4,3; 5,2; 6,2; 4,4; 4,3; 5,8; 3,7.$$

Знайти для даної вибірки розподільчий ряд.

Р о з в' я з о к. Упорядкуємо дану вибірку в порядку зростання. Отримаємо

$$3,0; 3,1; 3,3; 3,5; 3,6; 3,7; 3,8; 3,9; 4,0; 4,0; 4,1; 4,2; 4,3; 4,3; 4,4; 4,4; 4,5; 4,5; 4,6; 4,7;$$
$$4,7; 4,7; 4,8; 4,9; 4,9; 5,0; 5,0; 5,1; 5,1; 5,2; 5,2; 5,2; 5,2; 5,3; 5,3; 5,4; 5,4; 5,5; 5,5; 5,6;$$
$$5,6; 5,8; 5,8; 5,9; 5,9; 6,1; 6,1; 6,2; 6,4; 6,4.$$

При розмірі вибірки $n = 50$, $[\sqrt{50}] = 7$, а тому число класів $k = 7$. Знаходимо $x_{min} = 3,0, x_{max} = 6,4$. Отже $R = 3,4$, $R/k \approx 0,49$. Приймаємо значення довжини класу b=0,5. Оскільки в цьому випадку точність $\alpha = 0,1$, то за ліву межу першого класу приймаємо $l_1 = x_{min} - 0,05 = 2,95$. Отже, перший клас має вигляд [2,95; 3,45). Позаяк, лише 3 елементи з варіаційного ряду містяться в першому класі, то $n_1 = 3$. Продовжуючи таким чином далі, отримаємо результати, представлені в наступній таблиці.

№ класу	Класи	Середини класів	Розмір класу	Частоти
1	[2,95;3,45)	3,2	3	0,06
2	[3,45;3,95)	3,7	5	0,1
3	[3,95;4,45)	4,2	8	0,16
4	[4,45;4,95)	4,7	9	0.18
5	[4,95;5,45)	5,2	12	0.24
6	[5,45;5,95)	5,7	8	0.16
7	[5,95;6,45)	6,2	5	0,1
			$\sum n_i = 50$	$\sum w_i = 1$

Рис. 2.3: Гістограма.

Отриманий розподільчий ряд можна також відобразити у вигляді *гістограми*. На горизонтальній вісі позначаються середини класів, або межі окремих класів, а на вертикальній вісі значення n_i

Можна також на вертикальній вісі відкласти значення w_i. Зєднавши точки з координатами $(\bar{x}_1 - b; 0), (\bar{x}_i, w_i)$, $i = 1, \ldots, k, (\bar{x}_k + b; 0)$, отримаємо *полігон частот*.

Рис. 2.4: Полігон частот

Після групування замість вибірки (1) будемо мати наступні пари чисел: $(n_i, \bar{x}_i), i = 1, 2, ..., k$. З практичної точки зору це є аналогічне тому, що значення вибірки (1), які містяться в одному і тому ж класі замінюються серединою класу. Цей факт потрібно взяти до уваги в поданих вище виразах для розрахунку числових характеристик для вибірки. Так, якщо n_i -число елементів i-того класу, k -кількість класів, а $[l_i; l_{i+1})$ межі класів, то для середнього арифметичного, дисперсії, моменту порядку l, а також центрального моменту порядку l відповідно будемо мати

$$x(n) = \frac{1}{n}\sum_{i=1}^{k}\bar{x}_i n_i, \qquad S^2(n) = \frac{1}{n}\sum_{i=1}^{k}\bigl(\bar{x}_i - x(n)\bigr)^2 n_i,$$
$$m_l(n) = \frac{1}{n}\sum_{i=1}^{k}\bar{x}_i^l n_i, \qquad \mu_l(n) = \frac{1}{n}\sum_{i=1}^{k}\bigl(\bar{x}_i - x(n)\bigr)^l n_i \tag{5}$$

Наведемо також вирази для розрахунку квартилів та моди.

$$Q_k(n) = l_{m(k)} + \frac{b(nk/4 - \sum_{i=1}^{m(k)-1} n_i)}{n_{m(k)}}, \qquad k = 1, 2, 3, \tag{6}$$

де $m(k)$ номер класу, в якому міститься k-тий квартиль.

Мода вираховується згідно наступного виразу

$$D(n) = l_m + \frac{b(n_m - n_{m-1})}{2n_m - n_{m-1} - n_{m+1}}, \tag{7}$$

де m - номер класу, в якому міститься мода.

Всі інші вирази для числових характеристик, про які ми вели мову в попередньому параграфі залишаються без змін.

Приклад 2. *Знайти числові характеристики (середнє, дисперсію та інші) для згрупованої вибірки з прикладу* 1 .

Р о з в' я з о к. Перш за все знайдемо середнє значення, дисперсію а також третій та четвертий моменти з вибірки з вибірки. Для цього використаємо формули (5) а конкретні числові значення візьмемо з таблиці на сторінці 165. Маємо

$$x(50) = \frac{1}{50}\sum_{i=1}^{7}\bar{x}_i n_i = \frac{1}{50}\Bigl(3,2\cdot 3 + 3,7\cdot 5 + 4,2\cdot 8 + 4,7\cdot 9 + 5,2\cdot 12 +$$
$$+ 5,7\cdot 8 + 6,2\cdot 5\Bigr) = 4,86;$$
$$S^2(50) = \frac{1}{50}\sum_{i=1}^{7}(\bar{x}_i - x(n))^2 n_i = \frac{1}{50}\Bigl((3,2-4,86)^2\cdot 3 + (3,7-4,86)^2\cdot 5 +$$
$$+ (4,2-4,86)^2\cdot 8 + (4,7-4,86)^2\cdot 9 + (5,2-4,86)^2\cdot 12 +$$
$$+ (5,7-4,86)^2\cdot 8 + (6,2-4,86)^2\cdot 5\Bigr) \approx 0,69;$$

$$\mu_3(50) = \frac{1}{50}\sum_{i=1}^{7}(\bar{x}_i - x(50))^3 n_i = \frac{1}{50}\ \ (3,2-4,86)^3\cdot 3 + (3,7-4,86)^3\cdot 5+$$
$$+ (4,2-4,86)^3\cdot 8 + (4,7-4,86)^3\cdot 9 + (5,2-4,86)^3\cdot 12+$$
$$+ (5,7-4,86)^3\cdot 8 + (6,2-4,86)^3\cdot 5 \ \ \approx -0,13;$$
$$\mu_4(50) = \frac{1}{50}\sum_{i=1}^{7}(\bar{x}_i - x(50))^4 n_i = \frac{1}{50}\ \ (3,2-4,86)^4\cdot 3 + (3,7-4,86)^4\cdot 5+$$
$$+ (4,2-4,86)^4\cdot 8 + (4,7-4,86)^4\cdot 9 + (5,2-4,86)^4\cdot 12+$$
$$+ (5,7-4,86)^4\cdot 8 + (6,2-4,86)^4\cdot 5 \ \ \approx 1,07.$$

Порахуємо тепер з формули (6) перший квартиль. Перш за все потрібно знайти значення для $m(1)$. Цим значенням буде перше число k для якого виконується співвідношення $n_1 + n_2 + ... + n_k > n/4$. В нашому випадку маємо $n/4 = 50/4 = 12,5$ а оскільки

$$n_1 = 3 < 12,5, \qquad n_1 + n_2 = 3 + 5 = 8 < 12,5,$$
$$n_1 + n_2 + n_3 = 3 + 5 + 8 = 16 > 12,5,$$

то $m(1) = 3$, тобто перший квартиль потрапив до третього класу [3,95;4,45) а тому з формули (6) маємо

$$Q_1(50) = l_3 + \frac{b(n/4 - \sum_{i=1}^{2} n_i)}{n_3} = 3,95 + \frac{0.5(50/4 - 3 - 5)}{8} = 4,2$$

Порахуємо тепер з формул (**??**), (**??**) другий та третій квартилі. Значення для $m(2)$ та $m(3)$ знаходимо подібно до того як $m(1)$. Тобто в якості $m(2)$ вибираємо перше число k для якого виконується співвідношення $n_1 + n_2 + ... + n_k > n/2$, а $m(3)$ буде перше число k для якого виконується співвідношення $n_1 + n_2 + ... + n_k > 3n/4$. Виконавши необхідні дії знаходимо $m(2) = 5, \ m(3) = 6$. Тобто другий квартиль потрапив до п'ятого класу $[4,95; 5,45)$ а третій до шостого $[5,45; 5,95)$. Тому з формул (**??**), (**??**) аналогічно попередньому маємо

$$M(50) = Q_2(50) = 4,95 + \frac{0.5(50/2 - 3 - 5 - 8 - 9)}{12} = 4,95;$$
$$Q_3(50) = 5,45 + \frac{0.5(50\frac{3}{4} - 3 - 5 - 8 - 9 - 12)}{8} = 5,5.$$

З цих обчислень та означення 7 отримуємо значення для відхилення четвертинного

$$Q(50) = \frac{Q_3(50) - Q_1(50)}{2} = \frac{5,5 - 4,2}{2} = 0,65.$$

Для моди потрібно користуватися формулою (7). В нашому випадку розмір п'ятого класу є найбільшим а тому $m = 5$ і для моди маємо

$$D(50) = l_m + \frac{b(n_m - n_{m-1})}{2n_m - n_{m-1} - n_{m+1}} = 4,95 + \frac{0,5(12 - 9)}{2\cdot 12 - 9 - 8} = 5,16.$$

Знайдемо тепер показник змінності, асиметрію та ексцес. Використовуючи формули з означень 9, 12 та 14, а також наведені вище обчислення отримуємо маємо

$$V(50) = \frac{S(n)}{x(n)} = \frac{0,83}{4,86} \approx 0.17, \quad \gamma_1(50) = \frac{\mu_3(n)}{S^3(50)} = \frac{-0,13}{0,83^3} \approx -0.23;$$
$$\gamma_2(50) = \frac{\mu_4(n)}{s^4(n)} - 3 = \frac{1,07}{0,83^4} - 3 \approx -0.78.$$

Знайдемо тепер інтервали в яких містяться типові значення характеристики яка відповідає нашій вибірці. З формули (3) та отриманих вище даних маємо

$$[x(n) - S(n); x(n) + S(n)] = [4,86 - 0,83; 4,86 + 0,83] = [4,03; 5,69].$$

Аналогічний інтервал але обрахований на підставі квартилі буде таким

$$[M(n) - Q(n); M(n) + Q(n)] = [4,95 - 0,65; 4,95 + 0,65] = [4,3; 5,6].$$

ЗАВДАННЯ

1. Вибірка I: $16, 13, 15, 16, 16, 15, 14, 12, 17, 16, 18, 14, 15, 17, 16$.

   Вибірка II: $27, 24, 28, 24, 25, 23, 29, 26, 29, 25$.

   Знайти R, m_0.

2. Для вибірки 1; 0; -2; 2; 5; 3; -1; 6; 2; 0; 4; -3; 5; 1; 2; 3; -1; 6; 3; 2. Знайти: $x(n)$, $s(n)$, $M(n)$, $D(n)$, полігон частот, типовий інтервал значень.

3. Легкоатлет А отримав протягом сезону наступні результати в стрибках: 6,82; 6,96; 7,23; 7,05; 7,80; 7,75. А легкоатлет В в тому ж самому сезоні отримав результати 6,72; 7,06; 7,3; 7,05; 7,70; 7,9, 7,0. Який з легкоатлетів отримав більш регулярні результати?

4. Середнє споживання молока для певної групи родин складає 15 л, а його стандартне відхилення 8. Аналогічні характеристик для числа осіб в родині складають відповідно: 3,4 особи і 1,3 особи. Обчислити коефіцієнт змінності для споживання молока і кількості осіб в родині. Які наслідки можна винести на підставі цих результатів?

5. В певній фірмі дослідження часу очікування на обслуговування дали наступні результати $M = 7$ хв., $Q_3 = 8$ хв., $Q_1 = 4$ хв. Прокоментувати.

   В якому інтервалі містяться типові зарплати в цій фірмі?

6. Дані про заробітну плату в певному закладі є наступними:

Зарплата(гр.)	до 800	800–1000	1000–1200	1200–1400	1400–1600
Кількість осіб, які отримують відповідну суму	10	30	10	50	30

   В якому інтервалі містяться типові зарплати в цій фірмі?

Лекція 3. Деякі розподіли макематичної статистики

3.1 Нормальний та пов'язані з ним розподіли ймовірностей.

В цьому і наступних підрозділах розглянемо параметричні сімейства ймовірносних розподілів, які часто з'являються в задачах математичної статистики. В подальшій частині цього посібника ці розподіли будуть використовуватися не лише для дослідження конкретних проблем, а і для ілюстрації понять, які будуть вводитися. Значна частина їх пов'язані з нормальним, а, зокрема, з стандартним нормальним розподілом. Тому зосередимо спочатку нашу увагу на цих розподілах.

3.1.1 Одновимірний нормальний розподіл

Цей розподіл був описаний в першій честині цього підручника, але беручи до уваги його важливість в статистиці математичний ми наведемо коротко його головні властивості тут. Нормальний розподіл з параметрами $m,\ \sigma^2$ позначається символом $\mathbb{N}(m, \sigma^2)$. Щільність цього розподілу має вигляд

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},\ -\infty < x < \infty, \tag{1}$$

де константи $-\infty < m < \infty$, $\sigma^2 > 0$ є параметрами. Якщо випадкова величина ξ має нормальний розподіл з параметрами $m,\ \sigma^2$, то цей факт будемо позначати так $\xi \in \mathbb{N}(m, \sigma^2)$. Якщо $\xi \in \mathbb{N}(m, \sigma^2)$, то

$$\mathbf{M}\xi = m, \qquad \mathbf{D^2}\xi = \sigma^2, \qquad \mathbf{M}e^{it\xi} = e^{imt-\sigma^2 t^2/2}.$$

Розподіл $\mathbb{N}(0, 1)$ називається стандартним нормальним розподілом. Він має щільність

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},\ -\infty < x < \infty.$$

Для $\xi \in \mathbb{N}(0,1)$ маємо (задача 2)

$$\mathbf{M}\xi^{2k+1} = 0, \quad \mathbf{M}\xi^{2k} = (2k-1)!!, k = 0, 1, \dots .$$

$$\Gamma(\alpha) = \int\limits_0^\infty x^{\alpha-1} e^{-x} dx \tag{2}$$

є відома з аналізу гамма-функція. Вона має наступні властивості.

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha), \quad \Gamma(n+1) = n!, \quad \Gamma(1/2) = \sqrt{\pi}. \tag{3}$$

Для нормального розподілу має місце наступне твердження, яке часто використовується в математичній статистиці.

Теорема 1. *Нехай випадкові величини ξ_k, $k = 1, \dots, n$ є незалежними і $\xi_k \in \mathbb{N}(m_k, \sigma_k^2)$. Тоді для довільних a_k, $k = 1, \dots, n$*

$$\sum_{k=1}^n a_k \xi_k \in \mathbb{N}(M, D^2), \quad M = \sum_{k=1}^n a_k m_k, \quad D^2 = \sum_{k=1}^n a_k^2 \sigma_k^2.$$

3.1.2 Розподіл хі-квадрат (χ^2)

Нехай $\xi_1, \dots, \xi_n$ є незалежними випадковими величинами, які мають стандартний нормальний розподіл.

Визначення 1. Розподіл випадкової величини $\chi_n^2 \overset{def}{=} \sum_{k=1}^n \xi_k^2$ називається *розподіл χ^2 (хі-квадрат) з n степенями свободи.*

Іншими словами, з цього визначення випливає, що розподіл хі-квадрат з n степенями свободи визначається наступним чином

$$\chi_n^2(x) = \mathbf{P}(\sum_{k=1}^n \xi_k^2 < x). \tag{4}$$

Той факт, що випадкова величина ξ має розподіл χ^2 з n степенями свободи записується так: $\xi \in \mathbb{H}_n$.

Теорема 2. *Розподіл випадкової величини χ_n^2 має щільність*

$$k_n(x) = \begin{cases} 0, & \text{для } x < 0, \\ \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & \text{для } x \geq 0, \end{cases} \tag{5}$$

і

$$\mathbf{M}\chi_n^2 = n, \qquad \mathbf{D^2}\chi_n^2 = 2n. \tag{6}$$

Доведення цього твердження ми залишаємо читачеві (задача 3).

Функція з правої частини (5) буде щільністю для всіх $n > 0$, а не лише для натуральних значень цього параметра (задача 4). Тому має сенс запис $\xi \in \mathbb{H}_\alpha, \quad \alpha > 0$, який означає, що випадкова величина ξ має розподіл зі щільністю

$$k_\alpha(x) = \begin{cases} 0, & \text{для } x < 0, \\ \frac{1}{2^{\alpha/2}\Gamma(\alpha/2)} x^{\alpha/2-1} e^{-x/2}, & \text{для } x \geq 0, \end{cases} \tag{7}$$

і вирази (6) залишаться справедливими (вочевидь, якщо n замінити на α). Лише тепер, якщо $\alpha > 0$ і не є числом натуральним, то вже не буде інтерпретації випадкової величини χ^2_α у вигляді (4) і не можна говорити про "степені свободи".

Розподіли з $\mathbb{H}_\alpha$ мають наступну властивість (задача 5).

Лема 1. *Якщо випадкові величини $\xi_i \in \mathbb{H}_{\alpha_i}$ і є незалежними, то*

$$\sum_{k=1}^{n} \xi_k \in \mathbb{H}_\alpha, \qquad \alpha = \sum_{k=1}^{n} \alpha_i.$$

З (6) і центральної граничної теореми маємо

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{\chi^2_n - n}{\sqrt{2n}} < x\right) = \Phi(x). \tag{8}$$

Якщо тепер нерівність $\sqrt{2\chi^2_n} - \sqrt{2n-1} < x$ записати у вигляді

$$\frac{\chi^2_n - n}{\sqrt{2n}} < \frac{x^2 + 2x\sqrt{2n-1} - 1}{2\sqrt{2n}},$$

то з (8) випливає, що

$$\lim_{n\to\infty} \mathbf{P}\left(\sqrt{2\chi^2_n} - \sqrt{2n-1} < x\right) = \Phi(x). \tag{9}$$

З (8), (9) маємо два вирази для наближеного розрахунку розподілу випадкової величини χ^2_n для великих n і x.

$$\mathbf{P}(\chi^2_n < x) \approx \Phi\left(\frac{x-n}{\sqrt{2n}}\right), \qquad \mathbf{P}(\chi^2_n < x) \approx \Phi(\sqrt{2x} - \sqrt{2n-1}).$$

Друге наближення є більш точним.

3.1.3 Розподіл Фішера

Нехай $\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_n$ є незалежними випадковими величинами, які мають стандартний нормальний розподіл.

Визначення 2. Розподіл випадкової величини $F_{m,n} \stackrel{def}{=} \frac{n\sum_{k=1}^m \xi_k^2}{m\sum_{k=1}^n \eta_k^2}$ називається *розподілом Фішера*[1] *з m, n степенями свободи*

Іншими словами з цього визначення випливає, що якщо χ_m^2 є випадковою величиною хі-квадрат з m степенями свободи, побудована з випадкових величин $\xi_1, \dots, \xi_m$, а $\widehat{\chi}_n^2$ є випадковою величиною хі-квадрат з n степенями свободи, побудована з випадкових величин $\eta_1, \dots, \eta_n$, то розподіл Фішера (інша назва -розподіл Фішера - Снедекора) з m, n степенями свободи визначається наступним чином:

$$F_{m,n}(x) = \mathbf{P}\left(\frac{n\sum_{k=1}^m \xi_k^2}{m\sum_{k=1}^n \eta_k^2} < x\right) = \mathbf{P}\left(\frac{n\chi_m^2}{m\widehat{\chi}_n^2} < x\right). \tag{10}$$

Той факт, що випадкова величина ξ має розподіл Фішера з m, n степенями свободи позначається так: $\xi \in \mathbb{F}_{m,n}$.

Теорема 3. *Щільності розподілу випадкової величини $F(m,n)$ задається формулою*

$$f_{mn}(x) = \frac{\Gamma((m+n)/2)m^{m/2}n^{n/2}}{\Gamma(m/2)\Gamma(n/2)} \frac{x^{m/2-1}}{(n+mx)^{(m+n)/2}} \quad \text{для} \quad x \geq 0. \tag{11}$$

Крім того,

$$\begin{array}{rcl} \mathbf{M}F(m,n) = & \frac{n}{n-2} & \text{для} \quad n > 2, \\ \mathbf{D^2}F(m,n) = & \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} & \text{для} \quad n > 4. \end{array} \tag{12}$$

Для $n = 1, 2$ математичне сподівання, а для $n = 1, 2, 3, 4$ дисперсія не існують.

Для доведення цього твердження треба скористатися виразом для щільності частки незалежних випадкових величин [5] і виразом для щільності випадкової величини χ^2 з Теореми 2. Це доведення також можна знайти в підручниках [1], [10].

Так само, як вже зазначалося вище, функція, яка стоїть в правій частині (11) буде щільністю для всіх $m, n > 0$, а не лише для натуральних

[1] Фішер(*Ronald Aylmer Fisher*) (17.02.1890– 29.07.1962) – британський генетик і статистик.

значень цих параметрів. Тому запис $\xi \in \mathbb{F}_{\alpha,\beta}, \quad \alpha, \beta > 0$, означає, що випадкова величина ξ має розподіл зі щільністю

$$f_{\alpha,\beta}(x) = \frac{\Gamma((\alpha+\beta)/2)\alpha^{\alpha/2}\beta^{\beta/2}}{\Gamma(\alpha/2)\Gamma(\beta/2)} \frac{x^{\beta/2-1}}{(\beta+\alpha x)^{(\alpha+\beta)/2}}, \quad \text{для} \quad x \geq 0,$$

і

$$\begin{array}{lll} \mathbf{M}F(\alpha,\beta) = & \frac{\beta}{\beta-2} & \text{для} \quad \beta > 2, \\ \mathbf{D^2}F(\alpha,\beta) = & \frac{2\beta^2(\alpha+\beta-2)}{\alpha(\beta-2)^2(\beta-4)} & \text{для} \quad \beta > 4. \end{array}$$

Тепер, якщо принаймні один з параметрів α, β не є натуральним числом, то вже не буде інтерпретації випадкової величини $F(\alpha, \beta)$ у вигляді(10) і тоді також не говоримо про "степені свободи".

3.1.4 Розподіл t Стьюдента.

Будемо говорити, що випадкова величина $t_\alpha, \quad \alpha > 0$ має розподіл *Стьюдента*[1], якщо щільність розподілу випадкової величини t_α має вигляд

$$s_\alpha(x) = \frac{1}{\sqrt{\pi\alpha}} \frac{\Gamma((\alpha+1)/2)}{\Gamma(\alpha/2)} \left(1 + \frac{x^2}{\alpha}\right)^{-(\alpha+1)/2}.$$

Легко переконатися (задача 1.1), що

$$\mathbf{M}t_\alpha = 0, \qquad \mathbf{D^2}t_\alpha = \frac{\alpha}{\alpha-2}, \quad \alpha > 2. \tag{13}$$

Той факт, що випадкова величина ξ має розподіл Стьюдента з параметром α записується так $\xi \in \mathbb{S}_\alpha$. Функцію розподілу випадкової величини t_α будемо позначати $S_\alpha(x) = \mathbf{P}(t_\alpha < x)$.

Якщо $\alpha = n$ є натуральним числом, то розподіл Стьюдента можна інтерпретувати подібно то того, як це було у випадку розподілів χ^2 і Фішера. Нехай $\xi, \xi_1, \ldots, \xi_n$ є незалежними випадковими величинами, які мають стандартний нормальний розподіл. Позначимо

$$t_n \stackrel{def}{=} \frac{\xi}{\sqrt{n^{-1}\sum_{k=1}^{n} \xi_k^2}}.$$

[1] Стьюдент (*William Sealy Gosset*) (псевдонім Student) (13.06.1876–16.10.1937)– британський статистик. Розподіл випадкової величини t_n був опублікований ним у 1908 г. Займався проблемами контролю якості (зокрема пива).

Якщо скористатися виразом для щільності частки двох випадкових величин, то нескладно довести (це доведення можна знайти в [10]), що щільності розподілу випадкової величини t_n задається формулою

$$s_n(x) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} .$$

Звідси випливає, що випадкова величина t_n має розподіл Стьюдента з параметром n, який в цьому випадку називають ступенем свободи. Оскільки згідно посиленого закону великих чисел $n^{-1}\sum_{k=1}^n \xi_k^2 \xrightarrow[n\to\infty]{\text{м.н.}} \mathbf{M}\xi_1^2 = 1$, то

$$\lim_{n\to\infty} S_n(x) = \lim_{n\to\infty} \mathbf{P}\Big(\frac{\xi}{\sqrt{\frac{1}{n}\sum_{k=1}^n \xi_k^2}} < x\Big) = \mathbf{P}(\xi < x) = \Phi(x).$$

В практичних застосуваннях зазвичай вважають, що для $n \geq 30$ розподіл Стьюдента можна замінювати нормальним.

А тепер наведемо так звану теорему Фішера, на якій ґрунтується багато досліджень в математичній статистиці.

Теорема 4. (Фішера) *Якщо випадкові величини $\xi_1, \xi_2, \ldots, \xi_n$ є незалежними з розподілом $\mathbb{N}(m, \sigma)$, тоді:*

а) *випадкові величини*

$$X_n = \frac{1}{n}\sum_{k=1}^n \xi_k, \qquad S_n^2 = \frac{1}{n}\sum_{k=1}^n (\xi_k - X_n)^2$$

незалежні;

б) *випадкова величина nS_n^2/σ^2 має розподіл хі– квадрат з $n-1$ степенями свободи;*

в) *випадкова величина*

$$\frac{X_n - m}{S_n}\sqrt{n-1}$$

має розподіл Стьюдента з $n-1$ ступенем свободи.

ЗАВДАННЯ

1. Довести, що $\Gamma(k+1/2) = \sqrt{\pi}(2k-1)!!/2^k$ для $k = 0, 1, 2, ...$

   Вказівка. *Використати рівність* $\Gamma(\alpha+1) = \alpha\Gamma(\alpha),\ \Gamma(1/2) = \sqrt{\pi}$.

2. Довчести, що (задача 2)

$$\mathbf{M}\xi^{2k+1} = 0, \quad \mathbf{M}\xi^{2k} = (2k-1)!!, k = 0, 1, ...$$

$$\Gamma(\alpha) = \int\limits_0^\infty x^{\alpha-1}e^{-x}dx \tag{14}$$

   для $\xi \in \mathbb{N}(0,1)$.

3. Довести формули (5), (6).

   Вказівка. *Скористаємося методом математичної індукції.*

4. Довести, що функція $k_\alpha(x)$, з виразу (7) є щільністю деякої невід'ємної випадкової величини ξ і

$$\mathbf{M}e^{-s\xi} = (1+2s)^{-\alpha/2}, \qquad Res > -1/2.$$

5. Довести лему 1.

6. Довести рівності (6) для довільного $n > 0$.

   Вказівка до завдань 5, 6. *Використати завдання* 4.

   1.1. Довести формули (13).

7. Довести, що для щільності розподілу Стьюдента s_n справедливе співвідношення $\lim_{n\to\infty} s_n(x) = \Phi'(x) = \varphi(x)$.

8. Довести, що якщо випадкові величини $\xi_1 \in \mathbb{G}(p_1, \lambda)$, $\xi_2 \in \mathbb{G}(p_2, \lambda)$ і є незалежними, то $\xi_1 + \xi_2 \in \mathbb{G}(p_1 + p_2, \lambda)$. (*Теорема про додавання.*)

9. Нехай $\xi_1,\ \xi_2$- незалежні випадкові величини, які мають розподіл Пуассона з параметрами $\lambda_1,\ \lambda_2$ відповідно.

   а) Довести, що $\xi_1 + \xi_2$ має розподіл Пуассона з параметром $\lambda_1 + \lambda_2$.

   б) Довести, що умовний розподіл величини ξ_1 при умові $\xi_1 + \xi_2 = n$ є біноміальний з параметрами n, $p = \lambda_1/(\lambda_1 + \lambda_2)$.

10. Довести, що якщо $\xi \in \mathbb{N}(0;1)$, то $\xi^2 \in \mathbb{G}(0,5;0,5)$.

11. Довести, що якщо випадкові величини $\xi_1, ..., \xi_n \in \mathbb{U}(0;1)$ і є незалежними, то величина $2\ln(\xi_1 \cdot \xi_2 \cdots \xi_n)$ має розподіл χ^2 з $2n$ степенями свободи.

Лекція 4. Теорія оцінювання. Статистики, оцінки та їх властивості

4.1 Статистики і оцінки. Попередні зауваження

В математичній статистиці традиційно вирізняють два класи задач:
A) оцінювання невідомих параметрів;
B) перевірка статистичних гіпотез.

Як побачимо пізніше, перевірка статистичних гіпотез теж часто пов'язана з визначенням невідомих параметрів, а тому, взагалі кажучи, поміж цими класами задач не існує принципової різниці. Деяка різниця все ж таки є, але більше на цю тема буде сказано в розділі 12. В цьому підрозділі зосередимося на задачах першого класу. Такі задачі виникають, коли потрібно на підставі вибірки оцінити деяку числову характеристику θ випадкової величини ξ (генеральна популяція) з невідомою функцією розподілу $F(x)$. Позаяк, маючи функцію розподілу $F(x)$, можна розрахувати всі числові характеристики для випадкової величини ξ, то оцінювання невідомої функції розподілу на підставі вибірки є однією з найважливіших задач теорії оцінювання. Формулювання цієї проблеми може бути наступним.

Функціональний вигляд функції F є відомим, але він залежить від невідомих параметрів. Цей факт часто записують наступним чином $F(x, \theta_1, ..., \theta_k)$, де θ_i є невідомі параметри або ще говорять, що розподіл генеральної популяції належить до деякої параметричної множини яку позначають так $\{F(x, \theta_1, ..., \theta_k),\ (\theta_1, ..., \theta_k) \in \Theta\}$. В цьому випадку завдання полягає в тому, щоб на підставі n - елементної вибірки,

$$x_1,\ x_2,\ ...,\ x_n, \tag{1}$$

яка взята з генеральної популяції ξ з функцією розподілу $F(x, \theta_1, ..., \theta_k)$, оцінити значення невідомих параметрів. Іншими словами, потрібно знайти значення

$$\theta_i(n) = \varphi_i(x_1, x_2, ..., x_n), \quad i = 1, ..., k, \tag{2}$$

які будемо використовувати замість невідомих параметрів θ_i. Зрозуміло, що функції φ_i в (1) повинні бути відомі, і значення $\varphi_i(x_1, x_2, ..., x_n)$ будемо називати статистиками або оцінками параметрів θ_i. Власне знаходження функції φ_i і є головною проблемою теорії оцінювання.

Якщо від початку відомо, що $(\theta_1, ..., \theta_k) \in \Theta \subset \mathbf{R}^k$, то зрозуміло, що для оцінок з (1) завжди має мати місце умова $(\theta_1(n), ..., \theta_k(n)) \in \Theta$ для довільних n.

Приклад 1. *Нехай вибірка вибрана з генеральної популяції показникового розподілу з невідомим параметром* $\lambda \geq 0$. *Тобто,* $F(x, \lambda) = 1 - e^{-\lambda x},\ x > 0$ *і потрібно знайти оцінку для* λ *на підставі вибірки* (1).

В цьому випадку $\Theta = [0, \infty)$.

Приклад 2. *Нехай вибірка вибрана з генеральної популяції, яка належить до сімейства* $\mathbb{N}(m, \sigma^2)$ *з невідомими параметрами* m, σ. *Отже*

$$F(x, m, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{x} e^{-\frac{(y-m)^2}{2\sigma^2}} dy$$

і потрібно знайти оцінку для m *і* σ *на підставі вибірки* (1).

В цьому випадку $\Theta = (-\infty, \infty) \times [0, \infty)$.

До іншої групи задач з класу **A)** належать задачі, в яких потрібно оцінити не функцію розподілу випадкової величини ξ, а деякий параметр θ цієї величини (або функції розподілу). Іншими словами, для заданого параметра θ, який є функцією від $F(x)$, тобто $\theta = \theta(F)$, потрібно знайти функцію від вибірки (1)

$$\theta(n) = \theta(x_1, x_2, ..., x_n),$$

яку можна буде застосувати як значення невідомого параметра θ.

Приклад 3. *На підставі вибірки* (1) *оцінити математичне сподівання генеральної популяції, тобто знайти оцінку для* $\mathbf{M}\xi = \int_{-\infty}^{\infty} x dF(x)$.

Як буде показано далі, в цій ситуації для оцінки $\mathbf{M}\xi$ можна взяти середнє арифметичне вибірки (1), тобто $n^{-1} \sum_{i=1}^{n} x_i$.

Для одного і того ж параметра можна запропонувати багато оцінок, і тому природно виникає проблема порівнювання оцінок. Ця задача є простішою якщо вибірка вибрана з генеральної популяції з деякою фіксованою (але невідомою) функцією розподілу і різні оцінки даного параметра порівнюємо при цьому розподілі. Але часто буває так, що функція розподілу належить до деякого параметричного сімейства $\{F(x, \theta), \theta \in \Theta\}$ (як в прикладах 1, 2), і тоді процедура порівнювання оцінок стає більш складною, оскільки хочемо вибрати оцінку щоб вона була "добра" для всіх можливих значень параметра, а розподіл генеральної популяції тепер вже не є фіксованим а змінюється разом з параметром. Надалі в разі потреби будемо говорити яка з цих ситуацій має місце.

4.1.1 Статистики і оцінки.

Нехай $x_1,\ x_2,\ ...,\ x_n$– n елементна вибірка спостережень за випадковою величиною ξ, яка задана на певному ймовірносному просторі і має функцію розподілу (невідому) $F(x)$. Як вже зазначалося вище, ця функція може залежати від деякого параметра θ, а тому потрібно було б писати так $F_\xi(x,\theta)$. Але в цьому підрозділі цей факт не буде для нас істотним, і тому будемо використовувати попередній запис.

Визначення 1. Для довільної функції n аргументів $g(u_1,...,u_n)$ значення $g(x_1,\ldots,x_k)$ будемо називати *статистикою.* .

Зауваження 1. В поданому визначенні не вимагається, аби функція $g(u_1,...,u_n)$ приймала значення з $\mathbf{R}^1=[-\infty,\infty)$. Ця функція може бути відображенням вигляду $g:\mathbf{R}^n\longrightarrow\mathbf{R}^k$, $1\le k\le n$. Наприклад, сама вибірка є статистикою, і тоді $g=(u_1,...,u_n)$. Але далі найчастіше будемо мати справу зі статистиками, які приймають значення в $\mathbf{R}^1$.

Зауваження 2. Якщо x_i трактувати як випадкові величини, тоді статистика також є випадковою величиною.

Приклад 4. Функції $\sum_{i=1}^n x_i$, $\sum_{i=1}^n i^2x_i$, $\max_i x_i$, $\sum_{i=1}^n \sin x_i$, $(x_1,\sum_{i=2}^n x_i)$ є статистиками.

Нехай для ξ потрібно оцінити деякий невідомий параметр θ.

Визначення 2. *Оцінкою* параметра θ будемо називати кожну статистику $g(x_1,\ldots,x_k)$, значення якої приймаємо за значення параметра θ.

Оцінку параметра θ будемо позначати $\theta(x_1,\ldots,x_n)$, або $\theta_n(\vec{x})$, де $\vec{x}=(x_1,\ldots,x_n)$, чи просто θ_n.

Зауваження 3. Зазначимо, що при означені статистики ми, взагалі кажучи, не вимагали щоб ця статистика була незалежна від параметра. Але лише статистики, які не залежать від параметра, можуть виступати в якості оцінок того параметра.

Приклад 5. *Нехай θ є математичним сподіванням. В якості оцінки для θ можна взяти $\theta_n^{(1)}=(x_1+x_2+\cdots+x_n)/n$. Нехай $n=3$, а $x_1=5,x_2=4,x_3=6$. За оцінку θ беремо $\theta_3^{(1)}=(x_1+x_2+x_3)/3$, таким чином значення $(5+4+6)/3=5$ вважаємо значенням параметра θ в подальших дослідженнях.*

Вочевидь, існує нескінченно багато функцій θ_n, які можна запропонувати як оцінки параметрів. Так в попередньому прикладі можна було взяти за оцінку, наприклад, статистику

$$\theta_3^{(2)} = (2x_1 + 3x_2 - x_3)/4, \quad \text{або} \quad \theta_3^{(3)} = (x_1 + 3x_2 + x_3)/4$$

і тоді значення θ оцінювались би числами 4 або 7 відповідно.

Існують певні вимоги до оцінок, виконання яких гарантує, що знайдені оцінки є "добрими" з точки зору їх використання на практиці. Визначення таких термінів, як "добра", "краща", "гірша" і т.д. по відношенню до оцінок вимагає введення до розгляду нових понять.

Визначення 3. Оцінку θ_n параметра θ будемо називати *не зміщеною,* якщо для кожного n має місце рівність $\mathbf{M}\theta_n = \theta$.

Ця властивість означає, що незміщена оцінка немає систематичної похибки і, що в "середньому" ми маємо справжнє значення параметра θ. У випадку, коли $\mathbf{M}\theta_n > \theta$, значення наших оцінок були б в середньому завищеними, а в ситуації, коли $\mathbf{M}\theta_n < \theta$ - заниженими. Так для оцінок $\theta_3^{(i)},\ i = 1, 2, 3$, які розглядалися вище маємо

$$\begin{aligned}
\mathbf{M}\theta_3^{(1)} &= \mathbf{M}\frac{x_1 + x_2 + x_3}{3} = \frac{1}{3}\big(\mathbf{M}x_1 + \mathbf{M}x_2 + \mathbf{M}x_3\big) = \frac{3\mathbf{M}x_1}{3} = \mathbf{M}x_1 = \theta, \\
\mathbf{M}\theta_3^{(2)} &= \mathbf{M}\frac{2x_1 + 3x_2 - x_3}{4} = \frac{1}{4}\big(2\mathbf{M}x_1 + 3\mathbf{M}x_2 - \mathbf{M}x_3\big) = \frac{4\mathbf{M}x_1}{4} = \theta, \\
\mathbf{M}\theta_3^{(3)} &= \mathbf{M}\frac{x_1 + 3x_2 + x_3}{4} = \frac{1}{4}\big(\mathbf{M}x_1 + 3\mathbf{M}x_2 + \mathbf{M}x_3\big) = \frac{5\mathbf{M}x_1}{4} = \frac{5\theta}{4}.
\end{aligned}$$

Отже, оцінки $\theta_3^{(1)}$, $\theta_3^{(2)}$ є незміщеними. Натомість оцінка $\theta_3^{(3)}$ є зміщена якщо $\theta \neq 0$. Таким чином, вимога незміщеності для оцінки виглядає досить логічною, але часом ця вимога є занадто сильною. Може так статися, що незміщеної оцінки для даного параметра або взагалі немає, або ця оцінка є непридатною для використання. Пояснимо це на наступному прикладі (див. [12]).

Приклад 6. *Нехай генеральна популяція має розподіл Пуассона з невідомим параметром $\lambda \geq 0$, тобто*

$$\mathbf{P}(\xi = n) = e^{-\lambda}\frac{\lambda^n}{n!}, \qquad n \geq 0.$$

Потрібно знайти незміщену оцінку для $\theta = \lambda^{-1}$ на підставі вибірки, яка містить лише одне спостереження за ξ.

Р о з в' я з о к. Нехай $x_1 = k$ буде нашою вибіркою. Доведемо, що не існує незміщеної оцінки параметра θ. Дійсно, якщо б $\theta_1(k)$ була такою оцінкою, тоді б мала місце рівність

$$\mathbf{M}\theta_1 = \sum_{k=0}^{\infty} \theta_1(k) \frac{\lambda^k e^{-\lambda}}{k!} = \frac{1}{\lambda}, \qquad \lambda > 0$$

або, що є рівносильне,

$$\sum_{k=0}^{\infty} \theta_1(k) \frac{\lambda^{k+1}}{k!} = e^{-\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}, \qquad \lambda > 0.$$

Але з останнього співвідношення випливає, що завжди $\theta_1(k) = \lambda^{-1}$. А оскільки оцінка параметра не може залежати від самого параметра, то отримана статистика не може бути вибрана в якості оцінки і отже бажаної оцінки не існує.

Як ми побачимо пізніше, одним з найважливіших критеріїв якості оцінки є середньоквадратичне відхилення оцінки від параметра, тобто $\mathbf{M}(\theta_n - \theta)^2$, і чим менше це відхилення тим краща оцінка. Може статися так, що це відхилення для зміщеної оцінки є меншим ніж для незміщеної. В задачі 10 подано приклад таких оцінок. Все це вказує на те, що нерозумно обмежуватися лише незміщеними оцінками.

Визначення 4. Оцінку θ_n параметра θ будемо називати *асимпотично незміщеною,* якщо

$$\lim_{n\to\infty} \mathbf{M}\theta_n = \theta.$$

Ця властивість означає, що зі зростанням розміру вибірки докладність нашої оцінки теж зростає.

Приклад 7. *В якості оцінки параметра $m = \mathbf{M}\xi$ пропонується статистика $m_n = \frac{2}{n^2}\sum_{i=1}^{n} ix_i$. Довести, що ця оцінка є асимптотично незміщеною.*

Р о з в' я з о к. Маємо

$$\mathbf{M}m_n = \frac{2}{n^2}\sum_{i=1}^{n} i\mathbf{M}x_i = \frac{2m}{n^2}\sum_{i=1}^{n} i = \frac{m(n+1)}{n} \neq m,$$

отже оцінка m_n є зміщена. Але, оскільки $\lim_{n\to\infty} \mathbf{M}m_n = \lim_{n\to\infty} \frac{n+1}{n} m = m$, то запропонована оцінка є асимптотично незміщеною.

Різниця $b_n(\theta) = \mathbf{M}\theta_n - \theta$ називається *зміщенням оцінки* θ_n.

Визначення 5. Оцінку θ_n параметра θ будемо називати *слабо консистентною*, (або просто *консистентною*), якщо для кожного $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}(|\theta_n - \theta| > \varepsilon) = 0, \tag{3}$$

або, в еквівалентному записі, $\theta_n \underset{n\to\infty p}{\longrightarrow} \theta$.

Визначення 6. Оцінку θ_n параметра θ будемо називати *сильно консистентною*, якщо

$$\mathbf{P}(\lim_{n\to\infty} \theta_n = \theta) = 1, \tag{4}$$

або, в еквівалентному записі, $\theta_n \underset{n\to\infty \text{м.н.}}{\longrightarrow} \theta$.

Іншими словами, оцінка θ_n є консистентною оцінкою параметра θ, якщо при $n \to \infty$ вона прямує до θ за ймовірністю, а якщо вона прямує до θ з ймовірністю 1, то говорять, що вона є сильно консистентною.

Властивість (3) означає, що в міру зростання об'єму вибірки ймовірність будь-якого малого відхилення оцінки від істотного значення параметра наближається до нуля. На практиці властивості (3), (4) означають, що для великих n можна вважати, що $\theta_n \approx \theta$.

Для того щоб перевірити чи є оцінка сильно консистентною чи ні часто застосовується закон великих чисел. Приклади використання цього закону ми наведемо в наступному підрозділі.

Визначення 7. Оцінка θ_n параметра θ називається *асимптотично нормальною* з параметром $\sigma^2 > 0$, якщо

$$\lim_{n\to\infty} \mathbf{P}(\frac{\sqrt{n}(\theta_n - \theta)}{\sigma} < x) = \Phi(x)$$

Як побачимо пізніше, більшість важливих оцінок є асимптотично нормальними, а це для практики означає, що маємо наближену рівність $\theta_n \approx \theta + \sigma\xi/\sqrt{n}$, де випадкова величина $\xi \in \mathbb{N}(0,1)$. Тобто похибка, яку ми робимо замінюючи параметр на його оцінку, є порядку $O(n^{-1/2})$.

Приклад 8. *В якості оцінки параметра $m = \mathbf{M}\xi$ для популяції генеральної ξ такої, що $\sigma^2 = \mathbf{D^2}\xi < \infty$ пропонується статистика $x(n) = \frac{1}{n}\sum_{i=1}^n x_i$. Довести, що ця оцінка є асимптотично нормальною.*

Р о з в' я з о к. Маємо

$$\frac{\sqrt{n}(x(n) - m)}{\sigma} = \frac{\sum_{i=1}^n x_i - nm}{\sqrt{n}\sigma}$$

Оскільни вибірку $x_1, ..., x_n$ ми трактуємо, як послідовність незалежних випадкових величин з таким самим розподілом як у випадкової величини ξ, то $\mathbf{M}\xi = m$, $\mathbf{D^2}\xi = \sigma^2$ і тепер на підставі центральної граничної теореми (див. 142) маємо

$$\lim_{n\to\infty} \mathbf{P}(\frac{\sqrt{n}(x(n) - m)}{\sigma} < x) = \lim_{n\to\infty} \mathbf{P}(\frac{\sum_{i=1}^n x_i - nm}{\sqrt{n}\sigma} < x) = \Phi(x).$$

Лекція 5. Порівняння оцінок

5.1 Формулювання зазачі та попередні зауваження

Як ми вже бачили, для даного параметра θ може існувати більше ніж одна оцінка. Нехай, наприклад, $\theta_n(1)$, $\theta_n^{(2)}$ будуть оцінками параметра θ. Питання, на яке хочемо дати відповідь в цьому розділі, звучить так: як порівняти оцінки $\theta_n(1)$, $\theta_n^{(2)}$? Іншими словами, як перевірити, котра з цих оцінок краща. В попередньому розділі були описані два критерії, які можуть служити для порівняння оцінок. Головною вадою цих критеріїв є те, що вони не описують кількісно відхилення значень $\theta_n^{(1)}$, $\theta_n(2)$ від справжнього значення параметра θ. А інтуїтивно зрозуміло, що, власне, це відхилення повинно служити для описання якості оцінки. Крім того, при порівнянні оцінок відразу виникає природнє питання:а яка оцнка буде "найкращою"для даного параметра. В цьому випадку нам потрібно визначитися з тим, що ми розуміємо під терміном "найкраща оцінка". Відповідь на ці питання залежить від багатьох факторів. Рзглянемо деякі з них.

Отже нехай вибірка $x_1, ..., x_n$ взята з генеральної сукупності ξ, розподіл якої має густину $f(x, \theta)$, де є θ невідомим параметром відносно якого будемо вважати, що $\theta \in \Theta = [a, b]$. Нехай, також $\mathbf{M}\xi^2 < \infty$. Ми хочемо порівняти дві оцінки $\theta_n^{(1)} = \theta_n^{(1)}(x_1, ..., x_n)$, $\theta_n^{(2)} = \theta_n^{(2)}(x_1, ..., x_n)$ параметра θ. Можливі дві ситуації:

а) ми порівнюємо наші оцінки для фіксованого (**але невідого**) значення параметра θ, або

б) ми порівнюємо наші оцінки для (**всіх**) значень параметра $\theta \in [a, b]$

.

Ситуація а) є рівноважна тому, що ми порівнюємо дві оцінки параметру при умові, що густина $f(x, \theta)$ є фіксованою (**але невідомою**) і в цьому випадку наступні означення є природніми

Визначення 1. Якщо

$$\mathbf{M}(\theta_n^{(1)} - \theta)^2 < \mathbf{M}(\theta_n^{(2)} - \theta)^2, \tag{1}$$

то будемо говорити, що $\theta_n(1)$ є *ефективнішою оцінкою* параметра θ, ніж оцінка $\theta_n^{(2)}$.

Надалі слова "ефективніша" і "краща" вживатимемо, як синоніми.

Зауваження 1. Нерівність (1) в розгорнутому вигляді виглядає наступним чином

$$\int\limits_{\mathbf{R}^n} (\theta_n^{(1)}(\vec{v}) - \theta)^2 L(\vec{v}) d\vec{v} < \int\limits_{\mathbf{R}^n} (\theta_n^{(2)}(\vec{v}) - \theta)^2 L(\vec{v}) d\vec{v}, \tag{2}$$

де $\vec{v} = (x_1, ..., x_n),\ L(\vec{v}) = \prod_{i=1}^n f(v_i, \theta)$.

Визначення 2. Якщо існує така оцінка θ_n^*, що

$$\mathbf{M}(\theta_n^* - \theta)^2 \leq \mathbf{M}(\theta_n - \theta)^2 \tag{3}$$

для будь якої оцінки θ_n, то будемо говорити, що оцінка θ_n^* є *ефективною оцінкою* параметра θ.

Цей підхід до порівняння оцінок називається *середньоквадратичним.* Його сенс очевидний: з двох оцінок кращою буде та, для якої середньоквадратичне відхилення від істинного значення параметра є меншим. Очевидно, що $\mathbf{M}(\theta_n^{(1)} - \theta)^2$ не є єдиною характеристикою, яка описує відхилення оцінки від параметра, наприклад, можна було б це відхилення характеризувати за допомогою $\mathbf{M}|\theta_n^{(1)} - \theta|$. Перевага першої характеристики полягає в тому, що функція x^2, на відміну від $|x|$, є диференційовною для всіх x. Цей факт часто є істотним в аналітичних дослідженнях конкретних оцінок.

Приклад 1. *Для параметру $m = \mathbf{M}\xi$ популяції генеральної ξ такої, що $\mathbf{D^2}\xi < \infty$ в прикладі 5, с. 181 (і відразу після нього) мали оцінки, $\theta_3^{(1)} = (x_1 + x_2 + x_3)/3$, $\theta_3^{(2)} = (2x_1 + 3x_2 - x_3)/4$ які є незміщеними. Маємо*

$$\mathbf{M}(\theta_3^{(1)} - \theta)^2 = \mathbf{D^2}\theta_3^{(1)} = \frac{1}{9}(\mathbf{D^2}x_1 + \mathbf{D^2}x_2 + \mathbf{D^2}x_3) = \frac{1}{3}\mathbf{D^2}\xi,$$

$$\mathbf{M}(\theta_3^{(2)} - \theta)^2 = \mathbf{D^2}\theta_3^{(2)} = \frac{1}{16}(4\mathbf{D^2}x_1 + 9\mathbf{D^2}x_2 + \mathbf{D^2}x_3) = \frac{7}{8}\mathbf{D^2}\xi,$$

і, отже, оцінка $\theta_3^{(1)}$ є ефективнішою.

А для оцінки $\theta_3^{(3)} = (x_1 + 3x_2 + x_3)/4$ з того ж прикладу маємо

$$\mathbf{M}(\theta_3^{(3)} - \theta)^2 = \mathbf{M}\ \ \frac{x_1 + 3x_2 + x_3}{4} - \frac{5}{4}\theta + \frac{1}{4}\theta\ \ ^2 = \mathbf{D^2}\ \ \frac{x_1 + 3x_2 + x_3}{4}\ \ + \frac{(\mathbf{M}\xi)^2}{16} =$$

$$= \frac{11\mathbf{D^2}\xi}{16} + \frac{(\mathbf{M}\xi)^2}{16}.$$

З цих відношень випливає, що оцінка $\theta_3^{(1)}$ є найкращою з-поміж $\theta_3^{(1)}$, $\theta_3^{(2)}$, $\theta_3^{(3)}$, причому оцінка $\theta_3^{(2)}$ буде кращою від $\theta_3^{(3)}$, якщо

$$\frac{7}{8}\mathbf{D^2}\xi < \frac{11\mathbf{D^2}\xi}{16} + \frac{(\mathbf{M}\xi)^2}{16}, \quad \text{або} \quad 3\mathbf{D^2}\xi < (\mathbf{M}\xi)^2.$$

А коли $3\mathbf{D^2}\xi > (\mathbf{M}\xi)^2$, то кращою буде оцінка $\theta_3^{(3)}$. Відмітимо, що в останньому випадку не дивлячись на те, що оцінка $\theta_3^{(3)}$ є зміщена якраз її потрібно надати перевагу перед незміщеною оцінкою $\theta_3^{(2)}$ оскільки вона гарантує менше відхилення від істинного значення параметра.

З цього прикладу випливає, що одна і та ж оцінка може буде кращою чи гіршою від деякої фіксованої оцінки в залежності від властивостей розподілу генеральної сукупності. Іншими словами, може статися так, що для одних значень параметра θ оцінка $\theta_n^{(1)}$ буде кращою від $\theta_n^{(2)}$ а для інших навпаки. В ситуації **а)** для задачу знаходження найкращої, або ефективної оцінки, важко, або навіть неможливо, розв'язати, оскільки для цього не існує відповідного математичного апарату.

Маючи на увазі ці дві обставини ми хочемо порівнюти оцінки для (**всіх**) значень параметра $\theta \in [a, b]$, тобто розглядаємо ситуацію **б)**.

В цьому випадку означення, наведені вище, можемо переписати наступним чином

Визначення 3. Якщо

$$\mathbf{M}(\theta_n^{(1)} - \theta)^2 \le \mathbf{M}(\theta_n^{(2)} - \theta)^2, \quad \text{для всіх } \theta \in [a, b] \tag{4}$$

і принаймі для одного $\theta \in [a, b]$ нерівність в (10) є строгою, то будемо говорити, що $\theta_n(1)$ є *ефективнішою оцінкою* параметра θ, ніж оцінка $\theta_n^{(2)}$.

Визначення 4. Якщо існує така оцінка θ_n^*, що

$$\mathbf{M}(\theta_n^* - \theta)^2 \le \mathbf{M}(\theta_n - \theta)^2 \qquad \text{для всіх } \theta \in [a, b] \tag{5}$$

для будь якої оцінки θ_n, то будемо говорити, що оцінка θ_n^* є *ефективною оцінкою* параметра θ.

Але виявляється, що ефективної оцінки в сенсі означення 4 не існує. Дійсно, припустимо, що для параметра θ існує найкраща оцінка θ_n^* в сенсі означення 4. Позначимо $d(\theta) = \mathbf{M}(\theta_n^* - \theta)^2$, $\theta \in [a, b]$– середньоквадратичне відхилення ефективної оцінки θ_n^* від параметра $\theta \in [a, b]$.

Нехай тепер $\theta = \alpha \in [a, b]$ буде деяким фіксованим (відомим) значенням параметра θ. Тоді для оцінки $\hat{\theta}_n \equiv \alpha$, очевидно, маємо

$$d(\alpha) = \mathbf{M}(\theta_n^* - \alpha)^2 \leq \mathbf{M}(\theta_n - \alpha)^2 = \mathbf{M}(\alpha - \alpha)^2 = 0. \qquad (6)$$

Отже $d(\alpha) = 0$. А оскільки $\alpha \in [a, b]$ довільне, то $d(\theta) = 0$ для всіх $\theta \in [a, b]$. Отже для найкращої оцінки маємо $\mathbf{M}(\theta_n^* - \theta)^2 = 0$ звідки $\mathbf{P}(\theta_n^* = \theta) = 1$. В цій ситуації вибірка однозначно визначає значення невідомого параметра. Але таке може трапитися лише у виняткових ситуаціях. Наприклад, якщо знаємо, що ξ набуває значень лише з відрізка $[\theta, 1 + \theta]$ і $\theta \in \{0, \pm 1, ...\}$, то одного значення для ξ вистачить, щоб однозначно визначити параметр θ. Тому в ситуації **б)** застосовуємо дещо інший підхід до порівняння оцінок, який буде описано в наступному парраграфі.

5.1.1 Середньоквадратичний підхід

Нехай, як і раніше, вибірка $x_1, ..., x_n$ взята з генеральної сукупності ξ, розподіл якої має густину $f(x, \theta)$, де є θ невідомим параметром відносно якого будемо вважати, що $\theta \in \Theta = [a, b]$.

Означимо клас $\mathbf{K}_b$ оцінок параметра θ наступним чином

$$\mathbf{K}_b = \{\theta_n : \mathbf{M}\theta_n - \theta = b(\theta)\}, \qquad (7)$$

де функція $b(\theta),\ \theta \in \Theta$ є заданою. Клас $\mathbf{K}_b$ є класом оцінок параметра θ, зміщення яких є заданим (і рівним $b(\theta)$).

Визначення 5. Якщо $\theta_n^{(1)},\ \theta_n^{(2)} \in \mathbf{K}_b$ і

$$\mathbf{M}(\theta_n^{(1)} - \theta)^2 < \mathbf{M}(\theta_n^{(2)} - \theta)^2, \qquad (8)$$

то будемо говорити, що оцінка $\theta_n(1)$ є *ефективнішою оцінкою* параметра θ, ніж оцінка $\theta_n^{(2)}$ в класі $\mathbf{K}_b$.

Визначення 6. Якщо існує така оцінка $\theta_n^* \in \mathbf{K}_b$, що

$$\mathbf{M}(\theta_n^* - \theta)^2 \leq \mathbf{M}(\theta_n - \theta)^2 \qquad (9)$$

для довільної оцінки $\theta_n \in \mathbf{K}_b$, то будемо говорити, що оцінка θ_n^* є *ефективною оцінкою* в класі $\mathbf{K}_b$.

Для ефективної оцінки з цього означення число

$$E_f(\theta_n) = \frac{\mathbf{M}(\theta_n^* - \theta)^2}{\mathbf{M}(\theta_n - \theta)^2} \tag{10}$$

називаємо *ефективністю оцінки* $\theta_n \in \mathbf{K}_b$.

І оцінка $\theta_n \in \mathbf{K}_b$ називається *асимптотично ефективною оцінкою*в класі $\mathbf{K}_b$ якщо $\lim_{n\to\infty} E_f(\theta_n) = 1$.

Зауваження 2. Означення 5, 6 фактично об'єднують ситуації **а)**, **б)** з попереднього підрозділу але лише для оцінок з класу $\mathbf{K}_b$. Так якщо в (8) значення θ фіксовано, то ми маємо нерівність (1), тобто ситуацію **а)**, а якщо $\theta \in [a, b]$, то маємо нерівність (4), тобто ситуацію - **б)**. Правда в останньому випадку ми нерівність " $\le$ " в (4) заміняємо на " $<$ " в (8).

Особливу роль грає клас оцінок, для яких $b(\theta) \equiv 0$, тобто клас $\mathbf{K}_0$ - клас незміщених оцінок. Оцінки, ефективні в цьому класі, будемо називати просто *ефективними.* В цій ситуації (11) виглядає так

$$E_f(\theta_n) = \frac{\mathbf{D^2}\theta_n^*}{\mathbf{D^2}\theta_n}.$$

Для ефективної оцінки з цього означення число

$$E_f(\theta_n) = \frac{\mathbf{D^2}\theta_n^*}{\mathbf{D^2}\theta_n} \tag{11}$$

називаємо *ефективністю оцінки* .

Зрозуміло, що $0 < E_f\theta_n \le 1$ і якщо $\lim_{n\to\infty} E_f\theta_n = 1$, то оцінка θ_n називається *асимптотично ефективною*

5.1.2 Асимптотичний підхід

В означенії 7 було подано визначення асимптотичної нормальності оцінки θ_n параметра θ. Ця оцінка була називана асимптотично нормальною оцінкою з параметром $\sigma^2 > 0$ для θ_n, якщо

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{\sqrt{n}(\theta_n - \theta)}{\sigma} < x\right) = \Phi(x),$$

де позначає стандартний нормальний розподіл. Було зауважено, що для практики це означає, що маємо наближену рівність $\theta_n \approx \theta + \sigma\xi/\sqrt{n}$, де випадкова величина $\xi \in \mathbb{N}(0, 1)$. Доданок $\sigma\xi/\sqrt{n}$ можна інтерпретувати як похибку оцінки θ_n або, іншими словами, "розсіяння"оцінки θ_n

довкола θ. Поняття асимптотичної нормальності можна використати до порівняння оцінок для великих вибірок.

Нехай оцінки $\theta_n^{(1)}$, $\theta_n^{(2)}$ параметра θ є асимптотично нормальними з параметрами $\sigma_1,\ \sigma_2 > 0$, відповідно.

Визначення 7. Якщо $\sigma_1 < \sigma_2$, то будемо говорити, що $\theta_n(1)$ є *асимптотично ефективнішою оцінкою параметра θ, ніж оцінка $\theta_n^{(2)}$.*

Сенс цього цього означення є дуже простий. Як було зауважено вище, якщо оцінки $\theta_n^{(1)}$, $\theta_n^{(2)}$ є асимптотично нормальними з параметрами σ_1^2, σ_2^2 відповідно, то якщо $\sigma_1 < \sigma_2$, то для великих n "розсіяння"оцінки $\theta_n^{(1)}$ довкола θ буде меншим, ніж оцінки $\theta_n^{(2)}$. Тоді натурально вважати, що оцінка $\theta_n^{(1)}$ є кращою. В цьому, власне, і полягає ідея асимптотичного порівняння оцінок.

Нехай $\mathbf{K}_\Phi$ означає клас оцінок, які є асимптотично нормальними

Визначення 8. Нехай $\theta_n^*,\ \theta_n \in \mathbf{K}_\Phi$ і

$$\frac{\sqrt{n}(\theta_n^* - \theta)}{\sigma_*} \underset{n\to\infty}{\Longrightarrow} \mathbb{N}(0,1), \qquad \frac{\sqrt{n}(\theta_n - \theta)}{\sigma} \underset{n\to\infty}{\Longrightarrow} \mathbb{N}(0,1).$$

Якщо $\sigma_* \le \sigma$ для всіх оцінок $\theta_n \in \mathbf{K}_\Phi$, то будемо говорити, що θ_n^* є *асимптотично ефективною* оцінкою параметра θ.

Лекція 6. Нерівність Рао-Крамера.Ефективні оцінки.

До порівняння оціцнок та знаходження ефективних оцінок часто служить так звана Нерівність Рао-Крамера, яка є предметом цього підрозділу.

6.1 Нерівність Рао-Крамера

Вже знаємо, що однією з числових характеристик, яка характеризує відхилення оцінки від істинного значення параметра θ, є функція $d(\theta) = \mathbf{M}(\theta_n - \theta)^2$. Наступна теорема описує нижню границю для середньоквадратичного відхилення оцінок фіксованим зміщенням. Тобто для оцінок, для якихо функція $b(\theta) = \mathbf{M}\theta_n - \theta$, де функція $b(\theta)$ є заданою. Якщо, наприклад, $b(\theta) = 0$, то будемо мати клас незміщених оцінокю.

Теорема 1. (*Нерівність Рао-Крамера*) *Нехай випадкова змінна ξ має щільність $f(x,\theta)$, де $\theta \in \Theta = [a,b]$, $-\infty \le a < b \le \infty$ є невідомим параметром і $b(\theta) = \mathbf{M}\theta_n - \theta$. Припустимо, що для всіх $\theta \in [a,b]$ виконуються наступні умови:*

а) *функція $\frac{\partial}{\partial\theta}\sqrt{f(x,\theta)}$ є неперервною відносно $\theta \in \Theta$;*

б) *функція $I^2(\theta) = \mathbf{M}\Big(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\Big)^2$ додатня і неперервна відносно $\theta \in \Theta$.*

Тоді

$$\mathbf{M}(\theta_n - \theta)^2 \ge \frac{\big(1 + b'(\theta)\big)^2}{n\mathbf{M}\Big(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\Big)^2} = \frac{\big(1 + b'(\theta)\big)^2}{n\int\limits_{-\infty}^{\infty}\big(\frac{\partial}{\partial\theta}\ln f(u,\theta)\big)^2 f(u,\theta)du} \tag{1}$$

для кожної оцінки θ_n параметра θ такої, що $\mathbf{M}\theta_n^2 < c < \infty$, $\theta \in [a,b]$ і рівність в (1) буде тоді і лише тоді, коли

$$\frac{\partial}{\partial\theta}\ln \Pi_{i=1}^{n} f(x_i,\theta) = c(\theta)\big(\theta_n(\vec{x}) - \theta\big), \tag{2}$$

для всіх x_i, для яких $f(x_i,\theta) > 0$, а $c(\theta)$ є сталою, можливо залежною від θ (але незалежною від $x_1, ..., x_n$).

Зауваження 1. Нерівність (1) в ситуації, коли генеральна сукупність ξ є дискретного типу і $\mathbf{P}(\xi = p_k) = p(x_k, \theta)$ має вигляд

$$\mathbf{M}(\theta_n - \theta)^2 \geq \frac{\big(1 + b'(\theta)\big)^2}{n \sum_{k=1}^{n} \left(\frac{\partial}{\partial\theta} \ln p(x_k,\theta)\right)^2 p(x_k,\theta)}.$$

Наслідок 1. *Якщо оцінка θ_n є незміщеною, то тоді нерівність* (1) *набуває вигляду*

$$\mathbf{D^2}\theta_n \geq \frac{1}{n\mathbf{M}\Big(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\Big)^2}.$$

Доведення нерівності Рао-Крамера опирається на наступну лему.

Лема 1. *Нехай для щільності $f(x,\theta)$ генеральної сукупності ξ виконуються умови теореми* 1. *Тоді для кожної статистики θ_n такої, що $\mathbf{M}\theta_n^2 < c(\theta) < \infty, \theta \in [a,b]$ має місце рівність*

$$\frac{\partial}{\partial\theta}\int_{\mathbf{R^n}} \theta_n(\vec{u}) L(\vec{u},\theta) d\vec{u} = \int_{\mathbf{R^n}} \theta_n(\vec{u}) \frac{\partial}{\partial\theta} L(\vec{u},\theta) d\vec{u}, \qquad \theta \in [a,b], \qquad (3)$$

де $L(\vec{u},\theta) = \prod_{i=1}^{n} f(u_i,\theta))$.

Іншими словами, в умовах теореми 1 можна змінити місцями інтегрування по $\vec{u} = (u_1, u_2, ..., u_n)$ і диференціювання по θ в лівій частині рівності (3). Ця лема є технічною і тому її доведення не подаємо. Більше інформації на тему, пов'язану з рівністю (3), читач може знайти в [1] с.440 або [13] с.105.

Якщо в (3) взяти $\theta_n \equiv 1$, то будемо мати

$$\int_{\mathbf{R^n}} \frac{\partial}{\partial\theta} L(\vec{u},\theta) d\vec{u} = \frac{\partial}{\partial\theta}\int_{\mathbf{R^n}} L(\vec{u},\theta) d\vec{u} = \frac{\partial}{\partial\theta} 1 = 0, \qquad \theta \in [a,b]. \qquad (4)$$

Д о в е д е н н я. теореми 1. Оскільки вибірку $x_1, x_2, ..., x_n$ ми можемо інтерпретувати як послідовність незалежних випадкових величин, то функція $L(\vec{u},\theta) = \prod_{i=1}^{n} f(u_i,\theta))$ є їх спільною густиною. Згідно з припущенням маємо $\mathbf{M}\theta_n = \theta + b(\theta)$ або, що те ж саме,

$$\int_{\mathbf{R}^n} \theta_n(\vec{u}) L(\vec{u},\theta) d\vec{u} = \theta + b(\theta). \qquad (5)$$

Як це випливає з (3), можемо диференціювати в рівності (5) по параметру θ під знаком інтеграла. Після такого диференціювання отримуємо

$$\int_{\mathbf{R}^n} \theta_n(\vec{u}) \frac{\partial L(\vec{u},\theta)}{\partial\theta} d\vec{u} = 1 + b'(\theta)$$

і якщо взяти до уваги (4), то остання рівність може бути переписана так

$$\int_{\mathbf{R}^n} (\theta_n(\vec{u}) - \theta) \frac{\partial L(\vec{u},\theta)}{\partial\theta} d\vec{u} = 1 + b'(\theta).$$

Отже маємо

$$\left(\int_{\mathbf{R}^n} (\theta_n(\vec{u}) - \theta) \frac{\partial \ln L(\vec{u},\theta)}{\partial\theta} L(\vec{u},\theta) d\vec{u}\right)^2 = \left(1 + b'(\theta)\right)^2.$$

З цієї рівності і нерівності Коші-Буняковського випливає

$$\int_{\mathbf{R}^n} (\theta_n(\vec{u}) - \theta)^2 L(\vec{u},\theta) d\vec{u} \int_{\mathbf{R}^n} \left[\frac{\partial \ln L(\vec{u},\theta)}{\partial\theta}\right]^2 L(\vec{u},\theta) d\vec{u} \geq \left(1 + b'(\theta)\right)^2, \tag{6}$$

звідки дістаємо

$$\mathbf{M}(\theta_n - \theta)^2 \geq \frac{\left(1 + b'(\theta)\right)^2}{\mathbf{M}\left(\frac{\partial \ln L(\mathbf{x},\theta)}{\partial\theta}\right)^2}.$$

Щоб довести рівність

$$\mathbf{M}\left(\frac{\partial \ln L(\vec{x},\theta)}{\partial\theta}\right)^2 = n\mathbf{M}\left(\frac{\partial \ln f(\xi,\theta)}{\partial\theta}\right)^2$$

зауважмо, що

$$\mathbf{M}\left(\frac{\partial \ln L(\vec{x},\theta)}{\partial\theta}\right)^2 = \mathbf{M}\left(\sum_{i=1}^{n} \frac{\partial \ln f(x_i,\theta)}{\partial\theta}\right)^2 = \mathbf{M}\sum_{i\neq j}^{n} \frac{\partial \ln f(x_i,\theta)}{\partial\theta} \frac{\partial \ln f(x_j,\theta)}{\partial\theta}$$
$$+\mathbf{M}\sum_{i=1}^{n}\left(\frac{\partial \ln f(x_i,\theta)}{\partial\theta}\right)^2 = \mathbf{M}\sum_{i=1}^{n}\left(\frac{\partial \ln f(x_i,\theta}{\partial\theta}\right)^2 = n\mathbf{M}\left(\frac{\partial \ln f(\xi,\theta}{\partial\theta}\right)^2.$$

Передостання рівність в цих співвідношеннях випливає з того, що випадкові змінні x_i є незалежними, а

$$\mathbf{M}\frac{\partial \ln f(x_i,\theta)}{\partial\theta} = \int_{-\infty}^{\infty} \frac{\partial \ln f(u,\theta)}{\partial\theta} f(u,\theta) du = \int_{-\infty}^{\infty} \frac{\partial f(u,\theta)}{\partial\theta} du = 0.$$

Рівність в (1) буде досягатися тоді і лише тоді, коли в (6) буде рівність, тобто в нерівності Коші-Буняковського мусить буде рівність, що, в свою чергу, можливе лише тоді, коли

$$c(\theta)(\theta_n(\vec{u}) - \theta)\sqrt{L(\vec{u},\theta)} = \frac{\partial \ln f_n(\vec{u},\theta)}{\partial\theta}\sqrt{L(\vec{u},\theta)}$$

для всіх значень $\vec{u} = (u_1, ..., u_n)$ для яких $L(\vec{u}) > 0$, а $c(\theta)$ є сталою, яка не залежить від $\vec{u}$. Якщо $\vec{u}$ таке, що $L(\vec{u},\theta) > 0$, то з останньої рівності дістаємо (2).

Визначення 1. Значення

$$nI^2(\theta) = n\mathbf{M}\Big(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\Big)^2 \tag{7}$$

називаємо *інформацією Фішера* .

Можна сказати, що $nI^2(\theta)$ це та кількість інформації відносно ξ яка міститься у вибірці $\vec{x}$. Деякі властивості цієї характеристики можна знайти в [1]. Для обчислення інформації Фішера часом є корисною наступна лема.

Лема 2. *Нехай виконуються наступні умови:*

а) $0 < \mathbf{M}\Big(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\Big)^2 < \infty$;

б) $\Big|\frac{\partial^i}{\partial\theta^i}\ln f(x,\theta)\Big| \le G_i(x)$, *для всіх* $\theta\in\Theta$ *і* $\int_{-\infty}^{\infty} G_i(x)dx < \infty$, $i = 1, 2$.

Тоді

$$\mathbf{M}\Big(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\Big)^2 = -\mathbf{M}\Big(\frac{\partial^2}{\partial\theta^2}\ln f(\xi,\theta)\Big). \tag{8}$$

Д о в е д е н н я. Якщо в тотожності

$$\frac{\partial^2}{\partial\theta^2}\ln f(\xi,\theta) = \frac{\partial^2 f(\xi,\theta)/\partial\theta^2}{f(\xi,\theta)} - \left(\frac{\partial f(\xi,\theta)/\partial\theta}{f(\xi,\theta)}\right)^2$$

перейти до математичного сподівання $\mathbf{M}$, то отримаємо

$$\mathbf{M}\frac{\partial^2}{\partial\theta^2}\ln f(\xi,\theta) = \int_{-\infty}^{\infty}\frac{\partial^2}{\partial\theta^2}f(u,\theta)du - \mathbf{M}\left(\frac{\partial f(\xi,\theta)/\partial\theta}{f(\xi,\theta)}\right)^2. \tag{9}$$

Умова б) леми дозволяє записати

$$\int_{-\infty}^{\infty}\frac{\partial^2}{\partial\theta^2}f(u,\theta)du = \frac{\partial^2}{\partial\theta^2}\int_{-\infty}^{\infty}f(u,\theta)du = \frac{\partial^2}{\partial\theta^2}1 = 0$$

і (8) випливає з (9).

В попередньому підрозділі ми означили ефективну оцінку як таку оцінку, яка має найменше середньоквадратичне відхилення від істинного значення параметра (Означення 6, с. 188). Але це означення не дає відповіді на питання, а як дослідити, чи деяка конкретна оцінка є ефективною чи ні? Нерівність Рао-Крамера дає інструмент для дослідження ефективності оцінок але лише для генеральних популяцій, для яких справедливі умови цієї нерівності.

Визначення 2. Якщо існує така оцінка $\theta_n^* \in \mathbf{K}_b$, що

$$\mathbf{M}(\theta_n^* - \theta)^2 = \frac{\big(1 + b'(\theta)\big)^2}{n\mathbf{M}\Big(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\Big)^2}, \qquad \theta \in \Theta, \tag{10}$$

то будемо говорити, що оцінка θ_n^* є *R-ефективною оцінкою* в класі $\mathbf{K}_b$.

Іншими словами, R-ефективна оцінка це така оцінка, для якої в нерівності Рао-Крамера маємо рівність. Ясно, що це означення може бути застосованим лише для таких оцінок, для яких виконуються умови теореми 1. Але з означення ефективної оцінки і нерівності Рао-Крамера випливає, що якщо маємо оцінку $\theta_n^* \in \mathbf{K}_b$ таку, що для неї в нерівності Рао- Крамера маємо рівність, то ця оцінка автоматично буде ефективною оскільки в цьому випадку поняття R-ефективності і ефективності співпадають. Отже щоб дослідити ефективність конкретної оцінки потрібно підрахувати обидві сторони в (1) і якщо результат буде однаковий, то наша оцінка автоматично буде ефективна. Приклад такого застосування нерівності Рао-Крамера ми вже бачили в підрозділі 7.1 при дослідженні властивостей оцінок середнього значення та дисперсії нормальної генеральної сукупності (теорема **??**, с. **??**). Якраз та обставина, що означення R-ефективної оцінки фактично вказує, що слід зробити, щоб дослідити, чи є деяка конкретна оцінка ефективною чи ні, робить поняття R-ефективності оцінки більш вживанішим.

Подібна ситуація і з поняттям асимптотично ефективної оцінки.

Визначення 3. Оцінку $\theta_n^* \in \mathbf{K}_b$ називаємо *асимптотично R-ефективною* оцінкою, якщо

$$\mathbf{M}(\theta_n^* - \theta)^2 = \frac{\big(1 + b'(\theta)\big)^2}{n\mathbf{M}\Big(\frac{\partial}{\partial\theta}\ln f(\xi,\theta)\Big)^2} + o(n^{-1}), \quad n \to \infty, \quad \theta \in \Theta.$$

Лекція 7. Аналіз деяких оцінок

7.1 Оцінки для математичного сподівання та дисперсії

Для
генеральної популяції ξ з вибіркою $x_1, \ldots, x_n$ позначимо $m = \mathbf{M}\xi$, $\sigma^2 = \mathbf{D^2}\xi$. Позначимо

$$x(n) = \frac{1}{n}\sum_{k=1}^{n} x_k,, \qquad S^2(n) = \frac{1}{n}\sum_{k=1}^{n}(x_k - x(n))^2, \tag{1}$$

$$\widehat{S}^2(n) = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - x(n))^2, \qquad \bar{S}^2(n) = \frac{1}{n}\sum_{k=1}^{n}(x_k - m)^2. \tag{2}$$

Теорема 1. *Нехай $\mathbf{M}\xi^2 < \infty$. Тоді:*

а) *статистика $x(n)$ є незміщеною і сильно консистентною оцінкою параметра m, а статистики $\widehat{S}^2(n)$, $\bar{S}^2(n)$ є незміщеними і сильно консистентними оцінками параметра σ^2;*

б) *статистика $S^2(n)$ є асимптотично незміщеною і сильно консистентною оцінкою параметра σ^2.*

Д о в е д е н н я. Для оцінки $x(n)$ будемо мати $\mathbf{M}x(n) = n^{-1}\sum_{k=1}^{n}\mathbf{M}x_k = m$, а оскільки з посиленого закону великих чисел випливає, що $x(n) = n^{-1}\sum_{k=1}^{n} x_k \xrightarrow[n\to\infty]{\text{м.н.}} \mathbf{M}x_1 = m$, то оцінка $x(n)$ є незміщеною і сильно консистентною.

Тепер вивчимо оцінку $S^2(n)$. Для цього нам буде потрібно наступне співвідношення (задача 8)

$$\sum_{k=1}^{n}(x_k - x(n))^2 = \sum_{k=1}^{n}(x_k - m)^2 - n(x(n) - m)^2. \tag{3}$$

Беручи до уваги (3), отримаємо

$$\mathbf{M}S^2(n) = \mathbf{M}\ \frac{1}{n}\sum_{k=1}^{n}(x_k - x(n))^2\ = \frac{1}{n}\sum_{k=1}^{n}\mathbf{M}(x_k - m)^2 - \mathbf{M}(x(n) - m)^2$$
$$= \sigma^2 - \mathbf{D^2}x(n) = \sigma^2 - \mathbf{D^2}\ \frac{1}{n}\sum_{k=1}^{n} x_k\ = \sigma^2 - \sigma^2/n.$$

Отже,

$$\mathbf{M}S^2(n) = \frac{n-1}{n}\sigma^2 \neq \sigma^2. \tag{4}$$

Це означає, що розглянута оцінка є зміщена, але оскільки

$$\mathbf{M}S^2(n) = \sigma^2 - \frac{\sigma^2}{n} \underset{n\to\infty}{\overset{\text{м.н.}}{\longrightarrow}} \sigma^2,$$

то ця оцінка є асимптотично незміщеною.

Аналогічно, з закону великих чисел та рівності (3) випливає

$$S^2(n) = \frac{1}{n}\sum_{k=1}^{n}(x_k - m)^2 - (x(n) - m)^2 \underset{n\to\infty}{\overset{\text{м.н.}}{\longrightarrow}} \mathbf{M}(x_1 - m)^2 = \sigma^2, \tag{5}$$

що й доводить пункт б).

Позаяк,

$$\widehat{S}^2(n) = \frac{n-1}{n}S^2(n),$$

то незміщенність та сильна консистентність цієї оцінки випливають з співвідношень (4), (5). Дослідження властивостей оцінки $\bar{S}^2(n)$ залишаємо читачеві.

На закінчення цього підрозділу розглянемо наступне, цілком логічним є наступне питання. Нехай оцінка θ_n є незміщена для параметра θ, а $f(\cdot)$ є деякою функцією. Чи буде тоді оцінка $f(\theta_n)$ незміщеною оцінкою для параметра $f(\theta)$? Як випливає з теореми 1 та наступного прикладу відповідь є негативною.

Приклад 1. *Оцінка $\bar{S}^2(n) = n^{-1}\sum_{k=1}^{n}(x_k - m)^2$ є незміщеною оцінкою для параметра σ^2 генеральної популяції $\xi \in \mathbb{N}(m, \sigma^2)$ при умові, що m є відомим. Доведемо, що $\bar{S}(n)$ не буде незміщеною оцінкою для σ.*

Позаяк $(\xi_i - m)/\sigma \in \mathbb{N}(0, 1)$, то випадкова величина $\eta = \sum_{k=1}^{n}(x_k - m)^2/\sigma^2$ має розподіл χ_n^2 з n степенями свободи, а оскільки $\bar{S}(n) = \sigma n^{-1/2}\eta^{1/2}$, то з урахуванням виразу щільності розподілу хі–квадрат з n степенями свободи (див. (5), с. 172) маємо

$$\mathbf{M}\bar{S}(n) = \frac{\sigma}{\sqrt{n}2^{n/2}\Gamma(n/2)}\int_0^\infty \sqrt{x}e^{-x/2}x^{n/2-1}dx = \frac{\sigma\sqrt{2}}{\sqrt{n}\Gamma(n/2)}\int_0^\infty e^{-x}x^{(n-1)/2}dx =$$
$$= \frac{\sigma\sqrt{2}\Gamma((n+1)/2)}{\sqrt{n}\Gamma(n/2)} \neq \sigma.$$

Отже, оцінка S_n є зміщеною.

7.2 Асимптотичний підхід

Для порівняння оцінок можемо застосовувати асимптотичний підхід, який виглядає так.

Приклад 1. *Для генеральної сукупності ξ з показниковим розподілом з параметром $\lambda > 0$ для λ маємо наступні оцінки*

$$\lambda_1(n) = \frac{1}{x(n)}, \qquad \lambda_2(n) = \sqrt{\frac{2n}{\sum_{i=1}^n x_i^2}}.$$

Дослідити асимптотичну ефективність цих оцінок.

Р о з в' я з о к.

Оскільки $\mathbf{M}x_i = 1/\lambda,\ \mathbf{M}x_i^2 = 2/\lambda^2$, то на підставі закону великих чисел маємо

$$\lambda_n(1) = \frac{1}{\frac{1}{n}\sum_{i=1}^n x_i} \underset{n\to\infty}{\overset{\text{м.н.}}{\longrightarrow}} \frac{1}{\mathbf{M}\xi_1} = \lambda, \qquad \lambda_n(2) = \sqrt{\frac{2}{\frac{1}{n}\sum_{i=1}^n x_i^2}} \underset{n\to\infty}{\overset{\text{м.н.}}{\longrightarrow}} \sqrt{\frac{2}{\mathbf{M}x_i^2}} = \lambda,$$

а, отже, ці оцінки є конзистентними. Маємо

$$\sqrt{n}(\lambda_n(1) - \lambda) = -\lambda\frac{\frac{\sum_{i=1}^n (x_i - \lambda^{-1})}{\sqrt{n}}}{n^{-1}\sum_{i=1}^n x_i} = -\frac{\frac{\sum_{i=1}^n (x_i - \lambda^{-1})}{\sqrt{n}\lambda^{-1}}}{n^{-1}\sum_{i=1}^n x_i}. \tag{6}$$

Оскільки $\mathbf{D^2}x_i = 1/\lambda^2$, то на підставі центральної граничної теореми

$$\frac{\sum_{i=1}^n (x_i - \lambda^{-1})}{\sqrt{n}\lambda^{-1}} \underset{n\to\infty}{\Longrightarrow} \mathbb{N}(0,1).$$

Тепер з (6) дістаємо $\sqrt{n}(\lambda_n(1)-\lambda) \underset{n\to\infty}{\longrightarrow} \mathbb{N}(0, |\mathbf{M}x_i|^{-2}) = \mathbb{N}(0,\lambda^2)$ і, отже, асимптотична дисперсія для $\sqrt{n}(\lambda_n(1) - \lambda)$ дорівнює λ^2. Аналогічно

$$\sqrt{n}(\lambda_n(2)-\lambda) = \frac{\sqrt{n}\ \ \sqrt{2}-\lambda\sqrt{n^{-1}\sum_{i=1}^n x_i^2}}{\sqrt{n^{-1}\sum_{i=1}^n x_i^2}} = \frac{\sqrt{n}\ \ 2-n^{-1}\lambda^2\sum_{i=1}^n x_i^2}{\sqrt{n^{-1}\sum_{i=1}^n x_i^2}\ \ \sqrt{2}+\lambda\sqrt{n^{-1}\sum_{i=1}^n x_i^2}} =$$

$$= -\sqrt{20}\frac{\frac{\sum_{i=1}^n (x_i^2 - 2\lambda^{-2})}{\sqrt{20n}\lambda^{-2}}}{\sqrt{n^{-1}\sum_{i=1}^n x_i^2}\ \ \sqrt{2}+\lambda\sqrt{n^{-1}\sum_{i=1}^n x_i^2}}. \tag{7}$$

Оскільки $\mathbf{D^2}x_i^2 = \mathbf{M}x_i^4 - (\mathbf{M}x_i^2)^4 = 24/\lambda^4 - 4/\lambda^2 = 20/\lambda^4$, то на підставі центральної граничної теореми маємо

$$\frac{\sum_{i=1}^n (x_i^2 - 2\lambda^{-2})}{\sqrt{20n}\lambda^{-2}} \underset{n\to\infty}{\Longrightarrow} \mathbb{N}(0,1).$$

Тепер з (7) випливає

$$\sqrt{n}(\lambda_n(2) - \lambda) \underset{n\to\infty}{\Longrightarrow} -\frac{\sqrt{20}}{\sqrt{2}\lambda^{-1}\ \ \sqrt{2}+\sqrt{2}}\mathbb{N}(0,1) = \mathbb{N}(0, \frac{5\lambda}{2}),$$

і, отже, асимптотична дисперсія для $\sqrt{n}(\lambda_n(2) - \lambda)$ дорівнює $5\lambda^2/4$. Звідси випливає, що оцінка $\lambda_n(1)$ є асимптотично ефективнішою ніж оцінка $\lambda_n(2)$.

Ще один приклад, але тепер для дискретних розподілів.

Приклад 2. *Маємо схему Бернуллі з параметрами n, p, де n відоме, а p є невідомим параметром. Кількість успіхів у цій схемі дорівнює k. Довести, що оцінка $p(n) = k/n$ є найефективнішою оцінкою параметра p.*

Р о з в' я з о к. Перш за все, запишемо так сформульовану задачу у вигляді, який вживається в формулюванні нерівності Рао-Крамера. Схему Бернуллі можна представити як послідовність n незалежних випробувань з ймовірністю успіху в одному випробуванні, рівною p. Генеральна популяція ξ має тоді розподіл типу 0–1. Значить тепер наша задача може бути записана так. Генеральна популяція ξ набуває значення 1 з ймовірністю p і 0 з ймовірністю $1 - p$, тобто $\mathbf{P}(\xi = j) = p^j(1-p)^{1-j}$, де $j = 0, 1$. Вибірка для характеристики ξ це є множина $k_1, k_2, ..., k_n$ де k_i набуває значення 0 або 1 і $\sum_{i=1}^n k_i = k$. Маємо $p(n) = k/n = \sum_{i=1}^n k_i/n$.

Оскільки $\mathbf{M}k_i = p$, $\mathbf{D^2}k_i = pq$, то, по-перше, $\mathbf{M}pp(n) = p$ і отже оцінка є незміщеною $p(n)$, а, по-друге,

$$\mathbf{D^2}p(n) = \frac{pq}{n}. \tag{8}$$

В нашій ситуації $p(k) = p^k q^{1-k}$ і значить $\ln p(\xi) = \xi \ln p + (1-\xi)\ln q$. Тому

$$I^2(p) = \mathbf{M}\left(\frac{\partial \ln p(\xi)}{\partial p}\right)^2 = \mathbf{M}\left(\frac{\xi}{p} - \frac{1-\xi}{q}\right)^2 = \mathbf{M}\left(\frac{\xi^2}{p^2} - 2\frac{\xi-\xi^2}{pq} + \frac{1-\xi}{q^2}\right) = \frac{1}{pq}$$

Остання рівність і (8) дають

$$\mathbf{M}(p(n) - p)^2 = \frac{1}{nI^2(p)}$$

що і означає R-ефективність оцінки $p(n)$.

ЗАВДАННЯ

1. Нехай x_1, x_2, x_3 – вибірка спостережень за випадковою величиною ξ з рівномірним розподілом на інтервалі $[0;1]$. Для параметра $m = \mathbf{M}\xi$ розглянемо наступні оцінки:

$$m_1 = \frac{x_1 + x_2 + x_3}{3}, \quad m_2 = m_e(3), \quad m_3 = \frac{x_{(1)} + x_{(3)}}{2}.$$

   Показати, що всі вони є незміщеними.

   (*Під* $m_e(3)$ *розуміємо медіану вибірки* (*див. означення* 6, *або формулу* (4).)

2. Генеральна сукупність ξ має рівномірний розподіл на відрізку $[\theta - 1; \theta + 1]$, де θ вважаємо параметром. Довести, що оцінка $\theta(n) = \max\{x_1, ..., x_n\} + \min\{x_1, ..., x_n\}$ параметра θ є незміщеною. Чи буде ця оцінка консистентною?

3. Нехай $x_1, ..., x_n$ - випадкова вибірка спостережень за нормальним розподілом $\mathbb{N}(m, \sigma^2)$. Показати, що оцінка

$$\theta_n = \frac{1}{n} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2$$

   є асимптотично незміщеною оцінкою дисперсії σ^2.

4. Нехай генеральна популяція має геометричний розподіл з невідомим параметром $\theta \in (0;1)$, тобто $\mathbf{P}(\xi = n) = \theta^n(1-\theta), \qquad n \geq 0$. Довести, що не існує незміщеної оцінки параметра θ на підставі вибірки, яка має лише одне спостереження.

5. Довести, що оцінка $\bar{S}(n)$ є асимптотично незміщеною для параметра σ розподілу генеральної сукупності $\xi \in \mathbb{N}(m, \sigma^2)$, де m є відомим.

   Вказівка. *Скористатися прикладом* 1, *с.* 197 *і співвідношенням*

$$\ln \Gamma(x) = (x - \frac{1}{x}) \ln x - x + \frac{1}{2} \ln 2\pi + O(x^{-1}), \quad x \to \infty.$$

6. Нехай генеральна сукупність $\xi \in \mathbb{N}(m_0, \sigma^2)$, де m_0 є відомим. Довести, що оцінка

$$\theta_n = \frac{1}{n} \sqrt{\frac{\pi}{2}} \sum_{i=1}^{n} |x_i - m_0|$$

   є незміщеною оцінкою параметра σ.

7. Довести, що для $\xi \in \mathbb{N}(m, \sigma^2)$, статистика

$$\theta_n = \frac{1}{n} \sum_{i=1}^{n} |x_i - x(n)|$$

має наступні властивості:

$$\mathbf{M}\theta_n = \sigma\sqrt{\frac{2}{\pi}(1-\frac{1}{n})}, \quad \mathbf{D^2}\theta_n = \frac{\sigma^2}{\pi n}\ 1-\frac{1}{n}\ \ \frac{\pi}{2}+\sqrt{n(n-2)}-n+\arcsin\frac{1}{n-1}\ .$$

А статистика $\theta_n\ \ \frac{2}{\pi}(1-\frac{1}{n})\ ^{-1/2}$ буде незміщеною оцінкою для σ. Знайти її дисперсію.

8. Довести тотожність (3).

9. Нехай генеральна сукупність ξ має розподіл гамма з параметрами p, λ, тобто $\xi \in \mathbb{G}(p, \lambda)$, де p є знане, а λ виступає як незнаний параметр. Довести, що оцінка $S = \frac{np-1}{nx(n)}$ параметра λ є незміщена та консистентна. Знайти ефективність цієї оцінки і показати, що вона є асимптотично ефективною.

   Вказівка. *Скористатися вправою* 8 *с.* 177.

10. Довести, що для генеральної сукупності $\xi \in \mathbb{N}(m, \sigma^2)$ незміщена оцінка $\widehat{S}^2(n) = \frac{1}{n-1}\sum_{k=1}^{n}\ x_k - x(n)\ ^2$ параметра σ^2 має більше середньоквадратичне відхилення ніж зміщена оцінка $\widetilde{S}^2(n) = \frac{1}{n+1}\sum_{k=1}^{n}\ x_k - x(n)\ ^2$.

11. Генеральна сукупність має щільність рівномірний розподіл на інтервалі $[0;\theta]$, де $\theta > 0$ є параметром, для якого запропоновано оцінки $\theta_n^{(1)} = \frac{n+1}{n}x_{(n)}$, $\theta_n^{(2)} = \frac{2}{n}\sum_{i=1}^{n} x_i$. Довести, що ці оцінки є незміщеними і перша з них є кращою.

Лекція 8. Емпірична функція розподілу

8.1 Емпірична функція розподілу

Нехай $\vec{x} = (x_1, ..., x_n)$ є n -елементною вибіркою спостережень за випадковою величиною ξ з функцією розподілу $F(x)$. Для кожного $x \in R^1$ позначимо через $\nu(x)$ число елементів вибірки $\vec{x}$, які потрапили в інтервал $(-\infty, x)$. Введемо до розгляду функцію

$$F_n(x) \overset{def}{=} \frac{\nu(x)}{n}, \qquad -\infty < x < \infty. \tag{1}$$

Вочевидь, функція $F_n(x)$ є неспадна, неперервна зліва і $F_n(-\infty) = 0, F_n(\infty) = 1$. Отже, $F_n(x)$ є деякою функцією розподілу для кожної фіксованої вибірки $\vec{x}$.

Визначення 1. Функція $F_n(x)$, яка визначена в (1), називається *емпіричною функцією розподілу* випадкової величини ξ.

Можна навести інше, більш наглядне, визначення емпіричної функції розподілу. Для цього зробимо припущення, що у вибірці $\vec{x}$ не має однакових елементів. Утворимо варіаційний ряд з нашої вибірки. Тоді будемо мати послідовність $x_{(1)} < x_{(2)} < ... < x_{(n)}$. Тепер, як неважко бачити, можемо записати

$$F_n(x) = \begin{cases} 0, & \text{для} \quad x \le x_{(1)}, \\ k/n, & \text{для} \quad x_{(k)} < x \le x_{(k+1)}, \quad k = 1, 2, .., n-1, \\ 1, & \text{для} \quad x > x_{(n)}. \end{cases} \tag{2}$$

Вочевидь, що функція $F_n(x)$, яка визначена в (2), має стрибок величини $1/n$ в кожній точці $x_{(i)}$. Якщо ж деякі елементи вибірки $\vec{x}$ є однаковими, то емпірична функція розподілу знову може бути визначена як в (2), але тепер, якщо, наприклад, для деяких k, i будемо мати $x_{(k)} < x_{(k+1)} = x_{(k+2)} = ... = x_{(k+i)} < x_{(k+i+1)}$, то $F_n(x) = k/n$ для $x_{(k)} < x \le x_{(k+1)}$ і $F_n(x) = (k+i)/n$ для $x_{(k+1)} < x \le x_{(k+i)}$. Тобто функція $F_n(x)$ в точці $x_{(k+1)}$ буде мати стрибок величини i/n.

Рис. 8.1: Емпірична функція розподілу.

Приклад 1. *Побудувати емпіричну функцію для вибірки* 0; 3; -2; 0; 1.

Р о з в' я з о к. Емпірична функція розподілу для цієї вибірки зображена нижче. В точці $x = 0$ ця функція має стрибок величини $2/5$, а в решті точок величина стрибка дорівнює $1/5$.

Наведемо ще один вираз для емпіричної функції розподілу, який буде нам потрібен при доведенні деяких тверджень.

Позначимо через $\chi_A(x)$ характеристичну функцію множини A, тобто $\chi_A(x) = 1$, якщо $x \in A$ і $\chi_A(x) = 0$, якщо $x \notin A$. Тоді, вочевидь, для емпіричної функції розподілу можемо записати

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\chi_{(-\infty,x)}(x_i). \tag{3}$$

Зауваження 1. Оскільки, випадкові величини x_i є незалежними і мають однаковий розподіл, то для кожного $x \in \mathbf{R}^1$ випадкові величини $\chi_{(-\infty,x)}(x_i), i = 1, 2, ..., n$ є також незалежними і також мають однаковий розподіл (типу $0-1$). Тому вираз (3) можемо трактувати як представлення емпіричної функції розподілу через суму n незалежних одинаково розподілених випадкових величин.

З цього зауваження зокрема випливає, що емпірична функція розподілу є статистикою. Легко можна знайти розподіл цієї статистики.

Теорема 1. *Якщо випадкова величина ξ має функцію розподілу $F(x)$, то*

$$\mathbf{P}(F_n(x) = \frac{m}{n}) = C_n^k(F(x))^m(1 - F(x))^{n-m}, \quad m = 0, 1, ..., n. \tag{4}$$

Д о в е д е н н я. Нехай x є фіксованим. Процес вибору вибірки можна зобразити наступним чином. Якщо при черговому виборі елемента маємо $x_i < x$, то говоримо, що відбувся "успіх". Позаяк, згідно з нашим загальним припущенням випадкова величина x_i має такий самий розподіл як ξ, тобто $F(x)$, то ймовірність "успіху" є $\mathbf{P}(x_i < x) = \mathbf{P}(\xi < x) = F(x)$, а процес вибору вибірки є класичною схемою Бернуллі. Вочевидь, подія $\{F_n(x) = \frac{m}{n}\}$ еквівалентна тому, що при виборі вибірки будемо мати рівно m "успіхів" і згідно з виразом ймовірності числа "успіхів" в схемі Бернуллі маємо формулу (4).

■

З цього твердження випливає, що $\mathbf{M}F_n(x) = F(x)$ а отже маємо

Наслідок 1. *Емпірична функція розподілу є незміщеною оцінкою теоретичної функції розподілу.*

Оскільки під знаком суми в (3) випадкові величини $\xi_i = \chi_{(-\infty,x)}(x_i)$, $i = 1, ..., n$ є незалежними і $\mathbf{M}\xi_i = \mathbf{P}(x_i < x) = \mathbf{P}(\xi < x) = F(x)$, то на підставі посиленого закону великих чисел будемо мати

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\xi_i \xrightarrow[n\to\infty]{\text{м.н.}} \mathbf{M}\xi_1 = F(x).$$

Отже, ми довели наступну теорему.

Теорема 2. *Для будь-якого x має місце рівність*

$$\mathbf{P}(\lim_{n\to\infty} |F_n(x) - F(x)| = 0) = 1 \tag{5}$$

Оскільки, $\mathbf{D^2}\chi_{(-\infty,x)}(x_i) = F(x)(1-F(x))$, то з центральної граничної теореми випливає

$$\mathbf{P}(\frac{\sqrt{n}(F_n(x) - F(x))}{\sqrt{F(x)(1 - F(x)}} < z) = \mathbf{P}(\frac{\sum_{i=1}^{n}\xi_i - nF(x)}{\sqrt{nF(x)(1 - F(x)}} < z) \xrightarrow[n\to\infty]{} \Phi(z),$$

а тому маємо такий результат.

Наслідок 2. *Для кожного x емпіричні функції розподілу $F_n(x)$ є асимптотично нормальні з параметрами $F(x)$, $F(x)(1 - F(x)/n$.*

В (5) ми отримали, що для кожного фіксованого x емпірична функція розподілу збігається до теоретичної функції розподілу з ймовірністю одиниця. Наступна теорема підсилює цей результат.

Теорема 3. (Глівенко[1] –Кантелі[2]) Має місце рівність

$$\mathbf{P}(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0) = 1. \tag{6}$$

Отже, з цієї теореми випливає, що в (5) збіжність є рівномірною відносно x.

Д о в е д е н н я. Ми обмежимося випадком коли функція розподілу $F(x)$ є неперервною. Доведення в загальному випадку можна знайти в [1], [21]. Нехай $\varepsilon > 0$ буде малим числом але таким, що число $N = 1/\varepsilon$ є цілим. З неперервності функції $F(x)$ випливає, що існують числа $-\infty \le x_0 < x_1 < .... < x_N \le \infty$ такі, що

$$F(x_0) = 0, F(x_1) = \varepsilon, ..., F(x_k) = k/N, ..., F(x_N) = 1.$$

Якщо $x_k \le x < x_{k+1}$, то $F(x_{k+1}) = F(x_k) + \varepsilon$ і

$$F_n(x) - F(x) \le F_n(x_{k+1}) - F(x_k) = F_n(x_{k+1}) - F(x_{k+1}) + \varepsilon. \tag{7}$$

Аналогічно

$$F_n(x) - F(x) \ge F_n(x_k) - F(x_{k+1}) = F_n(x_k) - F(x_k) - \varepsilon. \tag{8}$$

З (7) і (8) маємо

$$\sup_{-\infty\le x\le\infty} |F_n(x) - F(x)| \le \max_{0\le k\le N} |F_n(x_k) - F(x_k)| + \varepsilon. \tag{9}$$

Оскільки, згідно з попередньою теоремою $F_n(x_k) \xrightarrow[n\to\infty]{\text{м.н.}} F(x_k)$ для будь-якого $k = 0, 1, ..., N$, то існують множини елементарних подій A_k, для яких $\mathbf{P}(A_k) = 1$ і $\lim_{n\to\infty} F_n(x_k, \omega) = F(x_k)$ для всіх $\omega \in A_k$, $k = 0, 1, ..., N$. Нехай тепер $A = \cup_0^N A_k$. Тоді

$$\mathbf{P}(A) = 1 - \mathbf{P}(\overline{A}) = 1 - \mathbf{P}(\cap_{k=0}^N \overline{A}_k) \ge 1 - \sum_{k=0}^N \mathbf{P}(\overline{A}_k) = 1$$

і, отже, $\mathbf{P}(A) = 1$. Тепер маємо $\lim_{n\to\infty} F_n(x_k, \omega) = F(x_k)$ для всіх $\omega \in A$, $k = 0, 1, ..., N$, що в свою чергу означає, що можна знайти $N(\omega)$ такі, що

$$\max_{0\le k\le N} |F_n(x_k) - F(x_k)| \le \varepsilon, \quad n \ge N(\omega), \quad \omega \in A,$$

що разом з (9) дає

$$\sup_{-\infty\le x\le\infty} |F_n(x) - F(x)| \le 2\varepsilon, \quad n \ge N(\omega), \quad \omega \in A.$$

Позаяк, в останньому виразі ε можна взяти як завгодно малим, то теорема для неперервних $F(x)$ є доведеною.

---

[1] Гливенко Валерій Іванович (02.01.1897(Київ)–15.02.1940(Москва))–закінчив Московський університет (1925). Головні наукові заінтересування– основи математики та математична логіка, теорія функцій дійсної змінної та теорія ймовірностей. Одним з перших вивчав питання обґрунтування математики.

[2] Кантеллі (*Francesco Paolo Cantelli*) (20.12.1875(Палермо)–21.07.1966 (Рим)) – італійський математик. Сфери зацікавлення: астрономія, теорія ймовірностей, застосування математики в економії.

Теорема 4. *Нехай $k_n(p)$ буде квантиллю рівня $p \in (0;1)$ для вибірки $\vec{x} = (x_1, ..., x_n)$. Тоді має місце рівність*

$$k_n(p) = \min\{x : F_n(x) > p\}. \tag{10}$$

Доведення. Без втрати загальності можемо припустити, що у вибірці $x_1, ..., x_n$ не має однакових елементів, і нехай, як завжди, $x_{(1)} < ... < x_{(n)}$ буде варіаційним рядом для цієї вибірки. Для $k_n(p)$ маємо (див. визначення (6), с. 160)

$$k_n(p) = \min\{x_{(k)} : k/n > p\}. \tag{11}$$

Позаяк, $0 < p < 1$, а емпірична функція розподілу є функцією сходинковою зі стрибками в точках $x_{(i)}$, $i = 1, ..., n$, то завжди існує такий стрибок, в якому функція $F_n(x)$ долає рівень p. Можливі лише дві ситуації:

а) існує таке k, що $F_n(x_{(k)} + 0) > p$, а $F_n(x_{(k)}) < p$;

б) існує таке k, що $F_n(x_{(k)} + 0) > p$, а $F_n(x_{(k)}) = p$.

Якщо має місце а), тоді $k/n = F_n(x_{(k)} + 0) > p$, а $(k-1)/n = F_n(x_{(k)}) < p$ і, значить, згідно з (11) будемо мати $k_n(p) = x_{(k)}$. Оскільки, в цій ситуації рівність $\min\{x : F_n(x) > p\} = x_{(k)}$ є очевидною, то маємо (10).

Якщо має місце б), то аналогічно $k/n = F_n(x_{(k)}+0) > p$, а $(k-1)/n = F_n(x_{(k)}) = p$ і знову маємо $k_n(p) = x_{(k)}$. Тепер $F_n(x) = p$, $x_{(k-1)} < x \le x_{(k)}$ і $F_n(x_{(k)}+0) > p$, звідки знову отримаємо $\min\{x : F_n(x) > p\} = x_{(k)}$, що й закінчує наше доведення.

На закінчення цього підрозділа наведемо без доведення теорему, яка була доведена А.М.Колмогоровим[1] в 1933 році, і яка часто використовуються при перевірці статистичних гіпотез. Доведення цієї теореми можна знайти в [1] і [27].

Теорема 5. *Якщо функція розподілу $F(x)$ є неперервною, то*

$$\lim_{n\to\infty} \mathbf{P}(\sqrt{n} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| < z) = K(z) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2z^2}. \tag{12}$$

Функція $K(z)$ називається *функцією розподілу Колмогорова.* Ця функція розподілу часто зустрічається в математичній статистиці при перевірці статистичних гіпотез.

Границя в (12) не залежить від $F(x)$, за умови, що ця функція є неперервною. Якщо $F(x)$ не є неперервною, то границя в (12) теж існує, але тоді може бути залежною від теоретичної функції розподілу (тобто від $F(x)$) , а більш точно, від значень цієї функції в точках розривів. Більше інформації з цього питання читач може знайти в [34].

[1]Андрій Миколайович Колмогоров (25.04.1903–20.10.1987) – радянський математик, працював в різних розділах математики, фундатор сучасної теорії ймовірностей.

ЗАВДАННЯ

1. Довести, що для монотонної обмеженої функції $F(x)$ і кожного $\varepsilon > 0$ можна знайти пункти $-\infty \le x_0 < x_1 < ... < x_N \le \infty$ так, що $F(x_{k+1}) - F(x_k + 0) \le \varepsilon$.

2. Для генеральної популяції $\xi \in \mathbb{N}(0,1)$ отримано вибірку: $x_1 = -0,5; x_2 = 0; x_3 = 0,25; x_4 = 0,54; x_5 = 1; x_6 = -1; x_7 = 0,15; x_8 = 0,2; x_9 = -0,25; x_{10} = -0,6$.

   а) Знайти емпіричну функцію розподілу $F_{10}(x)$.

   б) Полічити $\mathbf{M}F_{10}(x)$, $\mathbf{D^2}F_{10}(x)$ для $x = 0,1$.

3. Для згрупованих вибірок

x_i	2	5	8	11	14
n_i	5	3	8	10	4

x_i	1	3	5	7	9
n_i	3	5	7	11	4

   побудувати полігон частот та емпіричні функції розподілу.

4. Маємо вибірку з рівномірного на $[0;10]$ розподілу:7,6; 2,5; 0,3;9,8; 8,3;3,1; 6,1; 6,9; 1,2; 9,9; 2,2; 7,0; 8,1; 8,6; 5,6. Побудувати варіаційний ряд та емпіричну функцію розподілу $F_n(x)$. Обчислити $\sup_{-\infty<x<\infty} |F(x) - F_n(x)|$, де $F(x)$ функція рівномірного на $[0;10]$ розподілу.

Лекція 9. Інтервальне оцінювання

9.1 Загальні зауваження

Методи оцінювання, якими ми займалися до цього часу, дають точкові оцінки невідомих параметрів, тобто якщо θ_n є оцінкою параметра θ, то взагалі кажучи, та оцінка не дозволяє сказати яким є справжнє значення параметра θ. І якщо нас цікавить відповідь на питання, яке справжнє значення приймає параметр θ, то єдине, що ми можемо стверджувати є те, що $\theta_n \approx \theta$ для великих n за умови, що оцінка є сильно консистентною. З точки зору практики в багатьох випадках таке оцінювання невідомого параметра є недостатнім. І не лише з тієї причини, що в практичних дослідженнях часто не маємо достатньої кількості спостережень щоб стверджувати, що значення оцінки близьке до істинного значення параметра. Пояснемо це на такому простому прикладі.

Приклад 1. *Нехай для деякої випадкової величини ξ нас цікавить значення параметра $\theta = \mathbf{M}\xi$. Якщо, наприклад, ξ споживання палива автомобілем на 100 км. пробігу, то θ є дуже суттєвою характеристикою для кожного власника авто. А якщо ξ врожай з одного гектара поля, то тоді θ середня врожайність господарства - теж дуже важливий параметр для кожного селянина. Процес оцінювання параметра θ є стандартним: спочатку потрібно зробити кілька незалежних спостережень ξ (генеральної популяції), а потім оцінити θ, використовуючи відповідну оцінку. Припустимо, що $\xi \in N(\theta, \sigma)$ (і таке припущення є часто сенсовним) і нехай було здійснено, наприклад, 10 спостережень. Отже, маємо вибірку $x_1, ..., x_{10}$. За оцінку θ беремо середнє арифметичне, тобто $x(10) = 0,1(x_1 + ... + x_n)$. Як ми вже встановили вище (див. Теорему* **??**), *ця оцінка є ефективною. Але що дає значення, скажімо, $x(n) = 8.4$ літрів для водія, або $x(n) = 4.5$ тонн для селянина, якщо він все одно не знає, яким є справжнє значення θ. Більш того, він ніколи цього не дізнаєтьcя на маючи скінчену кількості спостережень. А зробити так багато спостережень, щоб мати впевненість, що оцінка є близькою до параметра θ в практиці є неможливим. Для героїв нашого прикладу було б більш зручним, якби на основі цих десяти спостережень можна було б знайти інтервал $[a, b]$ такий, що $a < \theta < b$, тобто інтервал (a, b) накриває θ. Таке оцінювання для θ було б більш інформативним, ніж точкове оцінювання $x(n)$ за умови, звичайно, що цей інтервал не є аж занадто широким. Наприклад, таке оцінювання $7.5 < \theta < 8.5$ для водія і $4.3 < \theta < 4.8$ для селянина давало б більше інформації для подальшої поведінки в порівнянні з точковими оцінками. Зрозуміло, що значення a, b повинні бути підраховані на основі вибірки (оскільки лише таку інформацію маємо), а тому вони мають бути функціями вибірки, тобто $a = a(\vec{x})$, $b = b(\vec{x})$ і вигляд цих функцій ми маємо знати ще до виконання спостережень для того, щоб після отримання вибірки підставити значення $\vec{x}$ і підрахували конкретне значення для a, b. Аналогі-*

чна ситуація і для точкового оцінювання - вигляд статистики відомий заздалегідь. Зрозуміло, що для функцій $a(\vec{x})$, $b(\vec{x})$ має бути виконана умова $a(\vec{x}) < b(\vec{x})$ для всіх $\vec{x}$. Але якщо вибірку $\vec{x}$ розглядати як послідовність незалежних випадкових величин, то $a(\vec{x})$, $b(\vec{x})$ також будуть випадковими величинами, а це означає, що лише після проведення експерименту (спостережень x_i) ми маємо справжнє значення для a, b. І якщо, наприклад, функція розподілу $b(x)$ є $\mathbf{P}(b(\vec{x}) < x) = G(x)$ і має щільність $g(x)$, то нам потрібно щоб відбувалась подія $\{b(\vec{x}) > \theta)$ і ймовірність цієї події є затемнена область на графіку, який поданий нижче.

Рис. 9.1:

Але не виключено, що станеться подія $\{b(\vec{x}) < \theta\}$ і ймовірність цієї події є не затемнена область на тому ж самому графіку (не затемнена область між $g(x)$ та віссю $0X$). Тоді наш інтервал (a, b) не накриє θ. Аналогічним чином, може статися так, що відбудеться подія $\{a(\vec{x}) > \theta\}$ і, тоді інтервал (a, b) теж не накриє θ. З цих міркувань випливає, що ми не можемо розраховувати на те, що на 100% інтервал (a, b) накриє θ. І це є наслідком того, що скінчена вибірка не дає нам цілковитої інформації про генеральну популяція ξ. Як ми вже казали, з малюнку 9.1 випливає, що ймовірність небажаної для нас події $\{b(\vec{x}) < \theta\}$ є не затемнена область між $g(x)$ та віссю $0X$ і, отже, нам потрібно намагатися вибрати функцію $b(\cdot)$ так, щоб ця область була малою. Аналогічна ситуація є з вибором функції $a(\cdot)$. Тоді ми можемо очікувати,що не бажана для нас подія $\theta \notin (a(\vec{x}), b(\vec{x}))$ буде мати малу ймовірність, тобто на практиці буде відбуватися рідко.

Такі міркування виводять нас на наступне визначення надійного інтервалу.

Визначення 1. *Надійним інтервалом* для параметра $\theta \in \Theta$ з рівнем довіри $1 - \alpha$ $(0 < \alpha < 1)$ будемо називати інтервал (θ_1, θ_2), який задовольняє умовам:

а) його межі $\theta_1 = \theta_1(x_1, \ldots x_n)$, $\theta_2 = \theta_2(x_1, \ldots, x_n)$ є функціями випадкової вибірки і не залежать від параметра θ;

б) ймовірність покриття цим інтервалом невідомого параметра θ є не меншою ніж 1-α, тобто

$$\mathbf{P}(\theta_-(x_1, \ldots x_n) < \theta < \theta_+(x_1, \ldots, x_n)) \geq 1 - \alpha \tag{1}$$

для довільного $\theta \in \Theta$.

За α вибирють мале число і найчастіше одне з наступних: 0,1; 0,05; 0,01; 0,005; 0,001.

Зауваження 1. В (1) стоїть знак " $\geq$ ". Це є тому, що, як побачимо далі, в деяких ситуаціях (якщо, наприклад, генеральна популяція має дискретний розподіл) неможливо, взагалі кажучи, зробити вибір $\theta_{\pm}$ так, щоб в (1) була рівність. Якщо ж генеральна популяція має неперервний розподіл, то зазвичай можемо отримати в (1) рівність.

Вочевидь, є слушним прагнути до того, щоб інтервал, який накриває невідомий параметр, був як найменшої довжини і, тому умова

$$\theta_+(x_1, \dots x_n) - \theta_-(x_1, \dots, x_n) \to \min \tag{2}$$

завше буде братися до уваги при побудові надійного інтервалу.

Подивимось тепер, що нам дає той факт, що, наприклад, для середнього споживання бензину автівкою маємо надійний інтервал $(7,5; 8,5)$ з рівнем довіри $0,95$, тобто $\mathbf{P}(7,5 < \theta < 8,5) \geq 0,95$. Можемо сказати так: якщо зробимо 100 незалежних оцінювань параметра θ, то в середньому не більше ніж 5 з них не потраплять до інтервалу $(7,5; 8,5)$. На практиці це означає, що на 95% ми впевнені в тому, що справжнє значення параметра θ лежить в інтервалі $(7,5; 8,5)$. Але це не виключає ситуації, що ми помиляємося і насправді θ не належить до інтервалу $(7,5; 8,5)$. Але ймовірність такої ситуації є 5% і, тому можемо діяти так, ніби на 100% θ належить інтервалу $(7,5; 8,5)$. Таким є життя. Ми не так вже часто є впевненими в чомусь на 100%. Таку впевненість можуть забезпечити (вочевидь на словах) лише рекламні агенції. А здоровий глузд підказує, що на основі неповної інформації про ξ (тобто лише на підставі вибірки) не можемо одержати 100-процентні знання про ξ. В різних статистичних дослідженнях, які використовуються на практиці (соціології, економіці, медицині і т.д.), часто вибирають $\alpha = 0,05$ і кажуть про 5% рівні довіри.

Переходимо тепер до опису процедури побудови надійного інтервалу. Зауважимо на початку, що не існує усталеної методики побудови надійного інтервалу, яка б підходила до всіх параметрів і тому вирішення цієї проблеми залежить від конкретної ситуації. Але не зважаючи на це, можна описати деякі найуживаніші підходи при побудові надійних інтервалів. Зазвичай використовують два головних методи: побудова надійного інтервалу за допомогою статистик і за допомогою граничних теорем. Про другий підхід поведемо мову трохи пізніше, а зараз опишемо методику, яка базується на використанні статистик.

9.2 Побудова надійного інтервалу за допомогою статистик

Нехай на підставі вибірки $x_1, ..., x_n$ треба побудувати надійний інтервал для параметра θ. Розглянемо ситуацію лише скалярного параметра $\theta \in \Theta = [a, b], \quad -\infty \leq a < b \leq \infty$.

Теорема 1. *Нехай існує функція $T(\vec{z}, \theta) : \mathbf{R}^n \times \Theta \to \mathbf{R}^1$ така, що виконані наступні умови:*

а) *для кожного фіксованого $\vec{z}$ функція $T(\vec{z}, \theta)$ є неперервною і монотонною стосовно θ;*

б) *розподіл*

$$\mathbf{P}(T(\vec{x}, \theta) < x) \stackrel{\text{def}}{=} G(x) \tag{3}$$

відомий, не залежить від параметра θ і є неперервним.

Нехай $\alpha_1, \alpha_2 > 0$ є фіксованими числами такими, що $\alpha_1 + \alpha_2 = \alpha$ і $k(\alpha_1), k(1 - \alpha_2)$ - квантилі рівня $\alpha_1,\ 1 - \alpha_2$ розподілу $G(x)$. Тоді:

а) *якщо функція $T(\vec{z}, \theta)$ є зростаючою відносно θ для кожного $\vec{z}$, то*

$$\mathbf{P}(\theta_-(\alpha_1) < \theta < \theta_+(\alpha_2)) = 1 - \alpha, \tag{4}$$

де $\theta_-(\alpha_1) = T^{-1}(\vec{x}, k(\alpha_1)), \quad \theta_+(\alpha_2) = T^{-1}(\vec{x}, k(1 - \alpha_2))$;

б) *якщо функція $T(\vec{z}, \theta)$ є спадною відносно θ для кожного $\vec{z}$, то*

$$\mathbf{P}(\theta_-(\alpha_2) < \theta < \theta_+(\alpha_1)) = 1 - \alpha,$$

де $\theta_-(\alpha_2) = T^{-1}(\vec{x}, k(1 - \alpha_2)), \quad \theta_+(\alpha_1) = T^{-1}(\vec{x}, k(\alpha_1))$.

Д о в е д е н н я. Розглянемо ситуацію зростаючої функції $T(\vec{z}, \theta)$. Маємо

$$\mathbf{P}(z_1 < T(\vec{x}, \theta) < z_2) = G(z_2) - G(z_1). \tag{5}$$

Якщо в цій рівності взяти $z_1 = k(\alpha_1),\ z_2 = k(1 - \alpha_2)$, тоді, позаяк

$$G(k(\alpha_1)) = \alpha_1, \qquad G(k(1 - \alpha_2)) = 1 - \alpha_2, \tag{6}$$

отримаємо

$$\mathbf{P}(k(\alpha_1) < T(\vec{x}, \theta)) < k(1 - \alpha_2)) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha. \tag{7}$$

Оскільки, згідно припущення функція $T(\vec{x}, \theta)$ є неперервною і зростаючою відносно θ для кожного $\vec{x}$, тоді існує неперервна (і зростаюча) обернена функція $T^{-1}(\vec{x}, \theta)$ для кожного $\vec{x}$, і отже нерівність $a < T(\vec{x}, \theta) < b$ має розв'язок вигляду $T^{-1}(\vec{x}, a) < \theta < T^{-1}(\vec{x}, b)$ Тепер з (7) отримуємо

$$\mathbf{P}(\theta_-(\alpha_1) < \theta < \theta_+(\alpha_2)) = 1 - \alpha,$$

де числа $\theta_-(\alpha_1),\ \theta_+(\alpha_2)$ визначені в умові теореми. Це і дає нам потрібний надійний інтервал.

Зробимо тепер коментар до описаного підходу. Вочевидь, що числа α_1, α_2, для яких справедлива рівність $\alpha_1 + \alpha_2 = \alpha$, можемо вибрати багатьма способами. Але оскільки ми зацікавлені щоб довжини інтервалу була як найменша (див. (2), то в якості цих чисел належить вибрати $\alpha_1^*,\ \alpha_2^*$ так, щоб була виконана умова

$$\theta_+(\alpha_2) - \theta_-(\alpha_1) = \inf\bigl(\theta_+(\alpha_2) - \theta_-(\alpha_1)\bigr), \tag{8}$$

де inf береться по всіх $\alpha_1, \alpha_2 > 0$ таких, що $\alpha_1 + \alpha_2 = \alpha$.

Завдання (8) не є простим для розв'язання. Можна довести, що коли розподіл $G(x)$ є симетричним, тобто $G(x) = 1 - G(-x)$, тоді $\alpha_1 = \alpha_2 = \alpha/2$ буде розв'язком для (8). Але на практиці зазвичай вибирають $\alpha_1 = \alpha_2 = \alpha/2$ і для несиметричних розподілів. Тоді довжина інтервалу може не бути найменшою, але процедура побудови надійного інтервалу стає однозначною.

Якщо функція $G(x)$ не є неперервною, тоді може не існувати чисел $k(\alpha_1),\ k(1-\alpha_2)$, для яких мають місце рівності (6), які були важливими при побудові надійного інтервалу. Нехай тепер виконані всі умови теореми 1 за винятком того, що функція $G(x)$ є неперервною. Нехай як і раніше $k(\alpha_1),\ k(1-\alpha_2)$ є квантилями рівня $\alpha_1,\ 1-\alpha_2$ розподілу $G(x)$. Якщо функція $g(x)$ є неперервною в точках $k(\alpha_1),\ k(1-\alpha_2)$ то попередня конструкція не потребує жодних змін. В протилежному випадку замість (6) будемо мати:

$$\begin{aligned} &G(k(\alpha_1)) \le \alpha_1, && G(k(\alpha_1) + 0) \ge \alpha_1, \\ &G(k(1 - \alpha_2)) \le 1 - \alpha_2, && G(k(1 - \alpha_2) + 0) \ge 1 - \alpha_2. \end{aligned} \tag{9}$$

А замість (5) тепер маємо

$$\mathbf{P}(z_1 < T(\vec{x}, \theta) < z_2) = G(z_2) - G(z_1 + 0). \tag{10}$$

Якщо в цій рівності взяти $z_1 = k(\alpha_1),\ z_2 = k(1-\alpha_2)$, тоді зважаючи на (9), отримаємо

$$\begin{aligned}\mathbf{P}(k(\alpha_1) < T(\vec{x},\theta)) < k(1-\alpha_2)) &= G(k(1-\alpha_2)) - G(k(\alpha_1)+0)\\ &\le 1-\alpha_2-\alpha_1 = 1-\alpha,\end{aligned}$$

що, очевидно, нас не задовольняє. Ситуацію можна поправити наступним чином. Нехай $\varepsilon > 0$ буде достатньо мале число. Згідно з визначенням квантилі (див.(**??**)) маємо

$$G(k(\alpha_1)-\varepsilon+0) \le \alpha_1, \qquad G(k(1-\alpha_2)+\varepsilon) > 1-\alpha_2.$$

Якщо тепер в (10) взяти $z_1 = k(\alpha_1)-\varepsilon,\ z_2 = k(1-\alpha_2)+\varepsilon$, то отримаємо

$$\begin{aligned}&\mathbf{P}(k(\alpha_1)-\varepsilon < T(\vec{x},\theta)) < k(1-\alpha_2)+\varepsilon) =\\ &= G(k(1-\alpha_2)+\varepsilon) - G(k(\alpha_1)-\varepsilon+0) > 1-\alpha_2-\alpha_1 = 1-\alpha.\end{aligned}$$

Отже, для довільного $\varepsilon > 0$ маємо

$$\mathbf{P}(k(\alpha_1)-\varepsilon < T(\vec{x},\theta)) < k(1-\alpha_2)+\varepsilon) > 1-\alpha.$$

Зробивши все так само, як і в теоремі 1 отримаємо:

а) якщо функція $T(\vec{z},\theta)$ є зростаючою відносно θ для кожного $\vec{z}$, тоді

$$\mathbf{P}(\theta_-(\alpha_1,\varepsilon) < \theta < \theta_+(\alpha_2,\varepsilon)) > 1-\alpha, \tag{11}$$

де $\theta_-(\alpha_1,\varepsilon) = T^{-1}(\vec{x},k(\alpha_1)-\varepsilon),\ \ \theta_+(\alpha_2,\varepsilon) = T^{-1}(\vec{x},k(1-\alpha_2)+\varepsilon)$;

б) якщо функція $T(\vec{z},\theta)$ є спадною відносно θ для кожного x, тоді

$$\mathbf{P}(\theta_-(\alpha_2,\varepsilon) < \theta < \theta_+(\alpha_1,\varepsilon)) > 1-\alpha, \tag{12}$$

де $\theta_-(\alpha_2,\varepsilon) = T^{-1}(\vec{x},k(1-\alpha_2)+\varepsilon),\ \ \theta_+(\alpha_1,\varepsilon) = T^{-1}(\vec{x},k(\alpha_1)-\varepsilon)$.

Отже, для кожного $\varepsilon > 0$ така побудова дозволяє отримати надійні інтервали з рівнем довіри $1-\alpha$ і для розривної функції $G(x)$. Якщо тепер в (11), (12) перейти до границі при $\varepsilon \to 0$, то:

а) якщо функція $T(\vec{x},\theta)$ є зростаючою відносно θ для кожного $\vec{x}$, тоді

$$\mathbf{P}(\theta_-(\alpha_1) \le \theta \le \theta_+(\alpha_2)) \ge 1-\alpha, \tag{13}$$

де $\theta_-(\alpha_1) = T^{-1}(\vec{x},k(\alpha_1)),\ \ \theta_+(\alpha_2) = T^{-1}(\vec{x},k(1-\alpha_2))$;

б) якщо функція $T(\vec{x}, \theta)$ є спадною відносно θ для кожного $\vec{x}$, тоді

$$\mathbf{P}(\theta_-(\alpha_2) \le \theta \le \theta_+(\alpha_1)) \ge 1 - \alpha, \qquad (14)$$

де $\theta_-(\alpha_2) = T^{-1}(\vec{x}, k(1-\alpha_2)), \quad \theta_+(\alpha_1) = T^{-1}(\vec{x}, k(\alpha_1))$.

Тому, в ситуації, коли функція $G(x)$ не є неперервною, надійний інтервал можемо вибирати замкнутим, і тоді можна формально застосувати теорему 1. Але, для того, щоб залишити класичне визначення надійного інтервалу як інтервалу відкритого, можемо побудувати інтервал $[\theta_-, \theta_+]$, так, щоб виконувалась умова (13) (або (14)), а потім відступити від точки θ_- трохи вліво, а від точки θ_+ трохи вправо і отриманий інтервал вже буде класичним надійним інтервалом.

9.3 Побудова надійних інтервалів за допомогою граничних теорем

Важливим припущенням в побудові, яка була описана раніше, був факт, що існує така функція $T(\vec{z}, \theta)$, що розподіл випадкової величини $T(\vec{x}, \theta)$ відомий і не залежав від θ. Це припущення в практичних дослідженнях часто трудно забезпечити, особливо, якщо розподіл генеральної популяції є довільним. Розглянемо тепер другий підхід до проблеми побудови надійних інтервалів, який дозволяє подивитись на неї з іншого боку. І у випадку, коли ми нічого не знаємо про тип розподілу генеральної сукупності цей підхід є фактично єдиним, який дає можливість побудувати надійний інтервал для не знаного параметра.

Визначення 2. *Асимптотичним надійним інтервалом* для параметру $\theta \in \Theta$ з рівнем довіри $1 - \alpha$ $(0 < \alpha < 1)$ будемо називати інтервал $(\theta_1^{(n)}, \theta_2^{(n)})$ такий, що:

а) його межі $\theta_1^{(n)} = \theta_1(x_1, \ldots x_n)$, $\theta_2^{(n)} = \theta_2(x_1, \ldots, x_n)$ є функціями випадкової вибірки і не залежать від параметра θ;

б) має місце співвідношення

$$\liminf_{n\to\infty} \mathbf{P}(\theta_1^{(n)} < \theta < \theta_2^{(n)}) \ge 1 - \alpha, \qquad \theta \in \Theta. \qquad (15)$$

Як і раніше α є малим числом.

З цього визначення випливає, що фактично ми маємо справу з послідовністю надійних інтервалів $(\theta_1^{(n)}; \theta_2^{(n)})$, $n \ge 1$ і для малих n ймовірність

$\mathbf{P}(\theta_1^{(n)} < \theta < \theta_2^{(n)})$ може бути суттєво менше від $1-\alpha$ і лише для великих n наближається до $1-\alpha$, або більша цього числа. Тому на практиці вважаємо, що якщо виконане співвідношення (15), то $(\theta_1^{(n)}; \theta_2^{(n)})$ є надійним інтервалом з рівнем довіри $1-\alpha$ для параметра θ за умови, що n велике. Як і раніше ми завше прагнемо, щоб довжина асимптотичного надійного інтервалу була як найменшою.

Розглянемо тепер один з найчастіше вживаних підходів до побудови асимптотичних надійних інтервалів.

Припустимо, що існує функція $T(\vec{z}, \theta) : \mathbf{R}^n \times \Theta \to \mathbf{R}^1$, для якої справедливе наступне співвідношення

$$\lim_{n\to\infty} \mathbf{P}(T(\vec{x}, \theta) < x) = G(x), \tag{16}$$

де функція $G(x)$ є відомою і не залежить від параметра. Розглянемо лише випадок, коли функція $G(x)$ є неперервною.

Нехай $\alpha_1, \alpha_2 > 0$ є фіксованими числами такими, що $\alpha_1 + \alpha_2 = \alpha$ і $k(\alpha_1),\ k(1-\alpha_2)$ є квантилями рівня $\alpha_1,\ 1-\alpha_2$ розподілу $G(x)$. Тоді

$$\lim_{n\to\infty} \mathbf{P}(k(\alpha_1) < T(\vec{x}, \theta) < k(1-\alpha_2)) = G(k(1-\alpha_2)) - G(k(\alpha_1)) = 1-\alpha. \tag{17}$$

Якщо функція $T(\vec{z}, \theta)$ задовольняє умовам з теореми 1, тоді нерівність під символом $\mathbf{P}()$ в (17) може бути розв'язана відносно θ і ми отримаємо наступний результат

Теорема 2. *Нехай існує функція $T(\vec{z}, \theta) : \mathbf{R}^n \times \Theta \to \mathbf{R}^1$, для якої виконані наступні умови:*

а) *для кожного фіксованого $\vec{z}$ функція $T(\vec{x}, \theta)$ є неперервною і монотонною відносно θ;*

б) $\lim_{n\to\infty} \mathbf{P}(T(\vec{x}, \theta) < x) = G(x)$, *де функція $G(x)$ є відомою, незалежною від параметра і неперервною.*

Тоді:

а) *якщо функція $T(\vec{z}, \theta)$ є зростаючою відносно θ для кожного $\vec{z}$, то*

$$\lim_{n\to\infty} \mathbf{P}(\theta_-(\alpha_1) < \theta < \theta_+(\alpha_2)) = 1-\alpha,$$

де $\theta_-(\alpha_1) = T^{-1}(\vec{x}, k(\alpha_1)),\quad \theta_+(\alpha_2) = T^{-1}(\vec{x}, k(1-\alpha_2));$

б) *якщо функція* $T(\vec{z},\theta)$ *є спадною відносно* θ *для кожного* $\vec{z}$, *то*

$$\lim_{n\to\infty}\mathbf{P}(\theta_-(\alpha_2)<\theta<\theta_+(\alpha_1))=1-\alpha,$$

де $\theta_-(\alpha_2)=T^{-1}(\vec{x},k(1-\alpha_2)),\quad \theta_+(\alpha_1)=T^{-1}(\vec{x},k(\alpha_1))$.

На практиці в якості $G(x)$ найчастіше виступає стандартний нормальний розподіл. Якщо тепер $k(1-\alpha/2)$ є квантилем рівня $1-\alpha/2$ розподілу $\Phi(x)$, то з попередньої теореми маємо для параметра θ асимптотичний надійний інтервал з рівнем довіри $1-\alpha$ вигляду

$$\theta_-(\alpha)<\theta<\theta_+(\alpha_2),\qquad \theta_\pm(\alpha)=T^{-1}(\vec{x},\pm k(1-\alpha/2),\tag{18}$$

якщо функція $T(\vec{z},\theta)$ є зростаючою відносно θ і вигляду

$$\theta_-(\alpha)<\theta<\theta_+(\alpha),\qquad \theta_\pm(\alpha)=T^{-1}(\vec{x},\mp k(1-\alpha/2),\tag{19}$$

якщо функція $T(\vec{z},\theta)$ є спадною відносно θ.

Розглянемо тепер приклад побудови асимптотичних надійних інтервалів, який часто зустрічається в застосуваннях

Приклад 1. *Нехай для параметра* θ *існує оцінка* $\theta(\vec{x})$, *і послідовність чисел* $d_n(\vec{x})$ (*які можуть залежати від вибірки* $\vec{x}$, *але не від параметра*) *такі, що*

$$\mathbf{P}(\sqrt{n}\frac{\theta(\vec{x})-\theta}{d_n(\vec{x})}<x)\underset{n\to\infty}{\longrightarrow}\Phi(x)\in\mathbb{N}(0,1).$$

Побудувати асимптотичний надійний інтервал для θ.

Р о з в' я з о к. Якщо $k(1-\alpha/2)$ є квантилем рівня $1-\alpha/2$ для функції розподілу $\Phi(x)$, тобто $\Phi(k(1-\alpha/2))=1-\alpha/2$, тоді

$$\lim_{n\to\infty}\mathbf{P}(\ \frac{\sqrt{n}(\theta(\mathbf{x})-\theta)}{d_n(\vec{x})}\ <k(1-\alpha/2))=2\Phi(k(1-\alpha/2))-1=1-\alpha,$$

або, що є рівнозначним,

$$\lim_{n\to\infty}\mathbf{P}(\theta(\mathbf{x})-\frac{d_n(\vec{x})k(1-\alpha/2)}{\sqrt{n}}<\theta<\theta(\mathbf{x})+\frac{d_n(\vec{x})k(1-\alpha/2}{\sqrt{n}})=1-\alpha$$

З цього співвідношення випливає, що інтервал

$$\theta(\mathbf{x})-\frac{d_n(x)k(1-\alpha/2)}{\sqrt{n}};\theta(\mathbf{x})+\frac{d_n(x)k(1-\alpha/2)}{\sqrt{n}}$$

є асимптотичним надійним інтервалом з рівнем довіри $1-\alpha$ для параметра θ.

Якщо для вибірки $\vec{x}$ розміру n побудуємо надійний інтервал вигляду, скажімо (18), то його рівень довіри буде, взагалі кажучи, іншим ніж $1-\alpha$, але різниця для великих n буде малою. Але в цьому місці нас підстерігає небезпека, яку ми сформулюємо у вигляді зауваження і яке показує, що користуватися асиптотичним підходом при побудові надійних інтервалів потрібно дуже обережно.

Зауваження 2. Асимптотичні надійні інтервали мають два істотних недоліки: при фіксованому значені параметра θ швидкість збіжності в (17) є, як правило, порядку $O\bigl(n^{-1/2}\bigr)$, але задача оцінки помилки, яку ми робимо замінюючи реальний розподіл на асимптотичний, є досить складною і має задовільне вирішення лише для деяких класів розподілів. Але ще більшою проблемною є та обставина, що ця швидкість часто суттєво залежить від значення самого параметра. Об'єднання цих двох причин часто приводить до того, що формально побудований асимптотичний інтервал має таку саму цінність як гадання на кавовій гущі.

9.4 Приклади побудов надійних інтервалів

Перейдемо тепер до побудови конкретних надійних інтервалів.

На практиці найчастіше зустрічається ситуація, коли генеральна популяція має нормальний розподіл. З іншого боку, власне та частина теорії інтервального оцінювання, яка пов'язана з нормальним розподілом, є найбільш розвинутою. Ось чому ми зробимо побудову надійних інтервалів саме для нормальних розподілів і для характеристик, які найчастіше нас цікавлять: середнє значення і дисперсія. Побудова відповідних надійних інтервалів грунтується на теоремі 1, тобто спочатку треба знайти функцію $T(x, \theta)$, для якої виконані умови цієї теореми, а потім записати відповідний надійний інтервал. Зосередимося лише на двосторонніх надійних інтервалах.

9.4.1 Надійні інтервали для математичного сподівання

Приклад 1. *Генеральна популяція має розподіл $\mathbb{N}(m, \sigma^2)$ з відомим σ. Знайти надійний інтервал для невідомого значення математичного сподівання m на підставі n-елементної вибірки $x_1 \dots, x_n$.*

Р о з в' я з о к. В якості функції $T(\vec{z}, m)$ вибираємо $\sqrt{n}(z(n) - m)/\sigma$, де $z(n) = n^{-1}\sum_{i=1}^{n} z_i$. Доведемо тепер, що для

$$T(x(n), m) = \frac{\sqrt{n}(x(n) - m)}{\sigma},$$

виконані умови теореми 1. Той факт, що функція $T(\vec{z}, m)$ є спадна відносно m для кожного $\vec{z}$ є очевидним. Нехай $\eta_i = (x_i - m)/\sigma$. Оскільки $x_i \in \mathbb{N}(m, \sigma^2)$, то $\eta_i \in \mathbb{N}(0, 1)$, а тепер, на підставі теореми 1 маємо

$$T(x(n), m) = \frac{\sqrt{n}(x(n) - m)}{\sigma} = \sqrt{n}\frac{\sum_{i=1}^{n}(x_i - m)}{n\sigma} = \sum_{i=1}^{n}(\sqrt{n})^{-1}\eta_i \in \mathbb{N}(0, 1).$$

Тому $\mathbf{P}(T(x(n), m) < x) = \Phi(x)$ і, отже, статистика $T(x(n), m)$ має розподіл, який не залежить від параметра m. Можемо відразу скористатися пунктом б) теореми 1, але зробимо всі кроки процедури побудови по черзі. Насамперед зауважимо, що, позаяк, розподіл $\Phi(x)$ є симетричним, тобто $\Phi(x) = 1 - \Phi(-x)$, то найкоротший надійний інтервал буде для $\alpha_1 = \alpha_2 = \alpha/2$. Для нашої конкретної ситуації цей факт легко довести (задача 1). Нехай $k(1-\alpha/2)$ буде квантилем рівня $1-\alpha/2$ для $\Phi(x)$, тобто $\Phi(k(1-\alpha/2)) = 1-\alpha/2$. Якщо взяти до уваги рівність $\Phi(x) = 1 - \Phi(-x)$, то

$$\mathbf{P}\left(\frac{\sqrt{n}(x(n)-m)}{\sigma} < k(1-\alpha/2)\right) = 2\Phi(k(1-\alpha/2)) - 1 = 1-\alpha. \tag{20}$$

Нерівність під символом $\mathbf{P}()$ легко розв'язується відносно m, що дає

$$x(n) - \frac{k(1-\alpha/2)\sigma}{\sqrt{n}} < m < x(n) + \frac{k(1-\alpha/2)\sigma}{\sqrt{n}},$$

і тепер з (20) отримаємо

$$\mathbf{P}\left(x(n) - \frac{k(1-\alpha/2)\sigma}{\sqrt{n}} < m < x(n) + \frac{k(1-\alpha/2)\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

Отже, надійний інтервал для m з рівнем довіри $1-\alpha$ виглядає так

$$x(n) - \frac{k(1-\alpha/2)\sigma}{\sqrt{n}} < m < x(n) + \frac{k(1-\alpha/2)\sigma}{\sqrt{n}}. \tag{21}$$

Приклад 2. *Генеральна популяція має розподіл $\mathbb{N}(m,\sigma^2)$, де σ є невідомим. Знайти надійний інтервал для невідомого математичного сподівання m на підставі n-елементної вибірки $x_1 \ldots, x_n$.*

Р о з в' я з о к. Тепер в якості функції $T(\vec{z}, m)$ вибираємо

$$T(\vec{z}, m) = \sqrt{n-1}\,\frac{(z(n)-m)}{Q(n)},$$

де

$$z(n) = n^{-1}\sum_{i=1}^{n} z_i, \qquad Q^2(n) = n^{-1}\sum_{i=1}^{n}(z_i - z(n))^2.$$

Якщо $\vec{z} = \vec{x}$, тоді $z(n) = x(n)$ і $Q^2(n) = S^2(n)$-дисперсія з вибірки. То що функція $T(\vec{z}, m)$ спадною стосовно m для кожного фіксованого $\vec{z}$ є очевидним, а на підставі теореми Фішера (теорема 4, с.176) маємо

$$\mathbf{P}\left(\frac{\sqrt{n-1}(x(n)-m)}{S(n)} < x\right) = S_{n-1}(x),$$

де $S_{n-1}(x)$ є розподіл Стьюдента з $n-1$ степенями свободи і отже $T(\vec{x}, m)$ має розподіл, який не залежить від параметра m. Подальші кроки є такими ж самими, як і в попередньому прикладі. Тому наведемо одразу відповідь.

Надійний інтервал з рівнем довіри $1-\alpha$ для m має наступний вигляд

$$x(n) - \frac{k(1-\alpha/2, n-1)S(n)}{\sqrt{n-1}} < m < x(n) + \frac{k(1-\alpha/2, n-1)S(n)}{\sqrt{n-1}}, \tag{22}$$

де $S_{n-1}(k(1-\alpha/2, n-1)) = 1-\alpha/2$.

Приклад 3.*Генеральна популяція ξ має довільний розподіл. Знайти надійний інтервал для невідомого математичного сподівання m на підставі n-елементної вибірки $x_1 ..., x_n$.*

Р о з в' я з о к. В цій ситуації, без додаткових припущень стосовно генеральної популяції, неможливо знайти статистику, яка б мала відомий розподіл. Тому, в цьому випадку скористаємося асимптотичним підходом, тобто будемо шукати асимптотичний надійний інтервал. В якості статистики для конструкції цього інтервалу використаємо

$$T(\mathbf{x}, m) = \frac{\sqrt{n}(x(n) - m)}{S(n)},$$

де $x(n)$, $S^2(n)$ є відповідно середнє і дисперсія вибірки.

Нехай $\sigma^2 = \mathbf{D^2}\xi^2$. З центральної граничної теореми маємо

$$\frac{\sqrt{n}(x(n) - m)}{S(n)} = \frac{\sum_{i=1}^n x_i - nm}{\sqrt{n}\sigma} \frac{\sigma}{S(n)} \underset{n\to\infty}{\Longrightarrow} \xi \in \mathbb{N}(0, 1).$$

Отже

$$\lim_{n\to\infty} \mathbf{P}(\frac{\sqrt{n}(x(n) - m)}{S(n)} < x) = \Phi(x).$$

Тепер згідно з алгоритмом, який був описаний вище, будемо мати наступний асимптотичний надійний інтервал

$$x(n) - \frac{k(1 - \alpha/2)S(n)}{\sqrt{n}} < m < x(n) + \frac{k(1 - \alpha/2)S(n)}{\sqrt{n}} \qquad (23)$$

з рівнем довіри 1-α, де $k(1 - \alpha/2)$ є таким, що $\Phi(k(1 - \alpha/2)) = 1 - \alpha/2$.

9.4.2 Надійні інтервали для дисперсії

Побудова надійних інтервалів для дисперсії більш-менш подібна до конструкцій аналогічних інтервалів для середнього значення. Тому для дисперсії всі міркування зробимо в скороченому варіанті.

Приклад 4. *Генеральна популяція має розподіл $\mathbb{N}(m, \sigma^2)$, де m є відомим. Знайти надійний інтервал для невідомої дисперсії σ^2 на підставі n-елементної вибірки $x_1, \ldots, x_n$.*

Р о з в' я з о к. В якості статистики для конструкції надійного інтервалу в цій ситуації використовуємо

$$T(\mathbf{x}, \sigma) = \frac{\sum_{i=1}^n (x_i - m)^2}{\sigma^2}.$$

Позаяк, $\eta_i = (x_i - m)/\sigma \in \mathbb{N}(0, 1)$, то статистика $T(\vec{x}, m) = \sum_{i=1}^n \eta_i^2$ має розподіл χ_n^2, тобто $\mathbf{P}(T(\vec{x}, \sigma) < x) = \chi_n^2(x)$, і отже $T(\mathbf{x}, \sigma)$ має розподіл, який не залежить від параметра σ^2. Нехай тепер $k(\alpha/2, n)$, $k(1 - \alpha/2, n)$ є квантилями рівня $\alpha/2$ і $1 - \alpha/2$ для $\chi_n^2(x)$. Тоді

$$\mathbf{P}(k(\frac{\alpha}{2}, n) < \frac{\sum_{i=1}^n (x_i - m)^2}{\sigma^2} < k(1 - \frac{\alpha}{2}, n)) = \chi_n(k(1 - \frac{\alpha}{2}, n)) - \chi_n^2(k(1 - \frac{\alpha}{2}, n)) = 1 - \alpha.$$

Розв'язуючи нерівність під символом $\mathbf{P}()$ відносно σ^2, отримуємо потрібний надійний інтервал для σ^2 з рівнем довіри $1-\alpha$,

$$\frac{\sum_{i=1}^{n}(x_i-m)^2}{k(1-\alpha/2,n)} < \sigma^2 < \frac{\sum_{i=1}^{n}(x_1-m)^2}{k(\alpha/2,n)}, \tag{24}$$

де $k(1-\alpha/2,n)$, $k(\alpha/2,n)$ є квантилями рівня $1-\alpha/2$ та $\alpha/2$ розподілу хі -квадрат з n степенями свободи.

Зауваження 3. Отриманий інтервал не буде найкоротшим, але оскільки в цій ситуації побудова надійного інтервалу найменшої довжини (тобто розв'язання задачі (8)) є досить складним, то на практиці зазвичай використовують інтервал (24).

Приклад 5. *Генеральна популяція має розподіл $\mathbb{N}(m,\sigma^2)$, де m є невідомим. Знайти надійний інтервал для невідомої дисперсії σ^2 на підставі n-елементної вибірки $x_1,\ldots,x_n$.*

Р о з в' я з о к. В якості статистики для побудови надійного інтервалу в цій ситуації використовуємо

$$T(\vec{x},\sigma) = \frac{nS^2(n)}{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i-x(n))^2}{\sigma^2}.$$

На підставі теореми Фішера статистика $T(\vec{x},\sigma)$ має розподіл $\chi^2_{n-1}(x)$. Так само як і в попередньому прикладі, отримаємо надійний інтервал

$$\frac{\sum_{i=1}^{n}(x_i-x(n))^2}{k(1-\alpha/2,n-1)} < \sigma^2 < \frac{\sum_{i=1}^{n}(x_1-x(n))^2}{k(\alpha/2,n-1)}, \tag{25}$$

де $k(1-\alpha/2,n-1)$, $k(\alpha/2,n-1)$ є такими, що $\chi^2_{n-1}(k(\alpha/2,n-1)) = \alpha/2$, і $\chi^2_{n-1}(k(1-\alpha/2,n-1)) = 1-\alpha/2$.

Приклад 6. *Генеральна популяція ξ має довільний розподіл. Знайти надійний інтервал для невідомої дисперсії на підставі n-елементної вибірки $x_1,\ldots,x_n$.*

Р о з в' я з о к. В цьому випадку також неможливо знайти статистику, яка б мала відомий розподіл. Тому знову застосуємо асимптотичний підхід.

За статистику для побудови цього надійного інтервалу візьмемо

$$T(\vec{x},\sigma^2) = \sqrt{\frac{2nS^2(n)}{\sigma^2}} - \sqrt{2n-3} = \sqrt{\frac{2}{\sigma^2}\sum_{i=1}^{n}(x_i-x(n))^2} - \sqrt{2n-3}.$$

Можна довести, що ця статистика для великих n має наближено розподіл $\mathbb{N}(0,1)$ звідки вже стандартно отримаємо асимптотичний надійний інтервал для σ

$$\frac{\sqrt{2n}S(n)}{\sqrt{2n-3}+k(1-\alpha/2)} < \sigma < \frac{\sqrt{2n}S(n)}{\sqrt{2n-3}-k(1-\alpha/2)}, \tag{26}$$

де $k(1-\alpha/2)$ є квантилем рівня $1-\alpha/2$ розподілу $\mathbb{N}(0;1)$.

9.4.3 Надійний інтервал для показника структури популяції

Розглянемо схему Бернуллі з параметрами n, p. Це означає, що ми маємо послідовність з n незалежних однакових випробувань з ймовірністю успіху p і неуспіху $1-p$. В цій схемі ймовірність успіху в одному випробуванні називають *показником структури популяції*. Нехай в n випробуваннях було k успіхів. Потрібно знайти надійний інтервал для p.

На підставі Теореми **??** знаємо, що статистика

$$W(\mathbf{x}, p) = \frac{k - np}{\sqrt{np(1-p)}}$$

слабко збігається до випадкової величини, яка має стандартний нормальний розподіл $\mathbb{N}(0,1)$ і, значить,

$$\lim_{n\to\infty} \mathbf{P}(|\frac{k - np}{\sqrt{np(1-p)}}| < k(1-\alpha/2) = 1-\alpha/2,$$

де $k(1-\alpha/2)$ є таким, що $\Phi(k(1-\alpha/2)) = 1-\alpha/2$.

Якщо розв'язати нерівність

$$|\frac{k - np}{\sqrt{np(1-p)}}| < k(1-\alpha/2)$$

стосовно p, тоді отримаємо наступний асимптотичний надійний інтервал:

$$\frac{nA_-}{n + k^2(1-\alpha/2)} < p < \frac{nA_+}{n + k^2(1-\alpha/2)},$$

де

$$A_\pm = \frac{k}{n} + \frac{k^2(1-\alpha/2)}{2n} \pm \frac{k(1-\alpha/2)}{n}\sqrt{\frac{k(n-k)}{n} + \frac{k^2(1-\alpha/2)}{4}}.$$

Як показано в цей надійний інтервал не є добрим особливо якщо є блиським до і різні способи його покращення практично нічого не дають. Але в цій ситуації ми знаємо тип розподілу генеральної популяції. Це дає нам можливість використати підхід з попереднього підпідрозділу. Відповідна конструкція була зроблена в де вуло використано ту обставину що для статистики ми можемо записати

$$\mathbf{P}_p(\sum_{i=1}^{n} \le k) = \sum_{i=1}^{k} C_n^k p^i (1-p)^{1-i} = B(n-k, k+1, 1-p)$$

де розподіл з парвметрами . Але при фіксованих є неперрвною функцією і тому для довільного можемо визначити лвантиль рівня цього розподілу

$$(k(nx(n), n+1-n(x(n); \gamma/2); k(nx(n)+1, n-n(x(n); 1-\gamma/2) \tag{27}$$

9.4.4 Визначення мінімального розміру вибірки

Як вже була мова раніше, надійний інтервал для даного параметра вказує ту область, в якій з певною ймовірністю знаходиться справжнє значення параметра. Часом прагнемо, щоб цей інтервал був визначеної довжини. Це завдання не є легким для розв'язання, позаяк, може статися так, що довжина надійного інтервалу залежить від вибірки і, значить, не можемо сказати скільки спостережень треба розглянути, поки не зробимо ці спостереження. Ми розглянемо лише ситуацію, коли такої проблеми немає.

Приклад 7. *Випадкова величина має розподіл $\mathbb{N}(m, \sigma^2)$, де σ є відомою. Яким повинний бути мінімальний розмір вибірки, щоб при заданому рівні довіри $1-\alpha$ отримати надійний інтервал для m довжини не більшої ніж $2l$?*

Р о з в' я з о к. Надійний інтервал для m в цій ситуації наведений в (21) і довжина цього інтервалу є $2k(1-\alpha/2)\sigma/\sqrt{n}$. Тепер з нерівності $2k(1-\alpha/2)\sigma/\sqrt{n} \leq 2l$ отримаємо

$$n \geq \left(\frac{k(1-\alpha/2)\sigma}{l}\right)^2.$$

Приклад 8. *Поміряно діаметри партії* 15 *свердел і отримано результати:* $5,79; 5,67; 5,89; 6,11; 5,75; 5,76; 5,75; 5,74; 5,75; 5,89; 5,72; 5,72; 5,75; 5,6; 5,7.$

а) Вважаючи, що діаметр свердла має розподіл $\mathbb{N}(m; \sigma^2)$ знайти 95% надійний інтервал для середньої довжини сердла. Спочатку якщо дисперсія відома $\sigma^2 = 0,01$, а потім якщо вона невідома.

б) Вважаючи, що діаметр свердла має розподіл $\mathbb{N}(m; \sigma^2)$ знайти 95% надійний інтервал для дисперсії довжини свердла. Спочатку якщо середній діаметр відомий $m = 5,8$, а потім коли він невідомий.

в) Діаметр свердла має невідомий розподіл. Знайти 95% надійний інтервал для середньої довжини свердла та дисперсії довжини свердла.

Р о з в' я з о к. а) Маємо $x(15) = \sum_{i=1}^{15} x_i/15 = 5,77$. Оскільки $1-\alpha = 0,95$, то $\alpha = 0,05$, а отже $1-\alpha/2 = 0,975$ і тому для конструкції надійного інтервалу нам потрібно, згідно (21), знайти квантиль рівня 0,975 розподілу $\mathbb{N}(0,1)$. З таблиці 2 знаходимо $k(0,975) = 1,96$. Оскільки $\sigma^2 = 0,01$, то $\sigma = 0,1$ а тому з (21) отримуємо потрібний надійний інтервал

$$5,77 - \frac{1,96 \cdot 0,1}{\sqrt{15}} < m < 5,77 - \frac{1,96 \cdot 0,1}{\sqrt{15}} \quad \text{або} \quad 5,72 < m < 5,82.$$

Якщо дисперсія невідома, то, перш за все, порахуємо дисперсію для нашої вибірки $S^2(15) = \sum_{i=1}^{15}(x_i - x(n))^2/15 = 0,013$ і скористаємося з попередніх обчислень. Маємо $S(15) = 0,114$ і згідно (22) нам потрібно знайти квантиль рівня 0,975 розподілу Стьюдента з 14 степенями свободи. З таблиці 3 $k(0,975,14) = 2,145$. Тепер з (22) отримуємо потрібний надійний інтервал.

$$5,77 - \frac{2,145 \cdot 0,114}{\sqrt{14}} < m < 5,77 - \frac{2,145 \cdot 0,114}{\sqrt{14}} \quad \text{або} \quad 5,7 < m < 5,84.$$

б) Скористаємося надійним інтервалом з (24), а для цього підрахуємо $\sum_{i=1}^{15}(x_i - 5,8)^2 = 0,207$. Для квантилів $k(0,975;15)$, $k(0,025;15)$ розподілу хі -квадрат з 15 степенями свободи з таблиці 4 маємо $k(0,975;15) = 27,488$, $k(0,025;15) = 6,262$. Отже тепер з (24) отримуємо потрібний інтервал

$$\frac{0,207}{27,488} < \sigma^2 < \frac{0,207}{6,262} \quad \text{або} \quad 0,008 < \sigma^2 < 0,033.$$

Якщо середній діаметр невідомий, то підрахуємо $\sum_{i=1}^{15}(x_i - x(n))^2 = 0,196$ і скористаємося надійним інтервалом з (25). Для квантилів $k(0,975;14)$, $k(0,025;14)$ розподілу хі -квадрат з 14 степенями свободи з таблиці 4 маємо $k(0,975;14) = 26,119$, $k(0,025;14) = 5,629$. Отже тепер з (25) отримуємо потрібний інтервал

$$\frac{0,196}{26,119} < \sigma^2 < \frac{0,196}{5,629} \quad \text{або} \quad 0,007 < \sigma^2 < 0,035.$$

в) В цьому випадку будуємо асимптотичні надійні інтервали з (23) та (26), використовуючи при цьому обчислення, які ми робили вище. Так для сердньої величини діаметра з (23) маємо

$$5,77 - \frac{0,114 \cdot 1,96}{\sqrt{15}} < m < 5,77 - \frac{0,114 \cdot 1,96}{\sqrt{15}} \quad \text{або} \quad 5,71 < m < 5,83.$$

Агалогічно, для дисперсії величини діаметра з (26) дістаємо

$$\frac{\sqrt{30} \cdot 0,114}{\sqrt{27} + 1,96} < \sigma < \frac{\sqrt{30} \cdot 0,114}{\sqrt{27} - 1,96} \quad \text{або} \quad 0,008 < \sigma^2 < 0,151.$$

ЗАВДАННЯ

1. Довести, що надійний інтервал побудований в (21) є найкоротшим.
2. Генеральна популяція ξ має розподіл гамма з параметрами p, λ, тобто $\xi \in \mathbb{G}(p, \lambda)$, де p є знане. Нехай $k(\alpha/2)$, $k(1-\alpha/2)$ будуть квантилями рівня $\alpha/2$, $1-\alpha/2$ для розподілу $\mathbb{G}(np, 1)$. Довести, що для параметра $Q = \lambda^{-1}$ інтервал $\frac{nx(n)}{k(1-\alpha/2)}; \frac{nx(n)}{k(\alpha/2)}$ є надійним інтервалом з рівнем довіри $1-\alpha$.

   Вказівка. *Використати статистику* $x(n) = \lambda \sum_{k=1}^{n} x_k$ *та скористатися вправою* 8 *с.* 177.
3. Маємо вибірку для генеральної популяції ξ: 2,5; 3,1; 4,2; 2,6; 3,3; 5,2; 4,3; 2,5; 3,6; 4,0; 3,7; 3,9; 4,4; 4,5; 3,8.

   а) Побудувати надійний інтервал для m, якщо $\xi \in \mathbb{N}(m; 0,6)$.

   б) Побудувати надійний інтервал для σ^2, якщо $\xi \in \mathbb{N}(3,7; \sigma^2)$

   в) Побудувати надійний інтервал для m, σ^2 якщо тип розподілу ξ є невідомий.

   У всіх випадках рівень довіри взяти $0,95$.
4. Досліджено на міцність деякий будівельний матеріал і отримано результати (в кг/см^2) : 120; 102; 135; 115. Припускаючи, що міцність даного матеріалу має розподіл нормальний, побудувати надійний інтервал для середнього значення та дисперсії міцності взявши за рівень довіри 0,95.

5. Скільки штук певного виробу потрібно взяти щоб отримати значення його ваги з точністю до 0,25 г і рівнем надійності 0,99 якщо відомо, що дисперсія ваги виробу є $0,8$.

6. З партії товару вибрано навмання 120 штук і серед них було виявлено 17 штук з браком. Подати 95 % - надійний інтервал для частоти бракованих товарів в цій партії.

7. З партії бавовни вибрано навмання 70 волокен і заміряно їх довжини (в мм), а результати згруповано в розділчий ряд

N класу	1	2	3	4	5	6	7	8
середина класу	25	40	55	70	85	100	115	130
кількість	7	14	13	15	12	6	2	1

Побудувати 95 % надійний інтервал для дисперсії довжини волокен.

8. З партії бавовни вибрано навмання 60 волокен і полічено їх середню довжину $\bar{x} = 22,8$ мм, та середньоквадратичне відхилення $s = 6,3$ мм. Скільки штук волокен ще потрібно взяти щоб отримати 95 % надійний інтервал, з довжиною не більшою ніж 2 мм, для середньої довжини волокен?

Лекція 10. Методи знаходження оцінок

В цьому розділі хочемо дати відповідь на питання:

як на підставі вибірки $\vec{x} = (x_1, ..., x_n)$, взятої з генеральної сукупності ξ знайти оцінку для невідомого параметра θ?

Як вже зазначалося раніше, кожна статистика S, може виступати в якості оцінки невідомого параметра θ. Тому задача пошуку оцінки для θ виглядає наступним чином. Нехай генеральна сукупність ξ має розподіл $F(x, \theta)$, де $\theta \in \Theta$ є невідомим параметром (скалярним чи векторним). Потрібно знайти статистику S таку, що $S(\vec{x}) \in \Theta$ для кожної конкретної вибірки з генеральної сукупності ξ і значення цієї статистики будуть служити значенням невідомого параметра θ. Щоб підкреслити той факт, що статистика використовується як оцінка для θ будемо записувати $\theta(\vec{x})$, або (якщо залежність від вибірки не буде суттєвою) просто θ_n. Вочевидь, що нас цікавить не кожна статистика. Було б добре запропонувати метод, який дозволяв би знаходити статистики, які в свою чергу були б гарними оцінками. Наприклад, консистентним, ефективними і т.д. З деякими проблемами порівняння оцінок ми вже зустрічались у попередньому розділі. Більше на цю тему будемо говорити в наступному, а зараз зосередимося на процедурах, які дають можливість знайти оцінку для невідомого параметра. Можна виділити (умовно) три головні методи пошуку оцінок.

I. Метод підстановки (або емпіричний метод).

II. Метод максимальної вірогідності.

III. Метод моментів.

Ці методи використовують найчастіше але, як побачимо далі, вони не є альтернативними, оскільки, по суті, метод максимальної вірогідності і метод моментів можна теж розглядати як реалізації метода підстановки, але для нас така класифікація буде зручною, і тому ми будемо нею користуватись. Існують деякі інші методи пошуку оцінок (наприклад, баєсівський, мінімаксний), які не так часто використовуються і в цьому посібнику розглядатись не будуть. Опис цих методів можна знайти в [1]. Ми також проведемо дослідження деяких загальних властивостей оцінок, отриманих за допомогою методів I-III.

Історично метод моментів був першим методом пошуку оцінок невідомих параметрів і був запропонований К. Пірсоном[1]. Зараз цей метод не є занадто часто вживаним, позаяк, оцінки отримані за допомогою цього метода як правило мають меншу ефективність, ніж оцінки метода підстановки. Є також інші причини, про які будемо говорити дещо згодом.

Метод максимальної вірогідності є одним з найважливіших методів пошуку оцінок, і був запропонований П. Фішером в 1912 р. Як правило, цей метод вимагає деякої

[1] Карл Пірсон (*Karl Pearson*) (27.03.1857–27.04.1936)–видатний британський статистик. Працював також в біології і медицині, один з засновників журналу "Біометрика"(існує по сьогодні).

інформації про вигляд розподілу генеральної сукупності з точністю до невідомого параметра (або параметрів). Перший метод, тобто метод підстановки, взагалі кажучи, не вимагає такої інформації і є достатньо простим для застосувань. Тому метод підстановки має більш широке використання в теорії оцінювання. З опису цього методу й почнемо.

10.1 Метод підстановки

Нехай генеральна сукупність ξ має розподіл $F(x)$ (невідомий!) і нас цікавить деякий параметр θ, який залежить від цього розподілу. Це будемо записувати так

$$\theta = G(F). \tag{1}$$

В цьому випадку кажуть, що параметр θ є функціоналом розподілу F. Приклади таких параметрів легко знайти. Так моменти всіх порядків є такими параметрами. Також до цієї групи параметрів належать порядкові статистики. Параметри, які можемо отримати з параметрів вигляду (1) за допомогою скінченої кількості арифметичних дій теж будемо зараховувати до цієї групи. Емпіричний метод базується на тому, що якщо F_n є емпіричною функцією розподілу побудованою для вибірки $\vec{x}$ з генеральної сукупності ξ з розподілом $F(x)$, то в якості оцінки для параметра θ з (1) слід брати

$$\theta(\vec{x}) = G(F_n). \tag{2}$$

Часом може статися, що така підстановка є неможливою. Наприклад, якщо в виразі для $G(F)$ бере участь похідна F'. В таких ситуаціях завше є можливість модифікувати метод підстановки таким чином, щоб головна ідея цього методу залишилась чинною. Так замість щільності функції розподілу F можна підставити ломану частоти, або якусь іншу оцінку щільності.

Зрозуміло, що дослідити властивості оцінок в такому загальному вигляді, як в(2), без додаткових припущень щодо G є неможливим. Тому наслідуючи А.А. Боровкова[1] [1], визначимо дві групи параметрів, для яких найчастіше застосовують метод підстановки.

$$\mathbf{A}=\ \theta : \theta{=}h\ \ \mathbf{M}g(\xi)\ \ , \text{ де } h(\cdot), g(\cdot) \text{ є деякими функціями і } \mathbf{M}|g(\xi)| < \infty.$$

До цієї групи будемо також зараховувати параметри, які можуть бути записані в вигляді скінченої кількості арифметичних дій над параметрами групи $\mathbf{A}$.

Приклад 1. *Параметри* $\theta_1 = \mathbf{M}\xi$, $\theta_2 = \mathbf{M}\xi^2$, $\theta_3 = \mathbf{D^2}\xi = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2$ *є параметрами типу* $\mathbf{A}$. *Тут* $h(x) = x$ *і* $g(x) = x$ *для першого функціоналу і* $h(x) = x,\ g(x) = x^2$ *для другого, а третій є, вочевидь, різницею двох параметрів типу* $\mathbf{A}$.

Визначення 1. Якщо $\theta \in \mathbf{A}$, то величину

$$\theta(n) \ \stackrel{def}{=}\ h\ \ \frac{1}{n}\sum_{i=1}^{n} g(x_i) \tag{3}$$

вибираємо за оцінку параметра θ.

[1]Олександр Олексійович Боровков (нар. 6.03. 1931)– російський математик, учень А.М.Колмогорова, видатний фахівець в галузі теорії ймовірностей і математичної статистики.

Зауваження 1. На практиці, щоб знайти оцінку для параметра типу **B**, з формальної точки зору слід підрахувати відповідний функціонал $G(\cdot)$ для вибірки. Наприклад, для оцінювання медіани потрібно підрахувати медіану з вибірки.

Наступна теорема стверджує, що при досить загальних умовах оцінки емпіричного методу є сильно конзистентними.

Теорема 1. *Нехай функція $h(x)$ є неперервною в точці $a = \mathbf{M}g(\xi)$. Тоді*

$$\theta_n = h\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right) \underset{n\to\infty}{\overset{\text{м.н.}}{\longrightarrow}} h(\mathbf{M}g(\xi)) = \theta.$$

Якщо функція $h(x)$ є неперервною і обмеженою, тоді оцінка θ_n буде асимптотично незміщеною.

Д о в е д е н н я. Нехай $G_n = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$. Вочевидь, це є сума незалежних однаково розподілених випадкових величин $g(x_i)$ і на підставі посиленого закону великих чисел маємо

$$G_n \underset{n\to\infty}{\overset{\text{м.н.}}{\longrightarrow}} \mathbf{M}g(x_1) = \mathbf{M}g(\xi) = a,$$

але тоді (оскільки функція $h(x)$ є неперервною в точці a) $h(G_n) \underset{n\to\infty}{\overset{\text{м.н.}}{\longrightarrow}} h(\mathbf{M}\xi) = \theta$. Що доводить теорему.

Доведена теорема має велике значення в математичній статистиці, оскільки з неї випливає, що для значного класу статистик їх значення апроксимують “теоретичні значення” відповідних характеристик зі збільшенням об’єму вибірки. Зокрема, всі моменти і характеристики, які підраховуються для вибірки в параграфі 2.1.1 прямують до відповідних теоретичних характеристик зі збільшенням n.

10.2 Асимптотичні властивості оцінок методу підстановки

Для параметра виду $\theta = h(\mathbf{M}g(\xi))$ емпіричний метод дає оцінку

$$\theta_n = h\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right). \tag{4}$$

Деякі властивості цієї оцінки були вивчені раніше. Наступні теореми є більш глибоким дослідженням властивостей таких оцінок.

Теорема 2. *Нехай функція $h(x)$ є двічі неперервно диференційовною в околиці точки $a = \mathbf{M}g(\xi)$ і $D^2(g) = \mathbf{D^2}g(\xi) < \infty$. Тоді:*

а) $\frac{\sqrt{n}(\theta_n-\theta)}{D(g)} \underset{n\to\infty}{\Longrightarrow} h'(a)\xi$, *де ξ має стандартний нормальний розподіл;*

б) *якщо функція $h''(x)$ є обмеженою, тоді оцінка θ_n буде асимптотично незміщеною.*

Д о в е д е н н я. З припущень теореми випливає, що для достатньо малих $|x-a|$ виконується рівність

$$h(x) = h(a) + h'(a)(x-a) + \frac{1}{2}h''(a^*)(x-a)^2,$$

де $a^* \in [a; x]$. Замінимо тепер у цій рівності x на $n^{-1}\sum_{i=1}^n g(x_i)$. Можливість такої заміни випливає з факту, що на підставі посиленого закону великих чисел маємо

$$\frac{1}{n}\sum_{i=1}^n g(x_i) \xrightarrow[n\to\infty]{\text{м.н.}} \mathbf{M}g(x_1) = a \tag{5}$$

і, отже, для майже кожного ω, окрім, можливо, $\omega \in A$ з $\mathbf{P}(A) = 0$, існує $n(\omega)$ таке, що різниця $n^{-1}\sum_{i=1}^n g(x_i) - a$ буде малою для всіх $n \geq n(\omega)$. Можемо записати

$$h(n^{-1}\sum_{i=1}^n g(x_i)) = h(a) + h'(a)\left(n^{-1}\sum_{i=1}^n g(x_i) - a\right) + \frac{1}{2}h''(a^*)\left(n^{-1}\sum_{i=1}^n g(x_i) - a\right)^2, \tag{6}$$

для всіх $\omega \in \Omega\backslash A$ і $n \geq n(\omega)$. Рівність (6) можна переписати у вигляді

$$\frac{\sqrt{n}(\theta_n - \theta)}{D(g)} = h'(a)\frac{\sqrt{n}\left(n^{-1}\sum_{i=1}^n g(x_i) - a\right)}{D(g)} + h''(a^*)\frac{\sqrt{n}\left(n^{-1}\sum_{i=1}^n g(x_i) - a\right)^2}{2D(g)}. \tag{7}$$

Позаяк, випадкові величини $g(x_i)$, $i = 1, 2, ..., n$ є незалежними, однаково розподіленими і $\mathbf{M}g(x_i) = a$, $\mathbf{D^2}g(x_i) = D^2(g)$, то з центральної граничної теореми маємо

$$\frac{\sqrt{n}(n^{-1}\sum_{i=1}^n g(x_i) - a)}{D(g)} = \frac{\sum_{i=1}^n g(x_i) - na}{\sqrt{n}D(g)} \underset{n\to\infty}{\Longrightarrow} \xi \in \mathbb{N}(0, 1). \tag{8}$$

З (5), (8) отримуємо

$$\frac{\sqrt{n}\left(n^{-1}\sum_{i=1}^n g(x_i) - a\right)^2}{D(g)} = \frac{\sum_{i=1}^n g(x_i) - na}{\sqrt{n}D(g)}\left(n^{-1}\sum_{i=1}^n g(x_i) - a\right) \underset{n\to\infty}{\Longrightarrow} 0.$$

Тепер справедливість пункту а) випливає з рівності (7) і співвідношень (5), (8).

Для доведення пункту б) візьмемо математичне сподівання лівої та правої частини рівності (6)

$$\mathbf{M}\theta_n = \mathbf{M}h(n^{-1}\sum_{i=1}^n g(x_i)) = \theta + \frac{1}{2}\mathbf{M}\left(h''(a^*)(n^{-1}\sum_{i=1}^n g(x_i) - a)^2\right). \tag{9}$$

(Зауважимо, що значення $h''(a^*)$ є випадковою величиною і, тому, вона не може бути винесена з під символу $\mathbf{M}$.) Згідно з умовою теореми функція $h''(x)$ є обмеженою. Нехай $\sup_x |h''(x)| \leq C < \infty$. Тоді можемо записати

$$\left|\mathbf{M}\left(h''(a^*)(n^{-1}\sum_{i=1}^n g(x_i) - a)^2\right)\right| \leq C\mathbf{M}(n^{-1}\sum_{i=1}^n g(x_i) - a)^2 =$$

$$= C\mathbf{D^2}\left(n^{-1}\sum_{i=1}^n g(x_i)\right) = \frac{CD^2(g)}{n} \underset{n\to\infty}{\longrightarrow} 0.$$

Звідси а також з (9) випливає правдивість пункту б).

Зауваження 2. З доведеної теореми випливає, що для достатньо великих n маємо наближену рівність $\theta_n \approx \theta + D(g)h'(a)\xi/\sqrt{n}$, де $\xi \in \mathbb{N}(0; 1)$. Тому, якщо замість справжнього значення параметра θ використовуємо оцінку θ_n, тоді, взагалі кажучи, припускаємося помилки порядку $O(n^{-1/2})$. Як побачимо далі, така ситуація є типовою в математичній статистиці, тобто можемо сказати, що, якщо прагнемо отримати результати з докладністю α, то спостережень повинно бути порядку $1/\alpha^2$.

Отже, за умов попередньої теореми оцінка θ_n є асимптотично нормальною з параметром $h^2(a)D^2(g)$.

Якщо в попередній теоремі взяти $h(x) = x$, $g(x) = x^k$, $k \geq 1$, тоді отримаємо

Наслідок 1. *Якщо* $\mathbf{M}|\xi|^{2k} < \infty$, *то вибірковий момент порядку* $l \leq k$, $m_l(n) = n^{-1}\sum_{i=1}^n x_i^l$, *є асимптотично нормальним з параметром* $d_l^2 = \mathbf{D^2}\xi^l$ *і є незміщеною оцінкою для теоретичного моменту порядку* l, $m_l = \mathbf{M}\xi^l$. *Зокрема, для оцінки* $x(n) = n^{-1}\sum_{i=1}^n x_i$ *параметра* $\theta = \mathbf{M}\xi$ *генеральної сукупності* ξ *маємо*

$$\frac{\sqrt{n}(x(n) - m)}{\sigma} \underset{n\to\infty}{\Longrightarrow} \eta \in N(0,1), \qquad \sigma^2 = \mathbf{D^2}\xi. \tag{10}$$

10.3 Метод максимальної вірогідності

Метод максимальної вірогідності є найважливішим методом побудови оцінок (особливо з теоретичної точки зору). Елементи цього методу можемо знайти ще у Гауса, але як загальний метод для пошуку оцінок він був запропонований Р. Фішером в 1912 році. Фішер також виконав перші дослідження оцінок, отриманих за допомогою цього методу. Ми почнемо з розгляду ситуації, коли маємо справу з одним параметром.

Випадок одного параметра

Нехай функціональний вигляд розподілу генеральної сукупності ξ буде відомим і залежним від одного невідомого параметра $\theta \in \Theta$. Позначимо через $f(x,\theta)$ щільність цього розподілу. Ми прагнемо оцінити параметр θ на підставі n -елементної вибірки $x_1, \ldots, x_n$.

Визначення 2. Функція *вірогідності* для вибірки $x_1, \ldots, x_n$ з розподілом $f(x,\theta)$ визначається виразом

$$L(\mathbf{x},\theta) = f(x_1,\theta)f(x_2,\theta)\cdots f(x_n,\theta),$$

де $\mathbf{x} = (x_1, ..., x_n)$.

Метод максимальної вірогідності полягає в тому, що за оцінку параметра θ береться таке θ_n, для якого функція вірогідності приймає найбільше значення. Іншими словами, якщо для статистики θ_n виконане співвідношення

$$L(\mathbf{x},\theta_n) = \sup_{\theta\in\Theta} L(\mathbf{x},\theta), \tag{11}$$

то ця статистика і буде оцінкою параметра θ. Не виключено, що sup в (3) досягається в декількох точках і, отже, оцінка максимальної вірогідності, взагалі кажучи, не є єдиною. Також може статися так, що sup не досягається на Θ, і тоді оцінка максимальної вірогідності не існує. З іншого боку, якщо функція $f(x,\theta)$ є неперервною відносно θ, а множина Θ є замкнутою, тоді оцінка максимальної вірогідності завжди існує.

Оскільки функція $\ln L(\vec{x},\theta)$ досягає максимального значення для того ж самого θ, що і $L(\vec{x},\theta)$, то дослідження на максимальне значення $L(\vec{x},\theta)$ можна замінити аналогічними дослідженнями для $\ln L(\vec{x},\theta)$.

Підсумуємо це все в наступному визначенні.

Визначення 3. Оцінкою максимальної вірогідності (о.м.в.) для параметра θ будемо називати таке θ^*, для якого

$$\ln L(\vec{x}, \theta^*) = \sup_{\theta \in \Theta} \ln L(\vec{x}, \theta) \tag{12}$$

Зробимо тепер пояснення (на евристичному рівні), чому така ідея знаходження оцінки має сенс. Нехай $f(x, \theta^*)$ є справжньою щільністю генеральної сукупності ξ. Тобто θ^* є справжнім значенням θ, і нехай щільність має вигляд як представлено на рис. 10.1.

Рис. 10.1:

Тоді більша частина вибірки $x_1, ..., x_n$ буде зосереджена в області, де щільність є максимальною, тобто в області, позначеній на графіку через A (потовщена лінія), а менша частина елементів вибірки буде знаходитись в області, яка позначена на графіку через B. Отже, будемо мати в цій ситуації

$$L(\vec{x}, \theta^*) = \Pi_{i \in A} f(x_i, \theta^*) \Pi_{i \in B} f(x_i, \theta^*). \tag{13}$$

Нехай тепер $\theta = \theta^{**} \neq \theta^*$ і $f(x, \theta^{**}) \neq f(x, \theta^*)$. Тоді, взагалі кажучи, область, де щільність $f(x, \theta^{**})$ досягає максимального значення буде знаходитись в іншому місці. Нехай, наприклад, графік цієї щільності буде таким, як це представлено на рис. 10.1.

Тепер будемо мати

$$L(\mathbf{x}, \theta^{**}) = \Pi_{i \in A} f(x_i, \theta^{**}) \Pi_{i \in B} f(x_i, \theta^{**}). \tag{14}$$

Оскільки, місце знаходження вибірки не залежить від наших припущень стосовно параметра, то в (13), (14) беруть участь ті самі елементи. Позаяк, для $x_i \in A$ маємо $f(x_i, \theta^*) > f(x_i, \theta^{**})$, і більша частина вибірки зосереджена в області A, то зрозуміло, що, в загалі кажучи, буде виконуватись співвідношення $L(\vec{x}, \theta^*) > L(\vec{x}, \theta^{**})$. З таких евристичних міркувань приходимо до висновку, що коли ми відхиляємося від справжнього значення параметра, то функція вірогідності зменшується. Отже, є сенс взяти за оцінку цього справжнього значення власне те значення, де функція вірогідності досягає максимума.

Припустимо тепер, що функція вірогідності є диференційованою відносно θ. Якщо значення θ^*, де $\ln L$ має максимум є внутрішньою точкою області Θ, то θ^* є розв'язком рівняння

$$\frac{\partial}{\partial \theta} \ln L(\vec{x}, \theta) = 0. \tag{15}$$

Визначення 4. Рівняння (15) будемо називати *рівнянням вірогідності.*

Алгоритм пошуку точок θ, в яких функція $L(\vec{x},\theta)$ досягає максимального значення є добре відомим з математичного аналізу. Згодом ми доведемо, що при деяких умовах регулярності стосовно щільності генеральної сукупності рівняння (15) для великих n завжди має розв'язок.

Якщо генеральна популяція ξ є дискретною і $\mathbf{P}(\xi = x) = p(x,\theta)$, де $\theta \in \Theta$ є невідомим параметром, то метод максимальної вірогідності принципово не відрізняється від ситуації, описаної вище. Для вибірки $x_1,\ldots,x_n$ функція вірогідності виглядатиме тепер так

$$p(\vec{x},\theta) = p(x_1,\theta)p(x_2,\theta)\cdots p(x_n,\theta),$$

де $\vec{x} = (x_1,...,x_n)$. Далі дiємо так само, як і раніше.

Приклад 1. *На підставі n- елементної вибірки з сукупності, в якій випадкова величина має показниковий розподіл, знайти оцінку параметра λ.*

Р о з в' я з о к. Маємо

$$L(\vec{x},\lambda) = \Pi_{i=1}^{n}\lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda\sum_{i=1}^{n}x_i}.$$

Тоді $\ln L(\vec{x},\theta) = n\ln\lambda - \lambda\sum_{i=1}^{n}x_i$, і, отже, $\frac{d\ln\sum_{i=1}^{n}x_i}{d\lambda} = n/\lambda - \sum_{i=1}^{n}x_i$. Рівняння $n/\lambda - \sum_{i=1}^{n}x_i = 0$ має розв'язок $\lambda^* = n/\sum_{i=1}^{n}x_i$. В цій точці, як легко побачити, функція вірогідності досягає максимума. Отже, $n/\sum_{i=1}^{n}x_i = 1/x(n)$ є оцінкою максимальної вірогідності для параметра λ.

Приклад 2. *На підставі n-елементної вибірки з сукупності з розподілом Пуассона з параметром λ, знайти оцінку для цього параметра.*

Р о з в' я з о к. Маємо $p(k,\lambda) = \mathbf{P}_\lambda(\xi = k) = \frac{\lambda^k e^{-\lambda}}{k!}$. Тоді

$$p(\vec{k},\lambda) = p(k_1,\lambda)\cdot p(k_2,\lambda)\cdots p(k_n,\lambda) = \frac{\lambda^{k_1+k_2+\cdots+k_n}e^{-\lambda n}}{k_1!k_2!\cdots k_n!},$$

де $\vec{k} = (k_1,...,k_n)$, і отже,

$$\ln p(\vec{k},\lambda) = (k_1+k_2+\cdots+k_n)\ln\lambda - n\lambda - \ln(k_1!\cdots k_n!).$$

Беручи похідну, отримаємо

$$\frac{d}{d\lambda}\ln p(\vec{k},\lambda) = \frac{k_1+k_2+\cdots+k_n}{\lambda} - n = 0.$$

Звідки $\lambda = (k_1+\cdots+k_n)/n \stackrel{def}{=} \lambda_n^\star$. Оскільки,

$$\frac{d^2}{d\lambda^2}\ln p(\mathbf{k},\lambda) = -(k_1+\cdots+k_n)/\lambda^2 < 0,$$

то функція $\ln p(\vec{k},\lambda)$ приймає максимум в точці $\lambda_n^\star$ і, отже, $\lambda_n^\star$ є оцінкою, яку ми шукали.

Наведемо тепер два приклади, в яких функція вірогідності не буде диференційованою.

Приклад 3. *На підставі n-елементної вибірки з сукупності, в якій досліджувана випадкова величина має щільність вигляду:*

$$f(x,\theta)=\begin{cases}\frac{2x}{\theta^2}, & 0\le x\le\theta,\\ 0, & \textit{для інших } x.\end{cases}$$

Знайти оцінку для невідомого параметра $\theta>0$.

Р о з в' я з о к. Застосуємо метод максимальної вірогідності. Оскільки добуток $\Pi_{i=1}^{n}f(x_i,\theta)$ буде дорівнювати нулеві, якщо хоча б один елемент x_i буде поза інтервалом $[0;\theta]$, то маємо

$$L(\mathbf{x},\theta)=\begin{cases}\frac{2^n\Pi_{i=1}^{n}x_i}{\theta^{2n}}, & 0\le x_i\le\theta,\ i=1,2,...,n,\\ 0 & \text{для інших } x_i.\end{cases}$$

Умова $0\le x_i\le\theta, i=1,2,...,n$ є вочевидь еквівалентна $x_{(n)}\le\theta$.

Отже, маємо

$$L(\mathbf{x},\theta)=\begin{cases}\frac{2^n\Pi_{i=1}^{n}x_i}{\theta^{2n}}, & \theta\ge x_{(n)}\\ 0, & \text{для інших } \theta.\end{cases}$$

Графік цієї функції вірогідності представлений на рис. 10.2

Рис. 10.2: Графік для $L(\vec{x},\theta)$

Очевидно, функція $L(\mathbf{x},\theta)$ не є диференційованою в точці $\theta=x_{(n)}$, але теж є очевидним, що в цій точці вона досягає максимального значення, і, отже, $x_{(n)}$ є оцінкою максимальної вірогідності для параметра θ .

Приклад 4. *На підставі n- елементної вибірки з сукупності, в якій досліджувана випадкова величина має рівномірний розподіл на відрізку $[\theta\le x\le 1+\theta]$. Знайти оцінку для невідомого параметра $\theta>0$.*

Р о з в' я з о к. Міркуючи так само як і раніше, маємо

$$L(\vec{x},\theta)=\begin{cases}1, & \theta\le x_i\le 1+\theta, i=1,2,...,n,\\ 0, & \text{для інших } x_i.\end{cases}$$

Умова $\theta \le x_i \le 1+\theta, i = 1, 2, ..., n$ є вочевидь еквівалентна таким $\theta \le x_{(1)}$ і $x_{(n)} \le 1+\theta$. Звідки одержимо $x_{(n)} - 1 \le \theta \le x_{(1)}$. Отже, будемо мати

$$L(\vec{x}, \theta) = \begin{cases} 1, & x_{(n)} - 1 \le \theta \le x_{(1)}, \\ \\ 0, & \text{для інших } \theta. \end{cases}$$

Графік цієї функції вірогідності представлений на рис. 10.3

Рис. 10.3: Графік для $L(\vec{x}, \theta)$

Очевидно функція $L(\vec{x}, \theta)$ не є диференційованою в точках $x_{(n)} - 1$, $x_{(1)}$, але теж є очевидним, що ця функція досягає максимального значення для всіх $\theta \in [x_{(n)} - 1, x_{(1)}]$, і, отже, будь-яке число з інтервалу $\theta \in [x_{(n)} - 1, x_{(1)}]$ може грати роль оцінки максимальної вірогідності для параметра θ.

10.3.1 Випадок багатьох параметрів

В цій ситуації теж вважаємо, що функціональний вигляд розподілу випадкової величини ξ є відомим, але залежний від k невідомих параметрів $(\theta_1, ..., \theta_k) \in \Theta \subset R^k$. Позначимо через $f(x, \widehat{\theta})$, $\widehat{\theta} = (\theta_1, ..., \theta_k)$ щільність цього розподілу.

Нам треба оцінити тепер векторний параметр $\widehat{\theta}$ на підставі n -елементної вибірки $x_1, \ldots, x_n$. Принципової різниці в застосуванню методу максимальної вірогідності до розв'язання цієї проблеми немає. В цьому випадку функція вірогідності виглядає так

$$L(\vec{x}, \widehat{\theta}) = \Pi_{i=1}^{n} f(x_i, \widehat{\theta}), \qquad \mathbf{x} = (x_1, ..., x_n).$$

Значення $\widehat{\theta}^*$, для якого виконується рівняння,

$$L(\vec{x}, \widehat{\theta}^*) = \sup_{\widehat{\theta} \in \Theta} L(\vec{x}, \widehat{\theta})$$

вибираємо в якості оцінки для параметра $\widehat{\theta}$.

Як і в ситуації одного параметра, оцінка максимальної вірогідності може існувати і бути єдиною, може й не бути єдиною, а може й в загалі не існувати. Якщо множина Θ є замкненою, то оцінка максимальної вірогідності завжди існує.

Так само, як і раніше, замість $L(\mathbf{x}, \theta)$ можна досліджувати максимальне значення $\ln L(\mathbf{x}, \theta)$. Якщо ця функція є диференційованою відносно θ_i, і якщо значення $\hat{\theta}^*$, де $\ln L$ має максимум, є внутрішньою точкою області Θ, то компоненти вектора $\hat{\theta}^*$ будуть розв'язками системи рівнянь

$$\frac{\partial}{\partial \theta_i} \ln L(\mathbf{x}, \hat{\theta}) = 0, \quad i = 1, .., k. \tag{16}$$

Систему рівнянь (17) теж будемо називати рівнянням вірогідності.

Приклад 5. *На підставі n- елементної вибірки з генеральної сукупності, в якій випадкова величина має розподіл $\mathbb{N}(m, \sigma^2)$ знайти оцінку для параметрів m, σ^2.*

Р о з в' я з о к. Оскільки $f(x, m, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(\xi - m)^2}{2\sigma^2}\}$, то

$$L(\mathbf{x}, m, \sigma^2) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - m)^2}{2\sigma^2}}$$

і, отже, $\ln L(\mathbf{x}, m, \sigma^2) = -n \ln 2\pi/2 - \ln \sigma^2/2 - (x_i - m)^2/2\sigma^2$. Беручи похідну по m і σ^2 (звертаємо увагу, що нашим параметром є σ^2, а не σ, тому похідну беремо відносно σ^2), маємо рівняння для оцінок

$$-\frac{n}{2\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - m)^2 = 0, \quad \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m)^2 = 0.$$

Звідки отримуємо

$$m = \frac{1}{n} \sum_{i=1}^n x_i = x(n), \qquad \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - x(n))^2 = S^2(n).$$

Легко перевірити, що функція $\ln L(\mathbf{x}, m, \sigma^2)$ досягає в тих точках максимуму і, отже, $x(n)$, $S^2(n)$ є оцінками максимальної вірогідності для параметрів m, σ^2.

Нехай θ^* є о.м.в. для θ і $\varkappa = g(\theta)$, де $g(x)$-деяка функція. Чи буде тоді $g(\theta^*)$ о.м.в. для $\varkappa$? За досить широких припущень відповідь на це питання є позитивною. Випливає це з наступної теореми, відомої як *принцип інваріантності* оцінок максимальної вірогідності. Доведення цієї теореми є нескладним, і тому ми його не наводимо.

Теорема 3. *Нехай відображення $g : \Theta \to B$ є таким, що існує g^{-1}. Тоді, якщо θ^* є о.м.в. для параметра θ генеральної сукупності ξ, то $g(\theta^*)$, є о.м.в. для параметра $g(\theta) \in B$ тієї ж самої генеральної сукупності.*

Ми знаємо, що $S^2(n) = n^{-1} \sum_{i=1}^n (x_i - x(n))^2$ є о.м.в. для σ^2 генеральної сукупності з розподілом $\mathbb{N}(m, \sigma^2)$. Отже, S_n буде о.м.в. для стандартного відхилення σ.

10.4 Асимптотичні властивості оцінок максимальної вірогідності

В цьому параграфі покажемо, що за деяких умов регулярності стосовно щільності генеральної сукупності рівняння вірогідності

$$\frac{\partial}{\partial \theta} \ln L(\mathbf{x}, \theta) = 0 \tag{17}$$

завжди має розв'язок, а також проведемо дослідження цього розв'язку. Розглянемо лише випадок одного параметра. Для векторних параметрів відповідні результати можна знайти в стандартних підручниках з математичної статистики таких, як [1], [10], [14], [19] та інших.

Теорема 4. *Нехай генеральна сукупність ξ має щільність $f(x,\theta)$ і $\theta \in \Theta = [a;b]$, $-\infty \le a < b \le \infty$. Припустимо, що виконані наступні умови:*

а) *справжнє значення параметра $\theta = \theta^*$ є внутрішньою точкою інтервалу $[a;b]$, тобто $a < \theta^* < b$;*

б) *існують $\frac{\partial^i f(x,\theta)}{\partial\theta^i}$, $i = 1,2,3$;*

в) *для кожного $\theta \in \Theta$, $|\frac{\partial^i f(x,\theta)}{\partial\theta^i}| < F_i(x)$, $i = 1,2$ і $|\frac{\partial^3 \ln f(x,\theta)}{\partial\theta^3}| < F_3(x)$, причому функції F_1, F_2 є інтегровані на $(-\infty,\infty)$, а функція $F_3(x)$ не залежить від θ і $\int_{-\infty}^{\infty} F_3(x)f(x,\theta)dx < M < \infty$;*

г) $0 < I^2 \equiv I^2(\theta^*) \overset{def}{=} \int_{-\infty}^{\infty}(\frac{\partial \ln f(x,\theta)}{\partial\theta})^2 f(x,\theta)dx < \infty$.

Тоді для достатньо великих n рівняння вірогідності (17) *завжди має корінь θ_n і якщо θ^* є справжнім значенням параметра θ, то:*

1) $\theta_n \underset{n\to\infty}{\xrightarrow{\text{м.н.}}} \theta^*$;

2) $\sqrt{n}I(\theta_n - \theta^*) \underset{n\to\infty}{\Longrightarrow} \eta \in N(0,1)$;

3) *Оцінка θ_n параметра θ є асимптотично ефективною.*

10.5 Метод моментів

На початку розглянемо ситуацію, коли $\theta \in \Theta = [a,b]$ є скалярним параметром, і ξ буде генеральною сукупністю з розподілом $F(x,\theta)$. Припустимо, що функція $g(x)$ є такою, що

$$m(\theta) = \mathbf{M}g(\xi) = \int_{-\infty}^{\infty} g(x,\theta)dF(x,\theta) \tag{18}$$

є неперервною функцією і, крім того, існує обернена функція $m^{-1}(\cdot)$. Для цього достатньо, щоб функція $m(\theta)$ була монотонною. Згідно з ідеєю методу підстановки за оцінку для $m(\theta)$ треба взяти $n^{-1}\sum_{i=1}^n g(x_i)$. Ідея методу моментів полягає в тому, що в якості оцінки для θ беремо розв'язок рівняння $m(\theta) = n^{-1}\sum_{i=1}^n g(x_i)$, тобто,

$$\theta_n(\vec{x}) = m^{-1}\ n^{-1}\sum_{i=1}^n g(x_i)\ . \tag{19}$$

Визначення 5. Оцінку параметра θ (19) будемо називати *оцінкою методу моментів.*

Зрозуміло, що оцінка (19) є оцінкою методу підстановки, і тому, з результатів підрозділів 10.1, **??** маємо

Теорема 5. *Оцінка методу моментів* (19) *є сильно консистентною, а якщо функція $m(x)$ двічі неперервно диференційована в околиці точки $a = \mathbf{M}g(\xi)$ і $D^2(g) = \mathbf{D^2}g(\xi) < \infty$, тоді:*

а) $\frac{\sqrt{n}(\theta_n-\theta)\ m'(\theta)}{D(g)}^{\,2} \underset{n\to\infty}{\Longrightarrow} \eta \in \mathbb{N}(0,1)$;

б) *якщо також функція $m''(x)$ є обмеженою, то оцінка θ_n буде асимптотично незміщеною.*

Ідея методу моментів є дуже простою: треба прирівняти теоретичні величини з їхніми аналогами, підрахованими з вибірки. Отримані рівняння потрібно розв'язати відносно параметра, що й дасть нам потрібну оцінку. Сама назва "метод моментів" походить з того, що найчастіше в якості функцій $g(x)$ вибирають x^k, $k \geq 1$, а тоді значення $m(\theta)$ будуть моментами сукупності ξ. Перевагою цього методу є його прозорість і простота. Але знаходження функції $m(\theta)$ і, тим більше, розв'язання рівняння (19) може завдати клопоту. Очевидно, до цієї проблеми можна застосовувати різні обчислювальні методи, але ми не будемо торкатися цього питання. Неоднозначність методу моментів полягає ще в тому, що вибір функції $g(\cdot)$ обмежений лише умовою $\mathbf{M}g(\xi) < \infty$. А це в свою чергу дає нам різні оцінки для одного і того ж параметра.

Приклад 1. *Нехай генеральна популяція ξ має показниковий розподіл з параметром λ. Знайти оцінку для λ, користуючись методом моментів.*

Р о з в' я з о к. Нехай $g_1(x) = x,\ g_2(x) = x^2$. Тоді будемо мати

$$m_1(\lambda) = \mathbf{M}g_1(\xi) = \mathbf{M}\xi = 1/\lambda; \qquad m_2(\lambda) = \mathbf{M}g_2(\xi) = \mathbf{M}\xi^2 = 2/\lambda^2,$$

і можемо записати рівняння

$$x(n) = 1/\lambda, \qquad \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 2/\lambda^2,$$

звідки отримаємо дві оцінки для λ

$$\lambda_1(n) = \frac{1}{x(n)}, \qquad \lambda_2(n) = \sqrt{\frac{2n}{\sum_{i=1}^{n} x_i^2}}, \tag{20}$$

В підпідрозділі 5.1.2 побачимо, що перша з цих оцінок є асимптотично нормальною з параметром λ^2, натомість друга - з параметром $5\lambda^2/4$ і, отже, перша оцінка є кращою, оскільки, для великих n її відхилення від справжнього значення параметра буде меншим.

ЗАВДАННЯ

1. Нехай генеральна популяція ξ має розподіл $\Gamma(x)$ з щільністю

$$f(x, \lambda) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}, \qquad x \geq 0,$$

   де p є відомим, а λ є невідомим параметром. Знайти о.м.в. для λ.

2. Нехай генеральна сукупність ξ має розподіл Релея з щільністю

$$f(x, \lambda) = \lambda x e^{-\lambda x^2/2}, \qquad x \geq 0,$$

   де λ є невідомим параметром. Знайти о.м.в. для λ і показати, що ця оцінка є сильно консистентна.

3. Нехай генеральна сукупність ξ має такий самий розподіл як і в попередньому прикладі, але в якості параметра тепер виступає $\theta = 1/\lambda$. Знайти о.м.в. для θ і показати, що ця оцінка є незміщена, сильно консистентна та ефективна.

4. Генеральна сукупність ξ має рівномірний розподіл на відрізку $[\theta-1;\theta+1]$, де θ вважаємо параметром. Методом максимальної вірогідності знайти оцінку для параметра θ.

5. Генеральна сукупність ξ має рівномірний розподіл на відрізку $[\theta;1]$, де $\theta \leq 1]$ і виступає як параметр. Методом максимальної вірогідності знайти оцінку цього параметра.

6. Генеральна сукупність ξ має рівномірний розподіл на відрізку $[a;b]$, де $a \leq b$ є параметрами. Методом максимальної вірогідності знайти оцінку для цих параметрів.

7. Знайти о.м.в. для ймовірності "успіху" в схемі Бернуллі.

8. Генеральна сукупність ξ має розподіл $f(x,\lambda) = (2\lambda)^{-1}\exp\{-|x|/\lambda\}$. Методом максимальної вірогідності знайти оцінку для λ.

9. Нехай $x_1,...,x_n$ є вибіркою з генеральної сукупності ξ такої, що $\mathbf{M}\xi = \alpha > 0, \mathbf{D^2}\xi = \sigma^2 < \infty$. Довести, що оцінка $S(\vec{x}) = 1/x(n)$, $x(n) = n^{-1}\sum_{i=1}^n x_i$ для параметра $1/\alpha$ є асимптотично незміщеною і асимптотично нормальною з параметром σ^2/α^4.

10. Нехай генеральна популяція ξ є такою, що $\mathbf{M}\xi = m$ відоме, $\mathbf{D^2}\xi = \sigma^2$- невідомий параметр і, крім того, $\mathbf{M}\xi^4 < \infty$. Довести, що:

    а) статистика $\bar{S}^2(n) = n^{-1}\sum_{i=1}^n (x_i - m)^2$ є асимптотично нормальною з параметром $q^2 = \mathbf{M}(\xi - m)^4 - \sigma^4$ оцінкою для σ^2;

    б) статистика $\sqrt{\bar{S}^2(n)}$ є асимптотично нормальною з параметром $\frac{q^2}{4\sigma^2}$ оцінкою для σ.

    Вказівка. *Скористуватись теоремою 2, взявши* $g(x) = (x-m)^2$, *а* $h(x) = x$ *в першому випадку і* $h(x) = \sqrt{x}$ *в другому.*

11. Знайти оцінки методу моментів для параметрів наступних розподілів:

    а) бета $f(x,\beta) = \frac{\Gamma(\beta+2)}{2\Gamma(\beta)} x^1(1-x)^{1-\beta}$ $x \in [0,1]$;

    б) рівномірного на інтервалі $[0,\theta]$;

    в) Бернуллі.

12. Знайти оцінки методу моментів для параметрів розподілу з щільністю (розподіл гама) $\frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}$, $x \geq 0$.

    Вказівка. *В якості функції* g_i *взяти* $g_1(x) = x$, $g_2(x) = x^2$.

13. Нехай генеральна сукупність ξ має розподіл Релея з щільністю $f(x,\lambda) = \lambda x e^{-\lambda x^2/2}$, $x \geq 0$, де λ є невідомим параметром. Методом моментів знайти оцінку для λ і показати, що ця оцінка є сильно консистентна.

14. Нехай генеральна сукупність ξ має такий самий розподіл як і в попередньому прикладі, але в якості параметра тепер виступає $\theta = 1/\lambda$. Методом моментів знайти оцінку для θ і показати, що ця оцінка є асимптотично незміщена та сильно консистентна.

Лекція 11. Достатні статистики

11.1 Умовне математичне сподівання

В цьому підрозділі будемо використовувати поняття умовного математичного сподівання і тому спочатку пригадаємо його означення, яке зазвичай подається в курсі теорії ймовірності. Це поняття розглянемо за певних додаткових припущень, а загальний випадок можна знайти в стандартних підручниках (див., наприклад, [1], [4]).

Нехай ξ, η будуть двома випадковими змінними з щільностями $f_\xi(x)$, $f_\eta(x)$ та спільною щільністю $f(x,y)$. Умовна щільність змінної ξ відносно η визначається формулою

$$f(x/y) = \begin{cases} \frac{f(x,y)}{f_\eta(y)}, & \text{якщо } f_\eta(y) > 0, \\ 0, & \text{якщо } f_\eta(y) = 0. \end{cases}$$

Тоді умовне математичне сподівання для ξ за умови $\eta = y$ задається так

$$\mathbf{M}\{\xi/\eta = y\} = \int\limits_{-\infty}^{\infty} uf(u/y)du.$$

Визначення 1. *Умовним математичним сподіванням* випадкової змінної ξ відносно η називається випадкова величина $\varphi(\eta)$, яка позначається $\varphi(\eta) = \mathbf{M}\{\xi/\eta\}$ (або ще так $\varphi(\eta) = \mathbf{M}\eta\xi$) і яка при $\eta = y$ приймає значення

$$\varphi(\eta)|_{\eta=y} = \mathbf{M}\{\xi/\eta = y\}.$$

Як випливає з наведеного означення, умовне математичне сподівання $\mathbf{M}\{\xi/\eta\}$ є випадковою величиною, а тому можемо порахувати математичне сподівання (тепер вже звичайне) цієї величини. Маємо

$$\mathbf{M}\{\mathbf{M}\{\xi/\eta\}\} = \mathbf{M}\varphi(\eta) = \int\limits_{-\infty}^{\infty} \varphi(y)f_\eta(y)dy = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} uf(u/y)duf_\eta(y)dy =$$

$$= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} uf(u,y)dudy = \int\limits_{-\infty}^{\infty} uf_\xi(u)du = \mathbf{M}\xi.$$

Аналогічно можна показати, що для кожної обмеженої функції $\varphi(x)$ виконується рівність

$$\mathbf{M}\{\varphi(\eta)\xi/\eta\} = \varphi(\eta)\mathbf{M}\{\xi/\eta\}. \tag{1}$$

Отже ми довели наступні властивості умовного математичного сподівання, які нам будуть потрібні далі

Теорема 1. *Нехай* $\mathbf{M}|\xi| < \infty$. *Тоді*

а) $\mathbf{M}\{\mathbf{M}\{\xi/\eta\}\} = \mathbf{MM}\eta\xi = \mathbf{M}\xi$,

б) *для кожної обмеженої функції* $\varphi(x)$ *справедлива рівність* $\mathbf{M}\{\varphi(\eta)\xi/\eta\} = \varphi(\eta)\mathbf{M}\{\xi/\eta\}$.

11.2 Достатні статистики. Факторизаційна теорема.

Як вже було сказано вище, статистика є функцією вибірки, і з цього випливає, що перехід від вибірки $x_1, ..., x_n$ до статистики $S(\mathbf{x})$, взагалі кажучи, призводить до втрати частини інформації щодо генеральної сукупності ξ, яка у вибірці містилася. Так, наприклад, якщо маємо вибірку $-1, 3, -2, 4$, то можемо зробити висновок, що генеральна популяція приймає від'ємні значення. Разом з тим, якщо маємо статистику $x(n) = n^{-1}(x_1 + x_2 + x_3 + x_4)$, то для даної вибірки маємо $x(4) = 1$. Очевидно, що тепер на підставі значення статистики $x(4)$ не можемо робити висновку, що генеральна популяція приймає від'ємні значення. Отже може трапитися так, що коли в якості оцінки для параметра θ вживаємо статистику $S(\mathbf{x})$, то не використовуємо всієї інформації, яка містилася у вибірці. Очевидно, що в такому разі можемо втратити на точності нашої оцінки. Було б добре, якби статистика $S(\mathbf{x})$ була такою, щоб подібної втрати інформації не трапилося. Такі статистики називаємо достатніми. Точне означення виглядає так.

Визначення 2. Статистика $S(\mathbf{x})$ називається *достатньою* для параметра θ якщо умовний розподіл вектора $\mathbf{x} = (x_1, ..., x_n)$ за умови $S(\mathbf{x}) = t$ не залежить від θ.

Інакше кажучи, якщо умовна функція розподілу $\mathbf{P}(x_1 < z_1, ..., x_n < z_n/S(\vec{x}) = z)$ є лише функцією від $z_1, ..., z_n$ і z, то статистика $S(\vec{x})$ є достатньою для параметра θ.

Сенс цього означення можемо пояснити наступним чином. Якщо θ є невідомий параметр для генеральної сукупності ξ, то функція розподілу $F(\theta, z_1, ...z_n) = \mathbf{P}(x_1 < z_1, ..., x_n < z_n)$ (не умовна) залежить від цього параметра і, значить, в принципі, є невизначеною, або відома лише з точністю до цього параметра і для точного описання цієї функції розподілу потребуємо додаткової інформації про параметр θ. Тому, якщо функція розподілу $\mathbf{P}(x_1 < z_1, ..., x_n < z_n/S(\mathbf{x}) = z)$ не залежить від параметра θ, то умова $S(\mathbf{x}) = z$, власне, і дає ту недостатню інформацію. Іншими словами, вся інформація про параметр θ міститься в значенні статистики $S(\vec{x})$, тобто знання статистики $S(\vec{x})$ *достатньо* для побудови оцінки параметра θ.

Приклад 1. *Нехай генеральна популяція* ξ *має розподіл Пуассона з параметром* λ, *тобто,* $\mathbf{P}(\xi = k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, ...$ *Довести, що статистика* $S(\vec{k}) = \sum_{i=1}^n k_i$ *є достатньою для параметра* λ.

Р о з в' я з о к. Маємо

$$\mathbf{P}(k_1 = m_1, ..., k_n = m_n/S(\vec{k}) = k) = \frac{\mathbf{P}(k_1 = m_1, ..., k_n = m_n, S(\vec{k}) = k)}{\mathbf{P}(\sum_{i=1}^n k_i = k)}. \tag{2}$$

Якщо $k_1 = m_1, ..., k_n = m_n$, то тоді $\sum_{i=1}^n k_i = \sum_{i=1}^n m_i$ і тому, якщо $\sum_{i=1}^n m_i \neq k$, то чисельник в (2) дорівнюватиме нулеві. З іншого боку, якщо $\sum_{i=1}^n m_i = k$, то, оскільки випадкові змінні k_i, $i = 1, 2, ..., n$ є незалежними, маємо

$$\mathbf{P}(k_1 = m_1, ..., k_n = m_n, S(\mathbf{k}) = k) = \mathbf{P}(k_1 = m_1, ..., k_n = m_n) =$$
$$= \prod_{i=1}^n \mathbf{P}(k_i = m_i) = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^n m_i}}{m_1! \cdots m_n!} = e^{-n\lambda}\frac{\lambda^k}{m_1! \cdots m_n!}. \tag{3}$$

Оскільки k_i має розподіл Пуассона з параметром λ, то $\sum_{i=1}^n k_i$ має розподіл Пуассона з параметром $n\lambda$. Іншими словами,

$$\mathbf{P}(\sum_{i=1}^n k_i = k) = e^{-n\lambda}\frac{(n\lambda)^k}{k!}.$$

Остання рівність і (2), (3) дають

$$\mathbf{P}(k_1 = m_1, ..., k_n = m_n / S(\mathbf{k}) = k) = \frac{k!}{n^k m_1! \cdots m_n!}$$

Права частина цієї рівності не залежить від λ, що і означає, що статистика $S(\vec{k}) = \sum_{i=1}^n k_i$ є достатньою.

Наступна теорема, відома як *факторизаційна теорема Неймана*[1]*-Фішера*, дає необхідні і достатні умови для того, щоб статистика S_n була достатньою.

Нехай $x_1, ..., x_n$ вибірка взята з генеральної сукупності ξ зі щільністю $f(x,\theta)$. Як ми вже знаємо, функція $L(\vec{x},\theta) = \prod_{i=1}^n f(x_i,\theta)$ називається функцією вірогідності, а якщо ξ має дискретний розподіл $\mathbf{P}(\xi = z_k) = p(z_k,\theta)$, $k = 1, 2, ...$, то ця функція виглядає так $L(\vec{x},\theta) = \prod_{i=1}^n p(x_i,\theta)$.

Теорема 2. *Статистика $S(\vec{x})$ для параметра θ буде достатньою тоді і лише тоді, коли*

$$L(\vec{x},\theta) = \psi(S(\vec{x}),\theta) \cdot h(\vec{x}), \tag{4}$$

де функції $\psi(\cdot,\cdot) \geq 0$, $h(\cdot) \geq 0$.

Зауваження 1. Запис функції вірогідності у вигляді (4), яку ми називаємо *факторизацією* функції $f_\theta(\vec{x})$, не є однозначним. Так, якщо маємо (4) і $\varphi(\cdot)$ довільна додатня функція, то

$$L(\vec{x},\theta) = \widetilde{\psi}(S(\vec{x}),\theta) \cdot \widetilde{h}(\vec{x}),$$

де $\widetilde{\psi}(S(\vec{x}) = \psi(S(\vec{x}),\theta)/\varphi(S(\vec{x}))$, $\quad \widetilde{h}(\vec{x}) = h(\vec{x})\varphi(S(\vec{x}))$ буде іншою факторизацією цієї ж функції.

Д о в е д е н н я. Доведення цієї теореми наведемо лише для випадку, коли генеральна сукупність має дискретний розподіл. Доведення для загального випадку можна знайти в [1], [19]. Оскільки тепер нам важливо підкреслити залежність від параметра, то замість $\mathbf{P}$ будемо писати $\mathbf{P}_\theta$ і це означає, що ймовірність вираховується відносно розподілу, якому відповідає значення параметра θ.

Отже, нехай маємо рівність (4). Тоді для довільного $\vec{y} = (y_1, ..., y_n)$ і $t \in (-\infty,\infty)$ маємо

$$\mathbf{P}(\vec{x} = \vec{y} / S(\vec{x}) = t) = \frac{\mathbf{P}(\vec{x} = \vec{y}, S(\vec{x}) = t)}{\mathbf{P}(S(\vec{x}) = t)} = \frac{\mathbf{P}(\vec{x} = \vec{y}, S(\vec{y}) = t)}{\mathbf{P}(S(\vec{x}) = t)} \tag{5}$$

[1]Нейман (*Splawa-Neyman Jerzy*) (16.04.1894–5.08. 1981)– американський математик польського походження, займався математичною статистикою і її застосуванням. Випускник Харківського університету. В 1917-1938 роках працював у Польщі, а з 1938р. мешкав і працював у Сполучених Штатах Америки.

за умови, що $\mathbf{P}(S(\vec{x}) = t) > 0$. Якщо $S(\vec{y}) = t$, то

$$\frac{\mathbf{P}(\vec{x} = \vec{y}, S(\vec{y}) = t)}{\mathbf{P}(S(\vec{x}) = t)} = \frac{\mathbf{P}(\vec{x} = \vec{y})}{\mathbf{P}(S(\vec{x}) = t)} = \frac{L(\vec{y}, \theta)}{\sum\limits_{\{\vec{z}:S(\vec{z})=t\}} L(\vec{z}, \theta)} =$$

$$= \frac{\psi(S(\vec{y}), \theta)h(\vec{y})}{\sum\limits_{\{\vec{z}:S(\vec{z})=t\}} \psi(S(\vec{z}), \theta)h(\vec{z})} = \frac{\psi(t, \theta)h(\vec{y})}{\sum\limits_{\{\vec{z}:S(\vec{z})=t\}} \psi(t, \theta)h(\vec{z})} = \frac{h(\vec{y})}{\sum\limits_{\{\vec{z}:S(\vec{z})=t\}} h(\vec{z})}.$$

Якщо $S(\vec{y}) \neq t$, то права частина в (5) рівна 0. Тому маємо

$$\mathbf{P}(\vec{x} = \vec{y}/S(\vec{x}) = t) = \begin{cases} \frac{h(\vec{y})}{\sum\limits_{\{\vec{z}:S(\vec{z})=t\}} h(\vec{z})}, & \text{якщо} \quad S(\vec{y}) = t, \\ \\ 0, & \text{якщо} \quad S(\vec{y}) \neq t \end{cases}$$

і права частина в цій рівності не залежить від θ.

Нехай тепер ймовірність $\mathbf{P}(\vec{x} = \vec{y}/S(\vec{x}) = t)$ для фіксованих $\vec{y}$, t не залежить від θ. Позначимо цю ймовірність через $h(\vec{x})$. Для фіксованого $\vec{y}$ нехай $t \stackrel{def}{=} S(\vec{y})$. Маємо

$$\begin{aligned} L(\vec{x}, \theta) &= \mathbf{P}(\vec{x} = \vec{y}) = \mathbf{P}(\vec{x} = \vec{y}, S(\vec{x}) = S(\vec{y})) = \mathbf{P}(\vec{x} = \vec{y}, S(\vec{x}) = t) \\ &= \mathbf{P}(S(\vec{x}) = t)\mathbf{P}(\vec{x} = \vec{y}/S(\vec{x}) = t) = \psi(S(\vec{x}), \theta)h(x), \end{aligned}$$

де $\psi(S(\vec{x}), \theta) \stackrel{def}{=} \mathbf{P}(S(\vec{x}) = t)$, що завершує доведення.

Приклад 2. *Нехай генеральна популяція $\xi \in \mathbb{N}(m, \sigma^2)$ і $\theta = (m, \sigma^2)$ трактуємо як параметр.*

Для функції вірогідності маємо

$$L(\vec{x}, \theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{(x_i - m)^2}{2\sigma^2}\} = \sigma^{-n}(2\pi)^{-n/2} \exp\{-\frac{\sum_{i=1}^{n}(x_i - m)^2}{2\sigma^2}\} =$$

$$= \sigma^{-n}(2\pi)^{-n/2} \exp\{-\frac{\sum_{i=1}^{n} x_i^2 - 2m\sum_{i=1}^{n} x_i + nm^2}{2\sigma^2}\}.$$

Якщо в якості статистики взяти $S = (S_1, S_2)$, де $S_1 = \sum_{i=1}^{n} x_i$, $S_2 = \sum_{i=1}^{n} x_i^2$, то маємо (4) з

$$\psi(S, \theta) = \sigma^{-n} \exp\{-\frac{S_2 - 2mS_1 + nm^2}{2\sigma^2}\}, \ h(\vec{x}) = (2\pi)^{-n/2}.$$

Значить, статистика $S = (S_1, S_2)$ є достатньою для параметра $\theta = (m, \sigma^2)$. З усієї інформації, що міститься у вибірці, нам вистачить значень $\sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} x_i^2$ для оцінювання параметра θ.

Приклад 3. *Нехай генеральна популяція ξ має показниковий розподіл з параметром λ.*

Для функції вірогідності маємо $L(\vec{x}, \theta) = \lambda^n \exp\{-\lambda \sum_{i=1}^{n} x_i\}$. З теореми про факторизацію випливає, що наступні статистики

$$S_k = (\sum_{i=1}^{k} x_i, x_{k+1}, x_{k+2}, ..., x_n), \qquad k = 1, ..., n$$

є достатніми для параметра λ. Очевидно, статистика S_1 співпадає з вибіркою (а вибірка завжди є достатньою статистикою), а, наприклад, для $S_2 = (\sum_{i=1}^{2} x_i, x_3, x_4, ..., x_n)$ маємо (4) з

$$\psi(S_2, \lambda) = \lambda^n \exp\{-\lambda(\sum_{i=1}^{2} x_i + x_3 + x_4 + ... + x_n)\},\ h(\vec{x}) = 1.$$

11.3 Достатні оцінки та статистики

Достатні статистики можна використати для побудови оцінок невідомих параметрів. Виявляється, що якщо маємо деяку оцінку $\theta(\vec{x})$ для параметра θ, то за допомогою достатніх статистик ця оцінка може бути поліпшена. Це буде випливати з теореми, яка була незалежно доведена Блеквелом[1], Колмогоровим і Рао .

Розглянемо випадок скалярного параметра і пригадаємо, що через $\mathbf{K}_b$ позначаємо клас оцінок $\theta(\vec{x})$ для θ таких, що $\mathbf{M}\theta(\vec{x}) - \theta = b(\theta)$.

Теорема 3. *Нехай оцінка $\theta(\vec{x}) \in \mathbf{K}_b$ і $S(\vec{x})$ буде достатньою статистикою. Тоді для $\theta^*_S = \mathbf{M}\{\theta(\vec{x})/S(\vec{x})\}$ маємо:*

а) θ^*_S *є статистикою і залежить від вибірки $\vec{x}$ лише через $S(\vec{x})$;*

б) $\theta^*_S \in \mathbf{K}_b$;

в) $\mathbf{M}(\theta^*_S - \theta)^2 \leq \mathbf{M}(\theta(\vec{x}) - \theta)^2$ *для довільного $\theta \in \Theta$ і рівність можлива тоді і лише тоді, коли*

$$\theta^*_{S(\vec{x})} = \theta(\vec{x}) \quad \text{якщо} \quad L(\vec{x}, \theta) > 0. \tag{6}$$

Отже, якщо маємо оцінку $\theta(\vec{x})$ для θ і деяку достатню статистику для цього ж параметра, то перехід до умовного математичного сподівання відносно цієї статистики поліпшує оцінку $\theta(\vec{x})$.

Д о в е д е н н я. Для першого пункту маємо довести, що величина θ^*_S не залежить від параметра θ, а лише від вибірки $\vec{x}$, більше того, ця залежність реалізується лише через $S(\vec{x})$. Оскільки статистика S є достатньою, то умовний розподіл вибірки $\vec{x}$ за умови $S(\vec{x}) = y$ не залежить від θ, тобто

$$\mathbf{P}(x_1 < z_1, ..., x_n < z_n / S(\vec{x}) = y) = G(z_1, ..., z_n, y)$$

і функція G не залежить від θ. Тепер маємо

$$\mathbf{M}\{\theta(\vec{x})/S(\vec{x}) = y\} = \int_{\mathbf{R^n}} \theta(\vec{z}) dG(\vec{z}, y)$$

і функція в правій частині цієї рівності не залежить від θ. Крім того, залежність від вибірки реалізується лише через y. А згідно з тим, що було сказано про умовне математичне сподівання на початку цього підрозділу, то це і означає, що $\mathbf{M}\{\theta(\vec{x})/S(\vec{x})\}$ є функцією S і перший пункт доведено.

[1] Блеквел (*David Harold Blackwell*) (нар. 24.04. 1919, США, Ілінойс)– статистик. Перший (і поки єдиний) афроамериканець, що є членом Національної Академії Наук Сполучених Штатів (National Academy of Sciences), Президентом ASS (American Statistical Society,) і віце-президентом AMS (American Mathematics Society).

Пункт б) одразу випливає з першого пункту теореми 1, оскільки

$$\mathbf{M}\theta_S^* = \mathbf{M}\mathbf{M}\{\theta(\vec{x})/S(\vec{x}) = y\} = \mathbf{M}\theta(\vec{x}) = b(\theta) + \theta.$$

Для доведення третього пункту маємо

$$\begin{aligned} &\mathbf{M}(\theta(\vec{x}) - \theta)^2 = \mathbf{M}(\theta(\vec{x}) - \theta_S^* + \theta_S^* - \theta)^2 = \\ &= \mathbf{M}(\theta(\vec{x}) - \theta_S^*)^2 + \mathbf{M}(\theta_S^* - \theta)^2 + 2\mathbf{M}(\theta(\vec{x}) - \theta_S^*)(\theta_S^* - \theta). \end{aligned} \tag{7}$$

Величина θ_S^*, як це було доведено, є функцією S, θ є числом і тому з пункту б) теореми 1 маємо

$$\begin{aligned} &\mathbf{M}(\theta(\vec{x}) - \theta_S^*)(\theta_S^* - \theta) = \mathbf{M}\{\mathbf{M}(\theta(\vec{x}) - \theta_S^*)(\theta_S^* - \theta)/S\} = \\ &= \mathbf{M}\{(\theta_S^* - \theta)\mathbf{M}\{(\theta(\vec{x}) - \theta_S^*)/S\}\} = \mathbf{M}(\theta_S^* - \theta)(\mathbf{M}\{\theta(\vec{x})/S\} - \theta_S^*) = 0 \end{aligned}$$

і рівність (7) може бути переписана так

$$\mathbf{M}(\theta(\vec{x}) - \theta)^2 = \mathbf{M}(\theta(\vec{x}) - \theta_S^*)^2 + \mathbf{M}(\theta_S^* - \theta)^2 \tag{8}$$

звідки випливає нерівність в пункті в) нашої теореми. Очевидно, що в цій нерівності рівність буде тоді і лише тоді, коли в (8) будемо мати $\mathbf{M}(\theta(\vec{x}) - \theta_S^*)^2 = 0$ або

$$\int_{\mathbf{R^n}} (\theta(\vec{z}) - \theta^*_{S(\vec{z})})^2 L(\vec{z}, \theta) d\vec{z} = 0,$$

що, в свою чергу, можливо тоді і лише тоді, коли має місце (6).

Приклад 1. *Нехай генеральна сукупність має розподіл Пуассона з параметром λ. В якості оцінки цього параметра виберемо $\lambda(\vec{k}) = k_1$.*

Оскільки $\mathbf{M}\lambda k_1 = \lambda$, то ця оцінка є незміщеною, але, очевидно, вона не є консистентною (вона навіть не залежить від n). Нехай $S = \sum_{i=1}^n k_i$ і, як вже знаємо з прикладу **??**, с. **??** ця статистика є мінімальною і достатньою. Порахуємо

$$\mathbf{P}(k_1 = m/S = k) = \mathbf{P}(k_1 = m/\sum_{i=1}^n k_i = k) = \frac{\mathbf{P}(k_1 = m, \sum_{i=1}^n k_i = k)}{\mathbf{P}(\sum_{i=1}^n k_i = k)}$$

для $m \le k$ (для $m > k$ ця ймовірність, очевидно, дорівнює нулю). Маємо

$$\begin{aligned} &\mathbf{P}(k_1 = m, \sum_{i=1}^n k_i = k) = \mathbf{P}(k_1 = m, \sum_{i=2}^n k_i = k - m) = \mathbf{P}(k_1 = m)\mathbf{P}(\sum_{i=2}^n k_i = k - m) = \\ &= e^{-\lambda}\frac{\lambda^m}{m!} e^{-\lambda(n-1)}\frac{(\lambda(n-1))^{k-m}}{(k-m)!} = e^{-\lambda n}\frac{\lambda^m(\lambda(n-1))^{k-m}}{m!(k-m)!}. \end{aligned}$$

Тепер

$$\mathbf{P}(k_1 = m/S = k) = \frac{e^{-\lambda n}\frac{\lambda^m(\lambda(n-1))^{k-m}}{m!(k-m)!}}{e^{-\lambda n}\frac{(\lambda n)^k}{k!}} = C_k^m \ \frac{1}{n}\ ^m \ 1 - \frac{1}{n}\ ^{m-k}.$$

В правій частині цієї рівності стоїть біноміальний розподіл з параметрами $k, 1/n$ і, отже, $\mathbf{M}\{k_1/S(\vec{k}) = k\} = k/n$. Тому $\mathbf{M}\{k_1/S\} = S/n = k(n)$ і ми отримали оцінку, яка є незміщеною і сильно консистентною.

ЗАВДАННЯ

1. Знайти достатні статистики для параметрів сукупностей з наступними розподілами(в цьому завданні всі літери α, β, p трактуємо як невідомі параметри)

   а) біноміального $C_n^k p^k (1-p)^{n-k}$;

   б) рівномірного на відрізку $[\alpha; \beta]$, $0 \le \alpha < \beta$;

   в) Ерланга зі щільністю $\frac{\alpha^3}{2} x^2 e^{-\alpha x}$ $x \ge 0$;

   г) бета зі щільністю $\alpha x^{\alpha-1}$, $x \in [0, 1]$;

   д) Парето зі щільністю $\frac{2^\alpha \alpha}{x^{\alpha+1}}$, $x \ge 2$.

2. Знайти достатню статистику для параметра α, якщо генеральна популяція має щільність
$$f(x, \alpha) = \begin{matrix} \frac{2}{\alpha}(\alpha - x), & 0 \le x \le \alpha, \\ 0, & \text{для інших.} \end{matrix}$$

3. Показати, що для генеральної популяції з рівноміним на відрізку $[0; \theta]$ розподілом статистика $x_{(n)}$ є достатня для параметра θ.

4. Показати, що для генеральної популяції з рівноміним на відрізку $[\theta; \theta + 1]$ розподілом статистика $(x_{(1)}, x_{(n)})$ є достатня для параметра θ.

5. Довести, що для генеральної популяції з розподілом $2xe^{-(x/\theta)^2}/\theta^2$, $x > 0$ статистика $\sum_{i=1}^n x_i^2$ є достатня для параметра θ.

Лекція 12. Загальні питання теорії перевірки гіпотез

12.1 Гіпотези та статистичні випробування

Як уже було сказано, другим основним розділом математичної статистики є теорія перевірки статистичних гіпотез.

Визначення 1. Під *статистичною гіпотезою* розуміється довільне припущення щодо розподілу генеральної сукупності. Правдивість такого припущення перевіряється на підставі вибірки з цієї сукупності.

Щодо характеристики ξ звичайно маємо два різновиди інформації:
а) вибірку $x_1, \ldots, x_n$, тобто маємо n незалежних значень цієї характеристики;
б) певну інформацію, яка випливає з фізичної природи характеристики ξ.

Прикладами інформації типу б) можуть бути, наприклад, такі. Знаємо, що характеристика ξ є вага. Тоді очевидно $\xi \geq 0$ і, таким чином, функція розподілу такої характеристики $F(x)$, дорівнює нулю при $x < 0$. Тому, наприклад, якщо наша гіпотеза стосується вигляду цієї функції, то нормальний розподіл може бути відкинутий від самого початку як такий, що не може бути розподілом характеристики ξ.[1] Інший приклад. Знаємо, що характеристика ξ є дискретною. Тоді природно в якості $F(x)$ досліджувати лише розподіли дискретного типу.

Отже, інформації типу б) можуть *a priori* окреслити множину гіпотез $D = \{H_i\}$, які можуть бути можливими припущеннями щодо характеристики ξ. В математичній статистиці множину D називають *множиною допустимих гіпотез.* В дальшій частині цього посібника завжди будемо розглядати лише ті гіпотези, які належать до множини допустимих гіпотез, і будемо їх називати *допустимими гіпотезами* або просто *гіпотезами.*

Статистична гіпотеза часто пов'язується з числовими характеристиками (або, як говоримо, параметрами) генеральної сукупності) ξ. Така інтерпретація статистичної гіпотези приводить до наступної стандартної схеми, яка найчастіше використовується, як вихідний пункт теорії перевірки статистичних гіпотез.

Нехай генеральна сукупність ξ означена на ймовірносному просторі $(\Omega, \mathcal{B}, \mathbf{P}_\theta)$, де символ $\theta \in \Theta$ і трактується як параметр. Іншими словами, якщо маємо вибірку $x_1, ..., x_n$, то ця вибірка виникла в результаті вимірювання характеристики ξ з розподілом $F(x, \theta) = \mathbf{P}(\xi < x)$ з деяким конкретним (але нам не відомим) параметром

[1] Але варто зауважити, що в практиці ми часто використовуємо нормальний розподіл для моделювання додатних характеристик, але робимо це в основному з двох причин: по-перше, саме для таких розподілів отримано найбільш глибокі теоретичні результати, а, по-друге, нормальний розподіл часто добре наближає реальний розподіл навіть додатніх випадкових величин.

$\theta \in \Theta$. Будемо вважати, що параметр θ однозначно визначає ймовірність $\mathbf{P}_\theta$. Як ми вже знаємо, в цьому випадку зручно говорити, що розподіл генеральної сукупності походить з параметричної множини $\{F(x,\theta),\ \theta \in \Theta\}$.

Нехай гіпотези $H_0, H_1, H_2, ..., H_k$ означають, що $\theta \in \Theta_i$, $i = 0, 1, ..., r$, $\cup_{i=0}^{r}\Theta_i = \Theta$, $\Theta_i \cap \Theta_j = \emptyset$, $i \neq j$.

Приклад 1. *Гіпотеза H_0 означає, що генеральна сукупність має показниковий розподіл $1 - e^{-\theta x}$, $x > 0$, і $\theta \in [1; 2]$, а гіпотеза H_1 означає, що генеральна сукупність має нормальний розподіл $\mathbb{N}(\theta, 1)$, і $\theta < 0$.*

В цій ситуації $\Theta = (-\infty, 0) \cup [1; 2]$, $\Theta_0 = [1; 2]$, $\Theta_1 = (-\infty; 0)$.

Тоді прийняття, наприклад, гіпотези H_1 означало б, що випадкова змінна ξ має нормальний розподіл з від'ємним математичним сподіванням і одиничною дисперсією.

Гіпотези H_i, описані за такою схемою, звичайно називають *гіпотезами параметричними.* Іншими словами, якщо гіпотеза стосується виключно значення параметра (чи параметрів) розподілу генеральної сукупності, то вона є параметричною гіпотезою.

Кожну гіпотезу, яка не є параметричною, називаємо *гіпотезою непараметричною.* .

Ось приклади гіпотез.

1. Відомо, що характеристика має розподіл $\mathbb{N}(m, 4)$. Висуваємо гіпотезу, що $m = 5$. Ця гіпотеза є параметричною.
2. Відомо, що характеристика має розподіл $\mathbb{N}(m, 4)$. Висуваємо гіпотезу, що $1 \leq m \leq 5$. Ця гіпотеза також є параметричною.
3. Нехай маємо дві генеральні сукупності ξ, η. Тоді гіпотеза про незалежність ξ, η є непараметричною.

Визначення 2. Гіпотеза H_i називається *простою*, коли $\Theta_i = \{\theta_i\}$.

Іншими словами, якщо параметрична гіпотеза точно визначає значення невідомих параметрів розподілу досліджуваної характеристики, то її називають *простою*, а інакше маємо *складну* гіпотезу.

Так гіпотеза з пункту 1 є простою, а гіпотеза з пункту 2 є складною.

Задача перевірки статистичних гіпотез виглядає наступним чином. Нам потрібно запропонувати алгоритм, який на підставі вибірки $x_1, ..., x_2$ дасть можливість прийняти рішення, яка з гіпотез H_i, $i = 0, ..., r$ є вірною і прийняти її, натомість інші гіпотези відкинути. Тут доречно сказати про певну особливість, яка практично завжди присутня при перевірці гіпотез. Найкраще це проілюструвати на такому прикладі. Нехай гіпотеза H полягає в тому, що характеристика ξ має показниковий розподіл з параметром $\lambda \in [1; 2]$. Тоді, якщо маємо, наприклад, вибірку $-2; 3, 5; -4$, то гіпотеза H na 100% може бути відкинута, оскільки генеральна сукупність ξ у випадку правдивості гіпотези H не може набувати від'ємних значень. А якщо маємо вибірку $2; 3, 5; 4$, то не дивлячись на те, що всі значення вибірки є додатними, цей факт ще не означає, що гіпотеза H є вірною. Так, якщо справжній розподіл генеральної сукупності є стандартним нормальним $\mathbb{N}(0, 1)$, то ймовірність того, що чотири значення цієї характеристики будуть додатними є $1/2^4$, а отже вибірка $2; 3, 5; 4$ може походити з розподілу $\mathbb{N}(0, 1)$. Тому в такій ситуації ми не можемо сказати на 100% чи правильна гіпотеза H чи ні. Така ситуація в статистиці є звичайною і це пов'язано з тим, що скінчена вибірка ніколи не містить всієї інформації про генеральну сукупність. (За винятком деяких "екзотичних" прикладів.) Тому при перевірці гіпотез ми говоримо не "гіпотеза є вірною", а "немає підстав для відхилення гіпотези". Таке формулювання не виключає ситуації, що наше рішення про прийняття (чи неприйняття гіпотези) було помилковим. Нижче про це буде сказано докладніше.

Оскільки дії з перевірки гіпотез, про які говорилось вище, своїм вихідним пунктом завжди мають вибірку, то нам потрібно знайти функцію $\delta(\vec{v})$, яка означена для $\vec{v} \in \mathbf{R}^n$, і набуває значень в $\{0, 1, ..., k\}$, тобто $\delta : \mathbf{R}^n \to \{0, 1, ..., k\}$, і якщо для конкретної вибірки $x_1, ..., x_n$ маємо $\delta(\vec{x}) = i$, то приймаємо гіпотезу H_i. Функція $\delta(\cdot)$ називається *статистичним тестом*. Формалізуємо це поняття у вигляді такого означення.

Визначення 3. *Статистичним тестом* називаємо довільне відображення $\delta : \mathbf{R}^n \to \{0, 1, ..., k\}$ і якщо $\delta(\vec{x}) = i$ для вибірки $x_1, ..., x_n$, то приймаємо гіпотезу H_i.

Іноді таке означення статистичного тесту є недостатнім. Так, наприклад, при порівнянні тестів бажано мати як можна більше різних тестів з тим, щоб мати більше можливостей знайти найкращий тест. Наступне означення статистичного тесту узагальнює попереднє.

Означимо наступну підмножину w $\mathbf{R}^n$.

$$\mathcal{R}_+^k = \{\vec{y} : \vec{y} = \{\pi_0, \pi_1, ..., \pi_k\}, \quad \pi_i \geq 0, \ i = 1, ..., k, \quad \sum_{i=0}^{k} \pi_i = 1.\}$$

Очевидно, що числа $\pi_i, \ i = 0, 1, ..., k$, які присутні в означенні множини $\mathcal{R}_+^k$, можемо інтерпретувати як ймовірності.

Визначення 4. *Рандомізованим* статистичним тестом назвемо довільне відображення $\delta : \mathbf{R}^n \to \mathcal{R}_+^k$ і якщо $\delta(\vec{x}) = \{\delta_0, \delta_1, ..., \delta_k\}$ для вибірки $x_1, ..., x_n$, то ймовірність прийняття гіпотези H_i є δ_i.

Дії в разі, коли маємо рандомізований тест δ виглядають так. Для конкретної вибірки $x_1, ..., x_n$ вираховуємо $\delta(\vec{x}) = \{\delta_0, \delta_1, ..., \delta_k\}$. Далі робимо стохастичний експеримент S, результатом якого будуть числа $0, 1, .., k$ тобто, елементарна подія $\omega_i = i$. Припускаємо, що ймовірність i-го результату є δ_i. Такий експеримент легко сконструювати. Тоді, якщо в експерименті S отриманий результат i, то приймаємо гіпотезу H_i.

Якщо в означенні 4 всі δ_i є нулями за винятком одного (яке тоді є рівним 1), то фактично маємо статистичний тест з означення 3, а тому цей тест є частковим випадком рандомізованого тесту.

Тести з означення 3 будемо називати *нерандомізованими* тестами. В подальшій частині цього посібника ми, вживаючи слово "тест", не будемо акцентувати увагу на те рандомізований він чи ні. Як правило, це буде зрозуміло з контексту. Зазначимо лише, що в розважаннях теоретичних це, як правило будуть рандомізовані тест, а у застосуваннях нерандомізовані.

Якщо відносно генеральної сукупності ξ маємо лише дві гіпотези $H_0, \ H_1$, то одна з них називається *головною гіпотезою* і зазвичай позначаємо її H_0, а друга гіпотеза називається *альтернативна гіпотеза* або *конкурентна гіпотеза* і позначаємо її H_1.

12.2 Порівняння тестів

З означення статистичного тесту, поданого в попередньому підрозділі, випливає, що тестів для перевірки гіпотез H_i, $i = 0, 1..., k$ може бути багато, а тому природно постає питання: як знайти (наскільки це можливо) в якомусь розумінні "найкращий" тест? Тому маємо запропонувати процедуру порівняння тестів. Одним з можливих (і, на перший погляд, природнім) способом вирішення цього питання є такий. З кожним статистичним тестом можемо пов'язати множину чисел

$$\alpha_i(\delta, \theta) = \mathbf{P}(\delta(\vec{x}) \neq i/\theta \in \Theta_i), \quad i = 0, 1..., k.$$

Число $\alpha_i(\delta, \theta)$ можна інтерпретувати як ймовірність неприйняття гіпотези H_i за умови, що ця гіпотеза була вірною і його називаємо *похибка i-го роду*. При такому означенні похибок є природним назвати тест $\delta_1(\cdot)$ кращим від тесту $\delta_2(\cdot)$ якщо

$$\alpha_i(\delta_1, \theta) \leq \alpha_i(\delta_2, \theta), \qquad \theta \in \Theta, \quad i = 0, 1..., k, \tag{1}$$

і принаймні для одного θ ця нерівність є строгою. Виявляється, що такий підхід до порівняння тестів не завжди може бути реалізований. Іншими словами, множина Θ може містити занадто багато значень параметра θ, аби для всіх θ вимагати виконання (1). Крім того, вся проблема порівняння тестів полягає не лише в тому, що множина θ може бути "великою", але сама процедура порівняння тестів залежить від кількості гіпотез H_i. Задача порівняння тестів стає простішою у випадку двох гіпотез. А тому в цьому посібнику ми обмежимося лише цим випадком.

Випадок більшої кількості гіпотез зацікавлений читач може знайти, наприклад, в [1], [14].

12.2.1 Порівняння тестів для двох простих гіпотез

В попередньому підрозділі ми означили нерандомізовані та рандомізовані тести і перші є підмножиною других. Введення рандомізованих тестів є логічним з погляду того, що якщо ми хочемо відшукати "найкращий" тест, то чим більша множина самих тестів тим краще, оскільки маємо з чого вибирати. З іншої сторони, як ми побачимо пізніше, "найкращі" тести, як правило є нерандомізованими. Тому може не варто розглядати рандомізовані тести взагалі? Але це не так. Справа в тому, що процедура порівняння тестів полягає на тому, що потрібно знайти деяку числову характеристику, свого роду "функцію якості", з допомогою якої можна порівнювати тести, а отже і вибрати "найкращий" з них. Як побачимо далі це краще робити використовуючи рандомізовані тести. Є ще інший бік цього питання. В цьому посібнику ми будемо використовувати лише один підхід до порівняння тестів, який найчастіше застосовується в практиці: це знаходження найпотужніших тестів. Але існують і інші підходи, наприклад, баєсівський та мінімаксний. В цих підходах рандомізовані тести теж грають важливу роль.

Припустимо тепер, що маємо справу з двома простими гіпотезами: головною гіпотезою $H_0 = \{\theta_0\}$, і гіпотезою альтернативною $H_1 = \{\theta_1\}$.

Це означає, що стосовно генеральної сукупності ξ висунемо припущення, що вона має розподіл $F_0(x)$ (гіпотеза H_0) або розподіл $F_1(x)$ (гіпотеза H_1).

Інакше кажучи, відносно генеральної сукупності висуваємо

головну гіпотезу $H_0 : \xi$ означена на просторі $(\Omega, \mathcal{B}, \mathbf{P}_0)$ і тоді
функція розподілу характеристики ξ є $F_0(x) = \mathbf{P}_0(\xi < x)$;

гіпотезу альтернативну $H_1 : \xi$ означена на просторі $(\Omega, \mathcal{B}, \mathbf{P}_1)$ і
тоді функція розподілу характеристики ξ є $F_1(x) = \mathbf{P}_1(\xi < x)$.

На підставі вибірки $x_1, ..., x_n$ потрібно перевірити, яка з цих ситуацій має місце. Іншими словами, потрібно запропонувати статистичний тест для перевірки цих гіпотез. Розпочнемо з нерандомізованих тестів.

Отже нехай тест δ є нерандомізованим, тобто коли $\delta(\vec{x}) = 0$, то приймаємо H_0, а коли $\delta(\vec{x}) = 1$, то приймаємо H_1.

Можемо діяти ще й так. Нехай

$$\mathbf{M_0}(n) = \{\mathbf{y} \in \mathbf{R}^n : \delta(\mathbf{y}) = 0\}, \quad \mathbf{M_1}(n) = \{\mathbf{y} \in \mathbf{R}^n : \delta(\mathbf{y}) = 1\}.$$

Ясно, що $\mathbf{R}^n = \mathbf{M_0}(n) \cup \mathbf{M_1}(n)$, $\mathbf{M_0}(n) \cap \mathbf{M_1}(n) = \emptyset$. Тепер, якщо після вимірювання характеристики ξ вектор значень вибірки $\vec{x} = (x_1, ..., x_n)$ потрапив до $\mathbf{M_0}(n)$, то приймаємо гіпотезу H_0, а якщо до $\mathbf{M_1}(n)$, то приймаємо гіпотезу H_1. В цій ситуації

$$\alpha(\delta) = \mathbf{P}_{\theta_0}(\delta(\vec{x}) \neq 0) = \mathbf{P}(\delta(\vec{x}) \neq 0/\theta = \theta_0)$$

називаємо *ймовірністю помилки першого роду,* а

$$\beta(\delta) = \mathbf{P}_{\theta_1}(\delta(\vec{x}) \neq 1) = \mathbf{P}(\delta(\vec{x}) \neq 1/\theta = \theta_1)$$

називаємо *ймовірністю помилки другого роду.*

Число $\alpha(\delta)$ ще називають *рівнем значимості тесту* (або просто *рівнем тесту* , а число $1 - \beta(\delta)$ – *потужністю тесту* δ.

Помилка першого роду означає, що ми відкинули вірну гіпотезу на користь гіпотези фальшивої, а помилка другого роду означає, що в якості правильної була прийнята гіпотеза, яка в дійсності є фальшивою. Є природним, що хочемо знайти такий тест, аби ймовірності помилок обох родів були мінімальними. Інакше кажучи, якщо $\mathcal{K}_{nr}$ позначає множину всіх можливих нерандомізованих тестів, то хочемо знайти тест $\delta^* \in \mathcal{K}_{nr}$ такий, щоб

$$\alpha(\delta^*) = \min_{\delta \in \mathcal{K}_{nr}} \alpha(\delta), \qquad \beta(\delta^*) = \min_{\delta \in \mathcal{K}_{nr}} \beta(\delta). \tag{2}$$

Якщо такий тест знайдено, то природно його потрібно називати "найкращим". (Було б ідеально, якби тест δ^* був таким, що $\alpha(\delta^*) = \beta(\delta^*) = 0$. Нижче буде наведено приклад для якого така можливість реалізується, але практичного значення такі ситуації не мають.) Існує три підходи до відшукання найкращих тестів. Це знаходження найпотужніших тестів, баєсівський підхід, мінімаксний підхід. Ми розглянемо лише перший з них, який найчастіше застосовується в практиці. Багато фактів відносно баєсівського та мінімаксного підходів можна знайти в [1], [14].

Зазначимо відразу, що проблема відшукання найкращих тесті подібна до проблеми відшукання найкращих оцінок. Як для конкретного параметра ми не завжди можемо знайти найкращу оцінку, так і для перевірки конкретної гіпотези ми далеко не завжди можемо знайти найкращий тест. Нижче ми побачимо, що в деяких випадках можна запропонувати задовільне вирішення задачі про відшукання найкращого тесту, але в більшості випадків приходиться задовільнятися тестами, які, скоріше всього, не є найкращими. Подібна ситуація була і при оцінці параметрів.

Означимо підмножину множини нерандомізованих тестів

$$\mathcal{K}_{nr}(\alpha) = \{\delta \in \mathcal{K}_{nr} : \alpha(\delta) \leq \alpha\}. \tag{3}$$

Тобто до $\mathcal{K}_{nr}(\alpha)$ належать всі нерандомізовані тести, для яких похибка першого роду не більша від α.

Визначення 5. Тест δ_α^* називається *найпотужнішим* тестом рівня α, якщо

$$\beta(\delta_\alpha^*) \leq \beta(\delta)$$

для довільного $\delta \in \mathcal{K}_{nr}(\alpha)$.

Інакше кажучи, потужність тесту δ_α^*, $1 - \beta(\delta_\alpha^*)$ буде максимальною серед всіх тестів з класу $\mathcal{K}_{nr}(\alpha)$.

Розглянемо тепер процедуру порівняння рандомізованих тестів і нехай $\mathcal{K}_r$ позначає множину таких тестів.

Для випадку двох гіпотез означення 4 може бути переписане так.

Визначення 6. *Статистичним тестом* називаємо відображення $\delta : \mathbf{R}^n \to [0; 1]$ при цьому $\delta(\mathbf{x})$ є ймовірність прийняття гіпотези H_1, а $1 - \delta(\mathbf{x})$ є ймовірність прийняття гіпотези H_0.

Згідно з цим означенням перевірка зазначених гіпотез для даного статистичного тесту $\delta(\cdot)$ відбувається наступним чином. Робимо вимірювання характеристики ξ в результаті чого маємо вибірку $\mathbf{x} = (x_1, ..., x_n)$. А далі ставимо експеримент типу 0-1 з ймовірністю успіху $\delta(\mathbf{x})$. Якщо в цьому експерименті маємо успіх - то приймаємо гіпотезу H_1, а інакше - гіпотезу H_0.

Якщо $\delta = 0$ або 1, то такий тест буде нерандомізованим.

(Очевидно, ситуація, коли $\delta(\mathbf{x}) \equiv 0$ (або $\equiv 1$) теж можлива, але тоді, фактично, наше завдання втрачає сенс, оскільки завжди приймаємо гіпотезу H_0 (або H_1)).

Згідно з нашою інтерпретацією вибірки, вона може розглядатися як n-вимірний вектор незалежних випадкових змінних $\mathbf{x} = (x_1, ..., x_n)$ з щільністю

$$L_0(\vec{x}) = \prod_{i=1}^{n} f_0(x_i), \qquad L_1(\vec{x}) = \prod_{i=1}^{n} f_1(x_i) \tag{4}$$

за умови, що має місце гіпотеза H_0 і H_1 відповідно. Для функції $g(\mathbf{x})$, $\mathbf{x} \in \mathbf{R}^n$ введемо позначення

$$\mathbf{M}_0 g(\mathbf{x}) = \mathbf{M}\{g(\mathbf{x})/H_0\} = \int\limits_{\mathbf{R}^n} g(\mathbf{y}) L_0(\mathbf{y}) \mathbf{dy},$$

$$\mathbf{M}_1 g(\mathbf{x}) = \mathbf{M}\{g(\mathbf{x})/H_1\} = \int\limits_{\mathbf{R}^n} g(\mathbf{y}) L_1(\mathbf{y}) \mathbf{dy}.$$

за умови, що інтеграли в правих частинах цих рівностей є скінченими. Так само запис $\mathbf{P}_0(A)$ означає ймовірність події A за умови, що має місце гіпотеза H_0, а $\mathbf{P}_1(A)$–за умови, що має місце гіпотеза H_1. Очевидно

$$\mathbf{M}_0 I_A(\omega) = \mathbf{P}_0(A) \qquad \mathbf{M}_1 I_A(\omega) = \mathbf{P}_1(A). \tag{5}$$

Тепер для похибки першого роду $\alpha(\delta)$ і для похибки другого роду $\beta(\delta)$ рандомізованого тесту δ маємо

$$\alpha(\delta) = \mathbf{M}_0 \delta(\mathbf{x}) = \int\limits_{\mathbf{R}^n} \delta(\mathbf{y}) L_0(\mathbf{y}) \mathbf{d}y, \tag{6}$$

$$\beta(\delta) = \mathbf{M}_1 \;\; 1 - \delta(\mathbf{x}) \;\; = 1 - \int\limits_{\mathbf{R}^n} \delta(\mathbf{y}) L_1(\mathbf{y}) \mathbf{d}y. \tag{7}$$

Як уже було сказано похибку першого роду називаємо ще рівнем значимості тесту δ, а число $1-\beta(\delta)=\mathbf{M}_1\delta(\mathbf{x})$ називається потужністю тесту.

Означення найпотужнішого рандомізованого тесту рівня α залишається практично таким самим, лише тепер для похибок першого і другого роду потрібно застосовувати формули (6), (7).

Отже, якщо означити підмножину множини рандомізованих тестів

$$\mathcal{K}_r(\alpha) = \{\delta \in \mathcal{K}_r : \alpha(\delta) \le \alpha\}, \tag{8}$$

то тест δ^*_α (взагалі кажучи, рандомізований) називається найпотужнішим тестом рівня α, якщо

$$\beta(\delta^*_\alpha) \le \beta(\delta)$$

для довільного $\delta \in \mathcal{K}_r(\alpha)$.

Отже, як у випадку рандомізованих так і нерандомізованих тестів, з кожним тестом δ можемо пов'язати точку $(\alpha(\delta), \beta(\beta)) \in [0;1]\times[0;1] \in \mathbf{R^2}$. Цікавіше є зворотне питання: чи кожній точці $(\alpha,\beta),\ 0 \le \alpha \le 1, 0 \le \beta \le 1$ відповідає деякий тест δ такий, що $\alpha(\delta) = \alpha,\ \beta(\delta) = \beta$? Щоб відповісти на це питання, потрібно дослідити структуру наступної множини в просторі $\mathbf{R}^2$

$$\mathcal{R} = \ \alpha(\delta);\beta(\delta)\ , \qquad \delta \in \mathcal{K}_r, \tag{9}$$

де, нагадуємо, $\mathcal{K}_r$ є всі рандомізовані тести. З цього означення випливає, що точка $(\alpha,\beta) \in \mathcal{R}$ тоді і лише тоді коли існує тест $\delta \in \mathcal{K}_r$ такий, що $\alpha(\delta) = \alpha,\ \beta(\delta) = \beta$. Наступна лема описує найпростіші її властивості.

Лема 1.

а) *Множина $\mathcal{R}$ є опуклою;*

б) $(0;1) \in \mathcal{R},\ (1;0) \in \mathcal{R}$*;*

в) *Якщо* $(\alpha,\beta) \in \mathcal{R}$, *то* $(1-\alpha, 1-\beta) \in \mathcal{R}$.

Д о в е д е н н я.

а) Нехай $(\alpha(\delta_1);\beta(\delta_1) \in \mathcal{R}$ і $(\alpha(\delta_2);\beta(\delta_2) \in \mathcal{R}$. Доведемо, що

$$(\lambda\alpha(\delta_1) + (1-\lambda)\alpha(\delta_2); \lambda\beta(\delta_1) + (1-\lambda)\beta(\delta_2)) \in \mathcal{R}$$

для всіх $0 \le \lambda \le 1$

Означимо тест $\delta = \lambda\delta_1 + (1-\lambda)\delta_2$. Зрозуміло, що $\delta \in [0;1]$ для всіх $0 \le \lambda \le 1$. Порахуємо похибки першого і другого роду для цього тесту. Маємо

$$\begin{aligned}
\alpha(\delta) &= \mathbf{M}_0\delta(\vec{x}) = \lambda\mathbf{M}_0\delta_1(\vec{x}) + (1-\lambda)\mathbf{M}_0\delta_2(\vec{x}) = \lambda\alpha(\delta_1) + (1-\lambda)\alpha(\delta_2),\\
\beta(\delta) &= \mathbf{M}_1\ 1-\delta(\vec{x})\ = \lambda\mathbf{M}_1\ 1-\delta_1(\vec{x})\ + (1-\lambda)\mathbf{M}_1\ 1-\delta_2(\vec{x})\ =\\
&= \lambda\beta(\delta_1) + (1-\lambda)\beta(\delta_2)
\end{aligned}$$

що завершує доведення.

б) Неважко переконатися, що для тесту $\delta(\vec{x}) \equiv 0$, $\alpha(\delta) = 0, \beta(\delta) = 1$, а для тесту $\delta(\vec{x}) \equiv 1$, $\alpha(\delta) = 1, \beta(\delta) = 0$.

в) Доведення цієї властивості є очевидним, оскільки якщо для тесту δ, $\alpha(\delta) = \alpha,\ \beta(\delta) = \beta$, то для тесту $\hat{\delta} = 1-\delta$ маємо $\alpha(\hat{\delta}) = 1-\alpha,\ \beta(\hat{\delta}) = 1-\beta$.

Рис. 12.1:

З цієї леми випливає, що головна діагональ квадрату $[0;1]\times[0;1]$ належить до $\mathcal{R}$. Повністю вигляд множини $\mathcal{R}$ досліджено в [14] і він є наступним

Тут використані позначення

$$\overline{\alpha} = \mathbf{M}_0 I\{L_1(\vec{x}) > 0\}, \qquad \overline{\beta} = \mathbf{M}_1 I\{L_0(\mathbf{v}) > 0\}. \tag{10}$$

а $L_0(\vec{x}),\ L_1(\mathbf{v})$ з (4).

З цього рисунку видно, що відрізок $[A;B]$ відповідає тестам, для яких похибка першого роду є α^*. Зрозуміло, що тест, який відповідає точці A, буде найпотужнішим (тобто найкращим) зі всіх тестів, для яких похибка першого роду є α^*. З цього ж рисунку видно, що цей тест буде найкращим і в класі $\mathcal{K}_r$. Зауважимо ще, що, як це буде виникати з результатів наступного підрозділу, тести, яким відповідають точки на границях множини $\mathcal{R}$, є нерандомізованими.

Виникає питання, а чи може статися так, що точка $(0;0)$ належить до множини $\mathcal{R}$. Очевидно, це буде означати існування тесту для якого похибки обох родів є нулями, тобто по вибірці ми безпомилково вирішуємо яка гіпотеза є правдивою. Виявляється, що така ситуація можлива (очевидно, що тоді множина $\mathcal{R}$ буде квадратом $[0;1]\times[0;1]$).

Приклад 1. *Нехай відомо, що розподіл генеральної сукупності ξ є рівномірним на відрізку $[0;1]$ (гіпотеза H_0) або на відрізку $[2;3]$ (гіпотеза H_1).*

Очевидно, в цій ситуації нам вистачить одного поміру для ξ щоб точно сказати, яка з гіпотез H_0, H_1 має місце. Так якщо, то $x_1 = 1/2$, то справедлива гіпотеза H_0, а якщо $x_1 = 7/3$, то справедлива гіпотеза H_1. Отже в цьому випадку по вибірці ми можемо точно сказати, яка з гіпотез має місце.

12.3 Порівняння тестів для складних гіпотез

Нехай тепер генеральна сукупність ξ означена на ймовірносному просторі $(\Omega, \mathcal{B}, \mathbf{P}_\theta)$, де $\theta \in \Theta$ і трактується як параметр. Іншими словами, якщо маємо вибірку $x_1, ..., x_n$, то ця вибірка виникла в результаті вимірювання характеристики ξ з розподілом $F_\theta(x) = \mathbf{P}(\xi < x)$ з якимось конкретним (нам невідомим!) $\theta \in \Theta$. Головна гіпотеза H_0 полягає в тому, що $\theta \in \Theta_0$ а гіпотеза альтернативна H_1 означає, що $\theta \in \Theta_1$, де $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_1 \cup \Theta_1 = \Theta$. Для простих гіпотез $\Theta = \{\theta_0, \theta_1\}$. Тому тепер припускаємо, що принаймні одна з множин Θ_0, Θ_1 не є одноелементною, тобто принаймні одна з гіпотез є складною.

Нехай для перевірки гіпотези H_0 відносно альтернативи H_1 маємо тест (взагалі кажучи, рандомізований) $\delta(\vec{x}) : \mathbf{R}^n \to [0, 1]$ і для вибірки $x_1, ..., x_n$, ймовірність прийняття гіпотези H_1 є $\delta(\vec{x})$, а ймовірність прийняття гіпотези H_0 є $1 - \delta(\vec{x})$.

Як і раніше, рандомізовані (нерандомізовані) тести будемо в позначати $\mathcal{K}_r$, ($\mathcal{K}_{nr}$).

Припускаємо, що генеральна сукупність має розподіл зі щільністю $f(x, \theta)$, інакше кажучи, відомо, що розподіл генеральної сукупності належить до родини $\{F(x, \theta) = \mathbf{P}(\xi < x),\ \theta \in \Theta\}$ і кожна функція розподілу $F(x, \theta)$ має щільність $f(x, \theta)$. Число

$$\alpha(\delta) = \sup_{\theta\in\Theta_0} \mathbf{M}\delta(\vec{x}) = \sup_{\theta\in\Theta_0} \int\limits_{\mathbf{R^n}} \delta(\vec{x}) L(\mathbf{v}, \theta)\mathbf{dv}, \tag{11}$$

називається *розміром тесту* або *ймовірністю помилки першого роду*, , а число

$$\gamma(\delta, \theta) = \mathbf{M}\delta(\vec{x})) = \int\limits_{\mathbf{R^n}} \delta(\mathbf{v}) f(\mathbf{v}, \theta)\mathbf{dv}, \quad \theta \in \Theta_1$$

називається *потужністю* тесту δ.

Число $1 - \alpha(\delta)$ називається *рівнем значимості тесту* δ, а число $1 - \gamma(\delta, \theta) = \mathbf{M}(1 - \delta(\vec{x})) = \beta(\delta, \theta)$- *ймовірністю помилки другого роду*. Для нерандомізованого тесту маємо

$$\alpha(\delta) = \sup_{\theta\in\Theta_0} \mathbf{P}(\delta(\vec{x}) = 1) = \sup_{\theta\in\Theta_0} \int\limits_{\mathbf{R^n}} I\{\delta(\mathbf{v}) = 1\} L(\mathbf{v}, \theta)\mathbf{dv},$$

$$\gamma(\delta, \theta) = \mathbf{P}(\delta(\vec{x}) = 0) = \int\limits_{\mathbf{R^n}} I\{\delta(\mathbf{v}) = 0\} L(\mathbf{v}, \theta)\mathbf{dv}, \quad \theta \in \Theta_1.$$

З метою полегшення конструкції оптимальних тестів зазвичай обмежуються класом тестів, який описується наступним чином

$$\mathcal{K}(\alpha) = \{\delta : \alpha(\delta) = \sup_{\theta\in\Theta_0} \mathbf{M}\delta(\vec{x}) \le \alpha\} \tag{12}$$

для деякого фіксованого $\alpha > 0$. Використання тесту $\delta \in \mathcal{K}_r(\alpha)$ означає, що коли тест δ буде застосований n разів, то кількість помилкових рішень буде в середньому не більшою за αn разів, якщо насправді була вірною гіпотеза H_0.

Визначення 7. Тест $\delta^* \in \mathcal{K}_r(\alpha)$ називаємо *рівномірно найпотужнішим* (р.н.т.) в класі $\mathcal{K}_r(\alpha)$, якщо для довільного тесту $\delta \in \mathcal{K}_r(\alpha)$

$$\gamma(\delta, \theta) \le \gamma(\delta^*, \theta), \qquad \theta \in \Theta_1. \tag{13}$$

Іншими словами, встановлюємо рівень (або помилку першого роду) шуканого тесту і серед таких тестів шукаємо такий, потужність якого найбільша (або похибка другого роду є найменшою).

Вибір рівня тесту α не нав'язується нам зарання і фактично залежить від нашого суб'єктивного відношення до досліджуваної гіпотези ще перед процедурою перевірки. Найкраще це можна побачити для нерандомізованого тесту δ. Тоді, якщо вибірка $\vec{x} \in \Omega_0 = \{\mathbf{z} : \delta(\mathbf{z}) = 0\}$, то приймаємо гіпотезу H_0, а якщо $\vec{x} \in \Omega_1 = \{\mathbf{z} : \delta(\mathbf{z}) = 1\}$, то приймаємо H_1. Якщо ми переконані, що гіпотеза H_0, яка нас цікавить, є вірною, то знадобляться потужні аргументи, щоб переконати нас, що ми помиляємося і наша гіпотеза є фальшивою. В такій ситуації α треба вибрати дуже малим. Тоді потрапляння в область Ω_1 буде практично неможливим, якщо в дійсності гіпотеза H_0 є вірною. Однак, якщо це трапилося (тобто, $\vec{x} \in \Omega_1$), то цей факт є дуже переконливим доводом того, що ми можемо помилятися.

Проблема відшукання р.н.т. є складною. По перше, такий тест може не існувати, а по друге, якщо такий тест існує, то як його знайти практично. В наступному розділі побачимо, що в деяких випадках проблема конструювання р.н.т. може бути ефективно розв'язана.

Лекція 13. Перевірка двох гіпотез

В цьому розділі буде показано, що для випадку двох простих гіпотез найпотужніший тест існує і, більше того, можемо запропонувати алгоритм його знаходження. Якщо принаймні одна з гіпотез є складною ситуація вже складніша. Може статися навіть так що найпотужнішого тесту не існує. Але розпочнемо з ситуації простих гіпотез.

Отже, нехай для генеральної сукупності ξ головна гіпотеза H_0 полягає в тому, що її розподіл має щільність $f_0(x)$, а гіпотеза альтернативна H_1 означає, що ця щільність є $f_1(x)$.

Іншими словами, гіпотеза H_0 означає, що ξ означена на просторі $(\Omega, \mathcal{B}, \mathbf{P}_0)$ і її функція розподілу $F_0(x) = \mathbf{P}_0(\xi < x)$ має щільність $f_0(x)$, а гіпотеза альтернативна H_1 означає, що ξ означена на просторі $(\Omega, \mathcal{B}, \mathbf{P}_1)$ і її функція розподілу $F_1(x) = \mathbf{P}_1(\xi < x)$ має щільність $f_1(x)$.

Зауваження 1. Якщо генеральна сукупність є дискретною, то замість щільності потрібно використовувати розподіл цієї сукупності.

Позначимо

$$z_0(\vec{x}) = \prod_{i=1}^{n} \frac{f_1(x_i)}{f_0(x_i)} = \frac{L_1(\vec{x})}{L_0(\vec{x})}, \qquad z_1(\vec{x}) = \prod_{i=1}^{n} \frac{f_0(x_i)}{f_1(x_i)} = \frac{L_0(\vec{x})}{L_1(\mathbf{x})}, \tag{1}$$

де для визначеності вважаємо, що $0/0 = 0$. Позначимо також (див. (10))

$$\overline{\alpha} = \mathbf{M}_0 I\{L_1(\vec{x}) > 0\} = \int\limits_{\mathbf{R^n}} I\{L_1(\mathbf{v}) > 0\} L_0(\mathbf{v}) \mathbf{dv}. \tag{2}$$

13.1 Фундаментальна лема Неймана-Пірсона

В цьому параграфі буде представлена так звана *фундаментальна лема Неймана -Пірсона*, на яку буде спиратись метод конструювання багатьох статистичних тестів. Ми сформулюємо цю лему не в самій загальній формі, а при деяких додаткових умовах. В застосуваннях ці умови, як правило, виконуються а загальний випадок читач може знайти в [1], [14].

Отже нехай виконується

Умова C: Випадкова змінна $f_1(\xi)/f_0(\xi)$ має неперервну функцію розподілу на $(0; \infty)$ при гіпотезі H_0.

Інакше кажучи, припускаємо, що функція $\mathbf{P}_0(f_1(\xi)/f_0(\xi) < t) = \int_{-\infty}^{\infty} I\{f_1(x)/f_0(x) < t\} f_0(x) dx$ є неперервною відносно $t > 0$.

Лема 1. *Якщо виконується умова* **C**, *то:*

а) *для довільного* $\alpha \in (0; \overline{\alpha})$ *тест* $\delta_{c(\alpha)} = I\{z_0(\vec{x}) > c(\alpha)\}$, *де* $c(\alpha)$ *є найменшим коренем рівняння*

$$\mathbf{P}_0(z_0(\vec{x}) \geq t) = \alpha, \tag{3}$$

має рівень α *і цей тест є найпотужнішим тестом рівня* α*;*

б) *якщо* δ^* *є іншим найпотужнішим тестом рівня* α*, то*

$$\mathbf{P}_0(\delta^*(\vec{x}) \neq \delta_{c(\alpha)}(\vec{x})) = \mathbf{P}_1(\delta^*(\vec{x}) \neq \delta_{c(\alpha)}(\vec{x})) = 0. \tag{4}$$

Д о в е д е н н я. Неважко зрозуміти, що умова **C** гарантує неперервність $\mathbf{P}_0(z_0(\vec{x}) < t)$ відносно $t > 0$, і, крім того,

$$\mathbf{P}_0(z_0(\vec{x}) = 0) = 1 - \int\limits_{\mathbf{R}^n} I\{z_0(\mathbf{v}) > 0\}\mathbf{f}_0(\mathbf{v})\mathbf{dv} = 1 - \int\limits_{\mathbf{R}^n} I\{\mathbf{f}_1(\mathbf{v}) > 0\}\mathbf{f}_0(\mathbf{v})\mathbf{dv} = 1 - \overline{\alpha},$$

$$\mathbf{P}_0(z_0(\vec{x}) < \infty) = \int\limits_{\mathbf{R}^n} I\{z_0(\mathbf{v}) < \infty\}\mathbf{f}_0(\mathbf{v})\mathbf{dv} = \int\limits_{\mathbf{R}^n} \mathbf{f}_0(\mathbf{v})\mathbf{dv} = 1. \tag{5}$$

Отже, $\mathbf{P}_0(z_0(\vec{x}) < t)$ є неперервною функцією розподілу з, можливо, атомом в нулі. Отже, число $c(\alpha)$ з (3) завше визначене. Для тесту $\delta_{c(\alpha)}$, який означений в лемі, маємо $\mathbf{M}_0\delta_{c(\alpha)} = \mathbf{P}_0(z_0(\vec{x}) > c(\alpha)) = \alpha$, тобто тест $\delta_{c(\alpha)}$ має рівень α.

Доведемо тепер, що цей тест є найпотужнішим з усіх тестів рівня α. (Далі замість $c(\alpha)$ будемо писати c.)

Нехай δ буде інший тест рівня α, тобто $\mathbf{M}_0\delta = \alpha$. Оскільки $\mathbf{M}_0(\delta_c - \delta \;\; = 0$, а $0 < c < \infty$, то

$$\beta(\delta) - \beta(\delta_c) = \mathbf{M}_1(\delta_c - \delta \;\; - c\mathbf{M}_0(\delta_c - \delta \;\; . \tag{6}$$

Оскільки

$$\begin{aligned}
&\mathbf{M}_1 \;\; \delta_c - \delta \;\; = \mathbf{M}_1 \;\; \delta_c - \delta \;\; I\{\mathbf{f}_0(\vec{x}) > 0\} + \mathbf{M}_1 \;\; \delta_c - \delta \;\; I\{\mathbf{f}_0(\vec{x}) = 0\} = \\
&= \int\limits_{\mathbf{R}^n} \;\; \delta_c(\mathbf{v}) - \delta(\mathbf{v}) \;\; I\{\mathbf{f}_0(\mathbf{v}) > 0\}\mathbf{f}_1(\mathbf{v})\mathbf{dv} + \mathbf{M}_1 \;\; \delta_c - \delta \;\; I\{\mathbf{f}_0(\vec{x}) = 0\} = \\
&= \int\limits_{\mathbf{R}^n} \;\; \delta_c(\mathbf{v}) - \delta(\mathbf{v}) \;\; z_0(\mathbf{v})\mathbf{f}_0(\mathbf{v})\mathbf{dv} + \mathbf{M}_1 \;\; \delta_c - \delta \;\; I\{\mathbf{f}_0(\vec{x}) = 0\} = \\
&= \mathbf{M}_0 \;\; \delta_c - \delta \;\; z_0(\mathbf{x}) + \mathbf{M}_1 \;\; \delta_c - \delta \;\; I\{\mathbf{f}_0(\mathbf{x}) = 0\},
\end{aligned}$$

то рівність (6) можемо переписати наступним чином

$$\begin{aligned}
\beta(\delta) - \beta(\delta_{c(\alpha)}) &= \mathbf{M}_0 \;\; \delta_c - \delta \;\; z_0(\mathbf{x}) + \mathbf{M}_1 \;\; \delta_c - \delta \;\; I\{\mathbf{f}_0(\vec{x}) = 0\} - c\mathbf{M}_0 \;\; \delta_c - \delta \\
&= \mathbf{M}_0 \;\; \delta_c - \delta \;\; z_0(\mathbf{x} \;\; - c \;\; + \mathbf{M}_1 \;\; \delta_c - \delta \;\; I\{\mathbf{f}_0(\vec{x}) = 0\}.
\end{aligned} \tag{7}$$

Маємо

$$\mathbf{M}_1\ \delta_c{-}\delta\ I\{\mathbf{f}_0(\vec{x})=0\}=\int\limits_{\mathbf{R}^n}\ \delta_c(\mathbf{v}){-}\delta(\mathbf{v})\ I\{\mathbf{f}_0(\mathbf{v}){=}0\}\mathbf{f}_1(\mathbf{v})\mathbf{dv}=$$

$$=\int\limits_{\mathbf{R}^n}\ \delta_c(\mathbf{v})-\delta(\mathbf{v})\ I\{\mathbf{f}_1(\mathbf{v})>0,\mathbf{f}_0(\mathbf{v})=0\}\mathbf{f}_1(\mathbf{v})\mathbf{dv}=$$

$$=\int\limits_{\mathbf{R}^n}\ 1-\delta(\mathbf{v})\ I\{\mathbf{f}_1(\mathbf{v})>0,\mathbf{f}_0(\mathbf{v})=0\}\mathbf{f}_1(\mathbf{v})\mathbf{dv}\geq 0$$

і з (7) дістаємо

$$\beta(\delta)-\beta(\delta_c)\geq \mathbf{M}_0\ \delta_c-\delta\quad z_0(\mathbf{x})-c\ . \tag{8}$$

Оскільки

$$\delta_{c(\alpha)}-\delta=\begin{array}{rll} 1-\delta, & \textit{якщо} & z_0(\vec{x})-c>0,\\ -\delta, & \textit{якщо} & z_0(\vec{x})-c<0,\end{array}$$

то завжди $(\delta_{c(\alpha)}-\delta\ \ (z_0(\mathbf{x})-c\ \ \geq 0$ і тепер з (8) отримуємо, що

$$\beta(\delta)-\beta(\delta_c)\geq 0\Longrightarrow \beta(\delta)\geq\beta(\delta_c).$$

Доведемо тепер пункт б) нашої леми. Нехай δ_* буде інший найпотужніший тест рівня α. Позначимо

$$S=\{\vec{x}:\delta_*(\vec{x})\neq\delta_c(\vec{x})\}.$$

Оскільки δ_* є найпотужнішим тестом рівня α, то $\beta(\delta_*)-\beta(\delta_c)\leq 0$, і точно так само як і вище маємо (треба в (7) замість δ записати δ_*)

$$\begin{aligned}0\geq\beta(\delta_*)-\beta(\delta_c)=\mathbf{M}_0\ \delta_c-\delta_*\quad z_0(\mathbf{x})-c\ \ +\mathbf{M}_1\ \delta_c-\delta_*\ I\{\mathbf{f}_0(\vec{x})=0\}=\\ =\mathbf{M}_0\ \delta_c-\delta_*)\ z_0(\mathbf{x})-c\ I\{S\}+\mathbf{M}_1\ \delta_c-\delta_*\ I\{S,\mathbf{f}_0(\vec{x})=0\}\end{aligned} \tag{9}$$

Як і вище

$$\mathbf{M}_1\ \delta_c-\delta_*\ I\{S,\mathbf{f}_0(\mathbf{v})=0\}=\int\limits_{\mathbf{R}^n}\ \delta_c(\mathbf{v})-\delta_*(\mathbf{v})\ I\{S,\mathbf{f}_0(\mathbf{v})=0\}\mathbf{f}_1(\mathbf{v})\mathbf{dv}$$

$$=\int\limits_{\mathbf{R}^n}\ \delta_c(\vec{x})-\delta_*(\mathbf{v})\ I\{S,\mathbf{f}_1(\mathbf{v})>0,\mathbf{f}_0(\mathbf{v})=0\}\mathbf{f}_1(\mathbf{v})\mathbf{dv}=$$

$$=\int\limits_{\mathbf{R}^n}\ 1-\delta(\vec{x})\ I\{S,\mathbf{f}_1(\mathbf{v})>0,\mathbf{f}_0(\mathbf{v})=0\}\mathbf{f}_1(\mathbf{v})\mathbf{dv}\geq 0.$$

Так само

$$\delta_c-\delta_*=\begin{array}{rll} 1-\delta_*, & \textit{якщо} & z_0(\vec{x})-c>0,\\ -\delta_*, & \textit{якщо} & z_0(\vec{x})-c<0,\end{array} \tag{10}$$

і тому $(\delta_c-\delta_*)\ \ z(\mathbf{x})-c\ \ \geq 0$. Тепер з (9) дістанемо

$$0\geq\mathbf{M}\ \delta_c-\delta_*\quad z_0(\mathbf{x})-c\ I\{S\}+\mathbf{M}_1\ \delta_c-\delta_*\ I\{S,\mathbf{f}_0(\vec{x})=0\}\geq 0,$$

звідки

$$\mathbf{M}_0\ \delta_c - \delta_* \quad z(\mathbf{x}) - c\ I\{S\} = 0, \quad \mathbf{M}_1\ \delta_c - \delta_*\ I\{S, \mathbf{f}_0(\mathbf{x}) = 0\} = 0. \tag{11}$$

Але згідно з означенням множини S і (10) маємо $\delta_c - \delta_*)\ z(\mathbf{x}) - c\ > 0$ на множині S, а тоді перша рівність в (11) є можливою лише тоді, коли $\mathbf{P}_0(S) = 0$. Отже,

$$\mathbf{P}_0(\delta_*(\vec{x}) \neq \delta_c(\vec{x})) = \mathbf{P}_0(S) = 0. \tag{12}$$

З другої рівності в (11) маємо

$$0{=}\mathbf{M}_1\ \delta_c{-}\delta_*\ I\{S, \mathbf{f}_0(\vec{x}){=}0\}{=}\int\limits_{\mathbf{R}^n} \delta_c(\mathbf{v}){-}\delta_*(\mathbf{v})\ I\{S, \mathbf{f}_0(\mathbf{v}){=}0\}\mathbf{f}_1(\mathbf{v})\mathbf{dv}{=}$$

$$= \int\limits_{\mathbf{R}^n} \delta_c(\mathbf{v}) - \delta_*(\mathbf{v})\ I\{S, \mathbf{f}_1(\mathbf{v}) > 0, \mathbf{f}_0(\mathbf{x}) = 0\}\mathbf{f}_1(\mathbf{v})\mathbf{dv}. \tag{13}$$

На множині $S \cup \{\vec{x} : \mathbf{f}_1(\vec{x}) > 0, \mathbf{f}_0(\vec{x}) = 0\}$ маємо (це випливає з першої рівності в (10)) $\delta_c - \delta_* > 0$ і тому тепер з (13) випливає

$$0{=}\int\limits_{\mathbf{R}^n} I\{S, \mathbf{f}_1(\mathbf{v}) > 0, \mathbf{f}_0(\mathbf{v}) = 0\}\mathbf{f}_1(\mathbf{v})\mathbf{dv}{=}\int\limits_{\mathbf{R}^n} I\{S, \mathbf{f}_0(\mathbf{v}) = 0\}\mathbf{f}_1(\mathbf{v})\mathbf{dv}{=}\mathbf{M}_1 I\{S, \mathbf{f}_0(\mathbf{x}) = 0\}.$$

Звідси та з рівності $\mathbf{P}_0(S) = 0$ (яку ми це вже довели вище) дістаємо

$$\mathbf{P}_1(S) = \mathbf{M}_1 I_S(\vec{x}) = \mathbf{M}_0 I_S(\vec{x}) z_0(\vec{x}) + \mathbf{M}_1 I\{S, \mathbf{f}_0(\vec{x}) = 0\} = 0,$$

а отже $\mathbf{P}_1(\delta_*(\vec{x}) \neq \delta_{c(\alpha)}(\vec{x})) = 0$, що завершує доведення леми.

Зробимо кілька зауважень до доведеної леми.

Зауваження 2. В лемі 1 найпотужніший тест збудований для $0 < \alpha < \overline{\alpha}$, але це обмеження не є принциповим. Можна довести (див. [14]), що для $\alpha = 0$ тест $\delta^1_\infty(\vec{x}) = I\{z_0(\vec{x}) = \infty\}$ є найпотужнішим, а для $\alpha \in [\overline{\alpha}, 1]$ найпотужнішим тестом є

$$\delta_0(\vec{x}) = I\{z_0(\vec{x}) > 0\} + \frac{\alpha - \overline{\alpha}}{1 - \overline{\alpha}} I\{z_0(\vec{x}) = 0\}.$$

Отже лема Неймана -Пірсона справедлива для всіх можливих рівнів значимості тесту.

Зауваження 3. Якщо умова **C** не виконується (наприклад, розподіл генеральної сукупності ξ є дискретним), то $c(\alpha)$ з (3) може не існувати. В цьому випадку $c(\alpha)$ потрібно визначити наступним чином

$$c(\alpha) = \inf\{t \geq 0 : \mathbf{P}_0(z_0(\vec{x}) \geq t)\} \geq \alpha.$$

Нехай для так визначеного $c(\alpha)$ справедлива умова $\mathbf{P}_0(z_0(\vec{x}) = c(\alpha)) = 0$. Ця умова не є істотним обмеженням для конструкції найпотужнішого тесту, бо вона може порушуватись не більше ніж для зліченої кількості чисел $c(\alpha)$, а це означає, що завжди можна добитися виконання цієї умови змінивши в разі потреби α. Тоді найпотужніший тест для цієї ситуації буде такий самий як і у випадку виконання умови **C**.

13.1.1 Тест Неймана -Пірсона

З фундаментальної леми Неймана -Пірсона випливає, що коли виконується умова **C**, то для $0 < \alpha < \overline{\alpha}$ існує тест вигляду

$$\delta_{c(\alpha)}(\vec{x}) = I\{z_0(\vec{x}) > c(\alpha)\}, \tag{14}$$

де $z_0(\vec{x}) = \mathbf{f}_1(\vec{x})/\mathbf{f}_0(\vec{x})$, а стала $c \geq 0$ є найменшим коренем рівняння

$$\mathbf{P}_0(z_0(\vec{x}) \geq c(\alpha)) = \alpha, \tag{15}$$

і цей тест є найпотужнішим тестом рівня α. Якщо взяти до уваги зауваження 3, то цей тест буде найпотужнішим і без умови **C** (але, взагалі кажучи, не для всіх α).

Будемо називати цей тест *Тестом Неймана - Пірсона* або *тестом відношення вірогідності.* Отже, з леми Неймана-Пірсона та зауваження 3 маємо

Наслідок 1.

а) *Нехай має місце умова* **C** *і* $0 < \alpha < \overline{\alpha}$. *Найпотужніший тест на рівні* α *для перевірки гіпотези* H_0 *проти альтернативи* H_1 *має вигляд*

$$\delta_{c(\alpha)}(\vec{x}) = I\{z_0(\vec{x}) = \mathbf{f}_1(\vec{x})/\mathbf{f}_0(\vec{x}) > c(\alpha)\}, \tag{16}$$

де стала $c \geq 0$ *є найменшим коренем рівняння*

$$\mathbf{P}_0(z_0(\vec{x}) \geq c(\alpha)) = \int_{\mathbf{R^n}} I\{\mathbf{f}_1(\vec{x})/\mathbf{f}_0(\vec{x}) \geq c(\alpha)\}\mathbf{f}_0(\vec{x})\mathbf{d}\vec{x} = \alpha. \tag{17}$$

б) *Нехай умова* **C** *не виконується, то для* $0 < \alpha < \overline{\alpha}$ *найпотужніший тест на рівні* α *для перевірки гіпотези* H_0 *проти альтернативи* H_1 *знову має вигляд* (16), *де стала* $c(\alpha) \geq 0$ *така, що* $c(\alpha) = \inf\{t \geq 0 : \mathbf{P}_0(z_0(\vec{x}) \geq t)\} \geq \alpha$, *і* $\mathbf{P}_0(z_0(\vec{x}) = c(\alpha)) = 0$.

Отже, нам потрібно знайти $c(\alpha)$ потім по вибірці $x_1, ..., x_n$ вираховати $\mathbf{f}_1(\vec{x})/\mathbf{f}_0(\vec{x})$ і тоді процедура перевірки виглядає так

$$\frac{\mathbf{f}_1(\vec{x})}{\mathbf{f}_0(\vec{x})} \begin{cases} \leq c(\alpha), & \text{приймаємо} \quad H_0, \\ \\ > c(\alpha), & \text{відхиляємо} \quad H_0. \end{cases} \tag{18}$$

Подамо ще формули для похибок першого та другого роду цього тесту

$$\alpha = \mathbf{P}_0(z_0(\vec{x}) > c(\alpha)), \qquad \beta = \mathbf{P}_1(z_0(\vec{x}) \leq c(\alpha)). \tag{19}$$

Приклад 1. *Нехай нульова гіпотеза* H_0 *означає, що вибірка* $x_1, ..., x_n$ *була зроблена з генеральної сукупності* $\xi \in \mathbb{N}(m_0, \sigma^2)$, *а альтернативна гіпотеза* H_1 *означає, що* $\xi \in \mathbb{N}(m_1, \sigma^2)$, *де параметр* σ *відомий. Побудувати тест Неймана -Пірсона для перевірки гіпотези* H_0 *проти альтернативи* H_1.

Р о з в' я з о к. Нехай для визначеності $m_1 > m_0$. Маємо

$$z_0(\vec{x}) = \prod_{i=1}^{n} \exp\{\frac{m_1 - m_0}{2\sigma^2}(2x_i - m_0 - m_1)\} = \exp\{\frac{m_1 - m_0}{\sigma^2} \sum_{i=1}^{n} x_i - \frac{n(m_0 + m_1)}{2} \}$$

$$= \exp\{\frac{\sqrt{n}(m_1 - m_0)}{\sigma^2} \ \eta(n) + \frac{\sqrt{n}(m_0 - m_1)}{2} \}, \tag{20}$$

де $\eta(n) = n^{-1/2}\sum_{i=1}^{n}(x_i - m_0)/\sigma = \sqrt{n}(x(n) - m_0))/\sigma$. В даній ситуації, як це неважко перевірити, $\overline{\alpha} = 1$. При гіпотезі H_0 випадкова змінна $\eta(n)$ має розподіл $\mathbb{N}(0,1)$, тобто,

$$\mathbf{P}(\eta(n) < x/H_0) = \mathbf{P}_0(\eta(n) < x) = \Phi(x).$$

Умова **C** виконується, оскільки функція

$$\mathbf{P}\{f_1(\xi)/f_0(\xi) < t/H_0\} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} I\{e^{\frac{(m_1-m_0)(2x-m_0-m_1)}{2\sigma^2}} < t\}e^{-\frac{(x-m_0)^2}{2\sigma^2}}dx =$$
$$= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} I\{2x - m_0 - m_1 < \frac{2\sigma^2 \ln t}{m_1 - m_0}\}e^{-\frac{(x-m_0)^2}{2\sigma^2}}dx =$$
$$= \frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{\frac{m_0+m_1}{2}+\frac{2\sigma^2 \ln t}{2(m_1-m_0)}} e^{-\frac{(x-m_0)^2}{2\sigma^2}}dx$$

є неперервною відносно $t > 0$. Тепер з (20) маємо

$$\mathbf{P}_0(z_0(\vec{x}) < t) = \mathbf{P}_0(\eta(n) < \frac{\sigma^2 \ln t}{\sqrt{n}(m_1 - m_0)} - \frac{\sqrt{n}(m_0 - m_1)}{2\sigma})$$
$$= \Phi\left(\frac{\sigma^2 \ln t}{\sqrt{n}(m_1 - m_0)} - \frac{\sqrt{n}(m_0 - m_1)}{2\sigma}\right),$$

а отже рівняння для $c(\alpha)$ виглядає так

$$\Phi\left(\frac{\sigma^2 \ln c}{\sqrt{n}(m_1 - m_0)} - \frac{\sqrt{n}(m_0 - m_1)}{2\sigma}\right) = 1 - \alpha,$$

звідки

$$c(\alpha) = \exp\{\frac{\sqrt{n}(m_1 - m_0)}{\sigma^2}\left(k(1-\alpha) + \frac{\sqrt{n}(m_0 - m_1)}{2\sigma}\right)\}.$$

де $k(1-\alpha)$ є квантиль рівня $1-\alpha$ для розподілу $\mathbb{N}(0;1)$. Згідно з (18) приймаємо гіпотезу H_0 якщо

$$e^{\frac{\sqrt{n}(m_1-m_0)}{\sigma^2}\left(\eta(n)+\frac{\sqrt{n}(m_0-m_1)}{2\sigma}\right)} \le e^{\frac{\sqrt{n}(m_1-m_0)}{\sigma^2}\left(k(1-\alpha)+\frac{\sqrt{n}(m_0-m_1)}{2\sigma}\right)},$$

що рівнозначне $\eta(n) \le k(1-\alpha)$. Отже маємо

$$\sqrt{n}\frac{x(n) - m_0}{\sigma}\begin{cases} \le k(1-\alpha), & \text{приймаємо} \quad H_0, \\ > k(1-\alpha), & \text{відхиляємо} \quad H_0. \end{cases}$$

13.2 Тест Неймана-Пірсона для середнього значення та дисперсії

13.2.1 Тест для середнього значення

Приклад 1. *Досліджувана характеристика ξ має розподіл $\mathbb{N}(m,\sigma^2)$ при відомому σ. Нехай головна гіпотеза і альтернативна будуть такими*

$$H_0 : m = m_0, \qquad H_1 : \ m = m_1 > m_0.$$

Найпотужніший тест для перевірки цих гіпотез було побудовано вище (приклад 1) і має він наступну форму

$$\sqrt{n}\frac{x(n)-m_0}{\sigma}\begin{cases}\leq k(1-\alpha), & \text{приймаємо} \quad H_0,\\ > k(1-\alpha), & \text{відхиляємо} \quad H_0.\end{cases}$$

де $k(1-\alpha)$-квантиль рівня $1-\alpha$ розподілу $\mathbb{N}(0,1)$.

Точно так само можемо розглянути наступну ситуацію

Приклад 2. *Досліджувана характеристика ξ має розподіл $\mathbb{N}(m,\sigma^2)$ при відомому σ. Головна гіпотеза і альтернативна є такими*

$$H_0 : m = m_0, \qquad H_1 : \ m = m_1 < m_0.$$

Найпотужніший тест для перевірки гіпотези H_0 проти альтернативи H_1 тепер буде таким

$$\sqrt{n}\frac{x(n)-m_0)}{\sigma}\begin{cases}\geq -k(1-\alpha), & \text{приймаємо} \quad H_0,\\ < -k(1-\alpha), & \text{віхиляємо} \quad H_0.\end{cases}$$

де $k(1-\alpha)$-квантиль рівня $1-\alpha$ розподілу $\mathbb{N}(0,1)$.

13.2.2 Тест для дисперсії

Приклад 3.

Досліджувана характеристика ξ генеральної сукупності має розподіл $\mathbb{N}(m,\sigma^2)$ при відомому m. Гіпотези головна і альтернативна є такими

$$H_0 : \ \sigma^2 = \sigma_0^2, \qquad H_1 : \ \sigma^2 = \sigma_1^2 > \sigma_0^2.$$

Перевірка гіпотези про дисперсію відбувається в спосіб, подібний до того, що і для значення середньої. Для відношення вірогідності маємо

$$z_0(\vec{x}) = \frac{\mathbf{f}(\vec{x},\sigma_1)}{\mathbf{f}(\vec{x},\sigma_0)} = \ \frac{\sigma_0}{\sigma_1}\ {}^n \exp\{\sum_{i=1}^n (x_i-m)^2\frac{\sigma_1^2-\sigma_0^2}{2\sigma_1^2\sigma_0^2}\}$$

і тест Неймана-Пірсона тепер виглядає так

$$\delta^* = I\{\ \frac{\sigma_0}{\sigma_1}\ {}^n \exp\{\sum_{i=1}^n (x_i-m)^2\frac{\sigma_1^2-\sigma_0^2}{2\sigma_1^2\sigma_0^2}\} > c\}, \tag{21}$$

де для сталої $c \geq 0$ маємо рівняння

$$\mathbf{P}(\ \frac{\sigma_0}{\sigma_1}\ {}^n \exp\{\sum_{i=1}^n (x_i-m)^2\frac{\sigma_1^2-\sigma_0^2}{2\sigma_1^2\sigma_0^2}\} > c \ \ H_0) = \alpha. \tag{22}$$

Означимо статистику $U(n) = \sum_{i=1}^n \ \frac{x_i-m}{\sigma_0}\ {}^2$, яка за умови, що має місце гіпотеза H_0, має розподіл хі-квадрат з n степенями свободи. Після простих перетворень з (21), (22) маємо

$$\delta^* = I\{U(n) > c(\alpha)\}, \qquad \mathbf{P}(U(n) > c(\alpha) \ \ H_0) = \alpha, \tag{23}$$

де $c(\alpha) = \quad \ln c + n \ln \frac{\sigma_1}{\sigma_0} \quad \frac{2\sigma_1^2}{\sigma_1^2 - \sigma_0^2}$. Оскільки, як уже було сказано, статистика $U(n)$ має при гіпотезі H_0 розподіл хі-квадрат з n степенями свободи, то з (23) отримуємо наступний тест для перевірки гіпотези H_0 проти альтернативи H_1

$$\frac{\sum_{i=1}^{n}(x_i - m)^2}{\sigma_0^2} \begin{cases} \leq k(1-\alpha, n), & \text{приймаємо} \quad H_0, \\ > k(1-\alpha, n), & \text{відхиляємо} \quad H_0, \end{cases}$$

де $k(1-\alpha, n)$-квантиль рівня $1-\alpha$ розподілу хі-квадрат з n степенями свободи.

І з конструкції випливає, що цей тест буде найпотужнішим тестом рівня α.

Точно так само можемо розглянути наступну ситуацію

Приклад 4.

Досліджувана характеристика ξ генеральної сукупності має розподіл $N(m, \sigma^2)$ при відомому m. Гіпотези головна і альтернативна є такими

$$H_0: \ \sigma^2 = \sigma_0^2, \qquad H_1: \ \sigma^2 = \sigma_1^2 < \sigma_0^2.$$

Тест Неймана-Пірсона для перевірки гіпотези H_0 проти альтернативи H_1 виглядає тепер так

$$\frac{\sum_{i=1}^{n}(x_i - m)^2}{\sigma_0^2} \begin{cases} \geq k(\alpha, n), & \text{приймаємо} \quad H_0, \\ < k(\alpha, n), & \text{відхиляємо} \quad H_0, \end{cases}$$

де $k(\alpha, n)$-квантиль рівня α розподілу хі-квадрат з n степенями свободи.

ЗАВДАННЯ

1. Довести, що умова **C** гарантує, що змінна $f_0(\xi)/f_1(\xi)$ має неперервну функцію розподілу на $(0; \infty)$ при гіпотезі H_1.
2. Маємо послідовність незалежних експериментів типу 0-1 з ймовірнстю "успіху" p. Побудувати тест Неймана- Пірсона для перевірки головної гіпотези H_0 : $p = 0,05$ проти альтернативи $H_1 : p = 0,02$ і визначити найменшу кількість експериментів, яка б гарантувала, що похибки першого та другого роду не будуть більшими ніж 0.01.
3. Побудувати тест Неймана- Пірсона для генеральної сукупності з розподілом $1 - e^{-\lambda x}, x \geq 0$, якщо гіпотеза головна $H_0 : \lambda = 5$, а альтернативна $H_1 : \lambda = 6$.
4. Побудувати тест Неймана- Пірсона та перевірити гіпотезу, що вибірка 0.45; -0,104; -0,315; 0,186 -0,164 взята з розподілу рівномірного на відрізку [-0,5; 0,5] проти альтернативи, що вона походить з розподілу $\mathbb{N}(0; 0,005)$.

Лекція 14. Статистичні тести

В цьому розділі наша увага буде концентрована на описані конкретних тестів для перевірки головної гіпотези (або нульової) H_0 проти гіпотези альтернативної H_1. Ці гіпотези стосуються генеральної сукупності ξ (вона може бути векторною), з якої маємо вибірку $x_1, ..., x_n$. Подібно до ситуації з оцінкою параметрів, де навіть для одного параметра можемо запропонувати різні оцінки, для перевірки тих самих гіпотез можемо застосовувати різні тести. Тому кількість різних тестів завелика, щоб подати навіть їх короткий опис в одному підручнику. З цієї причини головна частина цього розділу буде присвячена тестам, які з нашої точки зору, є найважливішими оскільки ці тести найчастіше використовуються в застосуваннях. Для інших тестів буде подано лише їх короткий опис і посилання на літературу, де зацікавлений читач може знайти більше інформації на цю тему.

Наш опис почнемо з тестів *значимості.*

В практиці часто маємо справу з ситуацією, в якій на підставі значень характеристики хочемо перевірити гіпотезу, що деякий параметр цієї характеристики, наприклад, її середнє значення, дисперсія тощо, набувають деяких конкретних значень. В інших ситуаціях маємо кілька незалежних вибірок для різних характеристик, відносно яких нас цікавить деякий параметр (наприклад, середнє значення), і ми хочемо перевірити гіпотезу, що цей параметр для тих різних характеристик є такий самий. Тому в цьому параграфі головна гіпотеза H_0 завжди буде стосуватися значення параметра, і найчастіше це буде проста гіпотеза тобто, вона буде стосуватися конкретного значення параметра. Тести, які призначені для перевірки таких гіпотез, називають *тестами значимості.*

Як побачимо далі, для деяких, важливих з точки зору практики, параметрів, можемо сконструювати найкращі тести (тобто найпотужніші) на підставі загальної теорії, розвинутої в попередньому розділі. Однак може трапитися так, що конструкція найпотужнішого тесту є складною або для даної альтернативної гіпотези такого тесту просто немає. В такій ситуації будуємо критичну множину таким чином, щоб за припущення вірності гіпотези H_0 ймовірність потрапляння тестової статистики до цієї множини була $\leq \alpha$, де α рівень значимості тесту. Тоді похибку другого роду зазвичай не вираховуємо і, якщо тестова статистика не потрапить до критичної множини, то немає підстав для відкидання гіпотези H_0 тобто, приймаємо H_0 (що, очевидно, ще не означає, що ця гіпотеза є вірною).

Конструювання тестів значимості (і не лише їх) є простішим, коли вдасться знайти тестову статистику $S(\vec{x})$ з відомим (за умови, що гіпотеза H_0 вірна) розподілом. Інакше майже єдиним способом конструювання таких тестів будуть граничні теореми. Так, якщо $\mathbf{P}(S(\vec{x}) < z \;\, H_0) = G(z)$, то тоді найчастіше в якості критичної множини вибираємо множину вигляду $[a, b]$, $-\infty \leq a < b \leq \infty$ так, щоб $G(b) - G(a) = 1 - \alpha$ (нехай для визначеності функція розподілу $G(z)$ буде неперервною). Тоді $\mathbf{P}(S(\vec{x}) \in [a, b] \;\, H_0) = 1 - \alpha$, що означає, що коли α достатньо мале, то для конкретної вибірки $\vec{x} = (x_1, ..., x_n)$ число $S(\vec{x})$ з великою ймовірністю потрапляє до відрізка $[a, b]$ за умови, що гіпотеза H_0 є вірною. Тому, якщо для конкретної вибірки значення статистики $S(\vec{x})$ лежать в $[a, b]$, то немає підстав не вірити, що гіпотеза H_0 є фальшива і тому приймаємо її. І навпаки, якщо $S(\vec{x}) \notin [a, b]$, то трапилася подія малої ймовірності α, що при вірній гіпотезі H_0 було б майже неможливим. Тобто, в цій ситуації маємо значні підстави для сумнівів щодо вірності гіпотези H_0. Як вже говорили, знаходження похибки другого роду при такому підході часто є завданням важким і тому зазвичай вона не вираховується.

Почнемо описання тестів значимості з найважливіших параметрів. Тести будемо подавати, як це прийнято, в наступному вигляді: тестова статистика S і критична множина K. І тоді якщо для конкретної вибірки $\vec{x} = (x_1, x_2, ..., x_n)$ обчисливши $S(\vec{x})$ маємо $S(\vec{x}) \notin K-$ то немає підстав не прийняти гіпотезу H_0 (і ми її приймаємо). Іншими словами, дані з вибірки не суперечать гіпотезі H_0 і у нас немає підстав не вірити, що вона дійсно правдива. А якщо $S(\vec{x}) \in K$, то вважаємо, що дані з вибірки суперечать гіпотезі H_0 і ми її відкидаємо. В цій другій ситуації, як ми вже знаємо, часто говорять так "відкидаємо гіпотезу H_0 на користь H_1". В практиці це ще не означає що відкинувши H_0 ми приймаємо H_1. Справа в тому, що в якості H_0 ми як правило вибираємо гіпотезу, правдивість якої для нас є бажаною. А тому якщо дані суперечать нашій гіпотезі, то ми найчастіше формулюємо нову головну гіпотезу і повторюємо все спочатку.

14.1 Тести значимості.

14.1.1 Тести значимості для середнього значення.

В цій ситуації головна гіпотеза є $H_0 : \ \{\mathbf{M}\xi = m = m_0\}$, тобто значення математичного сподівання досліджуваної характеристики ξ дорівнює даному числу m_0, а в якості альтернативної гіпотези H_1 зазвичай приймається одна з наступних

$$\text{а) } m \neq m_0, \qquad \text{б) } m > m_0, \qquad \text{в) } m < m_0.$$

Для конструювання тесту використовуємо статистику $x(n) = n^{-1} \sum_{i=1}^{n} x_i$, і нам буде зручно розглядати наступні моделі.

Модель M 1. *Досліджувана характеристика ξ має розподіл $\mathbb{N}(m, \sigma^2)$ при відомому σ.*

В цій ситуації тестова статистика буде такою

$$U(n) = \sqrt{n}\frac{x(n) - m_0}{\sigma}, \tag{1}$$

яка при гіпотезі H_0 має розподіл $N(0, 1)$.

В прикладах **??**, **??** с. **??** були розглянуті ситуації альтернативної гіпотези б) і в) і було доведено, що в ситуації б) критична множина має вигляд $(k(1-\alpha), \infty)$, де $\Phi(k(1-\alpha) = 1-\alpha$ і $\Phi(x)$ є стандартний нормальний розподіл а в ситуації в) критична множина має вигляд $(-\infty, -k(1-\alpha))$ і ці тести є найпотужнішими.

Якщо альтернативна гіпотеза має вигляд а), то, як уже говорилося, в цій ситуації найпотужнішого тесту не існує (див. текст після висновку **??** с. **??**). Тому діємо так, як було сказано на початку цього підрозділа. Маємо

$$\mathbf{P}(\ \sqrt{n}\frac{x(n) - m_0}{\sigma} \ \le x \ \ H_0) = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$$

і якщо взяти в цій рівності замість x квантиль рівня $1-\alpha/2$ розподілу $\mathbb{N}(0,1)$, то отримаємо

$$\mathbf{P}(\ \sqrt{n}\frac{x(n)-m_0}{\sigma}\ \le k(1-\alpha/2)\ \ H_0) = 1-\alpha$$

і, отже, ймовірність, що при вірній гіпотезі H_0 статистика $U(n)$ потрапить в інтервал $(-\infty, -k(1-\alpha/2) \cup (x(1-\alpha/2), \infty)$, дорівнює α. Тому процедура перевірки гіпотези H_0 відносно альтернативи H_1 виглядає так

$$\sqrt{n}\frac{x(n)-m_0)}{\sigma}\begin{cases} \in [-k(1-\alpha/2), k(1-\alpha/2)], & \text{приймаємо} \quad H_0, \\ \notin [-k(1-\alpha/2), k(1-\alpha/2)], & \text{відхиляємо} \quad H_0, \end{cases}$$

де $k(1-\alpha/2)$- квантиль рівня $1-\alpha/2$ розподілу $\mathbb{N}(0,1)$.

Отже, для моделі **M 1** маємо тестову статистику з (2), а критичною множиною у випадку альтернативної гіпотези а) є $\ -\infty, -k(1-\alpha/2)\ \cup\ k(1-\alpha/2), \infty\ ]$, у випадку б) — $\ k(1-\alpha), \infty\ $, а у випадку в)— $\ -\infty, -k(1-\alpha)\ $, де $k(\gamma)$ є квантиль рівня γ стандартного нормального розподілу. У випадках б), в) тести є найпотужнішими.

Модель M 2. *Досліджувана характеристика ξ має розподіл $\mathbb{N}(m, \sigma^2)$ при невідомому σ.*

В цій ситуації тестова статистика буде такою

$$U(n) = \sqrt{n-1}\frac{x(n)-m_0}{S(n)}, \tag{2}$$

де $S^2(n) = n^{-1}\sum_{i=1}^n (x_i - x(n))^2$ є варіацією з вибірки. Тепер відношення вірогідності вже не буде монотонним відносно даної статистики, а тому підхід з моделі **M 1** не може бути застосованим. Але з теореми Фішера випливає, що статистика (2) при гіпотезі H_0 має розподіл Стьюдента з $n-1$ степенями свободи, і тому можемо застосувати такий підхід до конструювання тесту, який застосовувався в моделі **M 1** для альтернативної гіпотези а), тобто відповідні критичні множини будуть визначені за допомогою відношень $\mathbf{P}(|U(n)| > x\ \ H_0) = \alpha$, $\mathbf{P}(U(n) > x\ \ H_0) = \alpha$, $\mathbf{P}(U(n) < x\ \ H_0) = \alpha$ для альтернативноих гіпотез а) б) та в) відповідно.

Отже, тестова статистика для моделі **M 1** задається формулою (2), а критичною множиною у випадку альтернативної гіпотези а) є $\ -\infty, -k(1-\alpha/2, n-1)\ \cup\ k(1-\alpha/2, n-1), \infty\ $, у випадку б) — $\ k(1-\alpha, n-1), \infty\ $, а у випадку в) — $\ -\infty, -k(1-\alpha, n-1)\ $, де $k(\gamma, n-1)$ є квантиль рівня γ розподілу Стьюдента з $n-1$ степенями свободи.

Модель M 3. *Досліджувана характеристика ξ має довільний розподіл з невідомим значенням середньої m і зі скінченою, але невідомою, дисперсією σ^2.*

В цьому випадку ми, як правило, не маємо статистики з відомим розподілом. А тому застосуємо підхід, що спирається на граничні теореми. З цього випливає, що вибірка повинна бути великою ($n \ge 30$).

Для перевірки цієї гіпотези використовуємо статистику з моделі **M** 1 в якій як невідоме значення σ приймається оцінка для σ. Тому розглянемо статистику

$$U(n) = \sqrt{n}\frac{x(n)-m_0}{S(n)}. \tag{3}$$

Так само як і в прикладі 3 с.219 показуємо, що за умови правдивості гіпотези H_0, асимтотичний розподіл цієї статистика є нормальний стандартний, і отже для великих n

$$\mathbf{P}(\sqrt{n}\frac{x(n) - m_0}{S(n)} < x \;\; H_0) \approx \Phi(x).$$

Останнє відношення служить для конструювання тесту, який тепер виглядає практично так само, як і в першій моделі. Тестова статистика задається формулою (3) і у випадку альтернативної гіпотези а) критичною множиною є $-\infty, -k(1-\alpha/2) \cup k(1-\alpha/2), \infty$, у випадку б)— $k(1-\alpha), \infty$, а у випадку в)— $-\infty, -k(1-\alpha)$, де $k(\gamma)$ є квантиль рівня γ стандартного нормального розподілу.

Приклад 1. *Вибрано незалежним чином* 10 *фермерських господарств в певному регіоні і отримана наступні дані про урожайність пшениці в цих господарствах:* 44, 5; 45, 0; 47, 4; 43, 4; 46, 0; 43, 8; 48, 1; 43, 0; 46, 4; 47, 7 (*ц/г*). *Припускаємо, що урожайність має нормальний розподіл* $\mathbb{N}(m, \sigma^2)$. *Відомо що середня урожайності за минулий рік яка становила* 45, 1 (*ц/г*). *Чи можна стверджувати, що цьогорічна середня урожайність в цілому регіоні є вища? Рівень значимості взяти* $\alpha = 0.005$.

Р о з в' я з о к. Ясно що в цьому випадку ми маємо справу з моделлю **M 2**. Гіпотеза головна є $H_0 : m = 45,1$, а гіпотеза альтернативна буде наступною $H_1 : m > 45,1$.

Статистика тестова подана в (2), а критична множина є такою $k(0,995; 9), \infty$, де $k(0,995; 9)$ є квантиль рівня 0, 995 розподілу Стьюдента з 9–ма степенями свободи.

Маємо $x(10) = 45,53$, $S^2(10) = 3,11$, $S(10) = 1,76$. Порахуємо тепер статистику тестову (2). Маємо $U(10) = \sqrt{9}(45,53 - 45,1)/1,76 = 0,73$. З таблиць знаходимо квантиль $k(0,995; 9) = 3,25$. Оскільки $U(10) = 0,73 < 3,25$, то належить прийняти гіпотезу H_0, а це означає, що отримані врожаї не дають підстав (на 95%) стверджувати, що цьогорічна врожайність вища ніж минулорічна.

14.1.2 Перевірка гіпотез про рівність середніх значень двох популяцій

В дослідженнях з погляду на певну характеристику в двох сукупностях часто виникає потреба перевірки гіпотези про рівність значень середньої цієї характеристики або про однаковий (чи різний) ступінь розсіяння значень досліджуваної характеристики.

Розглянемо такий приклад.

Приклад 2.

Серед учнів другого класу двох шкіл вибрано відповідно по 15 *та* 18 *осіб і пораховано середню з оцінок, отриманих у семестрі для кожного з учнів. Отримано такі результати:*

для учнів першої школи: 3, 45; 4, 13; 2, 95; 3, 06; 4, 05; 4, 09; 4, 20; 3, 90; 3, 11; 2, 90; 2, 84; 4, 75; 4, 10; 3, 87; 3, 92;

для учнів другої школи: 3, 54; 3, 13; 3, 94; 3, 16; 3, 05; 4, 39; 4, 22; 2, 90; 3, 31; 2, 95; 2, 74; 4, 65; 4, 17; 3, 80; 3, 95; 3, 82; 4, 2; 3, 4.

Припускаючи, що середні результати оцінок мають нормальні розподіли, перевірити на рівні значимості $\alpha = 0,01$ *гіпотезу, що середні значення оцінок учнів з тих шкіл є однакові проти альтернативної гіпотези, що вони різні.*

Цю задачу розв'яжемо пізніше, а зараз зауважимо, що задачі такого вигляду, тобто коли потрібно порівнювати значення середніх двох характеристик, часто зустрічаються в медицині, сільському господарстві, метрології та інших областях. В цій частині опишемо типові підходи до дослідження таких проблем. Оскільки для нас поняття "характеристика" і "генеральна популяція" є синонімами і є очевидним, що порівнювати можемо лише однакові характеристики (наприклад температуру з температурою, урожай з урожаєм тощо), то в цій частині завжди вважаємо, що коли говоримо "дві характеристики ξ і η", то одразу припускаємо, що ці характеристики мають однакову природу і значення цих характеристик вимірювались незалежним способом.

В цьому підрозділі розглянемо середнє значення характеристик, а в наступному дисперсії. Як і вище, зручно подавати різні ситуації з допомогою формальних математичних моделей.

Отже, нехай маємо дві генеральні сукупності ξ, η і $x_1, ..., x_n$ є вибіркою з першої сукупності, а $y_1, ..., y_m$ з другої. Нехай $m_1 = \mathbf{M}\xi$ і $m_2 = \mathbf{M}\eta$. Потрібно перевірити гіпотезу $H_0: \ m_1 = m_2$, тобто, що середні значення обох популяцій є рівними, проти однієї з альтернатив

$$\text{а) } H_1: \ m_1 \neq m_2, \quad \text{б) } H_1: \ m_1 > m_2, \quad \text{в) } H_1: \ m_1 < m_2.$$

Почнемо з найпростішої ситуації, яка представлена наступною моделлю.

Модель ММ 1. *Досліджувані характеристики ξ, η мають відповідно розподіли $\mathbb{N}(m_1, \sigma_1^2)$, $\mathbb{N}(m_2, \sigma_2^2)$ з відомими σ_1, σ_2.*

Статистика

$$U(n, m) = \frac{x(n) - y(m)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}, \tag{4}$$

де $x(n) = n^{-1}\sum_{i=1}^n x_i, \quad y(m) = m^{-1}\sum_{i=1}^m y_i$, за умови справедливості гіпотези H_0 має стандартний нормальний розподіл. Це випливає з теореми 1 с. 172. І тепер у вже звичний спосіб маємо тест для перевірки гіпотези H_0 відносно альтернативи H_1. Тестова статистика описана в (4), а у випадку альтернативної гіпотези а) критичною множиною є $-\infty - k(1-\alpha/2) \ \cup \ k(1-\alpha/2), \infty$, у випадку б)— $k(1-\alpha), \infty$, а у випадку в)— $-\infty, -k(1-\alpha)$, де $k(1-\alpha/2)$, $k(1-\alpha)$ є відповідно квантилями рівнів $1-\alpha/2$ і $1-\alpha$ стандартного нормального розподілу.

Модель ММ 2. *Досліджувані характеристики ξ, η мають відповідно розподіли $\mathbb{N}(m_1, \sigma^2)$, $\mathbb{N}(m_2, \sigma^2)$ з невідомими (але однаковими для обох характеристик) σ.*

В такій моделі статистика

$$U(n) = \frac{x(n) - y(m)}{\sqrt{\frac{nS^2(x) + mS^2(y)}{n+m-2}\,\frac{n+m}{nm}}}, \tag{5}$$

де $S^2(x) = n^{-1}\sum_{i=1}^n (x_i - x(n))^2, \quad S^2(y) = m^{-1}\sum_{i=1}^m (y_i - y(m))^2$, за припущення справедливості гіпотези H_0 має розподіл Стьюдента з $n+m-2$ степенями свободи і, отже тест для перевірки гіпотези H_0 відносно альтернативи H_1 є наступним: тестова статистика подана в (5), а у випадку альтернативної гіпотези а) критичною множиною є $-\infty - k(1-\alpha/2, n+m-2) \ \cup \ k(1-\alpha/2, n+m-2), \infty$, у випадку б)— $k(1-\alpha, n+m-2), \infty$, а у випадку в)— $-\infty, -k(1-\alpha, n+m-2)$, де $k(\gamma, n+m-2)$ є квантиль рівня γ розподілу Стьюдента з $n+m-2$ степенями свободи.

Модель ММ 3. *Досліджувані характеристики* ξ, η *мають відповідно розподіли* $\mathbb{N}(m_1, \sigma_1^2)$, $\mathbb{N}(m_2, \sigma_2^2)$ *з невідомими* σ_1, σ_2.

Тест, який опишемо нижче, зазвичай застосовується для малих вибірок ($n, m \leq 30$). Інакше можемо застосувати модель **ММ 1**, де замість невідомих σ_1^2, σ_2^2 вставимо їх оцінки $S^2(x)$, $S^2(y)$.

Для перевірки гіпотези H_0 застосуємо тест, що спирається на статистику (**Кохрана і Коха**)

$$U = \frac{x(n) - y(m)}{\sqrt{\frac{S^2(x)}{n-1} + \frac{S^2(y)}{m-1}}}, \tag{6}$$

де $x(n)$, $y(m)$, $S^2(x)$, $S^2(y)$ є такими ж, як і вище.

Розподіл цієї статистики невідомий, оскільки він залежить від відношення σ_1/σ_2, але для числа

$$z(\alpha/2, n, m) = \frac{\frac{S^2(x)}{n-1}k(1-\alpha/2, n-1) + \frac{S^2(y)}{m-1}k(1-\alpha/2, m-1)}{\frac{S^2(x)}{n-1} + \frac{S^2(y)}{m-1}}, \tag{7}$$

де $k(1-\alpha/2, n-1)$, $k(1-\alpha/2, m-1)$, є квантилями рівня $1-\alpha/2$ розподілу Стьюдента з відповідно $n-1$, $m-1$ степенями свободи, маємо

$$\mathbf{P}(|U| > z(\alpha/2, n, m) \;\; H_0) \approx \alpha.$$

Якщо в цих відношеннях замінити $\alpha/2$ на α, то буде $\mathbf{P}(U > z(\alpha, n, m) \;\; H_0) \approx \alpha$ при альтернативі $m_1 > m_2$ і $\mathbf{P}(U < z(\alpha, n, m) \;\; H_0) \approx \alpha$ при альтернативі $m_1 < m_2$. З цього випливає, що тест для перевірки гіпотези H_0 відносно альтернативи H_1 виглядає так. Тестова статистика описана в (6) і у випадку альтернативної гіпотези а) критичною множиною є $\;-\infty, -z(\alpha/2, n, m) \;\cup\; z(\alpha/2, n, m), \infty\;$, у випадку б)— $z(\alpha, n, m), \infty)$, а у випадку в) — $(-\infty, -z(\alpha, n, m))$.

Повернемося до прикладу 2.

Зрозуміло, що в цій ситуації ми маємо справу з моделлю ММ 3, а тому порахуємо перш за все значення $z(\alpha/2, n, m)$=$z(0,005; 15; 18)$. Зробивши необхідні підрахунки маємо (значення з буквою x відносяться до першої вибірки, а з y-до другої.)

$$x(15) = 3,688, \qquad y(18) = 3,629, \qquad S^2(x) = 0,323 \qquad S^2(y) = 0,311.$$

З таблиці 3 в кінці посібника знаходимо $k(0,995; 14) = 2,977$, $k(0,995; 17) = 2,898$ і отже

$$z(0,005, 15, 18) = \;\; \frac{0,323}{14} \times 2,977 + \frac{0,311}{17} \times 2,898 \;\; : \;\; \frac{0,323}{14} + \frac{0,311}{17} \;\; = 2,942.$$

Тепер підрахуємо значення статистики з (6)

$$U = \frac{3,688 - 3.629}{\sqrt{\frac{0,323}{14} + \frac{0,311}{17}}} = 0,290.$$

Оскільки $0,29 \in (-2,942; 2,942)$, то у нас немає підстав стверджувати, що різниця між середніми з оцінок в цих школах є істотною.

ЗАВДАННЯ

У всіх поданих нижче задачах за рівень тесту брати $\alpha = 0,05$.

1. Відомо, що генеральна сукупність має розподіл $\mathbb{N}(m, \sigma^2)$. На підставі вибірки 1,2; 1,6; 1,0;1,9;1,4;1,7;1,0;1,5; 1,4 перевірити:

   а) гіпотезу $H_0 : m = 2$ проти альтернативи $H_1 : m \neq 2$ якщо додатково відомо, що $\sigma = 2$;

   б) гіпотезу $H_0 : m = 2$ проти альтернативи $H_1 : m < 2$, якщо σ невідоме.

2. Заміряно пробіг 100 автомобільних скатів певного типу, які були зняті з екплуатації, і отримано наступні результати

пробіг (тис.км)	20-25	25-30	30-35	35-40	40-45	45-50
кількість скатів	5	10	17	20	38	10

   Перевірити гіпотезу, що середній час пробігу скатів є 43 т.км проти альтернативи, що цей час є меншим від 43 т.км.

3. Поміряно максимальну ємність 20-ти конденсаторів і отримано результати (в pF) : 55,1; 67,3; 54,6; 52,2; 58,4; 50,4; 70,1; 55,3; 57,6; 62,5; 65,2; 68,4; 54,5; 56,7; 53,5; 61,6; 59,6; 49,0; 63,7; 58,1. Припускаючи, що ємність має рзподіл $\mathbb{N}(m, \sigma^2)$ перевірити гіпотезу $H_0 : \sigma^2 = 6,25$ проти альтернативи $H_1 : \sigma^2 < 6,25$, якщо:

   а) відомо, що $m = 60$; б) m теж є невідоме.

4. Для перевірки ефективності нових ліків проведено експеримент на двох групах мишей. Тривалість життя мишей з першої групи (6 мишей), якій давали ці ліки, виявилась наступною (в роках): 2,2; 3.2; 0,9; 4,1; 2,7; 1,9. Другій групі (5 мишей) ці ліки не подавалися і їх тривалість життя була така: 1,8; 0,7; 4,9; 3,5; 1,2. Важаючи, що тривалість життя всіх мишей має розподіл нормальний з однаковою дисперсією, перевірити гіпотезу, що нові ліки збільшують середню тривалість життя миші.

5. Для продукції певного виробу запропоновано дві технології. Для кожної з них заміряли кількість сировини, використану на одиницю кінцевого продукту і отримали такі результати: для першої технології $4,1; 3,9; 2,9; 3,1; 4,0$, а для другої– $4,1; 1,8; 5,2; 1,7$. Перевірити гіпотези, що для обох технологій:

   а) дисперсії використання сировини такі самі;

   б) середня кількість використаної сировини така сама

6. Два спортсмени показали наступні результати в стрибках в довжину (в метрах):

   перший спортсмен: 7,06; 7,13; 7,51; 6,98;7,28; 7,68; 7,8; 6,95;

   другий спортсмен: 8,13; 8,07; 7,27; 8,18; 7,72; 7,93; 8,02; 8,03; 7,68.

   Перевірити гіпотезу, що результати цих спортсменів однаково стабільні (тобто дисперсії їхніх результатів істотно не відрізняються).

7. Зроблено контрольний замір ваг порційних ліків, які дозуються двома автоматами і отримано результати (в гр.) для першого автомату: 5,23; 5,27; 5,31; 5,44; 5,21; 54,3; 5,34; 5,46; 5,43; 5,41; для другого: 6,21; 6,37; 6,31; 6,44; 6,25; 6,43; 6,34; 6,46; 6,33; 6,31. Відомо, що вага ліків розподілена нормально. Використовуючи тести Фішера та тест t- Стьюдента перевірити гіпотезу про те що обидва автомати працюють з однаковою докладністю, тобто що дисперсії однакові.

а) Середні ваги ліків відома і рівна 5,35 для першого автомата і 6,34 для другого

б) Середні ваги ліків теж невідомі.

Лекція 15. Тести згідності

В цьому підрозділі ми розглянемо тести трохи іншої природи ніж в попередньому. Там на підставі вибірки ми шукали підтвердження (чи не підтвердження) нашого припущення відносно значення деякого параметра (чи параметрів). Ми перевіряли "істотність" або "значимість" нашого припущення тобто, іншими словами, чи популяція генеральна з таким значенням параметра може генерувати дану вибірку. В цьому підрозділі головна гіпотеза буде полягати в тому, що генеральна популяція має певну властивість (позначимо її $\mathbf{A}$) і завдання буде полягати в тому щоб дати відповідь на таке питання: чи дана вибірка може бути породжена популяцією, яка має властивість $\mathbf{A}$? Іншими словами чи наша вибірка не суперечить тому (тобто згідна з тим), що генеральна популяцією має властивість $\mathbf{A}$. Тести, які дають відповідь на такі питання називаються тестами *згідності.* .

Одним з найважливіших тестів згідності є тест про вигляд функції розподілу. Нехай головна гіпотеза буде такою

$$\begin{aligned}&H_0 : \{\text{функція розподілу досліджуваної характеристики є } F_0(x),\\ &\quad \text{тобто, } \mathbf{P}(\xi < x) = F_0(x)\},\end{aligned}$$

де $F_0(x)$ є відома функція розподілу, а альтернативна гіпотеза є такою

$$\begin{aligned}&H_1 : \{\text{функція розподілу досліджуваної характеристики не є } F_0(x),\\ &\quad \text{тобто, } \mathbf{P}(\xi < x) \neq F_0(x)\}.\end{aligned}$$

Тест, який буде використовуватися для перевірки цих гіпотез, повинен дослідити згідність між розподілом множини значень у вибірці і розподілом теоретичним. Це ще одне пояснення чому такі тести називаємо тестами згідності. Зауважимо також, що, як це було сказано вище, припускаємо, що теоретична функція розподілу $F_0(x)$ є відомою. Але можлива ситуація, коли ця функція розподілу залежить від невідомих параметрів $\theta_1, ..., \theta_k$ тобто, не є відомою. Коротке описання тестів згідності для такої ситуації подамо також, але головна увага буде зосереджена на випадку відомої $F_0(x)$. Робимо це не лише тому, що в цьому випадку дослідження будуть простішими, а також тому, що на практиці частіше маємо справу саме з такою ситуацією. Найбільш відомими тестами згідності є тест χ^2- Пірсона, тест Колмогорова, тест Смірнова та тест Шапіро -Вілька.

15.1 Тест χ^2-Пірсона

Для деякого фіксованого натурального $k \geq 2$ поділимо вісь $(-\infty, +\infty)$ на k множин S_i, $i = 1, ..., k$, які не перетинаються, тобто, $\mathbf{R}^1 = \cup_{i=1}^k S_i,\ S_i S_j = \emptyset,\ i \neq j$. Нехай для визначеності множини S_i будуть інтервалами вигляду $S_i = [a_i, b_i)$. Порахуємо

$$\mathbf{P}(\xi \in S_i / H_0) = \mathbf{P}(a_i \leq \xi < b_i / H_0) = F_0(b_i) - F_0(a_i) \overset{def}{=} p_i.$$

Ясно, що $p_1 + ... + p_k = 1$. Завжди будемо вважати, що множини S_i вибрані так, що $p_i > 0$, $i = 1, ..., k$. Для фіксованого i процес здійснення вибірки можемо представити як схему Бернуллі: якщо значення потрапить до класу S_i, то говоримо, що маємо успіх, а інакше невдачу. Тоді, за умови, що гіпотеза H_0 є справедливою, ймовірність успіху буде p_i і середня кількість успіхів у такій схемі буде дорівнювати np_i. Ці числа для $i = 1, 2, ..., k$ називаємо частотами гіпотетичними. Число

$$n_i = \sum_{j=1}^{k} \chi_{S_i}(x_j) \tag{1}$$

є справжня кількість значення з вибірки $x_1, ..., x_n$, які потрапили до множини S_i і ці числа для $i = 1, 2, ..., k$ називаємо частотами експериментальними. Ідея Пірсона полягає в тому, що при справедливій гіпотезі H_0 різниця між частотами гіпотетичними і експериментальними не може бути великою. І в якості міри розбіжності ним була запропонована величина

$$\zeta(n) = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}. \tag{2}$$

Наступна теорема, відома як теорема Пірсона, описує асимптотичну поведінку випадкової змінної $\zeta(n)$ для $n \to \infty$.

Теорема 1. *Справедливе співвідношення*

$$\lim_{n\to\infty} \mathbf{P}(\zeta(n) < x/H_0) = \chi^2_{k-1}(x), \tag{3}$$

де $\chi^2_{k-1}(x)$, є функцією розподілу хі–квадрат з $k-1$ степенями свободи.

На підставі цієї теореми тест для перевірки гіпотези H_0 відносно альтернативи H_1 виглядає так: тестова статистика є

$$\zeta(n) = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$$

і сам тест тепер виглядає так

$$\sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \begin{cases} \leq k(1-\alpha, n-1), & \text{приймаємо} \quad H_0, \\ > k(1-\alpha, n-1), & \text{приймаємо} \quad H_1, \end{cases} \tag{4}$$

де $\chi^2_{n-1}(k(1-\alpha, n-1)) = 1-\alpha$, а α є рівнем значимості тесту.

Зауваження 1. На практиці, зазвичай, застосовуємо процедуру групування вибірки $x_1, ..., x_n$ тобто, створюємо розподільчий ряд і якщо $[a_i, b_i)$, $i = 1, ..., k$ класи, то тоді $p_1 = F_0(b_1)$, $p_i = F_0(b_i) - F_0(a_i)$, $i = 2, ..., k-1$, $p_k = 1 - F_0(a_k)$, а n_i - об'єми класів.

Як зазначено вище, застосування тесту (4) виглядає так. На підставі вибірки $x_1, x_2, \ldots x_n$ створюємо розподільчий ряд, як це було описано в підрозділі 2.2. Тоді в якості множин S_i виступають класи $[g_{i-1}, g_i)$, $i = 1, ..., k$. Але щоб виконувалась умова $\cup_{i=1}^{k}[g_{i-1}; g_i) = \mathbf{R}^1$, то як перший клас потрібно взяти $(-\infty, g_1)$, а останнім класом буде $[g_k; \infty)$. Розподільчий ряд, модифікований таким чином, виглядає тепер так

№ класу	Клас	Частоти (гіпотез.)	Ймовірн. (гіпотез.)	Частоти (експерим.)
1	$[-\infty; g_1)$	n_1	p_1	np_1
2	$[g_1; g_2)$	n_2	p_2	np_2
.....	....	....	......	.....
k	$[g_{k-1}; \infty)$	n_k	p_k	np_k
		$\sum n_i = n$	$\sum p_i = 1$	

$$p_1 = F_0(g_1), \quad p_k = 1 - F_0(g_{k-1}),$$
$$p_i = F_0(g_i) - F_0(g_{i-1}), \quad 2 \le i \le k-1.$$

Далі обраховуємо $\sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$ і застосовуємо тест (4).

Приклад 1. *В якості прикладу перевіримо гіпотезу, що вибірка з прикладу* 1 *с.* 165 *була зроблена з нормального розподілу* $\mathbf{N}(5; 0, 8)$.

Р о з в' я з о к. Розподільчий ряд для цієї вибірки представлений на с. 1. Зробимо модифікацію цього ряду так, як це було описано вище, проведемо необхідні обрахунки, результати яких зручно подати у формі наступної таблиці.

№ класу	Класи	Частоти (експерим.)	Ймовірн. (теорет.)	Частоти (гіпотет.)	$\frac{(n_i-np_i)^2}{np_i}$
1	$(-\infty; 3,45)$	3	0,027	1,35	2.017
2	[3,45;3,95)	5	0,068	3,4	0.753
3	[3,95;4,45)	8	0.151	7,55	0.027
4	[4,45;4,95)	9	0,229	11.45	0.524
5	[4,95;5,45)	12	0,238	11,9	0.001
6	[5,45;5,95)	8	0,169	8.45	0.024
7	$[5,95; \infty$	5	0.118	5.9	0.137
			$\sum p_i = 1$		$\sum = 3.483$

З таблиць беремо квантиль рівня $1 - 0,05 = 0,95$ розподілу $\chi^2_6(x)$ і маємо $k(1 - 0,05; 6) = 1,635$

Оскільки $3,483 > 1,635$, то гіпотеза H_0 мусить бути відкинута.

15.1.1 Тест Колмогорова

Тест Колмогорова.

Цей тест опирається на ідею порівняння теоретичного розподілу $F_0(x)$ з емпіричною функцією розподілу, сконструйованою на підставі вибірки $x_1, ..., x_n$. Ця функція була описана в підрозділі 8.1 і може бути записана у вигляді

$$Q_n(x) = \frac{1}{n} \sum_{i=1}^{n} \chi_{(-\infty,x)}(x_i). \tag{5}$$

З теореми Гливенка-Кантеллі с. 205 випливає, що

$$\mathbf{P}(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |Q_n(x) - F_0(x)| = 0/H_0) = 1,$$

і отже, малі значення статистики $D_n = \sup_{-\infty<x<\infty} |Q_n(x) - F_0(x)|$, будуть свідчити на користь гіпотези H_0, а великі значення D_n будуть підставою для відкидання гіпотези H_0. Якщо функція розподілу $F_0(x)$ є неперервною, то з теореми 5 с. 206 отримуємо $\lim_{n\to\infty} \mathbf{P}(\sqrt{n}D_n < z/H_0) = K(z$, де $K(z)$ є функцією розподілу Колмогорова.

Тест згідності Колмогорова виглядає так

$$\sqrt{n}D_n \begin{cases} \le k(1-\alpha), & \text{приймаємо} \quad H_0, \\ \\ > k(1-\alpha), & \text{відкидаємо} \quad H_0, \end{cases} \tag{6}$$

де $K(k(1-\alpha)) = 1 - \alpha$, а α є рівнем значимості.

На практиці труднощі застосування тесту Колмогорова пов'язані з підрахуванням статистики D_n. Зазвичай робимо так:

1) впорядковуємо вибірку за величиною $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$;
2) обраховуємо

$$D_n = \max \ \max_{1\le i\le n} |\frac{i}{n} - F_0(x_{(i)})|, \max_{1\le i\le n} |\frac{i-1}{n} - F_0(x_{(i)})|$$

і потім застосовуємо тест (6).

15.2 Тести незалежності

На практиці часто стикаємося з ситуацією, коли дослідженню підлягає пара випадкових змінних (ξ, η). Перше (і найважливіше) питання, яке тоді постає, є наступне: залежні ці випадкові змінні чи ні? Відповідь на це питання часто вирішує вибір підходів і методів подальших досліджень. Тому в цьому параграфі хочемо описати кілька тестів, які на підставі вибірки (x_i, y_i), $i = 1, ..., n$ з генеральної сукупності вигляду (ξ, η) служать для перевірки гіпотези

$$H_0 : \{\text{випадкові змінні } \xi,\ \eta \text{ є незалежними}\},$$

проти альтернативи

$$H_1 : \{\text{випадкові змінні } \xi,\ \eta \text{ є залежними}\}.$$

Одним з найчастіше вживаним тестів для розв'язування такого типу завдань є тест χ^2.

15.2.1 Тест незалежності χ^2.

Ідея цього тесту подібна до тесту Пірсона, описаного в параграфі 15.1.

Для деяких фіксованих натуральних $k \ge 2$, $m \ge 2$ поділимо вісь $OX = (-\infty, +\infty)$ на k множин S_i^x, $i = 1, ..., k$, які не перетинаються, а вісь $OY = (-\infty, +\infty)$ на m множин S_i^y, $i = 1, ..., m$, які теж не перетинаються. Нехай

$$\mathbf{P}(\xi \in S_i^x/H_0) \stackrel{def}{=} p_i, \qquad \mathbf{P}(\eta \in S_j^y/H_0) \stackrel{def}{=} q_i,$$
$$\mathbf{P}((\xi,\eta) \in (S_i^x, S_j^y)/H_0)) \stackrel{def}{=} p_{ij}.$$

І, якщо гіпотеза H_0 справедлива, то $p_{ij} = p_i q_j$. Нехай n_{ij} означає кількість пар (x_k, y_k), які потрапили в $S_i^x \times S_j^y$. Якби ймовірності p_i, q_j, p_{ij} були відомі, то, по аналогії з тестом Пірсона, статистика

$$\zeta(n) = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(n_{ij} - np_i q_j)^2}{np_i q_j} \tag{7}$$

могла б слугувати для перевірки гіпотези H_0 проти альтернативи H_1. Оскільки ці ймовірності нам невідомі, то замінимо їх оцінками вигляду

$$\widehat{p_i} = \frac{n_{i\cdot}}{n}, \qquad \widehat{q_j} = \frac{n_{\cdot j}}{n},$$

де $n_{i\cdot} = \sum_{j=1}^{m} n_{ij}$, $\quad n_{\cdot j} = \sum_{i=1}^{k} n_{ij}$. Тоді замість (7) дістаємо

$$\widehat{\zeta}(n) = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(n_{ij} - n_{i\cdot} n_{\cdot j}/n)^2}{n_{i\cdot} n_{\cdot j}/n} = n \left(\sum_{i=1}^{k} \sum_{j=1}^{m} \frac{n_{ij}^2}{n_{i\cdot} n_{\cdot j}} - 1 \right). \tag{8}$$

Теорема 2. *Справедливе співвідношення*

$$\lim_{n\to\infty} \mathbf{P}\left(n \left(\sum_{i=1}^{k} \sum_{j=1}^{m} \frac{n_{ij}^2}{n_{i\cdot} n_{\cdot j}} - 1 \right) < x / H_0\right) = \chi^2_{(k-1)(m-1)}(x),$$

де $\chi^2_{(k-1)(m-1)}(x)$ *є функцією розподілу* χ^2 *з* $(k-1)(m-1)$ *степенями свободи.*

Дана теорема випливає з теореми **??** с. **??**. Дійсно, ми маємо km вимірювань і $k + m - 2$ незалежних параметрів (Маємо k параметрів p_i, але оскільки має місце рівність $\sum_{i=1}^{k} p_i = 1$, то з цих параметрів лише $k - 1$ є незалежних. Так само є лише $m - 1$ незалежних параметрів q_i,) а тому граничний хі-квадрат розподіл має $km - (k + m - 2) - 1 = (k - 1)(m - 1)$ степенів свободи.

На підставі цієї теореми тест для перевірки гіпотези H_0 відносно альтернативи H_1 виглядає так

$$n \left(\sum_{i=1}^{k} \sum_{j=1}^{m} \frac{n_{ij}^2}{n_{i\cdot} n_{\cdot j}} - 1 \right) \begin{cases} \leq k(1-\alpha), & \text{приймаємо} \quad H_0, \\ > k(1-\alpha), & \text{відкидаємо} \quad H_0, \end{cases} \tag{9}$$

де $\chi^2_{(k-1)(m-1)}(k(1-\alpha)) = 1 - \alpha$, а α є рівнем значимості тесту.

Зауваження 2. Як і в тесті Пірсона, на практиці застосовуємо процедуру групування вибірок $x_1, ..., x_n$ і $y_1, ..., y_n$. І якщо $[a_i, b_i)$, $i = 1, ..., k$ класи для першої вибірки, а $[c_i, d_i)$, $i = 1, ..., m$ для другої, то n_{ij} є кількість пар, (x_l, y_l) які потрапили до прямокутника $[a_i, b_i) \times [c_j, d_j)$. Ясно, що мусить виконуватися рівність $n = \sum_{i=1}^{k} \sum_{j=1}^{m} n_{ij}$.

Більше на тему цього тесту можна знайти в [10], [12].

Приклад 1. *У відповідь на питання референдуму з групи в* 200 *жителів деякого міста "так" відповіло* 134 *особи а "ні"* 66 *осіб. На це ж питання з групи в* 150 *осіб жителів передмістя цього міста "так" відповіло 95 осіб а "ні"* 55 *осіб. Тестом хі-квадрат перевірити на рівні* $\alpha = 0,05$ *гіпотезу про те, що немає залежності між місцем проживання і відповіддю на питання референдуму.*

Р о з в' я з о к. Будемо вважати, що змінна ξ приймає два "значення": **житель міста** або **житель передмістя**, а змінна η теж приймає два значення: **так** або **ні**. У нашому випадку дані вже погруповано і подамо їх у вигляді наступної таблиці.

		η			
		так	ні		
ξ	житель міста	134	66	200	$n_{\cdot 1} = 200$
	житель передмістя	95	55	150	$n_{\cdot 2} = 150$
		$n_{1\cdot} = 229$	$n_{2\cdot} = 121$	$n = 350$	

Оскільки $k = m = 2$, то обраховуючи статистику з (9) маємо

$$n \left(\sum_{i=1}^{k} \sum_{j=1}^{m} \frac{n_{ij}^2}{n_{i\cdot} n_{\cdot j}} - 1 \right) = 350 \left(\frac{134^2}{229 \cdot 200} + \frac{66^2}{121 \cdot 200} + \frac{95^2}{229 \cdot 150} + \frac{55^2}{121 \cdot 150} - 1 \right) = 0,509.$$

Квантиль розподілу $\chi^2_{(2-1)(2-1)} = \chi^2_1$ рівня 0,95 знаходимо в таблиці 4 і він рівний 3.841. Отже, оскільки 0,95 $<$ 3,841, то згідно з тестом (9) у нас немає підстав вважати, що голосування залежить від місця проживання.

15.2.2 Тест нормальності Шапіро-Вілька

Легко звернути увагу на те, що багато результатів математичної статистики пов'язана з припущенням, що генеральна популяція має нормальний розподіл. Наступний тест, запропонований Семюелем Шапіро і Мартіном Вільком в 1965 р і опублікований в їхній статті [35], служить для перевірки гіпотези, що вибірка походить з генеральної сукупності з нормальним розподілом. Достоїнством цього тесту є те, що його можемо застосовувати для малих вибірок.

Отже, для вибірки $x_1, ..., x_n$ хочемо перевірити гіпотезу, що взята вона з нормального розподілу.

Тестова статистика для перевірки цієї гіпотези виглядає так

$$W = \frac{\left(\sum_{i=1}^{[n/2]} a_i(n)(x_{(n-i-1)} - x_{(i)}) \right)^2}{\sum_{i=1}^{n} (x_i - x(n))^2}, \tag{10}$$

де $a_i(n)$ є сталими числами, значення яких подано в Таблиці 12.

Гіпотезу про нормальність відхиляємо якщо значення статистики не потрапило до інтервалу $\left(w(\alpha/2, n), w(1-\alpha/2, n) \right)$, де квантилі $w(\gamma, n)$ розподілу статистики W подані в Таблиці 13.

Приклад 2.

Для вибірки $12,5; 14,1; 15,2; 15,7; 16,1; 19,2; 17,1; 21,2; 20,5; 13,2$ *на рівні* $\alpha = 0,1$ *перевірити гіпотезу про те, що походить вона з розподілу нормального.*

Р о з в' я з о к. Перепишемо, перш за все, нашу вибірку в зростаючому порядку 12,5; 13,2; 14,1; 15,2; 15,7; 16,1; 17,1; 19,2; 20,5; 21,2, а числа $a_i(n)$ візьмемо з таблиці 12. Тепер підрахунки зручно зобразити у вигляді наступної таблиці

i	$x_{(11-i)} - x_{(i)}$	$a_i(10)$	$a_i(10)\ x_{(10-i)} - x_{(i)}$
1	8,7	0,5739	4,9929
2	7,3	0,3291	2,4024
3	5,1	0,2141	1,0919
4	1,9	0,1224	0,2325
5	0,4	0,0399	0,0159
			$\Sigma = 8,7356$

Тепер можемо порахувати значення статистики з (10)

$$W = \frac{\sum\limits_{i=1}^{5} a_i(n)(x_{(n-i-1)} - x_{(i)})^{\ 2}}{\sum\limits_{i=1}^{n}(x_i - x(n))^2} = \frac{(8,7356)^2}{71,661} = 1,065,$$

З таблиці 13 знаходимо $w(0,05;10) = 0,842,\ w(0,95;10) = 0,978$. І оскільки у нашому випадку $W = 1,065 \notin (0,842;0,978)$, то гіпотезу про нормальний розподіл генеральної сукупності належить відхилити.

2.1. Довести формулу (**??**).

2.2. Довести формули **??**.

2.3. В деякому інституті провели тест з математики для трьох груп студентів по 100 осіб в кожній. Першу групу складали студенти, які проживають в гуртожитках, другу-студенти, які проживають в найманих мешканнях, а до третьої групи включено студентів які прожмвають дома. Отримано наступні результати.

Студенти першої групи. Оцінки "2", "3", "4", "5" отримали відповідно 10, 14, 50, 26 осіб.

Студенти другої групи. Оцінки "2", "3", "4", "5" отримали відповідно 13, 11, 52, 24 осіб.

Студенти третьої групи. Оцінки "2", "3", "4", "5" отримали відповідно 9, 16, 55, 20 осіб.

Тестом хі-квадрат перевірити на рівні $\alpha = 0,05$ чи існує залежність між міцем проживання та успішністю студента.

2.4. Навести приклад залежності між ξ, і η, щоб коефіцієнт рангової кореляції Спірмана для цих випадкових змінних був: а) 1, б)-1 .

2.5. Заміряно значення деяких характеристик (ξ, η) і отримано результати:

$$(1;41);(2;40);(3;35);(2;37);(5;35);(2;44);(4;43);(7;42);(8;43);(8;38);(10;36);(11;45);$$
$$(4;35);(9;37);(10;30);(14;26);(2;40);(5;36);(14;26);(8,42).$$

Перевірити чи існує залежність між ξ та η. Тест, а також його рівень значимості вибрати самостійно.

ЗАДАЧІ

2.6. Анкетування працівників деякого міністерства дало наступні результати: з 300 працівників адміністрації 180 належить до деякої організації А , а з 500 працівників, які не є адміністративними працівниками, 400 є членами цієї організації. На вірогідному рівні $\alpha = 0,05$ перевірити гіпотезу, що процент осіб, які належать до організації А є таким самим для цих двох груп працівників.

2.7. Поміряно час ослуговування одного покупця при касі в супермакеті і отримано наступні результати (в сек.) 23; 60; 54; 180; 47; 54; 125; 71; 36; 20; 58; 33; 67; 59; 89.

Тестом Шапіро-Вілка перевірити чи можна вважати, що з вірогідним рівнем $\alpha = 0,05$ цей час має нормальний розподіл?

2.8. З трьох відділень другого року навчання певного закладу випадково вибрано три групи студентів по 30 осіб і досліджено їх оцінки. Для кожної групи пораховано дисперсії оцінок і отримано: $S_1^2 = 1,24$, $S_2^2 = 1,20$ $S_1^2 = 1,31$. Припустивши, що розподіли оцінок є нормальними, перевірити тестами Бартлета та Хартлі з рівнем значимості $\alpha = 0,05$ гіпотезу про рівність дисперсій оцінок для всіх студентів цих трьох відділень.

2.9. Вибрано навмання 350 квартир в деякії місцевості і в 62 випадках це були квартири з стаціонарними телефонами. На рівні $\alpha = 0,05$ перевірити гіпотезу, що 40% квартир в цій місцевості мають стаціонарні телефони.

Лекція 16. Послідовні тести

В попередніх параграфах при перевірці головної гіпотези H_0 проти альтернативи H_1 для деякої характеристики ξ ми мали від початку вибірку $x_1, ..., x_n$ і на підставі цієї вибірки належало прийняти рішення про прийняття чи відкидання головної гіпотези. В цьому параграфі опишемо інший підхід до перевірки гіпотез, який полягає в тому, що значення характеристики ξ вибиратимемо поступово, по одному на кожному кроці. Так, на першому кроці маємо лише значення x_1 і на підставі цієї інформації мусимо прийняти одне з наступних рішень.

а) Приймаємо гіпотезу H_0;

б) Відхиляємо гіпотезу H_0 на користь альтернативи H_1;

в) Відкладаємо рішення про прийняття H_0 або H_1 і вибираємо наступне значення x_2 для ξ.

Отже у випадках а), б) перевірка закінчена. Зате в ситуації в) фактично стверджуємо, що ми не в стані прийняти однозначне рішення і для такого рішення потребуємо більше інформації. Після виконання наступного вимірювання маємо вибірку x_1, x_2 і знову мусимо прийняти одне з рішень а), б), в), але тепер на підставі вибірки x_1, x_2. Якщо випаде рішення в), то робимо чергове вимірювання x_3 і все повторюємо, маючи тепер вибірку x_1, x_2, x_3. Цей процес продовжуємо аж до моменту, коли буде прийняте одне з рішень а) або б). Така процедура перевірки гіпотез називається *послідовною.* Важливим припущенням (принаймні в цьому підручнику) буде факт, що всі вимірювання робимо за однакових умов і незалежним способом. Тому, як і раніше, значення $x_1, x_2, ...$ можемо трактувати як незалежні випадкові змінні з однаковим розподілом (таким самим, як і розподіл характеристики ξ.)

16.1 Послідовний тест Вальда

Тест, який буде описаний нижче, був запропонований і досліджений А. Вальдом[1] в його роботах [37], [40]. Ми розглянемо лише випадок простих гіпотез.

Головна гіпотеза H_0 : генеральна популяція ξ має функцію розподілу зі щільністю $f_0(x)$.

Альтернативна гіпотеза H_1 : генеральна популяція ξ має функцію розподілу зі щільністю $f_1(x)$.

Нехай $0 < A < 1 < B$ будуть два фіксованих числа, а $x_1, ..., x_n$ будуть значеннями з генеральної сукупності. Послідовний тест Вальда виглядає так: якщо на n-тому

[1]Вальд (*Abraham Wald*) (31.10.1902– 13.12.1950)– видатний статистик. Народився в Колошварі, Угорщина (тепер Клуж, Румунія). Закінчив університет у Відні. В 1938 р. емігрував до Сполучених Штатів. Загинув (разом з дружиною) в авіакатастрофі.

кроці

$$\prod_{i=1}^{n} \frac{f_1(x_i)}{f_0(x_i)} \begin{cases} \leq A & \text{приймаємо гіпотезу } H_0; \\ \geq B & \text{приймаємо гіпотезу } H_1; \\ \in (A,B) & \text{вибираємо } n+1- \text{ значення і} \\ & \text{переходимо до } n+1- \text{ого кроку.} \end{cases}$$

Нам буде зручно запровадити "нульовий" крок, тобто, тест починається з моменту $n = 0$. Тому за означенням вважаємо, що $\prod_{i=1}^{0} = 1$. Крім того, в наведеному тесті буде зручніше перейти до функції ln, що приводить до наступного рівнозначного вигляду послідовного тесту Вальда.

$$\sum_{i=1}^{n} \ln \frac{f_1(x_i)}{f_0(x_i)} \begin{cases} \leq -a & \text{приймаємо гіпотезу } H_0; \\ \geq b & \text{приймаємо гіпотезу } H_1; \\ \in (-a,b) & \text{вибираємо } n+1- \text{ значення і} \\ & \text{переходимо до } n+1- \text{ого кроку.} \end{cases} \tag{1}$$

де $a = -\ln A$, $b = \ln B$. Надалі будемо досліджувати послідовний тест Вальда у вигляді (1).

Означимо дві послідовності незалежних випадкових змінних з однаковим розподілом (в кожній послідовності)

$$z_i = \frac{f_1(x_i)}{f_0(x_i)}, \qquad \xi_i = \ln \frac{f_1(x_i)}{f_0(x_i)}, \qquad i = 1, \dots .$$

Ясно, що випадкові змінні ξ_i мають такий самий розподіл, як випадкова змінна $\ln \frac{f_1(\xi)}{f_0(\xi)}$. Момент завершення тесту Вальда може бути означений наступним чином

$$\tau = \inf\{n \geq 0 : \sum_{i=1}^{n} \xi_i \notin (A,B)\} = \inf\{n \geq 0 : \prod_{i=1}^{n} z_i \notin (-a,b)\}. \tag{2}$$

Очевидними питаннями, які виникають при застосуванні тесту (1), є такі:

1. Чи не трапиться так, що рішення про прийняття гіпотез H_0, H_1 ніколи не буде прийняте? Іншими словами, чи можлива ситуація, що $\mathbf{P}_i(\tau = \infty) > 0$, $i = 0, 1$?
2. Як вибирати сталі a, b (або, що те ж саме, A, B)?

Нижче доведемо, що за розумних припущень відповідь на перше питання є позитивною, а значення для a, b (чи для A, B) визначаються похибками першого (α) і другого (β) роду.

Теорема 1. *Якщо* $\mathbf{P}_i(|\xi_1|{>}0){>}0$, $i{=}0, 1$, *то* $\mathbf{P}_i(\tau = \infty){=}0$, $i = 0, 1$.

З цієї теореми випливає, що тест Вальда не закінчиться лише тоді, коли $f_0(x) = f_1(x)$, тобто коли $\xi_i \equiv 0$ з ймовірністю 1. Але така ситуація, очевидно, не є цікавою.

Д о в е д е н н я. Нехай $m > k$ будуть фіксовані натуральні числа і $r = [m/k]$.

Означимо випадкові величини

$$S_1 \overset{def}{=} \sum_{i=1}^{k} \xi_i, \qquad S_l \overset{def}{=} \sum_{i=(l-1)k+1}^{lk} \xi_i, \quad l = 2, 3, ..., r.$$

З означення випливає, що ці випадкові величини є незалежними і мають однаковий розподіл. Якщо $\tau > m$, то ясно, що $-a < \sum_{i=1}^{n} \xi_i < b$ для всіх $n = 1, ..., m$, тому

$$|S_l| = |\sum_{i=1}^{lk} \xi_i - \sum_{i=1}^{(l-1)k} \xi_i| \le |\sum_{i=1}^{lk} \xi_i| + |\sum_{i=1}^{(l-1)k} \xi_i| \le b + a = c, \quad l = 1, .., r.$$

З останнього співвідношення випливає, що подія $\{\tau > m\}$ тягне за собою подію $\cap_{l=1}^{r}\{|S_l| < c\}$, а оскільки події $\{|S_l| < c\}$, $l = 1, ..., r$ є незалежними, то маємо

$$\mathbf{P}_i(\tau > m) \le \mathbf{P}_i(\cap_{l=1}^{r}\{|S_l| < c\}) = \left(\mathbf{P}_i(|S_1| < c)\right)^r, \quad i = 0, 1. \tag{3}$$

З умови $\mathbf{P}_i(|\xi_1| > 0) > 0$, $i = 0, 1$ випливає, що завжди можемо знайти $h > 0$ таке, що $\mathbf{P}_i(\xi_1 \ge h) > 0$ або $\mathbf{P}_i(\xi_1 \le -h) > 0$. Нехай від початку $k > c/h$, а значить $c/k < h$. Ясно, що має місце включення

$$\{\xi_i \ge \frac{c}{k}, i = 1, 2, ..., k\} \cup \{\xi_i \le -\frac{c}{k}\} \subset \{|\xi_1 + ... + \xi_k| \ge c\}, \quad i = 1, 2, ..., k$$

і події в правій частині цієї рівності не перетинаються. Маємо

$$\begin{aligned}
&\mathbf{P}_i(|S_1|<c)=1-\mathbf{P}_i(|S_1|\ge c)=1-\mathbf{P}_i(|\xi_1+...+\xi_k|\ge c) \le \\
&\le 1-\mathbf{P}_i(\{\xi_i\ge\frac{c}{k}, i = 1, 2, ..., k\} \cup \{\xi_i\le-\frac{c}{k}, i = 1, 2, ..., k\})= \\
&= 1-\mathbf{P}_i(\xi_i\ge\frac{c}{k}, i = 1, 2, ..., k)-\mathbf{P}_i(\xi_i\le-\frac{c}{k}, i = 1, 2, ..., k)= \\
&= 1-\left(\mathbf{P}_i(\xi_1 \ge \frac{c}{k})\right)^k - \left(\mathbf{P}_i(\xi_1 \le -\frac{c}{k})\right)^k \le \\
&\le 1-\left(\mathbf{P}_i(\xi_1 \le h)\right)^k - \left(\mathbf{P}_i(\xi_1 \le -h)\right)^k \overset{def}{=} q < 1.
\end{aligned}$$

Тепер з (3) дістаємо $\mathbf{P}_i(\tau > m) \le q^r$, $i = 0, 1$. Ясно, що $r \to \infty$ коли $m \to \infty$, а тому $\mathbf{P}_i(\tau = \infty) = \lim_{m\to\infty} \mathbf{P}_i(\tau > m) \le \lim_{r\to\infty} q^r = 0, \quad i = 0, 1,$що завершує доведення теореми.

Нехай

$$T_n = \sum_{i=1}^{n} \xi_i = \sum_{i=1}^{n} \ln \frac{f_1(x_i)}{f_0(x_i)}, \qquad Z_n = \prod_{i=1}^{n} z_i = \prod_{i=1}^{n} \frac{f_1(x_i)}{f_0(x_i)}. \tag{4}$$

Якщо α, β є похибки першого і другого роду послідовного тесту Вальда, то з (2) випливає, що

$$\alpha = \mathbf{P}_0(T_\tau \ge b) = \mathbf{P}_0(Z_\tau \ge B), \quad \beta = \mathbf{P}_1(T_\tau \le -a) = \mathbf{P}_1(Z_\tau \le A).$$

Означимо множини

$$W_n = \{(u_1, ..., u_n) : A < \prod_{i=1}^{k} \frac{f_1(u_i)}{f_0(u_i)} < B,\ k = 1, 2, ..., n-1,\ \frac{f_1(u_n)}{f_0(u_n)} \geq B\} =$$
$$= \{(u_1, ..., u_n) : A < Z_k < B,\ k = 1, 2, ..., n-1,\ Z_n \geq B\},$$

$$V_n = \{(u_1, ..., u_n) : A < \prod_{i=1}^{k} \frac{f_1(u_i)}{f_0(u_i)} < B,\ k = 1, 2, ..., n-1,\ \frac{f_1(u_n)}{f_0(u_n)} \leq A\} =$$
$$= \{(u_1, ..., u_n) : A < Z_k < B,\ k = 1, 2, ..., n-1,\ Z_n \leq A\}.$$

Якщо $\vec{x}_n = (x_1, ..., x_n)$, $n \geq 1$ є вибірка, то легко зрозуміти, що

$$\{Z_\tau \geq B\} = \cup_{n=1}^{\infty}\{\vec{x}_n \in W_n\}, \quad \{Z_\tau \leq A\} = \cup_{n=1}^{\infty}\{\vec{x}_n \in V_n\}.$$

А тому

$$\alpha = \mathbf{P}_0(Z_\tau \geq B) = \sum_{n=1}^{\infty} \mathbf{P}_0(\vec{x}_n \in W_n) = \sum_{n=1}^{\infty} \int\limits_{W_n} f_0(\mathbf{u}_n)\mathbf{du}_n, \tag{5}$$

$$\beta = \mathbf{P}_1(Z_\tau \leq B) = \sum_{n=1}^{\infty} \mathbf{P}_1(\vec{x}_n \in V_n) = \sum_{n=1}^{\infty} \int\limits_{V_n} f_1(\mathbf{u}_n)\mathbf{du}_n, \tag{6}$$

де $f_i(\mathbf{u}_n) = \prod_{i=1}^{n} f_i(u_i)$, $\mathbf{u}_n = (u_1, ..., u_n)$. Зауважмо, що

$$\mathbf{P}_i(\tau = n) = \int\limits_{W_n} f_i(\mathbf{u}_n)\mathbf{du}_n + \int\limits_{V_n} f_i(\mathbf{u}_n)\mathbf{du}_n, \quad i = 0, 1$$

і отже для $i = 0, 1$ маємо

$$1 = \sum_{n=1}^{\infty} \mathbf{P}_i(\tau = n) = \sum_{n=1}^{\infty} \int\limits_{W_n} f_i(\mathbf{u}_n)\mathbf{du}_n + \sum_{n=1}^{\infty} \int\limits_{V_n} f_i(\mathbf{u}_n)\mathbf{du}_n. \tag{7}$$

Тепер з (5) отримуємо

$$\alpha = \mathbf{P}_0(Z_\tau \geq B) = \sum_{n=1}^{\infty} \int\limits_{W_n} f_0(\mathbf{u}_n)\mathbf{du}_n = \sum_{n=1}^{\infty} \int\limits_{W_n} \frac{1}{Z_n} f_1(\mathbf{u}_n)\mathbf{du}_n \leq$$
$$\leq \frac{1}{B} \sum_{n=1}^{\infty} \int\limits_{W_n} f_1(\mathbf{u}_n)\mathbf{du}_n \overset{(7)}{=} \frac{1}{B}\left(1 - \sum_{n=1}^{\infty} \int\limits_{V_n} f_1(\mathbf{u}_n)\mathbf{du}_n\right) = \frac{1-\beta}{B}. \tag{8}$$

Точно так само з(6) маємо

$$\beta = \mathbf{P}_1(Z_\tau \leq A) = \sum_{n=1}^{\infty} \int\limits_{V_n} f_1(\mathbf{u}_n)\mathbf{du}_n = \sum_{n=1}^{\infty} \int\limits_{V_n} Z_n f_0(\mathbf{u}_n)\mathbf{du}_n \leq$$
$$\leq A \sum_{n=1}^{\infty} \int\limits_{V_n} f_0(\mathbf{u}_n)\mathbf{du}_n \overset{(7)}{=} A\left(1 - \sum_{n=1}^{\infty} \int\limits_{W_n} f_0(\mathbf{u}_n)\mathbf{du}_n\right) = A(1-\alpha). \tag{9}$$

Наступна теорема, яка випливає з (8), (9) дає нерівності для A, B.

Теорема 2. *Якщо α, β є похибки першого і другого роду послідовного тесту Вальда, то*

$$A \geq \frac{\beta}{1-\alpha}, \qquad B \leq \frac{1-\beta}{\alpha}. \tag{10}$$

Зауваження 1. Легко зауважити, що нерівності в (8), (9) виникли тоді, коли замість Z_n підставили найменше значення B (у випадку (8)) і найбільше значення A (у випадку (9)). Фактично така заміна не враховує перестриб випадковою змінною Z_n рівня B (у випадку (8)) і рівня A (у випадку (9)). Якби такого перестрибу не було, то в (8), (9), а значить і в (8) мали б місце рівності. Врахувати цей перестриб в загальній ситуації (і, значить, отримати точнішу оцінку для A, B) є дуже трудною проблемою і тому в застосуваннях тесту Вальда цим перестрибом нехтуємо, що дає нам наступні значення для A, B:

$$A = \frac{\beta}{1-\alpha}, \qquad B = \frac{1-\beta}{\alpha}. \tag{11}$$

Оскільки тривалість тесту Вальда τ є випадковою змінною, то виникає питання про середню тривалість цього тесту. Вивчимо ці питання.

Середня тривалість тесту Вальда. Тут нам буде корисною наступна лема, в якій рівність (12) називається *тотожністю Вальда.*

Лема 1. *Якщо* $\mathbf{M}i|\xi_1| = \mathbf{M}i|\ln\frac{f_1(\xi)}{f_0(\xi)}| < \infty$, *то*

$$\mathbf{M}iT_\tau = \mathbf{M}i\tau\mathbf{M}i\xi_1, \qquad i = 0, 1, \tag{12}$$

а T_n є означеним в (4).

Зауваження 2. Зауважмо, що якби в (12) замість τ було довільне натуральне число n (не випадкове!), то ця рівність була б очевидною, оскільки з властивості значення математичного сподівання випливає, що

$$\mathbf{M}iT_n = \mathbf{M}i\sum_{j=1}^{n}\xi_j = \sum_{j=1}^{n}\mathbf{M}i\xi_j = n\mathbf{M}i\xi_1.$$

Д о в е д е н н я. Означимо випадкову змінну η_i таким чином: $\eta_1 = 1$ а для $n \geq 2$

$$\eta_n = \begin{cases} 1, & \text{якщо на підставі значень } x_1, ..., x_{n-1} \text{ рішення не прийняте,} \\ 0, & \text{на підставі значень } x_1, ..., x_{n-1} \text{ рішення прийняте.} \end{cases}$$

Легко зрозуміти, що випадкова змінна η_n, $n \geq 1$ не залежить від x_k, $k \geq n$ (а значить і від ξ_k, $k \geq n$) і

$$\mathbf{P}_i(\eta_n = 1) = \mathbf{P}_i(\tau \geq n), \qquad T_\tau = \xi_1\eta_1 + \xi_2\eta_2 + ...$$

Маємо

$$\mathbf{M}iT_\tau = \mathbf{M}i\sum_{n=1}^{\infty}\xi_n\eta_n = \sum_{n=1}^{\infty}\mathbf{M}i\xi_n\eta_n = \sum_{n=1}^{\infty}\mathbf{M}i\xi_n\mathbf{M}i\eta_n = \mathbf{M}i\xi_1\sum_{n=1}^{\infty}\mathbf{M}i\eta_n =$$

$$= \mathbf{M}i\xi_1\sum_{n=1}^{\infty}\mathbf{M}i\eta_n = \mathbf{M}i\xi_1\sum_{n=1}^{\infty}\mathbf{P}_i(\tau \geq n) = \mathbf{M}i\xi_1\mathbf{M}i\tau.$$

З (12) маємо формулу для середньої тривалості тесту Вальда

$$\mathbf{M}i\tau = \frac{\mathbf{M}iT_\tau}{\mathbf{M}i\xi_1}, \qquad i = 0, 1.$$

Проблема лише в тому, що значення $\mathbf{M}iT_\tau$, тобто, середнє значення суми $\sum_{i=1}^{n} \xi_i$ в перший момент, коли ця сума лежить поза інтервалом $(-a, b)$, в загальній ситуації не є відомим. Тому робимо подібно тому, як було сказано в зауваженні 2. Ясно, що $T_\tau \geq b$ або $T_\tau \leq -a$. Тому

$$\begin{aligned}\mathbf{M}_0 T_\tau &= \mathbf{M}_0\{T_\tau/T_\tau \geq b\}\mathbf{P}_0(T_\tau \geq b) + \mathbf{M}_0\{T_\tau/T_\tau \leq -a\}\mathbf{P}_0(T_\tau \leq -a) = \\ &= \mathbf{M}_0\{T_\tau/T_\tau \geq b\}\alpha + \mathbf{M}i\{T_\tau/T_\tau \leq -a\}(1-\alpha). \qquad (13)\end{aligned}$$

Якщо знехтувати перескоком змінною T_τ рівнів b, $-a$, то тоді $\mathbf{M}i\{T_\tau/T_\tau \geq b\} \approx b$, $\mathbf{M}i\{T_\tau/T_\tau \leq -a\} \approx -a$, і тоді з (13) маємо $\mathbf{M}_0 T_\tau \approx b\alpha - a(1-\alpha)$ і отже

$$\mathbf{M}_0\tau \approx \frac{(b+a)\alpha - a}{\mathbf{M}_0\xi_1}. \qquad (14)$$

Точно так само можемо вивести наближену формулу

$$\mathbf{M}_1\tau \approx \frac{b - (b+a)\beta}{\mathbf{M}_1\xi_1}. \qquad (15)$$

В [14] можна знайти доведення наступної теореми.

Теорема 3. *Нехай $\mathbf{P}_i(\xi_1 \neq 0) > 0$, $\mathbf{M}i|\xi_1| < \infty$ і α, β є похибками першого і другого роду тесту Вальда. Тоді*

$$\mathbf{M}_0\tau \geq -\frac{H(\alpha,\beta)}{\mathbf{M}_0\xi_1}, \qquad \mathbf{M}_1\tau \geq -\frac{H(\beta,\alpha)}{\mathbf{M}_1\xi_1},$$

де $H(x,y) = x\ln\frac{x}{y} + (1-x)\ln\frac{1-x}{1-y}$.

Більше про тест Вальда можна знайти в , [6],[8], [14], [44], [40].

16.1.1 Послідовний тест для показника структури популяції

Приклад 1. *Генеральна популяція ξ має розподіл типу нуль-один з невідомим параметром p, про який знаємо, що $p = p_0$ або $p = p_1 \neq p_0$. Побудувати послідовний тест Вальда перевірки головної гіпотези $H_0: \ p = p_0$, відносно альтернативи $H_1: \ p = p_1 \neq p_0$.*

Р о з в' я з о к. Нехай, для визначеності, $p_0 < p_1$. При гіпотезі H_0 популяція ξ має розподіл $\mathbf{P}_0(\xi = x) = p_0^x q_0^{1-x}$, а при гіпотезі H_1 - ξ має розподіл $\mathbf{P}_1(\xi = x) = p_1^x q_1^{1-x}$, де $0 \leq p_i \leq 1$, $q_i = 1 - p_i$, $i = 0, 1$, а $x = 0$ або 1.

В цій ситуації відношення функції вірогідності виглядає так

$$z_n = \frac{p_1^{\sum_{i=1}^n x_i} q_1^{n-\sum_{i=1}^n x_i}}{p_0^{\sum_{i=1}^n x_i} q_0^{n-\sum_{i=1}^n x_i}} = \frac{p_1^k q_1^{n-k}}{p_0^k q_0^{n-k}},$$

де k є кількість одиниць в n дослідах (n-тий етап дій). Отже маємо

$$T_n = \ln\frac{p_1^k q_1^{n-k}}{p_0^k q_0^{n-k}} = k\ln\frac{p_1}{p_0} + (n-k)\ln\frac{q_1}{q_0}.$$

Якщо α, β є похибки першого і другого роду, то $-a = \ln A = \ln\frac{\beta}{1-\alpha}, \quad b = \ln B = \ln\frac{1-\beta}{\alpha}$. Тепер, згідно з (1), маємо наступні дії на n-тому кроці:

а) $k\ln\frac{p_1}{p_0} + (n-k)\ln\frac{q_1}{q_0} \le -a$ - приймаємо гіпотезу H_0;

б) $k\ln\frac{p_1}{p_0} + (n-k)\ln\frac{q_1}{q_0} \ge b$ - приймаємо гіпотезу H_1;

в) $k\ln\frac{p_1}{p_0} + (n-k)\ln\frac{q_1}{q_0} \in (-a, b)$ - вибираємо наступне значення.

Нехай

$$\begin{aligned} D &= -\ln\frac{1-p_1}{1-p_0}\ \ \ln\frac{p_1}{p_0} - \ln\frac{1-p_1}{1-p_0}\ {}^{-1} \\ B_0 &= \ln\frac{\beta}{1-\alpha}\ \ \ln\frac{p_1}{p_0} - \ln\frac{1-p_1}{1-p_0}\ {}^{-1}, \\ B_1 &= \ln\frac{1-\beta}{\alpha}\ \ \ln\frac{p_1}{p_0} - \ln\frac{1-p_1}{1-p_0}\ {}^{-1} \end{aligned}$$

Якщо розв'язати нерівності в пунктах а)-в), то тест описаний вище можемо переписати в рівнозначному вигляді

а) $k \le Dn + B_0$ - приймаємо гіпотезу H_0;

б) $k \ge Dn + B_1$ - приймаємо гіпотезу H_1;

в) $Dn + B_0 < k < Dn + B_1$ - вибираємо наступне значення.

ЗАДАЧІ

1.1. Знайти середню тривалість тесту Вальда з прикладу 1, якщо $p_0 = 1/3$ а $p_1 = 1/2$.

1.2. Сконструювати тест Вальда, якщо генеральна популяція ξ має нормальний розподіл $N(m, \sigma^2)$ з невідомою σ і відомим m. Головна гіпотеза $H_0: \ \sigma = \sigma_0$, а альтернативна $H_1: \ \sigma = \sigma_1 > \sigma_0$. Похибки першого та другого родів взяти наступними $\alpha = 0,05, \ \beta = 0,01$

1.3. Сконструювати тест Вальда, якщо генеральна популяція ξ має показниковий розподіл з параметром λ. Головна гіпотеза $H_0: \ \lambda = 2$, а альтернативна $H_1: \ \lambda = 3$.

Похибки першого та другого родів взяти наступними $\alpha = 0,05, \ \beta = 0,01$

1.4. Вибираючи навмання товар з певної партії і контролюючи його якість отримали наступні результати: 0;0;1;0;1;0;0;1;0;0;0;0;0;1;0;0;0;1;0;0;0;0;0;1, де 1 означає брак, а 0-добрий товар. Нехай p буде процент браку в цій партії товару. Послідовним тестом перевірити гіпотезу $p = 4\%$ проти альтернативи $p = 8\%$.

Похибки першого та другого родів взяти наступними $\alpha = 0,01, \ \beta = 0,05$.

ВИПАДКОВІ ПРОЦЕСИ

Випадкові процеси мають, як відомо, багато застосувань в різних галузях науки і техніки. Фундаментом теорії випадкових процесів є теорія ймовірностей.

Під час роботи над цим посібником автори ставили собі за мету насамперед те, щоби:

a) матеріал був доступним для студентів прикладних спеціальностей;
b) посібник був написаний в стислий, але не спрощений спосіб, як це часто можна спостерігати в підручниках для технічних вузів, і представляв найважливіші класи випадкових процесів, методи їх досліджень, а також сфери застосувань.

Ці дві умови мали вирішальний вплив на вибір матеріалу для лекцій і спосіб їхнього подання.

Виховання математичної культури є неможливим без пізнання і зрозуміння доведення тверджень. Ось чому в посібнику крім визначень важливих класів випадкових процесів та їх властивостей є наведеними також їхні доведення. Вони повинні бути зрозумілими для читачів, які мають базові знання з математичного аналізу та лінійної алгебри.

У представленому посібнику не використовуються поняття ймовірносної міри, міри та інтегралу Лібега, умовного математичного сподівання стосовно σ-алгебри, строга марковська властивість, момент зупинки процесу та деякі інші поняття, які зустрічаються в загальній теорії випадкових процесі.

Звісно, навіть короткий огляд всіх напрямків теорії випадкових процесів був би неможливим. Через це автори зосередились на тих класах процесів, які мають широке застосування на практиці.

Це насамперед ланцюги Маркова з дискретним або неперервним часом й скінченною чи зліченною множиною станів, гіллясті процеси, процеси народження та загибелі, процес Пуассона та процес Вінера. Значна частина матеріалу присвячена застосуванням цих процесів у теорії ризику і теорії масового ослуговування. У двох лекціях мова йде про основи теорії відновлення. Теорія випадкових процесів початково була частиною теорії ймовірностей. Допіру зацікавленість новим поняттям, тобто поняттям траєкторії процесу і проблемами, які з нею пов'язані, спричинила відокремлення теорії випадкових процесів від теорії ймовірностей. Ось чому в першій лекції висвітлюються проблеми, які стосуються властивостей траєторій.

Лекція 1. Елементи загальної теорії випадкових процесів

1.1 Визначення і скінченно вимірні розподіли стохастичного процесу

Нехай $(\mathbf{\Omega}, \mathcal{B}, \mathbf{P})$ є ймовірносним простором, а T - множиною або підмножиною дійсних чисел.

Визначення 1. *Випадковий процес* $\xi(t)$ є сімейством випадкових величин $\xi(t) = \xi(t, \omega)$, $t \in T$, $\omega \in \mathbf{\Omega}$, залежних від параметру t і визначених на ймовірносному просторі $(\mathbf{\Omega}, \mathcal{B}, \mathbf{P})$.

Зміст цього визначення є наступним, не кожна числова функція $\xi(t, \omega)$ двох аргументів $t \in T$, $\omega \in \mathbf{\Omega}$ є стохастичним процесом, а тільки така, в якої для кожного фіксованого $t_0 \in T$, $\xi(t_0, \omega)$ є випадковою величиною на ймовірносному просторі $(\mathbf{\Omega}, \mathcal{B}, \mathbf{P})$ в сенсі поданого вище визначення випадкової величини.

За множину T найчастіше розглядають дві множини: T=$[0, \infty)$–у цьому випадку будемо говорити про стохастичний процес з *неперервним часом* або про процес з *дискретним часом*, якщо T=$\{0, 1, 2, ...\}$. В останньому випадку будемо мати справу з послідовністю випадкових величин $\xi_n(\omega) = \xi(n, \omega)$.

Якщо $\omega_0 \in \Omega$ є зафіксованим, тоді $\xi(t, \omega_0)$ можна розглядати як звичайну функцію змінної t. Ця функція називається *траєкторією* або *реалізацією* процесу $\xi(t)$. Отже, кожному $\omega \in \Omega$ відповідає певна траєкторія.

Розглянемо кілька прикладів стохастичних процесів.

Приклад 1. *Нехай $\alpha_k(\omega), \beta_k(\omega), k = 1, ..., n$ є випадковими величинами, визначеними на ймовірносному просторі $(\mathbf{\Omega}, \mathcal{B}, \mathbf{P})$. Рівність*

$$\xi(t, \omega) = \sum_{k=1}^{n} \alpha_k(\omega) \sin kt + \beta_k(\omega) \cos kt \ , \qquad t \geq 0$$

визначає певний стохастичний процес. Якщо $\omega_0 \in \mathbf{\Omega}$ є фіксованим, тоді реалізація $\xi(t, \omega_0)$ є тригонометричним многочленом.

Приклад 2. *Розглянемо випадковий процес $\xi(t)$, який з однаковими ймовірностями на інтервалі $[n, n+1)$ незалежно від значень $\xi(t)$ на попередніх інтервалах приймає значення 0 або 1. Будову цього процесу можна представити наступним чином. Припустимо, що в кожний момент $n = 0, 1, 2, ...$ відбуваються незалежні випробування з двома результатами: A і $\overline{A}$. Нехай $\mathbf{P}(A) = p \in [0, 1]$ і $\mathbf{P}(\overline{A}) = 1 - p$. Для $t \in [n, n+1)$ покладемо*

$$\xi(t) = \begin{cases} 1, & \textit{в момент } n \textit{ сталася подія} \quad A, \\ \\ 0, & \textit{в момент } n \textit{ сталася подія} \quad \overline{A}. \end{cases}$$

У такий спосіб визначемо процес $\xi(t)$ для всіх $t \geq 0$.

Нехай для фіксованих $t_1, ...t_n \in T$, $n \geq 1$, $\xi(t_1), ..., \xi(t_n)$ є послідовністю випадкових величин на ймовірносному просторі $(\mathbf{\Omega}, \mathcal{B}, \mathbf{P})$. Припустимо також, що

$$F_{\vec{t}_n}(\vec{x}_n) = F_{t_1,t_2,...,t_n}(x_1, x_2, ..., x_n) \stackrel{def}{=} \mathbf{P}(\xi(t_1) \leq x_1, ..., \xi(t_n) \leq x_n), \tag{1}$$

де $\vec{t}_n = (t_1, ..., t_n), \vec{x}_n = (x_1, ..., x_n)$ і $x_1, ..., x_n \in \mathbf{R}^1$ є багатовимірною функцією розподілу цих випадкових величин. Для процесу $\xi(t)$ n-вимірна функція розподілу $F_{\vec{t}_n}(\vec{x}_n)$ називається *скінченно вимірним розподілом* цього процесу.

З теорії ймовірностей відомо ([**?**] стор. 120), що функція розподілу $F_{\vec{t}_n}(\vec{x}_n)$ має наступні властивості:

1. Якщо $\{i_1, ..., i_n\}$ є перестановкою чисел $1, ..., n$, тоді

$$F_{t_1,t_2,...,t_n}(x_1, x_2, ..., x_n) = F_{t_{i_1},t_{i_2},...,t_{i_n}}(x_{i_1}, x_{i_2}, ..., x_{i_n}). \tag{2}$$

2. Якщо $m < n$, тоді

$$F_{t_1,t_2,...,t_m,t_{m+1},...,t_n}(x_1, x_2, ..., x_m, \infty, ..., \infty) = F_{t_1,t_2,...,t_m}(x_1, x_2, ..., x_m). \tag{3}$$

Скінченно вимірний розподіл $F_{\vec{t}_n}(\vec{x}_n)$ з (1) визначається для довільних $n \geq 1$ і довільних $t_1, ..., t_n \in T$. Наступна множина

$$\mathcal{F}_\xi = \{F_{\vec{t}_n}(\vec{x}_n),\ n \geq 1,\ t_1, ..., t_n \in T,\ x_1, ..., x_n \in R^1\} \tag{4}$$

називається *сімейством скінченно вимірних розподілів* процесу $\xi(t)$. Часом кажуть, що $\mathcal{F}_\xi$ є сімейством скінченно вимірних розподілів, *згенерованих процесом* $\xi(t)$ або, яка *відповідає* процесу $\xi(t)$.

Вочевидь для кожного розподілу $F_{\vec{t}_n}(\vec{x}_n)$ з (4) виконуються умови (2), (3). Ці умови називаються *умовами узгодженості* сімейства $\mathcal{F}_\xi$. Маємо наступний результат.

Теорема 1. *Кожному процесу $\xi(t)$ відповідає сімейство скінченно вимірних розподілів, яке породжується цим процесом, і для якого виконані умови угодженості.*

Дуже важливим є також зворотній результат. Але для його формулювання нам потрібне наступне визначення.

Визначення 2. *Будемо говорити, що множина функцій*

$$\mathcal{G} = \{G_{t_1,t_2,...,t_n}(x_1, x_2, ..., x_n),\ t_1, ..., t_n \geq 0,\ x_1, ..., x_n \in R^1,\ n \geq 1\}$$

є сімейством скінченно вимірних функцій, якщо для кожних фіксованих $t_1, ..., t_n \geq 0$, $n \geq 1$ функція $G_{t_1,t_2,...,t_n}(x_1, x_2, ..., x_n)$ є n-вимірною функцією розподілу стосовно $x_1, ..., x_n$.

Теорема 2. *Якщо для сімейства скінченно вимірних розподілів виконуються умови узгодженості, тоді існує ймовірносний простір $(\mathbf{\Omega}, \mathcal{B}, \mathbf{P})$, а також визначений на ньому випадковий процес $\xi(t, \omega)$, для якого це сімейство є сімейством скінченно вимірних розподілів даного процесу.*

Це твердження є аналогом твердження з теорії ймовірностей про існування випадкової величини, яка має задану функцією розподілу – якщо функція $F(x), -\infty < x < \infty$ є неспадною, неперервною справа і такою, що $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$, тоді існує ймовірносний простір $(\mathbf{\Omega}, \mathcal{B}, \mathbf{P})$ та випадкова величина ξ, визначена на цьому просторі, що $F(x) = \mathbf{P}(\xi \leq x)$.

Визначення 3. Процеси $\xi_1(t)$, $\xi_2(t)$ будемо називати *стохастично еквівалентними* (або *еквівалентними*), якщо для довільного $t \geq 0$ має місце

$$\mathbf{P}(\xi_1(t) \neq \xi_2(t)) = 0.$$

Теорема 3. *Якщо процеси $\xi_1(t)$, $\xi_2(t)$ є стохастично еквівалентні, тоді $\mathcal{F}_{\xi_1} = \mathcal{F}_{\xi_2}$.*

Іншими словами, якщо процеси є стохастично еквівалентними, тоді їхні сімейства скінченно вимірних розподілів співпадають. Доведення цього твердження залишаємо читачеві (завдання 2.2.).

Визначення 4. Процеси $\xi_1(t), \xi_2(t)$ будемо називати *стохастично ідентичними* (або попросту *ідентичні*), якщо має місце рівність

$$\mathbf{P}(\xi_1(t) \neq \xi_2(t), \text{ і } t \geq 0) = 0. \tag{5}$$

Легко зрозуміти, якщо процеси є ідентичними, тоді вони одночасно є стохастично еквівалентніими. Зворотній висновок не є слушним (завдання 2.3).

1.2 Процеси з траекторіями без розривів другого роду

У теорії ймовірностей об'єктами досліджень є випадкові величини, функції розподілу тощо. Зі стохастичними процесами пов'язазано нове і дуже важливе поняття, а саме поняття траекторії цього процесу, тобто функції $\xi(t, \omega)$ аргументу $t \geq 0$ для кожного фіксованого $\omega \in \Omega$.

Дослідження властивостей траекторій випадкового процесу є однією з головних задач. Твердження 2 інформує тільки про існування випадкового процесу $\xi(t)$ з певним сімейством скінченно вимірних розподілів, але не надає відомостей про властивості його траекторій, а ці властивості є дуже суттєвими. Це можна побачити на наступному простому прикладі. Припустимо, що ми хочемо знайти ймовірність події

$$A = \{\omega : a \leq \xi(s, \omega) \leq b, \text{ dla wszystkich } s \in [c, d]\}.$$

Вочевидь

$$A = \cap_{c \leq s \leq d} \{\omega : a \leq \xi(s, \omega) \leq b\}. \tag{6}$$

Згідно з визначенням випадкового процесу множина $\{\omega : a \leq \xi(s, \omega) \leq b\}$ є подією, тобто є елементом σ-алгебри $\mathcal{B}$ для кожного фіксованого $s \in [c, d]$, отже ймовірність $\mathbf{P}(a \leq \xi(s, \omega) \leq b)$ є визначеною. З властивостей σ-алгебри $\mathcal{B}$ випливає, що добуток *зліченного* числа подій з $\mathcal{B}$ належить також до $\mathcal{B}$, тобто є подією, але множина A з (6) є добутком *незліченного* числа подій і може не належати до $\mathcal{B}$, тобто не бути подією, тоді ймовірність $\mathbf{P}(A)$ не буде визначеною.

Натомість, якби траекторії процесу $\xi(t, \omega)$ були неперервними, тоді множина A була би подією. Неважко зрозуміти, що

$$A = \cap_{s \in Q(c,d)} \{\omega : a \leq \xi(s, \omega) \leq b\},$$

де $Q(c,d)$ є *зліченною* множиною раціональних чисел, які містяться в інтервалі $[c,d]$. З цієї рівності випливає, що множина A належить до $\mathcal{B}$, отже ймовірність $\mathbf{P}(A)$ є визначеною.

З наведеного прикладу випливає, що бажано, аби пов'язані з процесом $\xi(t)$ події, можна було описати, беручи до уваги лише значення процесу $\xi(t)$ у не більш ніж зліченному числі моментів t. Аналізуючи останній приклад, робимо висновок, що саме так і може бути, якщо траєкторії процесу мають певні властивості регулярності. З іншої сторони спільні розподіли значень процесу $\xi(t)$ у не більш ніж зліченному числі t генерують $\mathcal{F}_\xi$- сімейство скінченно вимірних розподілів цього процесу. Позаяк кожному сімейству скінченно вимірних розподілів відповідає цілий клас еквівалентних випадкових процесів, тоді природньо виникає питання: чи є в цьому класі процес, траекторії якого мають певні властивості регулярності? За певних додаткових умов відповідь на це питання є позитивною. Для формулювання відповідного твердження нам будуть потрібні певні поняття.

Визначення 5. Процес $\xi(t)$ називається *стохастично неперервним* у точці t_0, якщо для довільного $\varepsilon > 0$ має місце властивість:

$$\lim_{t\to t_0} \mathbf{P}(|\xi(t) - \xi(t_0)| > \varepsilon) = 0.$$

Якщо процес є стохастично неперервним у кожній точці $t \in A \subset [0,\infty)$, тоді будемо говорити, що він є стохастично нереревним на A.

Стохастична неперервність процесу не означає, що траекторії цього процесу є неперервними. Випадковий процес може бути стохастично неперервним, але всі його траекторії не є неперервними (завдання 2.4).

Визначення 6. Випадковий процес $\xi(t)$ називається *стохастично обмеженим* на $A \subset [0,\infty)$, якщо має місце властивість:

$$\lim_{K\to\infty} \sup_{t\in A} \mathbf{P}(|\xi(t)| > K) = 0.$$

Поняття стохастичної неперервності і обмеженості є подібними до відомих з математичного аналізу понять неперервності та обмеженості функцій. Так само аналогічними є твердження, що стосуються цих понять. Наприклад, якщо звичайна функція є неперервною на замкненому інтервалі, тоді вона є обмеженою на цьому інтервалі. Подібне твердження є справедливим і для випадкових процесів.

Теорема 4. *Якщо випадковий процес $\xi(t)$ є стохастично неперервним на інтервалі $[a,b]$, тоді він є стохастично обмеженим на цьому інтервалі.*

Д о в е д е н н я. Нехай $\varepsilon > 0$ є довільним малим числом. Позаяк $\mathbf{P}(|\xi(t)-\xi(s)| > 1) \longrightarrow 0,\ t \to s$ для довільного $s \in [a,b]$, тоді для кожного $s \in [a,b]$ існують околиці $O(s)$ вигляду $(s-\delta, s+\delta)$, (для точок a, b вигляду $[a, a+\delta), (b-\delta, b]$) такі, що $\mathbf{P}(|\xi(t) - \xi(s)| > 1) \le \varepsilon,\ t \in O(s)$. Множини $O(s),\ s \in [a,b]$ створюють покриття замкненого інтервала $[a,b]$. З математичного аналізу відомо про існування скінченного покриття цього інтервалу, тобто існують точки $s_i,\ i = 1,2,..,n$ такі, що $[a,b] = \cup_{i=1}^n O(s_i)$ і

$$\mathbf{P}(|\xi(t) - \xi(s_i)| > 1) \le \varepsilon,\ t \in O(s_i),\ i = 1,2,...n. \tag{7}$$

Візьмемо $t \in [a, b]$. Позаяк $[a, b] = \cup_{i=1}^{n} O(s_i)$, тоді $t \in O(s_m)$ для визначеного $1 \le m \le n$, і згідно з (7) будемо мати (див. завдання 2.5.)

$$\mathbf{P}(|\xi(t)|>K)=\mathbf{P}(|\xi(t)-\xi(s_m)+\xi(s_m)|>K) \le \mathbf{P}(|\xi(t)-\xi(s_m)| > 1)+ \\ +\mathbf{P}(|\xi(s_m)|>K-1) \le \varepsilon + \max_{1\le i\le n} \mathbf{P}(|\xi(s_i)| > K-1).$$

Звідки

$$\sup_{t\in A} \mathbf{P}(|\xi(t)| > K) \le \varepsilon + \max_{1\le i\le n} \mathbf{P}(|\xi(s_i)| > K-1).$$

Позаяк $\mathbf{P}(|\xi(s_i)| > K-1) \to 0$, якщо $K \to \infty$ для кожного s_i, $i = 1, ..., n$, отже з останньої нерівності будемо мати:

$$\lim_{K\to\infty} \sup_{t\in A} \mathbf{P}(|\xi(t)| > K) \le \varepsilon,$$

що завершує доведення. ◀

Визначимо наступний клас функцій:

$$D_{[a,b]} = \{f(t) : \exists \lim_{t\uparrow t_0} f(t),\ \lim_{t\downarrow t_0} f(t) = f(t_0),\ a \le t_0 \le b\}.$$

Іншими словами $D_{[a,b]}$ є класом функцій, визначених на інтервалі $[a, b]$, які не мають розривів другого роду і неперервних справа .

Зазвичай в теорії випадкових процесів припускають, що траєкторії процесу належать $D_{[a,b]}$ для довільних $0 \le a < b < \infty$. Відомі різні достатні умови для виконання цієї умови. Наведемо без доведення деякі з них.

Теорема 5. *Нехай випадковий процес $\xi(t)$ є стохастично неперервним і таким, що*

$$\mathbf{M}|\xi(t)-\xi(u)|^p|\xi(u)-\xi(s)|^q \le C|t-s|^{1+r} \tag{8}$$

для визначених $C, p, q, r > 0$ і будь-яких $a \le s < u < t \le b$. Тоді існує випадковий процес $\widehat{\xi}(t)$, який є стохастично еквівалентним процесу $\xi(t)$, і його траєкторії належать $D_{[a,b]}$.

Це твердження має наступний сенс: якщо випадковий процес $\xi(t)$ є стохастично неперервним, і виконується умова (8), тоді серед усіх процесів, які є еквівалентними $\xi(t)$ (отже, мають однакові сімейства скінченно вимірних розподілів), можна знайти процес, траєкторії якого є неперервними справа і мають границі зліва.

Теорема 6. *Нехай випадковий процес $\xi(t)$ є таким, що*

$$\mathbf{M}|\xi(t)-\xi(s)|^p \le C(t-s)^{1+r} \tag{9}$$

для визначених $C, p, r > 0$ і довільних $a \le s < t \le b$. Тоді існує випадковий процес $\widehat{\xi}(t)$, який є стохастично еквівалентним випадковому процесу $\xi(t)$, при цьому його траєкторії є неперервними.

Зауважемо, що в даному твердженні не йде мова про стохастичну неперервність процесу $\xi(t)$, позаяк цей факт випливає з (9) і нерівності Маркова. Дійсно, для довільного $\varepsilon > 0$ будемо мати:

$$\mathbf{P}(|\xi(t)-\xi(t_0)| > \varepsilon) = \mathbf{P}(|\xi(t)-\xi(t_0)|^p > \varepsilon^p) \le \frac{\mathbf{M}|\xi(t)-\xi(t_0)|^p}{\varepsilon^p} \le \\ \le C\frac{|t-t_0|^{1+r}}{\varepsilon^p} \underset{t\to t_0}{\longrightarrow} 0.$$

Зауваження 1. Твердження 6 залишається в силі, якщо замість умови (9) взяти наступну: для кожного $\varepsilon > 0$ існують константи $C, p, r > 0$ такі, що

$$\mathbf{P}(|\xi(t_2) - \xi(t_1)| > \varepsilon) \leq C\frac{|t_2 - t_1|^{1+r}}{\varepsilon^p}$$

для довільних $a \leq t_1 < t_2 \leq b$.

Одначе умова (9) є більш зручною для використання.

Завдання

2.1. Довести, що σ-алгебра є замкнутою (!)стосовно добутку зліченного числа подій.

2.2. Довести твердження 3.

2.3. $\xi(t)$ є випадковим процесом, а випадкова величина τ має показниковий розподіл. Нехай $\eta(t) = \xi(t) + I\{\tau = t\}$. Довести, що процеси $\xi(t)$, $\eta(t)$ є стохастично еквівалентними, але не є ідентичними.

2.4. Нехай випадкова величина $\xi \geq 0$ має неперервну функцію розподілу (!). Випадковий процес $\xi(t)$ визначається наступним чином:

$$\xi(t) = I\{\xi > t\}.$$

Довести, що цей процес є стохастично неперервним.

2.5. Довести, що для довільних випадкових величин ξ, η і $a, b \geq 0$ виконується нерівність:

$$\mathbf{P}(|\xi + \eta| > a + b) \leq \mathbf{P}(|\xi| > a) + \mathbf{P}(|\eta| > b).$$

2.6. Нехай траєкторії випадкового процесу належать простору $D_{[0,\infty)}$. Довести, що має місце

$$\mathbf{P}(\sup_{a \leq t \leq b} \xi(t) \leq x) = \lim_{n\to\infty} \mathbf{P}(\xi(a + \frac{b-a}{2^n} i) \leq x, i = 0, 1, ..., 2^n)$$

для довільних $0 \leq a < b < \infty$.

2.7. Випадкова величина ξ має показниковий розподіл з параметром 1. Визначимо випадковий процес наступним чином:

$$\xi(t) = I\{\xi \leq t\}.$$

a) Намалювати траекторії цього процесу.

b) Знайти $\mathbf{P}(\xi(t) = 0)$, $\mathbf{P}(\xi(t) = 1)$, $\mathbf{P}(\xi(t_1) = 0, \xi(t_2) = 0)$, $\mathbf{P}(\xi(t_1) = 0, \xi(t_2) = 1), \mathbf{P}(\xi(t_1) = 1, \xi(t_2) = 1), t_1 < t_2$.

2.8. Випадкова величина η має рівномірний розподіл на інтервалі $[-1; 1]$. Визначимо випадковий процес наступним чином:

$$\xi(t) = \eta t^2.$$

a) Намалювати траекторії цього процесу.

b) Знайти $\mathbf{P}(\xi(t) < x)$.

2.9. Випадкова величина ξ має показниковий розподіл з параметром. Визначимо випадковий процес як

$$\xi(t) = \frac{t}{t - \xi}.$$

a) Намалювати траєкторії цього процесу.

b) Знайти його одновимірний розподіл.

Лекція 2. Ланцюги Маркова 1

2.1 Визначення і класифікація станів

Нехай $\xi_n,\ n \geq 0$ є послідовністю випадкових величин, визначених на ймовірносному просторі $(\mathbf{\Omega}, \mathcal{B}, \mathbf{P})$, які приймають значення $E = \{0, \pm 1, \pm 2, ...\}$.

Визначення 1. *Послідовність $\xi_n,\ n \geq 0$ називається ланцюгом Маркова, якщо виконується умова:*

$$\mathbf{P}(\xi_{n+1}{=}j/\xi_0{=}i_0, \xi_1{=}i_1, ..., \xi_{n-1}{=}i_{n-1}, \xi_n{=}i){=}\mathbf{P}(\xi_{n+1}{=}j/\xi_n{=}i) \tag{1}$$

для довільних $n \geq 0, i, j, i_0, ..., i_{n-1} \in E$.

Зауваження 1. У цьому визначенні передбачаємо, що $\mathbf{P}(\xi_0 = i_0, \xi_1 = i_1, ..., \xi_{n-1} = i_{n-1}, \xi_n = i) > 0$. Інакше умовна ймовірність w (1) є невизначеною.

Властивість (1) називається *властивістю Маркова* і має наступний сенс. Сімейство подій $\Big\{[\omega\ :\ \xi_0(\omega){=}i_0, ..., \xi_{n-1}(\omega){=}i_{n-1}],\ \ i_1, ..., i_{n-1} \in E\Big\}$ можна трактувати як історію послідовності $\xi_i,\ i \geq 0$ до моменту n. Тоді властивість (1) означає, що починаючи від моменту n, майбутня еволюція послідовності ξ_i залежить виключно від стану послідовності в момент n і не залежить від її еволюції до моменту n.

Якщо

$$\mathbf{P}(\xi_{n+1} = j/\xi_n = i) = \mathbf{P}(\xi_1 = j/\xi_0 = i),$$

тоді ланцюг Маркова називають *однорідним.*

Зауваження 2. У подальшому будемо вивчати тільки однорідні ланцюги Маркова.

Визначення 2. *Ймовірність*

$$p_{ij} = \mathbf{P}(\xi_1 = j/\xi_0 = i)$$

будемо називати ймовірністю переходу за один крок, , а матрицю $P = (p_{ij})$ - матрицею ймовірностей переходів за один крок або матрицею переходів.

Матриця $P = (p_{ij})$ має наступні властивості:

1. $0 \leq p_{ij} \leq 1,\ \forall i, j \in E$;
2. $\sum_{j \in E} p_{ij} = 1$.

Матриці з такими властивостями називають *стохастичними.*

Ймовірність $p_i = \mathbf{P}(\xi_0 = i)$ називають *початковим розподілом* ланцюга Маркова.

Ланцюг Маркова зручно представляти у вигляді графа, вершини якого інтерпретуються як стани. Потім з вершини i проводять стрілку до вершини j, якщо $p_{ij} > 0$.

Теорема 1. *Початковий розподіл* (p_i) *і матриця переходів* $P = (p_{ij})$ *однозначно визначають ланцюг Маркова у тому сенсі, що*

$$\mathbf{P}(\xi_0 = i_0, \xi_1 = i_1 ..., \xi_n = i_n) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \tag{2}$$

Д о в е д е н н я. Для $n = 0$ рівність (2) виконується. Якщо тепер має місце (2), тоді

$$\mathbf{P}(\xi_0{=}i_0, \xi_1{=}i_1 ..., \xi_n{=}i_n, \xi_{n+1}{=}i_{n+1}){=}\mathbf{P}(\xi_{n+1}{=}i_{n+1}/\xi_0{=}i_0, \xi_1{=}i_1 ..., \xi_n{=}i_n)\times$$
$$\times\,\mathbf{P}(\xi_0{=}i_0, \xi_1{=}i_1 ..., \xi_n{=}i_n){=}p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \mathbf{P}(\xi_{n+1} = i_{n+1}/\xi_n = i_n) =$$
$$= p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \mathbf{P}(\xi_1 = i_{n+1}/\xi_0 = i_n) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} p_{i_n i_{n+1}}$$

і справедливість твердження випливає з принципу математичної індукції. ◄

Визначення 3. *Ймовірність*

$$\mathbf{P}(\xi_n = j/\xi_0 = i) = p_{ij}^{(n)}$$

називають ймовірністю переходу за n *кроків, а матрицю*
$P^{(n)} = \ p_{ij}^{(n)}$ *−матрицею переходів за* n *кроків.*

Теорема 2. $P^{(n)} = P^n$.

Доведення цього твердження, яке легко провести методом математичної індукції, залишаємо читачеві (завдання 3.4.).

З твердження 2 випливає рівність $P^{(m+n)} = P^{(n)} P^{(m)}$, з якої отримаємо рівняння

$$p_{ij}^{(n+m)} = \sum_k p_{ik}^{(n)} p_{kj}^{(m)}, \qquad i, j \in E. \tag{3}$$

Ці рівняння називаються *рівняннями Чепмена-Колмогорова.* Вони мають дуже простий сенс. Перехід за $n + m$ кроків зі стану i до стану j відбувається в наступний спосіб: спочатку за n кроків переходимо до деякого стану k. Потім, не зважаючи на те, що було до стану k (властивість Маркова), ніби розпочинаємо все спочатку зі стану k (властивість однорідності) і за m кроків потрапляємо до стану j. Позаяк проміжний стан може бути довільним, маємо сумувати по всіх можливих станах.

Визначення 4. *Стан* i *будмо називати неістотним, якщо існують такий стан* j *і таке* n_0, *що* $p_{ij}^{(n_0)} > 0$, *але* $p_{ji}^{(n)} = 0$ *для всіх* $n \geq 1$. *Стан, який не є неістотним станом, будемо називати істотним.*

Визначення 5. *Стан* j *називається досяжним зі стану* i, *якщо існує* n, *таке, що* $p_{ij}^{(n)} > 0$. *Тоді можемо записати* $i \to j$. *Якщо* $i \to j$ *і* $j \to i$, *тоді будемо говорити, що стани* i, j *сполучаються, і позначати це як* $i \leftrightarrow j$.

Відношення $\leftrightarrow$ має наступні властивості:
1) (рефлексивність) $i \leftrightarrow i$;
2) (симетричність) jeњli $i \leftrightarrow j$, to $j \leftrightarrow i$;
3) (транзитивність), якщо $i \leftrightarrow j$, $j \leftrightarrow k$, тоді $i \leftrightarrow k$ (завдання 3.5.).

З цього випливає, що відношення $\leftrightarrow$ є відношенням еквівалентності.

Теорема 3. *$E = E_0 \cup E_1 \cup E_2 ....$, де E_0 є класом всіх неістотних станів, натомість E_i, $i \geq 1$ є класами істотних станів, і $E_i \cap E_j = \emptyset$, $i \neq j$.*

Якщо $\xi_0 = i_0 \in E_i,\ i \geq 1$, тобто ланцюг $\xi_n, n \geq 0$ стартує з класу істотних станів E_i, тоді $\xi_n \in E_i$ для всіх $n \geq 0$. Дійсно, якби була можливість потрапляння на певному кроці до класу істотних або неістотних станів $E_j,\ i \neq j$, тоді для певних $n_0 \geq 1$ і $j_0 \in E_j$ виконувалося б $p^{(n)}_{i_0 j_0} > 0$, але тоді повернення до стану i_0 було б неможливим, позаяк інакше існувало б m таке, що $p^{(m)}_{j_0 i_0} > 0$. Звідки випливало б, що стани i_0, j_0 сполучаються, а це є неможливим. Нехай тепер $\xi_0 = i_0 \in E_0$. Якщо на певному кроці ланцюг залишить клас неістотних станів E_0 і потрапить до якогось класу істотних станів E_j, тоді згідно властивості Маркова і міркуванням, представленим вище, він залишиться в цьому класі назавжди.

Отже, ланцюг Маркова має наступні можливості для своєї еволюції: $\xi_n, n \geq 0$:

1) якщо в початковий момент $n = 0$ ланцюг перебуває в якомусь з класів істотних станів, він завжди буде там перебувати й надалі;
2) якщо в початковий момент ланцюг перебуває в класі неістотних станів E_0, тоді:
   a) або він залишиться в цьому класі назавжди;
   b) або на певному кроці він залишить E_0, потрапить до якогось
      з класів істотних станів і залишиться там назавжди.

Визначення 6. *Якщо всі стани ланцюга Маркова є істотними і такими, що сполучаються, тоді такий ланцюг називають незвідним .*

Іншими словами ланцюг Маркова ξ_n є незвідним, якщо в твердженні 3 маємо $E_0 = \emptyset, E_i = \emptyset, i = 2, 3, ...$ а $E_1 = E$.

Зауваження 3. У подальшому будемо розглядати тільки незвідні ланцюги Маркова.

Якщо $i \leftrightarrow j$, тоді існують n_0, m_0 такі, що $p^{(n_0)}_{ij} = \alpha > 0,\ p^{(m_0)}_{ji} = \beta > 0$. Тепер для довільного $n \geq 0$ будемо мати:

$$p^{(m_0+n_0+n)}_{jj} = \sum_{k,l \in E} p^{(m_0)}_{jk} p^{(n)}_{kl} p^{(n_0)}_{lj} \geq p^{(m_0)}_{ji} p^{(n)}_{ii} p^{(n_0)}_{ij} = \alpha\beta p^{(n)}_{ii},$$

звідки

$$p^{(n)}_{ii} \leq a p^{(m_0+n_0+n)}_{jj}, \tag{4}$$

де $0 < a = (\alpha\beta)^{-1} < \infty$.

Якщо поміняти місцями i та j, тоді замість (4) отримаємо:

$$p^{(n)}_{jj} \leq b p^{(m_1+n_1+n)}_{ii}, \tag{5}$$

для визначених натуральних n_1, m_1 і $0 < b < \infty$. З (4), (5) легко отримати наступний результат.

Лема 1. *Якщо стани i, j сполучаються, тоді існують натуральні n_1, n_2 і константи $0 < c_1, c_2 < \infty$ такі, що*

$$c_1 p^{(n-n_1)}_{jj} \leq p^{(n)}_{ii} \leq c_2 p^{(n+n_2)}_{jj} \tag{6}$$

для всіх $n \geq n_1$.

Позаяк у цій лемі поможемо поміняти i та j, отже отримаємо наступний результат, яким будемо у подальшому користуватись.

Наслідок 1. *Якщо стани i, j є такими, що сполучаються, тоді асимптотичні властивості ймовірностей $p_{ii}^{(n)}, p_{jj}^{(n)}$, якщо $n \to \infty$, співпадають.*

2.2 Рекурентність ланцюгів Маркова

Визначимо ймовірність першого повернення до стану i на n-тому кроці

$$f_{ii}^{(n)} = \mathbf{P}(\xi_1 \neq i, \xi_2 \neq i, ..., \xi_n = i/\xi_0 = i).$$

Нехай

$$F_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)}.$$

F_{ii} є ймовірністю того, що ланцюг Маркова, який стартував зі стану i,, колись до нього повернеться.

Нехай $\tau(i,i)=\inf\{n \geq 1 : \xi_n=i/\xi_0=i\}$ є моментом (кроком) першого повернення до стану i. Вочевидь, що $f_{ii}^{(n)}=\mathbf{P}(\tau(i,i) = n)$ і $F_{ii}=\mathbf{P}(\tau(i,i) < \infty)$.

Визначення 7. Стан i будемо називати *рекурентним,* якщо $F_{ii} = 1$, і нерекурентним або *транзитивним,* якщо $F_{ii} < 1$.

Теорема 4.

1) *Стан i є рекурентним тоді і тільки тоді, коли $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.*
2) *Якщо $\sum_{n=1}^{\infty} p_{ii}^{(n)} = P_i < \infty$, тоді $F_{ii} = \frac{P_i}{1+P_i}$.*

Д о в е д е н н я. Застосовуючи властивость Маркова і формулу повної ймовірності по першому моменту повернення до стану i, отримаємо:

$$p_{ii}^{(n)} = \sum_{k=1}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)}, \quad n \geq 1, \tag{7}$$

де за означенням $p_{ii}^{(0)} = 1$. Для $0 < z < 1$ розглянемо генератрису для послідовності $p_{ii}^{(n)}$, $f_{ii}^{(n)}$:

$$P_i(z) = \sum_{i=1}^{\infty} z^n p_{ii}^{(n)}, \qquad F_i(z) = \sum_{i=1}^{\infty} z^n f_{ii}^{(n)}.$$

Домножимо ліву і праву частину рівняння (7) на z^n і просумуємо від 1 до ∞. Після перетворень отримаємо:

$$P_i(z) = F_i(z)(1 + P_i(z)). \tag{8}$$

Звідки

$$P_i(z) = \frac{F_i(z)}{1 - F_i(z)}. \tag{9}$$

Зрозуміло, що $\lim_{z\uparrow 1} F_i(z)=\sum_{n=1}^{\infty} f_{ii}^{(n)}=F_{ii}\leq 1$ і $\lim_{z\uparrow 1} P_i(z)=\sum_{n=1}^{\infty} p_{ii}^{(n)} = P_i \leq \infty$.

Якщо стан i є рекурентним, тоді $F_{ii} = 1$ і перехід до границі $z \uparrow 1$ в (9) дає

$$P_i = \frac{F_{ii}}{1 - F_{ii}} = \frac{1}{1-1} = \infty,$$

якщо ж $P_i = \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$, тоді аналогічно з (9) для $z \uparrow 1$ отримаємо:

$$\infty = \sum_{n=1}^{\infty} p_{ii}^{(n)} = \lim_{z\uparrow 1} \sum_{n=1}^{\infty} z^n p_{ii}^{(n)} = \lim_{z\uparrow 1} \frac{F_i(z)}{1-F_i(z)} = \frac{F_{ii}}{1-F_{ii}} \leq \frac{1}{1-F_{ii}}.$$

Звідки $F_{ii} = 1$. Якщо $P_i = \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$, тоді з (8) при $z \uparrow 1$ будемо иати $P_i = F_{ii}(1 + P_i)$, отже, $F_{ii} = P_i/(1 + P_i) < 1$. ◀

Наступний результат в очевидний спосіб є наслідком леми 1.

Наслідок 2. *якщо $i \leftrightarrow j$, тоді стани i, j е одночасно рекурентними або нерекурентними.*

Нехай

$$Q_{ii} = \mathbf{P}(\xi_{n_k} = i \ \text{ для необмеженого числа кроків } n_k/\xi_0 = i).$$

Q_{ii} є ймовірністю того, що ланцюг Маркова, стартуючи зі стану i, повернеться до цього стану необмежену кількість разів.

Теорема 5. *$Q_{ii} = 1$ або $Q_{ii} = 0$ в залежності від того, чи стан $\{i\}$ е рекурентним чи ні.*

Д о в е д е н н я. Нехай $Q_{ii}(N) = \mathbf{P}(\xi_{n_k} = i \ \text{ dla } n_1, ..., n_k,\ k \geq N/\xi_0 = i)$. Іншими словами $Q_{ii}(N)$ є ймовірністю того, що ланцюг Маркова, стартуючи зі стану i, повернеться до цього стану принаймні N разів. Зрозуміло, що $Q_{ii}(N) \to Q_{ii}$, якщо $N \to \infty$. Отже,

$$Q_{ii}(N) = Q_{ii}(N-1) \sum_{k=1}^{\infty} f_{ii}^{(k)} = Q_{ii}(N-1)F_{ii},$$

що дасть $Q_{ii}(N) = F_{ii}^N$. Звідки при $N \to \infty$ й випливає твердження. ◀

Позначимо через

$$f_{ij}^{(n)} = \mathbf{P}(\xi_1 \neq j, \xi_2 \neq j, ..., \xi_n = j/\xi_0 = i)$$

ймовірність першого потрапляння до стану j на n-тому кроці, якщо ланцюг стартує зі стану i. Тоді $F_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$ є ймовірністю того, що ланцюг Маркова, стартуючи зі стану i, колись потрапить до стану j.

Якщо $i = j$, тоді будемо мати перше повернення до стану i.

Нехай $\tau(i, j) = \inf\{n \geq 1 : \xi_n = j/\xi_0 = i\}$ буде моментом першого потрапляння зі стану i до стану j. Вочевидь, $f_{ij}^{(n)} = \mathbf{P}(\tau(i, j) = n)$ і $F_{ij} = \mathbf{P}(\tau(i, j) < \infty)$.

Теорема 6. *Якщо $i \leftrightarrow j$ і стани i, j е рекурентними, тоді $F_{ij} = 1$.*

Д о в е д е н н я. Нехай ${}^j p_{ji}^{(n)}$ є ймовірністю того, що ланцюг Маркова, стартуючи зі стану j, потрапить до стану i на n-тому кроці, не відвідуючи жодного разу стан j. Іншими словами, ${}^j p_{ji}^{(n)} = \mathbf{P}(\xi_1 \neq j, \xi_2 \neq j, ..., \xi_{n-1} \neq j, \xi_n = i/\xi_0 = j)$. Для певного n_0 будемо мати ${}^j p_{ji}^{(n_0)} > 0$ (завдання 3.6). Тепер

$$0 = \mathbf{P}(\tau(j,j) = \infty) \geq {}^j p_{ji}^{(n_0)}(1 - \sum_{n=1}^{\infty} f_{ij}^{(n)}) \Longrightarrow \sum_{n=1}^{\infty} f_{ij}^{(n)} = 1.$$

◀

Наслідок 3. *Якщо $i \leftrightarrow j$, і стани i, j є рекурентними, тоді існує нескінченна послідовність $n_1, ..., n_k, ...$ (випадкова!) така, що*

$$\mathbf{P}(\xi_{n_k} = j, k = 1, 2, .../\xi_0 = i) = 1.$$

Доведення є очевидним.

2.3 Блукання Бернуллі

Визначення 8. Нехай η_n, $n = 1, 2, ...$ є незалежними випадковими величинами з розподілом $\mathbf{P}(\eta_n = 1) = p \in [0, 1]$, $\mathbf{P}(\eta_n = -1) = q = 1 - p$. Послідовність $S_n^{(1)} = \xi_0 + \sum_{i=1}^n \eta_i$, $n \geq 1$ називається *блуканням Бернуллі на прямій* або просто *блуканням Бернуллі.*

Блукання Бернуллі є однорідним ланцюгом Маркова (завдання 3.8.).

Визначення 9. Нехай $\xi_n^{(i)}$, $n = 1, 2, ...$ $i = 1, 2$ будуть незалежними блуканнями Бернуллі на прямій. Тоді $S_n^{(2)} = (\xi_n^{(1)}, \xi_n^{(2)})$ будемо називати *блуканням Бернуллі* в R^2.

В подібний спосіб можна визначити блукання Бернуллі в R^3: $S_n^{(3)} = (\xi_n^{(1)}, \xi_n^{(2)}, \xi_n^{(3)})$, де $\xi_n^{(i)}$, $n = 1, 2, ...$ $i = 1, 2, 3$ є незалежними блуканнями Бернуллі на прямій.

Теорема 7. *Блукання Бернуллі $S_n^{(1)}, S_n^{(2)}$ є рекурентним тоді і тільки тоді, коли $p = q = 1/2$, а блукння $S_n^{(3)}$ є завжди нерекурентним.*

Д о в е д е н н я. Розглянемо блукання $S_n^{(1)}$. Неважко зрозуміти, якщо $0<p<1$, тоді всі стани ланцюга Маркова $S_n^{(1)}$ сполучаються і згідно з наслідком 2 достатньо перевірити на рекурентність стан $\{0\}$. Маємо: (завдання 3.9.)

$$p_{00}^{(n)} = \begin{cases} C_{2k}^k \ p(1-p) \ ^k, & n = 2k, \\ \\ 0, & n = 2k+1, \end{cases} \tag{10}$$

для $k = 1, 2, ...$. Використовуючи формулу Стірлінга [1] $n! = n^n \sqrt{2\pi n} e^{-n}(1 + \alpha_n)$, де $\alpha_n \underset{n\to\infty}{\longrightarrow} 0$, по виконанню необхідних перетворень отримаємо:

$$p_{00}^{(2k)} = C_{2k}^k \ p(1-p) \ ^k = \frac{(2k)!}{(k!)^2} \ p(1-p) \ ^k = \frac{4p(1-p) \ ^k}{\sqrt{\pi k}} \frac{1 + \alpha_{2k}}{(1 + \alpha_k)^2}.$$

[1] *James Stirling* (1692–1770) – шотландський математик, який займався теорією нескінченних рядів і теорією алгебраїчних кривих krzywych.

З останньої рівності випливає, що умови збіжності ряду $\sum_{k=1}^{\infty} p_{00}^{(2k)}$ є такими ж самими як умови збіжності ряду $\sum_{k=1}^{\infty} \frac{4p(1-p)^{k}}{\sqrt{\pi k}}$. Легко переконатись, що $4p(1-p) = \alpha < 1$, коли $p \neq 1/2$, і $4p(1-p) = 1$, для $p = 1/2$. Тепер будемо мати:

$$\sum_{k=1}^{\infty} \frac{4p(1-p)^{k}}{\sqrt{\pi k}} = \begin{cases} \sum_{k=1}^{\infty} \frac{\alpha^k}{\sqrt{\pi k}} < \infty, & p \neq 1/2, \\ \sum_{k=1}^{\infty} \frac{1}{\sqrt{\pi k}} = \infty, & p = 1/2, \end{cases}$$

отже блукання $S_n^{(1)}$ є рекурентним тоді і тільки тоді, коли $p = 1/2$.

Розглянемо тепер блукання $S_n^{(2)} = (\xi_n^{(1)}, \xi_n^{(2)})$. Позначимо через (i, j) стан цього блукання. Аналогічно випадку блукання $S_n^{(1)}$ достатньо дослідити рекурентність стану $(0, 0)$. Вочевидь, блукання $S_n^{(2)}$ буде рекурентним тоді і тільки тоді, коли рекурентними будуть блукання $\xi_n^{(1)}$ і $\xi_n^{(2)}$. З цього випливає, що достатньо розглянути ситуацію, за якої ймовірність переходу за $2k$ кроків зі стану 0 до стану 0, коли $k \to \infty$, є порядку $\frac{1}{\sqrt{\pi k}}(1 + \alpha_k)$, де $\alpha_k \underset{k\to\infty}{\longrightarrow} 0$. З незалежності блукань $\xi_n^{(1)}$, $\xi_n^{(2)}$ отримаємо:

$$p_{(0,0),(0,0)}^{(2k)} = \mathbf{P}(\xi_{2k}^{(1)} = 0/\xi_0^{(1)} = 0)\mathbf{P}(\xi_{2k}^{(2)} = 0/\xi_0^{(2)} = 0) = \frac{1}{\pi k}(1 + \alpha_k)^2.$$

Позаяк

$$\frac{1}{\pi}\sum_{k=1}^{\infty}\frac{1}{k} = \infty, \qquad \text{отже} \qquad \sum_{k=1}^{\infty} p_{(0,0),(0,0)}^{(2k)} = \infty,$$

звідки робимо висновок о рекурентності блукання $S_n^{(2)}$. ◀

Випадок блукання $S_n^{(3)}$ можна дослідити аналогічно, тому залишаємо його читачеві(завдання 3.10.).

ЗАВДАННЯ

3.1. Для ланцюга Маркова ξ_n з матрицею ймовірностей переходів

$$\begin{matrix} 0,6 & 0,4 \\ 0,2 & 0,8 \end{matrix}$$

знайти $\mathbf{P}(\xi_2 = 2)$, $\mathbf{P}(\xi_3 = 1)$, якщо початковий розклад має вигляд: $(1/3, 2/3)$.

3.2. Для ланцюга Маркова ξ_n з матрицею переходів як у попередньому завданні, знайти початковий розподіл, за якого $\mathbf{P}(\xi_2 = 2) = 0,716$.

3.3. Провести класифікацію станів ланцюга Маркова ξ_n матрицею переходів

$$\text{a)} \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/3 & 0 & 2/3 \end{pmatrix}, \quad \text{b)} \begin{pmatrix} 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 4/5 & 0 & 1/5 \end{pmatrix}.$$

3.4. Довести твердження 2.

3.5. Довести, якщо $i \leftrightarrow j$ і $j \leftrightarrow k$, тоді $i \leftrightarrow k$.

3.6. Нехай ${}^{j}p_{ji}^{(n)} = \mathbf{P}(\xi_1 \neq j, \xi_2 \neq j, ..., \xi_{n-1} \neq j, \xi_n = i/\xi_0 = j)$. Довести, що існує таке n_0, що ${}^{j}p_{ji}^{(n_0)} > 0$.

3.7. Для ланцюга Маркова ξ_n визначимо $\tau_i = \inf\{n \geq 1 : \xi_n \neq i/\xi_0 = i\}$. Довести, що τ мпє геометричний розподіл.

3.8. Показати, що блукання Бернуллі є однорідним ланцюгом Маркова і знайти його матрицю переходів.

3.9. Довести вираз (10).

3.10. Довести, що блукання Бернуллі $S_n^{(3)}$ є нерекурентним.

3.11. Для ланцюга Маркова ξ_n знайти $\mathbf{P}(\xi_0 = i/\xi_n = j)$.

3.12. До двох урн поклали m чорних і m білих куль так, що кожна містить m куль. У моменти $1, 2, ...$ з кожної урни одна куля перекладається до другої урни. Нехай ξ_n є числом білих куль у першій урні. Показати, що ξ_n є ланцюгом Маркова і знайти його матрицю переходів.

3.13. Показати, що у випадку, коли ланцюг Маркова має скінченну множину станів, всі його стани не можуть бути неістотними. Навести приклад ланцюга Маркова, всі стани якого є неістотними.

3.14. Нехай $\xi_n,\ n \geq 0$ є незвідним ланцюгом Маркова.

a. Позначимо через ${}_jf_{ik}$ ймовірність того, що $\xi_n,\ n \geq 0$, стартуючи зі стану i, потрапить вперше до стану k раніше ніж до стану $j \neq k$. Довести, що ${}_jf_{kk} < 1$.

b. Нехай $\tau(j, j)$ є часом першого повернення ланцюга ξ_n до стану j, $\xi_0 = j$, а μ_{jk} є числом потраплянь ланцюга в стан k протягом часу $\tau(j, j)$.

   Довести, що:

   (a) $\mathbf{P}(\mu_{jk} > n) = {}_jf_{jk} \cdot {}_jf_{kk}^n$;

   (b) $\mathbf{M}\mu_{jk} < \infty$;

   (c) якщо ланцюг ξ_n має скінченну множину станів, тоді $\mathbf{M}\tau(j, j) < \infty$.

3.15. Розглянемо послідовність $\eta_n, n = 1, 2, ...$ незалежних однаково розподілених випадкових величин $\mathbf{P}(\eta_n = 1)) = p = 1 - \mathbf{P}(\eta_n = -1),\ p \in (0, 1)$. Нехай $S_n = \eta_1 + \eta_2 + \ldots + \eta_n,\ n = 1, 2, \ldots$ Довести, що послідовності $Y_n = |S_n|$, $Z_n = \max_{k \leq n} S_k - S_n$ є ланцюгами Маркова.

Лекція 3. Ланцюги Маркова 2

3.1 Періодичні ланцюги Маркова

Нехай ξ_n, $n = 0, 1, 2, ...$ є ланцюгом Маркова з матрицею ймовірностей переходів P.

Визначення 1. *Якщо*

$$\text{НСД}\{n : p_{ii}^{(n)} > 0\} = d(i) > 1,$$

тоді стан i називається періодичним з періодом $d(i)$.

З цього визначення випливає, якщо $d(i)$ є періодом стану i, тоді нерівність $p_{ii}^{(n)} > 0$ виконується, коли n має вигляд $n{=}kd(i)$, $k = 1, 2, ...$ але необов'язково для всіх k (завдання 1) і $p_{ii}^{(n)} = 0$, якщо $n \neq kd(i)$, для всіх $k = 1, 2, ...$ (завдання 2).

Теорема 1. *Якщо $i \leftrightarrow j$, тоді $d(i) = d(j) \overset{def}{=} d$ і d називається періодом ланцюга ξ_n.*

Крім того, у ланцюга Маркова властивість періодичності є властивістю всіх станів, які утворюють певний клас станів: всі стани класу, які сполучаються, є або періодичними з тим самим періодом, або неперіодичними.

Д о в е д е н н я. Позаяк стани i, j сполучаються, отже $p_{ij}^{(n_0)} = \alpha > 0$, $p_{ji}^{(m_0)} = \beta > 0$ для певних n_0, m_0. Нехай $N(i) = \{n : p_{ii}^{(n)} > 0\}$. Зрозуміло, що $d(i)$ є найбільшим дільником чисел $N(i)$. Тепер для $m \in N(i)$ будемо мати:

$$p_{jj}^{(n_0+m+m_0)} = \sum_{k,l \in E} p_{jk}^{(n_0)} p_{kl}^{(m)} p_{lj}^{(m_0)} \geq p_{ji}^{(n_0)} p_{ii}^{(m)} p_{ij}^{(m_0)} = \alpha\beta p_{ii}^{(m)} > 0,$$

отже $n_0 + m + m_0 = c_1 d(j)$. Позаяк $p_{ii}^{(2m)} \geq p_{ii}^{(m)} p_{ii}^{(m)} > 0$, тоді аналогічно будемо мати $m_0 + 2m + n_0 = c_2 d(j)$. Віднявши від другої рівності першу, отримаємо $m = (c_2 - c_1) d(j)$, звідки випливає, що числа $N(i)$ діляться на $d(j)$, крім того $d(j) \leq d(i)$. Помінявши місцями i з j, маємо $d(i) \leq d(j)$, отже $d(i) = d(j)$. ◀

Нехай i_0 є довільним станом. Визначимо наступні множини станів:

$$E_k = \{j \in E_k, \text{ якщо існує } n, \text{ таке що } p_{i_0 j}^{(nd+k)} > 0\}, \quad k = 0, 1, ..., d-1.$$

Множини E_k є різними. Дійсно, достатньо довести, що коли $p_{i_0 j}^{(md+l)} > 0$, $0 \leq l \leq d-1$, і $j \in E_k$, $0 \leq k \leq d-1$, тоді $l = k$. Позаяк всі стани сполучаються, отже $p_{ji_0}^{(n)} > 0$ для певного n, а $p_{i_0 j}^{(\widehat{m}d+k)} > 0$ для певного $\widehat{m}$ за визначенножининям м E_k. Тому $p_{i_0 i_0}^{(md+l+n)} > p_{i_0 j}^{(md+l)} p_{ji_0}^{(n)} > 0$, що дасть $md + l + n = c_1 d$. Подібним чином $p_{i_0 i_0}^{(\widehat{m}d+k+n)} >$

$p_{i_0j}^{(\widehat{m}d+k)}p_{ji_0}^{(n)} > 0$, звідки $\widehat{m}d + k + n = c_2 d$, отже $(m - \widehat{m})d + l - k = (c_1 - c_2)d$. З цієї рівності випливає, що $l - k$ ділеться на d, що є можливим тільки, якщо $l = k$.

Доведення наступного твердження є достатньо простим, отже залишаємо його читачеві (завдання 3).

Теорема 2. *Нехай d є періодом ланцюга Маркова ξ_n. Тоді $E = E_0 \cup E_1 ... \cup E_{d-1}$, де $E_i \cap E_j = \emptyset$, $i \neq j$, і $\mathbf{P}(\xi_{n+1} \in E_{i+1}/\xi_n \in E_i) = 1$, $(E_d = E_0)$.*

Зауваження 1. Множини $E_k,\ k = 0, 1, ..., d-1$ називаються *циклічними підкласами* періодичного ланцюга Маркова. Легко зрозуміти, що вони визначаються однозначно з точністю до номеру.

З твердження 2 випливає, що можна перенумерувати стани періодичного ланцюга Маркова з періодом d таким чином, щоб його матриця переходів мала вигляд:

	E_0	E_1	E_2		E_{d-1}
E_0	O	P_1	O	....	O
E_1	O	O	P_2	....	O
E_2	O	O	O	....	O
	O	O	O	....	O
	O	O	O	....	P_{d-1}
E_{d-1}	P_d	O	O	....	O

Матриці P_i відповідають тут ненульовим елементам, а O є матрицями, які складаються з нулів. Матриця P_k - це матриця переходів за один крок зі станів E_{k-1} до станів E_k, $k = 1, 2, ...d - 1$, а матриця P_d відповідає переходам за один крок зі станів E_{d-1} до станів E_0.

3.2 Ергодичні властивості ланцюга Маркова

У цьому параграфі проаналізуємо поведінку ймовірностей $p_{ij}^{(n)}$, коли $n \to \infty$.

Для цього нам буде потрібен один результат, який спочатку наведемо без доведення. Доведення буде представлене згодом у частині, яка буде присвячена теорії відновлення. Інше доведення можна знайти в [**?**], твердження 1.1, стор. 73.

Нехай p_k, $k = 0, 1, 2...$ є ймовірносним розподілом, а b_k, $k = 0, 1, 2...$ – числовою послідовністю такою, що $\sum_{k=0}^{\infty} |b_k| < \infty$.

Теорема 3. *Нехай ймовірносний розподіл p_n, $n = 0, 1, 2, ...$ не є гратчастим розподілом* [1]. *Тоді:*

[1]Розподіл випадкової величини ξ називають *гратчастим* (*арифметичним*) *з кроком* $h > 0$, якщо $\sum_{k=-\infty}^{\infty} \mathbf{P}(\xi = a + kh) = 1$ для деякого a.

1) *рівняння*

$$u_n - \sum_{k=0}^{n} p_k u_{n-k} = b_n, \qquad n \geq 0 \tag{1}$$

має тільки один обмежений розв'язок ;

2) *Якщо* u_n *є обмеженим розв'язком рівняння* (1), *тоді*

$$\lim_{n\to\infty} u_n = \frac{\sum_{k=0}^{\infty} b_k}{\sum_{k=0}^{\infty} k p_k}. \tag{2}$$

Якщо стан j не є рекурентним, тоді згідно з критерієм рекурентності має місце: $\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$, отже $p_{jj}^{(n)} \underset{n\to\infty}{\longrightarrow} 0$. З рівності $p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}$ випливає, що $p_{ij}^{(n)} \underset{n\to\infty}{\longrightarrow} 0$ (дивись зауваження 2 нижче).

Зауваження 2. Легко довести, що коли $\sum_{k=0}^{\infty} |a_k| < \infty$ і $\lim_{k\to\infty} b_k{=}c$, тоді $\lim_{n\to\infty} \sum_{k=0}^{n} a_k b_{n-k} = c \sum_{k=0}^{\infty} a_k$.

З цих міркувань випливає, що коли ми хочемо, аби границя $\lim_{n\to\infty} p_{ij}^{(n)}$ була додатньою треба, щоб стани ланцюга Маркова були рекурентними.

Лема 1. *Нехай данцюг Маркова* ξ_n, $n \geq 0$ *є неперіодимним і рекурентним. Тоді*

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\sum_{n=1}^{\infty} n f_{jj}^{(n)}}. \tag{3}$$

Д о в е д е н н я. У попередній лекції ми отримали рівність (див. рівність (7) на стор. 297)

$$p_{ii}^{(n)} = \sum_{k=1}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)}, \quad n \geq 1,$$

за умови, що $p_{ii}^{(0)} = 1$. Якщо за визначенням $f_{ii}^{(0)} = 0$, тоді ця рівність може бути записана у вигляді:

$$p_{ii}^{(n)} - \sum_{k=0}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)} = \delta_{0n}, \quad n \geq 0. \tag{4}$$

З неперіодичності ланцюга ξ_n, $n \geq 0$ випливає, що розподіл $f_{ii}^{(n)}, n{=}0,1,2,...$ не є гратчастим (завдання 5). Тепер відношення (3) для $i = j$ випливає з твердження 3 і рівняння (4).

Зробимо припущення, що $i \neq j$. Будемо мати $p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}, \quad n \geq 1$, отже

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{\sum_{k=1}^{\infty} f_{ij}^{(k)}}{\sum_{n=1}^{\infty} n f_{jj}^{(n)}} = \frac{1}{\sum_{n=1}^{\infty} n f_{jj}^{(n)}},$$

позаяк для рекурентних станів має місце $\sum_{k=1}^{\infty} f_{ij}^{(k)} = 1$ (твердження 6, стор. 298). ◄

З (3) випливає, що для неперіодичного рекурентного ланцюга Маркова границя $\lim_{n\to\infty} p_{ij}^{(n)}$ буде додатньою тоді і тільки тоді, коли $\sum_{n=1}^{\infty} n f_{jj}^{(n)} = \mathbf{M}\tau(j,j) < \infty$, де $\tau(j,j)$ є моментом першого повернення до стану j.

Визначення 2. Рекурентний стан i називається *нульовим*, якщо $\sum_{n=1}^{\infty} nf_{ii}^{(n)} = \infty$, або *додатним*, якщо $\sum_{n=1}^{\infty} nf_{ii}^{(n)} < \infty$.

Зауваження 3. З (3) для $i{=}j$ випливає, що стан i є нульвим (додатним) тоді і тільки тоді, коли $\lim_{n\to\infty} p_{ii}^{(n)} = 0 (> 0)$.

Лема 2. *У незвідного ланцюга Маркова всі стани є або нульовими, або додатними.*

Доведення цього твердження легко провести, використовуючи попереднє зауваження та лему 1 на стор. 296.

Визначення 3. *Ланцюг Маркова будемо називати ергодичним, якщо для всіх його станів i, j існують границі $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ і $\sum_{j\in E} \pi_j = 1$.*

Теорема 4. *Якщо ланцюг Маркова ξ_n, $n \geq 0$ є неперіодичним і виконується умова $m_j = \mathbf{M}\tau(j,j) = \sum_{n=1}^{\infty} nf_{jj}^{(n)} < \infty$, $j \in E$, тоді*

1) *існує границя $\lim_{n\to\infty} p_{ij}^{(n)} = 1/m_j \stackrel{def}{=} \pi_j > 0$, яка не залежить від початкового стану ланцюга* ;
2) $\sum_{j\in E} \pi_j = 1$;
3) *для вектора $\vec{\pi} = (\pi_j, j \in E)$ має місце рівність $\vec{\pi}P = \vec{\pi}$;*
4) *вектор $\vec{\pi}$ є единим.*

Наслідок 1. Якщо для ланцюга Маркова виконуються умови попереднього твердження, тоді він є ергодичним.

Визначення 4. Сукупність $\{\pi_j,\ j \in E\}$ з твердження 4 називається *ергодичним* (або *фінальним*) розподілом ланцюга Маркова ξ_n.

D o w y d твердження 4. Перший пункт випливає безпосередньо з леми 1. Доведемо рівність $\sum_j \pi_j = 1$. Спочатку покажемо, що $\sum_j \pi_j \leq 1$. Нехай $S(m)$, $m = 1, 2, ...$ є такими підмножинами E, що $S(m)$ складаються з m елементів, $E_m \subset E_{m+1}$ і $S(m) \uparrow E$, якщо $m \to \infty$. Будемо мати $\sum_{j\in S(m)} p_{ij}^{(n)} \leq 1$, тоді після переходу до границі $n \to \infty$ отримаємо $\sum_{j\in S(m)} \pi_j \leq 1$. Перейшовши до границі по $m \to \infty$ в останній нерівності, дістанемо $\sum_j \pi_j \leq 1$.

Позаяк $p_{ij}^{(n+1)} \geq \sum_{k\in S(m)} p_{ik}^{(n)} p_{kj}$, тоді після переходу до границі $n \to \infty$ отримаємо $\pi_j \geq \sum_{k\in S(m)} \pi_k p_{kj}$. Якщо в цій нерівності взяти $m \to \infty$,тоді будемо мати:

$$\pi_j \geq \sum_k \pi_k p_{kj}. \tag{5}$$

Якби в (5) нерівність була строгою для певного j, тоді сумуючи сторони в (5) по $j \in E$, отримали б

$\sum_j \pi_j > \sum_j \sum_k \pi_k p_{kj} = \sum_k \pi_k \sum_j p_{kj} = \sum_k \pi_k$, що призводить до протиріччя. Отже в (5) мають місце рівності, що доводить пункт 3 твердження.

Далі будемо мати:

$$\pi_j = \sum_k \pi_k p_{kj} = \sum_k \sum_l \pi_l p_{lk} p_{kj} = \sum_l \pi_l \sum_k p_{lk} p_{kj} = \sum_l \pi_l p_{lj}^{(2)} = \ldots = \sum_k \pi_k p_{kj}^{(n)}. \tag{6}$$

Позаяк $\sum_j \pi_j \leq 1$, тоді в (6) можемо перейти до границі $n \to \infty$ під символом $\sum$, що дасть $\pi_j = \pi_j \sum_k \pi_k$, отже $\sum_k \pi_k = 1$.

Якщо існує інший набір чисел $\widehat{\pi}_j,\ j \in E$, для яких має місце 1)-3), тоді $\widehat{\pi}_k = \sum_i \widehat{\pi}_i p_{ik} = ... = \sum_i \widehat{\pi}_i p_{ik}^{(n)}$ і перехід до границі $n \to \infty$ дає $\widehat{\pi}_k = \sum_i \widehat{\pi}_i \pi_k = \pi_k$. ◀

Якщо скінченний ланцюг Маркова є незвідним, тоді всі його стани є додатними (див. зпвдання 14. b(c) стор. 301). Звідки маємо наступний наслідок.

Наслідок 2. *Для скінченного неперіодичного ланцюга Маркова завжди існує ергодичний розподіл.*

Зауваження 4. Ймовірносний розподіл $\vec{\rho} = \{\rho_i, i \in E\}$, для якого виконується рівняння $\vec{\rho}P = \vec{\rho}$, називається *стаціонарним* розподілом ланцюга Маркова.

Якщо $\vec{\rho} = \{\rho_i,\ i \in E\}$ є стаціонарним розподілом ланцюга Маркова ξ_n з матрицею переходів P, тоді, позаяк $P^{(n)} = P^n$, будемо мати $\vec{\rho}P^{(n)} = \vec{\rho}$ для довільного $n \geq 1$. Отже, якщо за початковий розподіл ланцюга ξ_n взяти розподіл $\vec{\rho}$, тоді $\mathbf{P}(\xi_n = j) = \sum_{k \in E} \rho_k p_{kj}^{(n)} = \rho_j$. Іншими словами тоді безумовна ймовірність перебування ланцюга Маркова в стані не залежить від часу. Такий ланцюг називається *стаціонарним.*

Якщо існує ергодичний розподіл, тоді він автоматично є також стаціонарним розподілом, але може статись, що стаціонарний розподіл існує, а ергодичний розподіл ні (завдання 13).

Проаналізуємо тепер фінальні ймовірності періодичних ланцюгів Маркова. Нехай $\xi_n,\ n \geq 0$ є періодичним ланцюгом Маркова з періодом d, крім того $E = E_0 \cup E_1 ... \cup E_{d-1}$ буде його представленням у вигляді циклічних підкласів. Припустимо, що виконується умова $m_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)} < \infty$. Якщо $i, j \in E_k$, to $p_{ij}^{(dn+l)} = 0$ для всіх $n \geq 1,\ 1 \leq l \leq d-1$. Позаяк ймовірності $p_{ij}^{(nd)}$ могуть бути додатними, тоді границя $\lim_{n\to\infty} p_{ij}^{(nd)}$ для $i, j \in E_k$ може бути додатньою. З цього випливає, що границя $\lim_{n\to\infty} p_{ij}^{(n)}$ для $i, j \in E_k$ може не існувати.

Нехай знову $i, j \in E_k$. Розглянемо ймовірності $p_{ij}^{(dn)},\ n = 1, 2, ...$ Вочевидь, що ці ймовірності відповідають ланцюгу Маркова $\xi_{nd},\ n = 0, 1, 2, ...$, який стартує зі стану $i \in E_k$. Якщо взяти до уваги еволюцію періодичного ланцюга Маркова, тоді можемо сказати, що ймовірності $p_{ij}^{(dn)}$ є ймовірностями переходу за n кроків наступного ланцюга Маркова: $\widehat{\xi}_n \stackrel{def}{=} \xi_{nd},\ n = 0, 1, 2, ...$ з множиною станів E_k. Цей ланцюг є неперіодичним. Якщо $\widehat{p}_{ij}^{(n)}$ є ймовірностями його переходів за n кроків, тоді будемо мати:

$$\lim_{n\to\infty} p_{ij}^{(nd)} = \lim_{n\to\infty} \widehat{p}_{ij}^{(n)} = \frac{1}{\mathbf{M}\widehat{\tau}(j, j)}, \tag{7}$$

де $\widehat{\tau}(j, j)$ є часом повернення до стану j ланцюга $\widehat{\xi}_n$. Вочевидь, $d\widehat{\tau}(j, j) = \tau(j, j)$, де $\tau(j, j)$ є часом повернення до стану j ланцюга ξ_n. Тепер будемо мати:

$$\lim_{n\to\infty} p_{ij}^{(nd)} = \frac{d}{\mathbf{M}\tau(j, j)} = \frac{d}{m_j}.$$

Цей аналіз підводить нас до формулювання ергодичяного твердження для періодичних ланцюгів Маркова.

Теорема 5. *Розглянемо ланцюг Маркова ξ_n, $n \geq 0$, який є періодичним з періодом d і таким, що $m_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)} < \infty$, і $E = E_0 \cup E_1 ... \cup E_{d-1}$ є його представленням у вигляді циклічних підкласів. Нехай $i \in E_k$, $j \in E_l$. Тоді*

1) $\lim_{n\to\infty} p_{ij}^{(nd+|k-l|)} = d/m_j \stackrel{def}{=} \pi_j > 0$, *і* $\sum_j \pi_j = 1$;

2) $\lim_{n\to\infty} p_{ij}^{(nd+m)} = 0$, *для кожного* $m \neq |k-l|$.

ЗАВДАННЯ

1. Довести, що ланцюг Маркова з матрицею переходів
$$\begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}$$
має період 2, $p_{33}^{(2k)} > 0$, $k = 2, 3, ...$, але $p_{33}^{(2)} = 0$.

2. Довести, що коли d є періодом стану i, тоді $p_{ii}^{(ld+k)} = 0$ для всіх $k = 1, 2, ..., d-1$, $l = 0, 1, 2, ...$

3. Довести твердження 2.

4. Довести твердження 5 у випадку $i \in E_k, j \in E_l$, $k \neq l$.

5. Довести, що у випадку, коли ланцюг Маркова ξ_n, $n \geq 0$ є неперіодичним, тоді розподіл $f_{ii}^{(n)}$, $n = 0, 1, 2, ...$ не є гратчастим.

6. Дослідити на періодичність та знайти циклічні підкласи ланцюга Маркова з матрицею переходів
$$\text{a)} \begin{pmatrix} 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & & 0 & 1 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \end{pmatrix}, \qquad \text{b)} \begin{pmatrix} 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

7. Нехай p_k, $k = 0, 1, 2, ...$ є розподілом ймовірностей. Визначимо послідовність p_n^{k*}, $k = 0, 1, 2, ...$ наступнм чином:
$$p_n^{0*} = \delta_{0n}, \qquad p_n^{k*} = \sum_{i=0}^{n} p_i p_{n-i}^{(k-1)*}, \qquad n \geq 1,$$
де δ_{0n} є символом Кронекера. Довести, що рівняння (1) має тільки один обмежений розв'язок, який може бути записаний у вигляді:
$$u_n = \sum_{k=0}^{n} b_{n-k} P_k, \qquad n \geq 0, \tag{8}$$
де $P_n = \sum_{k=0}^{\infty} p_n^{k*}$.

8. Для ланцюга Маркова з матрицею переходів

$$P = \begin{matrix} 1/5 & 4/5 \\ 1/3 & 2/3 \end{matrix} \qquad P = \begin{pmatrix} 0,3 & 0,5 & 0,2 \\ 0,6 & 0 & 0,4 \\ 0 & 0,4 & 0,6 \end{pmatrix}$$

знайти:

a) ергодичний розподіл;

b) середній час повернення до стану.

9. Матриця переходів ланцюга Маркова ξ_n з множиною станів $E = \{1, 2, 3, 4\}$ має вигляд:

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}.$$

a) Припустивши, що $\xi_0 = 3$, підрахувати математичне сподівання часу потрапляння до стану 2.

b) Знайти ергодичний розполіл.

c) Чи є ланцюг періодичним? Чи є він незвідним?

10. Ланцюг Маркова має матрицю переходів:

$$\begin{pmatrix} p_0 & p_1 & p_2 & p_3 & .... \\ 1 & 0 & 0 & 0 & .... \\ 0 & 1 & 0 & 0 & .... \\ 0 & 0 & 1 & 0 & .... \end{pmatrix}.$$

Визначити:

a) умови рекурентності;

b) умови існування і вигляд ергодичного розподілу.

11. Ланцюг Маркова має матрицю переходів:

$$\begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & .... \\ 2/3 & 0 & 1/3 & 0 & 0 & 0 & ... \\ 3/4 & 0 & 0 & 1/4 & 0 & 0 & .... \\ 4/5 & 0 & 0 & 0 & 1/5 & 0 & .... \\ .. & .. & .. & .. & .. & .. & .... \end{pmatrix}.$$

Довести, що $p_{ij}^{(n)} \underset{n\to\infty}{\longrightarrow} \frac{1}{j!(e-1)}$, $i, j = 1, 2, ...$

12. Ланцюг Маркова має матрицю переходів:

$$\begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & .... \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & ... \\ 1/3 & 0 & 0 & 2/3 & 0 & 0 & .... \\ 1/4 & 0 & 0 & 0 & 3/4 & 0 & .... \\ .. & .. & .. & .. & .. & .. & .... \end{pmatrix}.$$

Довести, що $p_{ij}^{(n)} \underset{n\to\infty}{\longrightarrow} 0$, $i, j = 1, 2, ...$

13. Навести приклад ланцюга Маркова, для якого існує стаціонарний розподіл, але не існує ергодичний розподіл.

Лекція 4. Процес Пуасона

4.1 Властивості показникового розподілу. Рахуючий процес

Якщо функцією розподілу випадкової величини ξ є $F(x)$=$\mathbf{P}(\xi \leq x) = 1 - e^{-\lambda x}$, $x \geq 0$ і $F(x) = 0,\ x < 0$, де $\lambda > 0$, тоді кажуть, що ξ має *показниковий розподіл* з параметром λ. Щільність цього розподілу має вигляд: $f(x) = \lambda e^{-\lambda x},\ x \geq 0$ і $f(x) = 0,\ x < 0$. Нескладно показати, що $\mathbf{M}\xi = \lambda^{-1}, \mathbf{D^2}\xi = \lambda^{-2}$.

Нехай $\xi_1, \xi_2, ...$ є незалежними випадковими величинами, які мають показниковий розподіл з параметром λ. Позначимо через $S_n = \sum_{i=1}^{n} \xi_i,\ n = 1, 2, ...$

Лема 1. *Випадкова величина* S_n *має розподіл:*

$$E_n(\lambda, x) \overset{def}{=} \mathbf{P}(S_n \leq x) = 1 - e^{-\lambda x}\ 1 + \frac{\lambda x}{1!} + ... + \frac{(\lambda x)^{n-1}}{(n-1)!}\ , \quad x \geq 0, \tag{1}$$

$$e_n(\lambda, x) \overset{def}{=} \frac{dE_n(\lambda, x)}{dx} = \lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x}, \qquad x \geq 0. \tag{2}$$

$E_n(\lambda, x)$ називається *розподілом Ерланга* з параметрами n, λ.

Доведення виразу (2) нескладно провести методом математичної індукції, використовуючи формулу для щільності суми додатніх незалежних випадкових величин

$$f_{\xi+\eta}(x) = \int_0^x f_\xi(x - y) f_\eta(y) dy.$$

Вираз (1) можна отримати з (2) шляхом інтегрування .

Представимо суми S_n, $n \geq 1$ у вигляді точок на чисельній вісі.

Мал. 1

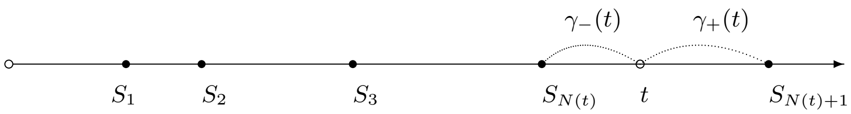
Позначимо

$$N(t) = \max\{n : S_n \leq t\}. \tag{3}$$

Зрозуміло, що $N(t)$ є число сум S_n, які помістилися в інтервалі $[0, t]$. Іншими словами, це є число точок $\bullet$ на мал. 1, які дежать в інтервалі $[0, t]$.
В теорії відновлення, яка буде представлена в наступних лекціях, ці точки називаються *точками регенерації.* Ця назва має наступне походження. Якщо випадкова

величина ξ описує час безаварійної роботи певного приладу, а в момент аварії зіпсутий прилад миттєво замінюється на новий, тоді точки • на мал. 1 є моментами заміни зіпсутого приладу – тобто, у цих точках відбувається відновлення роботи. Процес $N(t),\ t \geq 0$ називається *рахуючим процесом.* Цей процес підраховує число точок відновлення (регенерації) в інтервалі $[0, t]$.

Зрозуміло, що $S_{N(t)}$ є останнім відновлення до моменту t (може бути $S_{N(t)} = t$), а $S_{N(t)+1}$ є першим відновлення після моменту t. Випадкова величина $\gamma_+(t) = S_{N(t)+1} - t$ має назву *перескоку* моменту t послідовністю $S_n,\ n = 1, 2, ...$, а випадкова величина $\gamma_-(t) = t - S_{N(t)+1^-}$ є *недоскоком* до моменту t (див. мал. 2). Випадкова величина $\gamma(t) = \gamma_-(t) + \gamma_+(t) = S_{N(t)+1} - S_{N(t)}$ представляє довжину інтервала, який містить точку t.



Мал. 2

Наступне твердження визначає розподіл цих випадкових величин.

Теорема 1. *Мають місце наступні рівності:*

$$\mathbf{P}(\gamma_+(t) > x) = e^{-\lambda x}, \tag{4}$$

$$\mathbf{P}(\gamma_-(t) > x) = \begin{array}{ll} e^{-\lambda x}, & \textit{якщо} \quad 0 \leq x < t, \\ 0, & \textit{якщо} \quad x \geq t, \end{array} \tag{5}$$

$$\mathbf{P}(\gamma(t) > x) = \begin{array}{ll} e^{-\lambda x}(\lambda x + 1), & \textit{якщо} \quad 0 \leq x < t, \\ e^{-\lambda x}(\lambda t + 1), & \textit{якщо} \quad x \geq t. \end{array} \tag{6}$$

Д о в е д е н н я. Маємо

$$\mathbf{P}(\gamma_+(t) > x) = e^{-\lambda(t+x)} + \sum_{k=1}^{\infty} \mathbf{P}(S_k \leq t, S_{k+1} > t + x) = e^{-\lambda(t+x)} +$$

$$+ \sum_{k=1}^{\infty} \int_0^t \mathbf{P}(S_{k+1} > t + x / S_k = y) e_k(y) dy = \sum_{k=1}^{\infty} \int_0^t \mathbf{P}(\xi_{k+1} > t + x - y) e_k(y) dy =$$

$$+ e^{-\lambda(t+x)} = e^{-\lambda(t+x)} \ 1 + \lambda \int_0^t \sum_{k=1}^{\infty} \frac{(\lambda y)^{k-1}}{(k-1)!} dy \ = e^{-\lambda(t+x)} \ 1 + \lambda \int_0^t e^{-\lambda y} dy \ = e^{-\lambda x}.$$

З наведених міркувань випливає,що

$$\sum_{k=1}^{\infty} \mathbf{P}(S_k \leq t, S_{k+1} > t + x) = e^{-\lambda x} - e^{-\lambda(t+x)}. \tag{7}$$

Тепер будемо мати

$$\mathbf{P}(\gamma_-(t) > x) = I\{x < t\} \ e^{-\lambda t} + \sum_{k=1}^{\infty} \mathbf{P}(S_k \leq t - x, S_{k+1} > t) \ = I\{x < t\} e^{-\lambda x}.$$

Для (6) доведення проведемо у випадку, коли $t \le x$, випадок $t > x$ залищаємо читачеві (завдання 2.1.). Маємо

$$\mathbf{P}(\gamma(t) > x) = e^{-\lambda x} + \sum_{k=1}^{\infty} \mathbf{P}(S_k \le t, S_{k+1} > t, \xi_{k+1} > x) = e^{-\lambda x} +$$

$$+ \sum_{k=1}^{\infty} \int_0^t \mathbf{P}(S_k \le t, S_{k+1} > t, \xi_{k+1} > x / S_k = y) e_k(y) dy = e^{-\lambda x} +$$

$$+ \sum_{k=1}^{\infty} \int_0^t \mathbf{P}(\xi_{k+1} > t - y, \xi_{k+1} > x) e_k(y) dy = e^{-\lambda x} + \sum_{k=1}^{\infty} \int_0^t \mathbf{P}(\xi_{k+1} > x) e_k(y) dy =$$

$$= e^{-\lambda x} + e^{-\lambda x} \lambda \int_0^t dy = e^{-\lambda x}(t\lambda + 1). \qquad \blacktriangleleft$$

Можна поставити питання: яким є математичне сподівання інтервала типу $[S_k, S_{k+1}]$, який накриває момент t? Позаяк ця довжина співпадає з довжиною однієї з випадкових величин ξ_k, для якої $\mathbf{M}\xi_k = 1/\lambda$, тоді на перший погляд здається, що відповідь звучить $1/\lambda$, але з виразів (4), (5) (або з виразу (6)) нескладно підрахувати, що

$$\mathbf{M}\gamma(t) = \mathbf{M}\gamma_+(t) + \mathbf{M}\gamma_-(t) = \frac{2 - e^{-\lambda t}}{\lambda} > \frac{1}{\lambda}. \tag{8}$$

Це є так званий *парадокс часу очікування* (більше інформації можна знайти наприклад в [**?**], стор. 17). Виявляється, що випадкова величина ξ_k, яка реалізує накриття моменту t, має інший розподіл ніж випадкова величина ξ_k з визначеним k. Це є наслідком виразу (6).

Зауваження 1. Про парадокс часу очікування більш розлого поведемо розмову в лекціях, які присвячені теорії відновлення.

4.2 Визначення і властивості процесу Пуассона

Розглянемо послідовність незалежних однаково розподілених випадкових величин $\xi_1, \xi_2, ...$, які мають показниковий розподіл з параметром λ, і $S_n = \sum_{i=1}^n \xi_i,\ n = 0, 1, 2, ...$ Випадкову величину $N(t)$ за визначенням (3) ми трактували як число точок відновлення до моменту часу t і називали рахуючим процесом. За своїм визначенням $N(t)$ є функцією двох арнументів – ω і t. Отже, фактично $N(t)$ є випадковим процесом. Цей процес називається *процесом Пуасона.* Строге визначення має наступний вигляд.

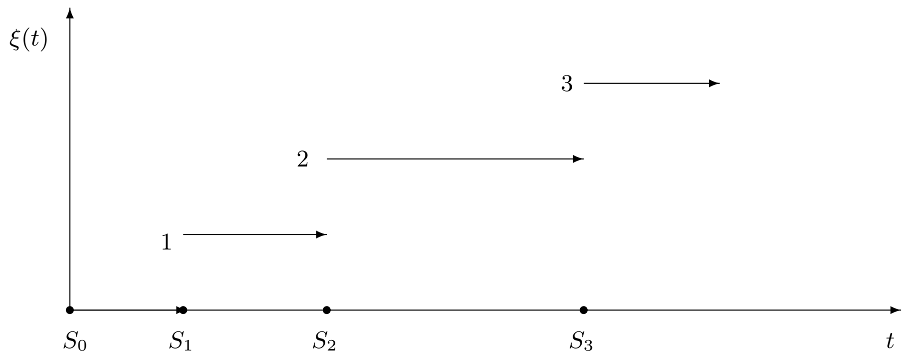
Визначення 1. *Випадковий процес* $\xi(t)$, *визначений як*

$$\xi(t) = \xi(0) + \max\{n \ge 0 : S_n \le t\}, \tag{9}$$

називається процесом Пуассона з параметром λ.

Зауваження 2. Зазвичай у визначенні процесу Пуассона припускають, що $\xi(0) = 0$, тобто процес розпочинається з нульового значення. Тоді процес Пуассона $\xi(t)$ є фактично рахуючим процесом $N(t)$, але умова $\xi(0) = 0$ не є обов'язковою і ситуація, коли $\xi(0) \ne 0$, є цілком можливою. Ба більше $\xi(0)$ може бути навіть випадковою величиною. У подальшому ми будемо припускати, що $\xi(0)$ є сталою величиною.

Наступний графік представляє собою типовий вигляд траєкторії процесу Пуассона (для $\xi(0) = 0$).



Мал. 3

Послідовність сум $S_n,\ n = 0, 1, 2, ...$ і значення $\xi(0)$ визначають процес $\xi(t)$ однозначно і навпаки. Така інтерпретація є зручною для доведення наступного твердження.

Лема 2. *Для довільного $t > 0$ будемо мати:*

$$\mathbf{P}(\xi(t) = k/\xi(0) = 0) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k \geq 0. \tag{10}$$

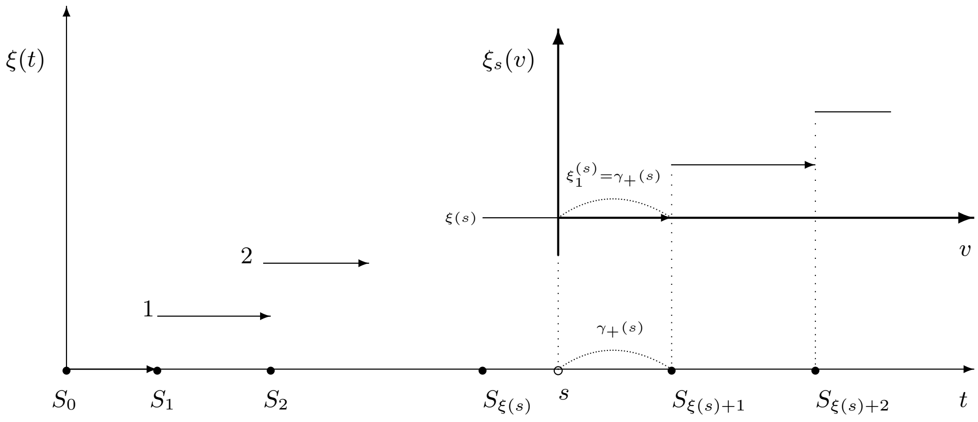
Доведення. Якщо $\{\xi(0) = 0\}$, тоді подія $\{\xi(t) = k\}$ є еквівалентною наступній $\{S_k \leq t, S_{k+1} > t\}$. Тому отримаємо:

$$\mathbf{P}(\xi(t)=k/\xi(0)=0)=\mathbf{P}(S_k \leq t, S_{k+1}>t)=\mathbf{P}(S_k \leq t)-\mathbf{P}(S_k \leq t, S_{k+1} \leq t) =$$
$$= \mathbf{P}(S_k \leq t) - \mathbf{P}(S_{k+1} \leq t) = E_k(\lambda, t) - E_{k+1}(\lambda, t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}. \quad \blacktriangleleft$$

Зауваження 3. Аналогічно можемо довести, що (завдання 2.2.)

$$\mathbf{P}(\xi(t) = j/\xi(0) = i) = e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}, \quad i \leq j, \qquad t \geq 0. \tag{11}$$

Нехай момент s є фіксованим і $\xi(0) = 0$. Тоді $S_{\xi(s)}$ є останньою сумою перед моментом s (включаючи s), а $S_{\xi(s)+1}$ буде першою сумою після моменту s. Визначимо наступні випадкові величини: $\xi_1^{(s)} = \gamma_+(s)$, $\xi_k^{(s)} = S_{\xi(s)+k} - S_{\xi(s)+k-1}$, $k = 2, 3, ...$ Легко довести (завдання 2.3.), що випадкові величини $\xi_k^{(s)}$, $k = 2, 3, ...$ є незалежними і мають той самий показниковий розподіл з параметром λ. Випадкова величина $\xi_1^{(s)} = \gamma_+(s)$ є незалежною від $\xi_k^{(s)}$, $k = 2, 3, ...$, а згідно з твердженням 1 має показниковий розподіл з параметром λ. З цього випливає, що від моменту s послідовність випадкових величин $\xi_k^{(s)}$, $k = 1, 2, 3, ...$ генерує у новій системі координат з часом v (див. мал. 4.) процес Пуассона $\xi_s(v)$.



Мал. 4

Цей процес не залежить від еволюції процесу $\xi(t)$ до моменту s. Вочевидь, що $\xi_s(v) = \xi(v+s) - \xi(s)$. З цієї побудови випливає наступна лема.

Лема 3. *Нехай $\xi(t)$ є процесом Пуассона з параметром λ. Для довільного фіксованого $s \geq 0$ процес $\xi_s(t) \overset{def}{=} \xi(t+s) - \xi(s)$ є процесом Пуассона з параметром λ, який не залежить від $\xi(u)$, $u \leq s$.*

З рівності $\xi_s(t) = \xi(t+s) - \xi(s)$ і леми 2 отримаємо:

$$\mathbf{P}(\xi(t+s) - \xi(s) = k) = \mathbf{P}(\xi_s(t) = k/\xi_s(0) = 0) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \quad k \geq 0.$$

Якщо тепер взяти до уваги (11), тоді будемо мати:

$$\mathbf{P}(\xi(t+s) = j/\xi(s) = i) = \mathbf{P}(\xi(t) - \xi(s) = j - i/\xi(s) = i) =$$
$$= \mathbf{P}(\xi(t) - \xi(s) = j - i) = e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!} = \mathbf{P}(\xi(t) = j/\xi(0) = i), i \leq j.$$

Отже,

$$\mathbf{P}(\xi(t+s) = j/\xi(s) = i) = \mathbf{P}(\xi(t) = j/\xi(0) = i), \quad i \leq j \tag{12}$$

для всіх $t, s \geq 0$.

Рівність (12) означає, що процес Пуассона є процесом *однорідним у часі* (або просто *однорідним*), тобто ймовірносні характеристики цього процесу на інтервалі $[s, t+s]$ залежать тільки від довжини цього інтервалу і не залежать від його розташування.

Якщо $0 \leq t_1 < t_2 < ... < t_n < s$, $t \geq 0$ і $i \leq j$, тоді згідно з попередніми міркуваннями отримаємо:

$$\begin{aligned}&\mathbf{P}(\xi(t+s)=j/\xi(t_1)=i_1,...,\xi(t_n)=i_n,\xi(s)=i)=\\&=\mathbf{P}(\xi(t+s)-\xi(s)=j-i/\xi(t_1)=i_1,...,\xi(t_n)=i_n,\xi(s)=i)=\\&=\mathbf{P}(\xi(t+s)-\xi(s)=j-i)=e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}=\mathbf{P}(\xi(t)=j/\xi(0)=i)=\\&=\mathbf{P}(\xi(t+s)=j/\xi(s)=i).\end{aligned}$$

Отже, для довільного натурального n і довільних $t \geq 0, 0 \leq t_1 < t_2 < ... < t_n < s$, $i \leq j$ будемо мати:

$$\mathbf{P}(\xi(t+s)=j/\xi(t_1)=i_1,...,\xi(t_n)=i_n,\xi(s)=i)=\mathbf{P}(\xi(t+s)=j/\xi(s)=i). \qquad (13)$$

З (13) випливає, що починаючи з моменту s, еволюція процесу $\xi(t)$ залежить тільки від значення цього процесу до моменту s і не задежить від попередньої історії. Ця властивість називається *властивістю Маркова.* Процеси, які мають таку властивість, називають *марковськими процесами.* Ми поведемо мову про такі процеси в подальших розділах підручника.

Отже, сформулюємо властивості процесу Пукассона в наступному твердженні.

Теорема 2. *Процес Пуассона має такі властивості:*

1) *він є однорідним* ;
2) $\mathbf{P}(\xi(t+s)=j/\xi(s)=i) = e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}$, $\quad i \leq j$, *зокрема* $\mathbf{P}(\xi(t)=k/\xi(0)=0) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$, $\quad k \geq 0$;
3) $\mathbf{P}(\xi(t+s)=j/\xi(t_1)=i_1,...,\xi(t_n)=i_n,\xi(s)=i)=\mathbf{P}(\xi(t+s)=j/\xi(s)=i)$

   для довільних натуральних n *і довільних* $0 \leq t$, $0 \leq t_1 < t_2 < ... < t_n < s$, $i_1 \leq i_2 \leq ... \leq i_n \leq i \leq j$;
4) *для довільних* $0 = t_0 \leq t_1 < t_2 < ... < t_n$ *випадкові величини* $\xi(t_i) - \xi(t_{i-1})$, $i = 1, ..., n$ *є незалежними* ;
5) $\mathbf{M}e^{-\mu(\xi(t)-\xi(0))} = e^{\lambda t(e^{-\mu}-1)}$, $\qquad \mathbf{M}z^{\xi(t)-\xi(0)} = e^{\lambda t(z-1)}$.

Як зазначалось раніше, властивість 3) називають властивістю Маркова, а властивість 4) означає, що процес Пуассона є процесом з незалежними приростами. Доведення цієї властивості для $n = 2$ випливає з леми 3. Для $n = 3$ будемо мати:

$$\begin{aligned}&\mathbf{P}(\xi(t_1)=i,\xi(t_2)-\xi(t_1)=j,\xi(t_3)-\xi(t_2)=k)=\\&=\mathbf{P}(\xi(t_3)-\xi(t_2)=k/\xi(t_1)=i,\xi(t_2)-\xi(t_1)=j)\mathbf{P}(\xi(t_1)=i,\xi(t_2)-\xi(t_1)=j)=\\&=\mathbf{P}(\xi(t_3)-\xi(t_2)=k/\xi(t_1)=i,\xi(t_2)=j+i)\mathbf{P}(\xi(t_1)=i)\mathbf{P}(\xi(t_2)-\xi(t_1)=j)=\\&=\mathbf{P}(\xi(t_1)=i)\mathbf{P}(\xi(t_2)-\xi(t_1)=j)\mathbf{P}(\xi(t_3)-\xi(t_2)=k).\end{aligned}$$

Подібні міркування можна провести для $n = 4, 5...$ Також можна застосувати метод математичної індукції.

Доведення властивості 5) залишаємо читачеві (завдання 2.4.).

ЗАВДАННЯ

2.1. Довести (6) у випадку $t \geq x$.

2.2. Довести вираз (11).

2.3. Показати, що для довільного фіксованого s випадкові величини $\xi_k^{(s)} = S_{\xi(s)+k} - S_{\xi(s)+k-1}$, $k = 2, 3, ...$ (див. мал. 4) є незалежними і мають показниковий розподіл з параметром λ.

2.4. Довести властивість 5) твердження 2.

2.5. Довести, що $\mathbf{M}\xi(t) = \lambda t$ і $\mathbf{D^2}\xi(t) = \lambda t$.

Позачимо через $\xi(t)$ процес Пуассона з параметром λ і $\xi(0) = 0$.

2.6. Знайти $\mathbf{P}(\xi(s) = k, \xi(t) = m)$.

2.7. Знайти $\mathbf{P}(\xi(t) = k/\xi(s) = n), \qquad t < s,\ k < n.$

2.8. Показати, що коли $\xi_1(t)$ є процесом Пуассона з параметром λ_1, а $\xi_2(t)$ є процесом Пуассона з параметром λ_2, тоді $\xi_1(t) + \xi_2(t)$ є процессом Пуассона з параметром $\lambda_1 + \lambda_2$.

2.9. Довести, що:

a) $\frac{\xi(t)}{t} \underset{t\to\infty}{\longrightarrow} \lambda$ за ймовірністю;

b) $\lim_{t\to\infty} \mathbf{P}(\frac{\xi(t)-\lambda t}{\sqrt{\lambda t}} < x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-v^2/2} dv.$

2.10. Довести, що процес Пуассона є стохастично неперервним.

Лекція 5. Узагальнений процес Пуассона і процес ризику

5.1 Узагальнений процес Пуассона

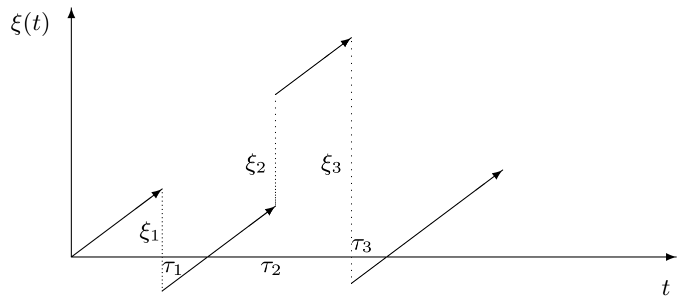
Розглянемо послідовність $\xi_1, \xi_2, ...$ незалежних однаково розподілених випадкових величин з функцією розподілу $F(x)$ і незалежний від $\xi_1, \xi_2, ...$ процес Пуассона $N(t)$ з параметром λ.

Визначення 1. *Випадковий процес* $\xi(t)$, *визначений як*

$$\xi(t) = ct + \sum_{i=1}^{N(t)} \xi_i, \tag{1}$$

називається узагальненим процесом Пуассона. Параметр c називається коефіцієнтом зсуву.

Якщо $\tau_k,\ k = 1, 2, ...$ є послідовними моментами стрибків процесу $N(t)$ ($N(t)$ є процесом Пуассона з параметром λ і $N(0) = 0$), тоді стрибки процесу $\xi(t)$ відбуваються так само в моменти τ_k. Типова траекторія цього процесу для $c > 0$ представлена на мал. 1.



Мал. 1

Зауваження 1. В альтернативному варіанті процес вигляду (1) називається *узагальненим процесом Пуассона з зсувом* при $c \neq 0$. Натомість для $c = 0$ процес називають о *узагальненим процесом Пуассона.*

Зауваження 2. У зазначеному вище визначенні робимо припущення, що $\xi(0) = 0$, тобто процес стартує з нуля, але ситуація, коли $\xi(0) = u$, де u може бути навіть випадковою величиною є можливою. Тоді $\xi(t) = u + ct + \sum_{i=1}^{N(t)} \xi_i$. У подальшому будемо передбачати, що у випадку, коли $\xi(0) = u$, тоді u є не випадковим значенням.

З стохастичної неперервності процесу Пуассона $N(t)$ (завдання 2.1., стор. 314) випливає стохастична непервність узагальненого процесу Пуассона.

Наступне твердження описує властивості узагальненого процесу Пуассона.

Теорема 1. *Узагальнений процес Пуассона має наступні властивості:*

1) *для довільних* $s, t > 0,\ x, y \in R^1$ *має місце*
$$\mathbf{P}(\xi(t+s) \le y/\xi(s)=x) = \mathbf{P}(\xi(t) \le y/\xi(0)=x);$$

2) *для довільного натурального* n *і довільних* $0 \le s_1 < s_2 < ... < s_n < s,\ t > 0,\ x_i, x, y \in R^1, i = 1, ..., n,$ *виконується*
$$\mathbf{P}(\xi(t+s) \le y/\xi(s_1)=x_1, ..., \xi(s_n)=x_n, \xi(s)=x) = \mathbf{P}(\xi(t+s) \le y/\xi(s)=x);$$

3) *для довільних* $0 \le t_1 < t_2 < ... < t_n$ *випадкові величини* $\xi(t_i) - \xi(t_{i-1})$, $i = 1, ..., n$ *є незалежними*;

4)
$$\mathbf{M}e^{i\mu(\xi(t)-\xi(0))} = e^{tk(\mu)}, \quad \mu \in R^1, \tag{2}$$
де $k(\mu) = i\mu c + \lambda(\mathbf{M}e^{i\mu\xi_1} - 1) = i\mu c + \lambda \int\limits_{-\infty}^{\infty} (e^{i\mu x} - 1)dF(x).$

Д о в е д е н н я. Властивість 1) означає, що узагальнений процес Пуассона є однорідним. Властивість 2) свідчить про те, що процес має марковську властивість, тобто є марковським процесом, а властивість 3) визначає наш процес як процес з незалежними приростами. Позаяк випадкові величини ξ_i і процес Пуассона $N(t)$ є незалежними, тоді перші три властивості випливають з першої, другої і третьої властивості процесу Пуассона, які наведені в твердженні 2 (стор. 314) . Виходячи з цього, доведемо тільки властивість 4).

Нехай $\varphi(\mu) = \mathbf{M}e^{i\mu\xi_1}$. Будемо мати:

$$\mathbf{M}e^{i\mu\ \xi(t)-\xi(0)} = \mathbf{M}e^{i\mu\ ct+\sum_{i=1}^{N(t)}\xi_i} = e^{i\mu ct}\sum_{k=1}^{\infty}\mathbf{M}\{e^{i\mu\sum_{i=1}^{N(t)}\xi_i}/N(t)=k\}\times$$
$$\times\,\mathbf{P}(N(t)=k) = e^{i\mu ct-\lambda t}\sum_{k=1}^{\infty}\frac{(\lambda t)^k}{k!}\mathbf{M}e^{i\mu\sum_{i=1}^{k}\xi_i} = e^{i\mu ct-\lambda t}\sum_{k=1}^{\infty}\frac{(\lambda t)^k}{k!}\ \mathbf{M}e^{i\mu\xi_1}\ ^k =$$
$$= e^{i\mu ct-\lambda t}\sum_{k=1}^{\infty}\frac{(\lambda\varphi(\mu)t)^k}{k!} = e^{t\ i\mu c+\lambda(\varphi(\mu)-1)} = e^{tk(\mu)}. \qquad ◄$$

Визначення 2. *Функція* $k(\mu) = i\mu c + \lambda(\varphi(\mu) - 1)$ *називається кумулянтою процесу.*

З однорідності процесу і влпстивості 4) робимо висновок, що

$$\mathbf{M}e^{i\mu(\xi(t)-\xi(s))} = e^{(t-s)k(\mu)}, \qquad 0 \le s \le t.$$

Позаяк для довільних $0 < t_1 < t_2 < ... < t_n$ і $\mu_i \in R^1,\ i = 1, ..., n$ має місце рівність:

$$\sum_{j=1}^{n} \mu_j \xi(t_j) = \xi(t_1) \sum_{i=1}^{n} \mu_i + \ \xi(t_1) - \xi(t_0)\ \sum_{i=2}^{n} \mu_i + \cdots + \ \xi(t_n) - \xi(t_{n-1})\ \mu_n,$$

а прирости процесу є незалежними, отже

$$\mathbf{M}\exp\Big\{ i \sum_{j=1}^{n} \mu_j \xi(t_j) \Big\} =$$

$$= \mathbf{M}\exp\Big\{ i\xi(t_1) \sum_{i=1}^{n} \mu_i + i(\xi(t_1) - \xi(t_1)) \sum_{i=2}^{n} \mu_i + \cdots + i(\xi(t_n) - \xi(t_{n-1}))\mu_n \Big\} =$$

$$= e^{t_1 k \ \sum_{i=1}^{n} \mu_i} \ e^{(t_2 - t_1)k \ \sum_{i=2}^{n} \mu_i} \ \cdots e^{(t_n - t_{n-1})k(\mu_n)}. \tag{3}$$

З визначення 1 випливає, що кожному узагальненому процесу Пуассона відповідає трійка параметрів $c, \lambda, F(x)$, де c – коефіцієнт зсуву, λ – параметр процесу Пуассона і $F(x)$ – функція розподілу стрибків процесу.

З іншої сторони позаяк характеристична функція однозначно визначає розподіл, отже з (3) випливає, що скінченно вимірні розподіли процесу $\xi(t)$ теж однозначно визначаються функцією $k(\mu)$. З визначення функції $k(s)$ випливає, що її однохначно можна задати за допомогою трьох параметрів: c – коефіцієнта зсуву, λ – параметра процесу Пуассона і $F(x)$ – функції розподілу стрибків процесу, отже кожній трійці параметрів $c, \lambda, F(x)$, де $c \in R^1, \lambda > 0$ і $F(x)$ є функцією розподілу на вісі $(-\infty, \infty)$ відповідає узагальнений процес Пуассона вигляду (1).

5.2 Процес ризику

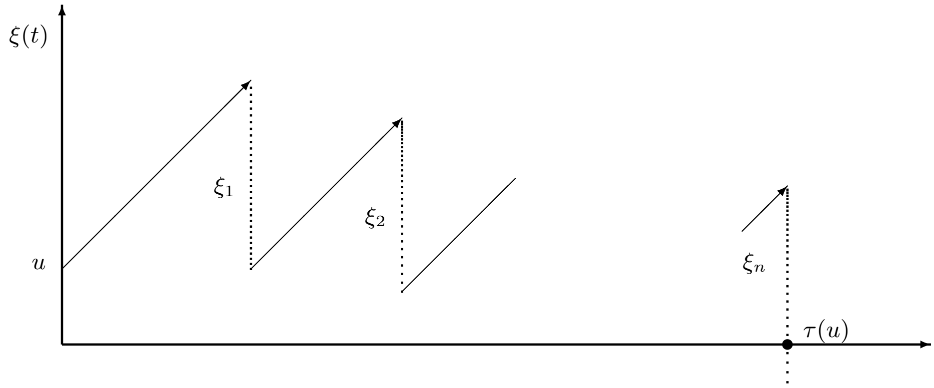
Нехай $\xi_1, \xi_2, ...$ є незалежними однаково розподіленими додатними випадковими величинами зі спільною функцією розподілу $F(x)$, а $N(t),\ N(0) = 0$ процесом Пуассона з параметром λ. Передбачаємо, що $N(t)$ є незалежним від $\xi_1, \xi_2, ...$

Визначення 3. *Випадковий процес*

$$\xi(t) = u + ct - \sum_{i=1}^{N(t)} \xi_i,$$

де $u \geq 0$, $c > 0$ описує класичну модель (або *процес*) *ризику.*

Наведений нижче графік представляє траєкторію процесу.



Малюнок. 2

У цій моделі u відповідає початковому капіталу страхової компанії, натомість c є інтенсивністю сплати внесків. Процес Пуассона $N(t)$ описує число страхових виплат (відшкодувань), які відбулися до моменту t. Випадкова величина ξ_i є страховою сумою, яка виплачена i-тому клієнту.

Така математична модель еволюції капіталу страхової компанії вперше була запропонована шведським математиком Філіпом Лундбергом [1] у 1903 році.

Вочевидь, процес ризику є узагальненим процесом Пуассона з додатним зсувом і від'ємними стрибками. Позаяк у визначенні процесу ризику випадкові величини ξ_i приймають невід'ємні значення, їх розподіл зручніше характеризувати не характеристичними функціями, а перетвореннями Лапласа. Отже, замість виразу(2) будемо мати:

$$\mathbf{M}e^{s(\xi(t)-\xi(0))} = \mathbf{M}e^{s\ \ ct-\sum_{i=1}^{N(t)}\xi_i} \ \ = e^{sct-\lambda t}\sum_{k=1}^{\infty}\frac{(\lambda t)^k}{k!}\mathbf{M}e^{-s\sum_{i=1}^{k}\xi_i} =$$

$$= e^{sct-\lambda t}\sum_{k=1}^{\infty}\frac{(\lambda t\mathbf{M}e^{-s\xi_1})^k}{k!} = e^{tk(s)}, \qquad s \geq 0, \tag{4}$$

де функція $k(s) = sc + \lambda\ \ \mathbf{M}e^{-s\xi_1} - 1\ \ = sc + \lambda\int_0^\infty (e^{-sx}-1)dF(x)$ також називається кумулянтою процесу $\xi(t)$.

Найважливішою характеристикою процесу ризику є момент τ, коли вперше станеться подія $\xi(\tau) < 0$. Така ситуація означає, що страховова компанія не має капіталу, який є достатнім для виплати клієнту. Іншими словами τ є моментом банкрутства компанії. Більшість проблем, які стосуються процесу ризику, пов'язані з дослідженням випадкової величини τ, яка формально визначається як $\tau = \inf\{t : \xi(t) < 0\}$ (див. мал. 2). Нам буде зручно в цьому визначенні вказувати також початковий стан процесу. Отже, дамо більш розлоге визначення моменту банкрутства з урахуванням початкового капіталу u: $\tau(u) = \inf\{t : \xi(t) < 0/\xi(0) = u\}$.

Розглянемо функцію $\psi(u) = \mathbf{M}e^{-\mu\tau(u)},\ \mu \geq 0$.

[1] *Філіп Лундберг (Filip Lundberg)* (2.07.1876- 31.12.1965), шведський математик, фундатор математичної теорії ризику.

Теорема 2. *Функція $\psi(u)$ є розв'язком наступної крайової задачі*

$$\begin{cases} c\frac{d\psi(u)}{du} + \lambda \int\limits_0^\infty \ \psi(u-y) - \psi(u) \ dF(y) - \mu\psi(u) = 0, & u \geq 0, \\ \psi(u) = 1, & u < 0. \end{cases} \tag{5}$$

Д о в е д е н н я наведемо за припущення, що функція розподілу $F(x)$ має щільність $f(x)$. Від самого початку зауважимо, що коли $\xi(0) = u < 0$, тоді згідно з визначенням випадкової величини $\tau(u)$ маємо $\tau(u) = 0$, отже $\psi(u) = 1$ для таких u, що дає виконання крайової умови в (5).

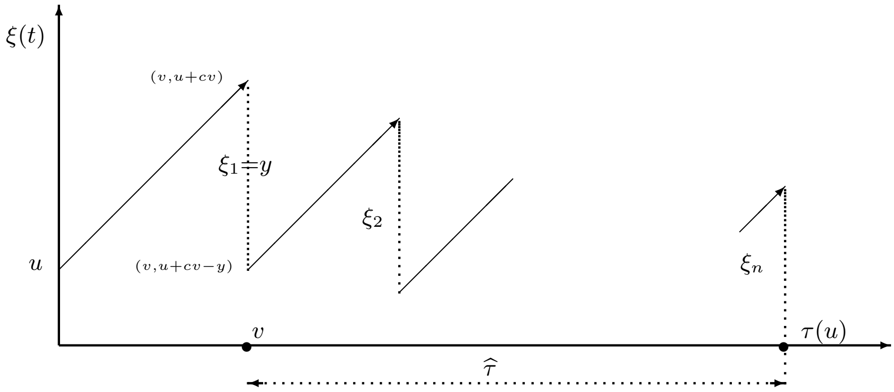
Нехай випадкова величина ζ є моментом першого стрибка процесу $\xi(t)$. З визначення процесу ризику випливає, що ζ і величина першого стрибка ξ_1 є незалежними, отже їх спільна щільність дорівнює $\lambda e^{-\lambda v} f(x)$. Використовуючи формулу повної ймовірності, з урахуванням першого стрибка процесу отримаємо:

$$\mathbf{M}e^{-\mu\tau(u)} = \lambda \int\limits_0^\infty e^{-\lambda v} \int\limits_0^\infty \mathbf{M} \ e^{-\mu\tau(u)}/\zeta{=}v, \xi_1{=}y \ f(y) dv dy. \tag{6}$$

Якщо $\ \xi(0){=}u, \zeta{=}v, \xi_1{=}y\ $, будемо мати $\xi(v) = u + cv - y$ (див. мал. 3). Тепер, якщо $y > u + cv$, тоді $\xi(v) = u + cv - y < 0$. Таким чином банкрутство сталося в момент часу v. Отже, $\tau(u) = v$ і (6) можна переписати у вигляді:

$$\begin{gathered} \mathbf{M}e^{-\mu\tau(u)}{=}\lambda \int\limits_0^\infty e^{-\lambda v} \int\limits_0^{u+cv} \mathbf{M} \ e^{-\mu\tau(u)}/\zeta{=}v, \xi_1{=}y \ f(y) dv dy + \\ +\lambda \int\limits_0^\infty e^{-\lambda v} \int\limits_{u+cv}^\infty e^{-\mu v} f(y) dv dy {=} \lambda \int\limits_0^\infty e^{-\lambda v} \int\limits_0^{u+cv} \mathbf{M} \ e^{-\mu\tau(u)}/\zeta{=}v, \xi_1{=}y \ f(y) dv dy + \\ + \lambda \int\limits_0^\infty e^{-(\lambda+\mu)v} \ 1 - F(u+cv) \ dv. \end{gathered} \tag{7}$$

Якщо $y \leq u + cv$, тоді за умови, що $\ \xi(0){=}u, \zeta{=}v, \xi_1{=}y\ $, будемо мати $\xi(v) = u{+}cv{-}y$ і $\tau(u) = v + \widehat{\tau}$ (див. мал. 3), де випадкова величина $\widehat{\tau}$ є відстанню від моменту v до першого моменту, в якому процес $\xi(t)$ прийме від'ємне значення, за умови $\xi(v) = u{+}cv{-}y$.



Мал. 3

Позаяк процес $\xi(t)$ є марковським, у момент $t=v$ можемо забути про попередню траєкторію цього процесу, тобто про подію $\xi(0)=u, \zeta=v, \xi_1=y$. Подальша поведінка процесу $\xi(t)$, а отже і випадкова величина $\widehat{\tau}$, будуть залежати виключно від значення цього процесу в момент v, тобто від $\xi(v) = u+cv-y$. Процес $\xi(t)$ є однорідним, тоді розподіл випадкової величини $\widehat{\tau}$ буде такий самий як розподіл випадкової величини $\tau(u+cv-y) = \inf\{t : \xi(t) < 0/\xi(0) = u+cv-y\}$.

Отже, для $y \leq u + cv$ отримаємо:

$$\mathbf{M}\ e^{-\mu\tau(u)}/\zeta=v, \xi_1=y\ = \mathbf{M}\ e^{-\mu(v+\widehat{\tau})}/\zeta=v, \xi_1=y\ = e^{-\mu v}\mathbf{M}\ e^{-\mu\widehat{\tau}}/\zeta=v, \xi_1=y\ =$$
$$= e^{-\mu v}\mathbf{M}e^{-\mu\tau(u+cv-y)} = e^{-\mu v}\psi(u+cv-y).$$

Підставляємо це до (7) і отримаємо рівняння для $\psi(u,)$ коли $u \geq 0$.

$$\psi(u)=\lambda\int\limits_0^\infty e^{-(\lambda+\mu)v}\int\limits_0^{u+cv}\psi(u+cv-y)f(y)dvdy+\lambda\int\limits_0^\infty e^{-(\lambda+\mu)v}\ 1-F(u+cv)\ dv.$$

Звідки після заміни $u + cv = x$ будемо мати:

$$\psi(u) = \frac{\lambda}{c}\int\limits_u^\infty e^{-\frac{(\lambda+\mu)(x-u)}{c}}\int\limits_0^x \psi(x-y)f(y)dxdy + \frac{\lambda}{c}\int\limits_u^\infty e^{-\frac{(\lambda+\mu)(x-u)}{c}}\ 1-F(x)\ dx. \quad (8)$$

З останнього рівняння випливає, що функція $\psi(u)$ є диференційованою по $u > 0$. Беручи похідну (8) по u, отримаємо:

$$\frac{d\psi(u)}{du} = -\frac{\lambda}{c}\int\limits_0^u \psi(u-y)f(y)dy - \frac{\lambda}{c}\ 1-F(u)\ +$$
$$+\frac{\lambda+\mu}{c}\Big[\frac{\lambda}{c}\int\limits_u^\infty e^{-\frac{(\lambda+\mu)(x-u)}{c}}\int\limits_0^x \psi(x-y)f(y)dxdy+\frac{\lambda}{c}\int\limits_u^\infty e^{-\frac{(\lambda+\mu)(x-u)}{c}}\ 1-F(x)\ dx\Big].$$

Згідно з (8) вираз у дужках [ ] дорівнює $\psi(u)$. Отже, будемо мати:

$$\frac{d\psi(u)}{du} = -\frac{\lambda}{c}\int_0^u \psi(u-y)f(y)dy - \frac{\lambda}{c}\ \ 1 - F(u)\ \ + \frac{\lambda+\mu}{c}\psi(u).$$

Звідки

$$c\frac{d\psi(u)}{du} + \lambda\int_0^u \psi(u-y)f(y)dy + \lambda\ \ 1 - F(u)\ \ - (\lambda+\mu)\psi(u) = 0, \quad u > 0. \tag{9}$$

Якщо взяти до уваги умову $\psi(u) = 1, u < 0$, тоді

$$\lambda\int_0^u \psi(u-y)f(y)dy = \lambda\int_0^\infty \psi(u-y)f(y)dy - \lambda(1-F(u)).$$

Тепер легко зрозуміти, що рівняння (9) є еквівалентне (5), що й завершує доведення твердження. ◀

ЗАВДАННЯ

1. Нехай $\xi(t)$ є узагальненим процесом Пуассона виду (1) і $\mathbf{M}\xi_1 = m$, $\mathbf{M}\xi_1^2 = \sigma^2$. Довести, що $\mathbf{M}\xi(t) = t(c+\lambda m)$ і $\mathbf{D^2}\xi(t) = t\lambda\sigma^2$.
2. Довести, що коли $\xi_1(t)$, $\xi_2(t)$ є незалежними узагальненими процесами Пуассона з параметрами $(c_1, \lambda_1, F_1(x))$ і $(c_2, \lambda_2, F_2(x))$ відповідно, тоді $\xi_1(t) + \xi_2(t)$ також є узагальненим процесом Пуассона. Знайти трійку параметрів, які відповідають цьому процесу.
3. Нехай $\xi(t)$ є узагальненим процесом Пуассона виду(1) і $\mathbf{M}\xi_1 = m$, $\mathbf{M}\xi_1^2 = \sigma^2$. Довести, що
   a) $\frac{\xi(t)}{t} \underset{t\to\infty}{\longrightarrow} c + \lambda m$ за ймовірністю;
   b) $\lim_{t\to\infty} \mathbf{P}(\frac{\xi(t)-t(c+\lambda m)}{\sigma\sqrt{\lambda t}} \le x) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-v^2/2}dv.$
4. Для процесу ризику $\xi(t)$ з початковим капіталом $u \ge 0$ випадкова величина $\gamma(u) = -\xi(\tau(u))$ описує дефіцит капіталу при виплаті клієнту в момент банкрутства $\tau(u)$. Записати крайову задачу для функції $\psi(u) = \psi(\mu, v, u) = \mathbf{M}e^{-\mu\tau(u)-v\gamma(u)}$, $\mu, v \ge 0$.

Лекція 6. Процес ризику. Ймовірність банкрутства

Нашим наступним завданням буде дослідження ймовірності банкрутства, тобто $\Phi(u) = \mathbf{P}(\tau(u) < \infty)$, яке ми проведемо за додаткової умови:

A) $s_0 = \inf\{s < 0 : \int\limits_0^\infty e^{-sx} dF(x) < \infty\} < 0$ і $k(s_0) > 0$.

Перша частина умови **A**), тобто факт $s_0 < 0$, є еквівалентна тому, що $1 - F(x) = o(e^{(s_0+\varepsilon)x})$ для $x \to \infty$ і довільного $\varepsilon > 0$ (завдання 1). Як побачимо далі, якщо $\mathbf{M}\ \xi(1) - \xi(0)\ \ \le 0$, тоді умова $k(s_0) > 0$ буде виконуватись автоматично.

Згідно з (4) (стор. 319) маємо:

$$\mathbf{M}e^{s(\xi(t)-\xi(0))} = e^{tk(s)}, \qquad s \ge 0, \tag{1}$$

де

$$k(s) = sc + \lambda\ \mathbf{M}e^{-s\xi_1} - 1\ = sc + \lambda \int\limits_0^\infty (e^{-sx} - 1) dF(x).$$

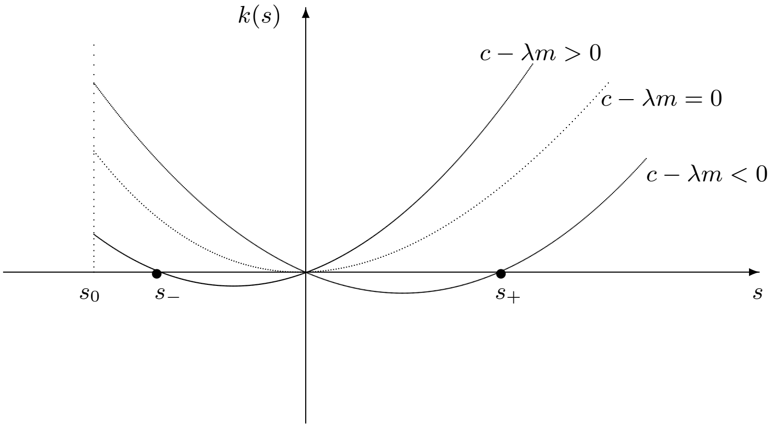
Позаяк

$$k''(s) = \lambda \int\limits_0^\infty e^{-sx} x^2 dF(x) > 0,$$

отже функція $k(s)$ є випуклою для $s > s_0$. Якщо продиференцюємо (1) по s, а потім підставимо $s = 0, t = 1$, тоді отримаємо:

$$\mathbf{M}\ \xi(1) - \xi(0)\ = k'(0) = c - \lambda \mathbf{M}\xi_1 = c - \lambda m.$$

Зрозуміло, що $k(s)/s \to c > 0$, якщо $s \to \infty$. Наступний малюнок представляє графіки функції $k(s)$ в залежності від значення $c - \lambda m$.



Мал. 1

З цього малюнку випливає, що припущення $k(s_0) > 0$ в умові **A**) для $c - \lambda m \leq 0$ виконується автоматично. Наступну лему теж легко отримати, використовуючи мал. 1.

Лема 1. *Рівняння $k(s) = \mu$, $0 < \mu \leq \delta$ має для достатньо малих $\delta > 0$ два розв'язки $s_\pm(\mu)$ такі, що $s_-(\mu) < 0 < s_+(\mu)$ і*

1) *якщо $c - \lambda m > 0$, тоді $\lim_{\mu\to 0} s_+(\mu) = 0$, $\lim_{\mu\to 0} s_-(\mu) = s_-$, де $s_- < 0$ є розв'язком рівняння $k(s) = 0$ на піввісі $s < 0$;*
2) *якщо $c - \lambda m < 0$, тоді $\lim_{\mu\to 0} s_+(\mu) = s_+$, $\lim_{\mu\to 0} s_-(\mu) = 0$, де $s_+ > 0$ є розв'язком рівняння $k(s) = 0$ на піввісі $s > 0$;*
3) *якщо $c - \lambda m = 0$, тоді $\lim_{\mu\to 0} s_\pm(\mu) = 0$.*

6.1 Ймовірність банкрутства

Теорема 1. *Має місце вираз:*

$$\int_0^\infty e^{-su}\mathbf{M}e^{-\mu\tau(u)}du = \frac{1}{s} + \frac{\mu}{s_+(\mu)}\frac{s_+(\mu)-s}{s(k(s)-\mu)}, \qquad s > 0. \tag{2}$$

Д о в е д е н н я. З доведення твердження 2, стор. 320, випливає, що задача (5) (стор. 320) для функції $\psi(u) = \mathbf{M}e^{-\mu\tau(u)}$ еквівалентна розв'язанню рівняння (9) на стор. 322

$$c\frac{d\psi(u)}{du}+\lambda\int_0^u\psi(u-y)dF(y)+\lambda\ 1-F(u)\ -(\lambda+\mu)\psi(u)=0, \quad u>0. \tag{3}$$

Для розв'язку цього рівняння застосуємо метод перетворень Лапласа. З цією метою введемо наступне позначення $\Psi(s) = \int\limits_0^\infty e^{-su}\psi(u)du$. Позаяк

$$\int\limits_0^\infty e^{-su}\frac{d\psi(u)}{du}du = \int\limits_0^\infty e^{-su}d\psi(u) = e^{-su}\psi(u)\Big|_{u=0}^{u=\infty} + s\int\limits_0^\infty e^{-su}\psi(u)du = s\Psi(s) - \psi(0),$$

$$\int\limits_0^\infty e^{-su}\int\limits_0^u \psi(u-y)dF(y)du = \Psi(s)\int\limits_0^\infty e^{-sy}dF(y) = \Psi(s)\varphi(s),$$

$$\int\limits_0^\infty e^{-su}\ 1-F(u)\ du = -\ s^{-1}\ \ e^{-su}\ \ 1-F(u)\ \Big|_{u=0}^{u=\infty} + \int\limits_0^\infty e^{-su}dF(u)\ = \frac{1-\varphi(s)}{s},$$

де $\varphi(s) = \mathbf{M}e^{-s\xi_1}$, отже застосовуючи перетворення Лапласа до рівняння (3), після нескладних перетворень отримаємо:

$$\Psi(s) = \frac{cs\psi(0) - \lambda(1-\varphi(s))}{s(k(s)-\mu)} = \frac{cs\psi(0) - cs + k(s)}{s(k(s)-\mu)}, \qquad s > 0. \tag{4}$$

Функція з лівої сторони рівняння (4) є скінченною для всіх $s > 0$, отже права частина рівняння також повинна бути скінченною. Позаяк для $s = s_+(\mu) > 0$ маємо $k(s_+(\mu)) - \mu = 0$, отже нуль у знаменнику правої сторони (4) має бути компенсований нулем у чисельнику. У зв'язку з чим отримаємо: $cs_+(\mu)\psi(0) - cs_+(\mu) + k(s_+(\mu)) = 0$, звідки $c\psi(0) = (cs_+(\mu) - \mu)/s_+(\mu)$. Тепер будемо мати

$$\Psi(s) = \frac{(cs_+(\mu) - \mu)s/s_+(\mu) - cs + k(s)}{s(k(s)-\mu)} = \frac{k(s) - \mu s/s_+(\mu)}{s(k(s)-\mu)} =$$
$$= \frac{1}{s} + \frac{\mu}{s_+(\mu)}\frac{s_+(\mu) - s}{s(k(s)-\mu)},$$

що завершує доведення. ◀

6.2 Асимптотичні властивості ймовірності банкрутства

Представлення (2) дає можливість знаходження ймовірності банкрутства, тобто $\mathbf{P}(\tau(u) < \infty)$. Позаяк $\mathbf{P}(\tau(u) < \infty) = \lim_{\mu\to 0}\mathbf{M}e^{-\mu\tau(u)}$, достатньо перейти до границі $\mu \to 0$ в (2).

Має місце наступне твердження.

Теорема 2.

$$\mathbf{P}(\tau(u) < \infty) = \begin{cases} 1, & \text{якщо} \quad c - \lambda m \le 0, \\ \frac{e^{s_-u}(\lambda m - c)}{k'(s_-)} + o\ e^{s_-u}\ , & \text{якщо} \quad c - \lambda m > 0. \end{cases} \tag{5}$$

Д о в е д е н н я. Якщо $c - \lambda m < 0$, тоді згідно з пунктом 2) з леми 1 маємо: $\lim_{\mu\to 0} s_+(\mu) = s_+ > 0$. Одночасно з (2) при $\mu \to 0$ отримаємо

$$\int\limits_0^\infty e^{-su}\mathbf{P}(\tau(u) < \infty)du = s^{-1},$$

звідки $\mathbf{P}(\tau(u) < \infty) = 1$.

Нехай тепер $c - \lambda m \geq 0$. Згідно з пунктами 1) і 3) леми 1 будемо мати: $\lim_{\mu \to 0} s_+(\mu) = 0$, отже

$$\lim_{\mu \to 0} \frac{\mu}{s_+(\mu)} = \lim_{\mu \to 0} \frac{k(s_+(\mu))}{s_+(\mu)} = k'(0) = c - \lambda m. \tag{6}$$

Якщо $c - \lambda m = 0$, тоді в той самий спосіб як раніше отримаємо $\mathbf{P}(\tau(u) < \infty) = 1$. Якщо $c - \lambda m > 0$, тоді з (2) при $\mu \to 0$ і (6) будемо мати:

$$\int\limits_0^\infty e^{-su} \mathbf{P}(\tau(u) < \infty) du = \frac{1}{s} - \frac{c - \lambda m}{k(s)}, \qquad s > 0. \tag{7}$$

Функція $1/k(s)$ є регулярною на півплощині $\mathrm{Re}\ s > s_0$ і має полюси першого роду в точках $s = 0, s = s_-$. З теорії аналітичних функцій відомо, що тоді $1/k(s)$ може бути записана у вигляді:

$$\frac{1}{k(s)} = \frac{1}{(c - \lambda m)s} + \frac{1}{k'(s_-)(s - s_-)} + \int\limits_0^\infty e^{-su} K(u) du$$

Інтеграл у цьому виразі є скінченним для всіх $s > s_- - \varepsilon$ для фіксованого $\varepsilon > 0$.

Тепер рівність (7) можемо записати у вигляді:

$$\int\limits_0^\infty e^{-su} \mathbf{P}(\tau(u) < \infty) du = \frac{\lambda m - c}{k'(s_-)(s - s_-)} + (\lambda m - c) \int\limits_0^\infty e^{-su} K(u) du,$$

звідки

$$\mathbf{P}(\tau(u) < \infty) = \frac{\lambda m - c}{k'(s_-)} e^{s_- u} + (\lambda m - c) K(u).$$

Можна довести (завдання 2), що $K(u) = O(e^{(s_- - \varepsilon)u})$, якщо $u \to \infty$ для фіксованого $\varepsilon > 0$. Це завершує доведення нашого твердження. ◀

ЗАВДАННЯ

1. Довести, що коли для функціїї розподілу $F(x)$, $x > 0$ виконується умова $\int\limits_0^\infty e^{\alpha x} dF(x) < \infty$, де $\alpha > 0$, тоді $1 - F(x) = o(e^{-\alpha x})$ для $x \to \infty$.
2. Функція $g(x)$, $x > 0$ є такою, що $|g(x)| \to 0$, якщо $x \to \infty$. Довести, що коли $\int\limits_0^\infty e^{\alpha x} |g(x)| dx < \infty$, де $\alpha > 0$, тоді $|g(x)| = o(e^{-\alpha x})$ для $x \to \infty$.
3. Нехай $c - \lambda m > 0$, де $m = \mathbf{M}\xi_1$. Визначимо функцію розподілу $G(u)$ (показати, що $G(u)$ є функцією розподілу!) наступним чимном:

$$G(u) = \frac{1}{m} \int_0^u (1 - F(v))\, \mathrm{d}v.$$

Використовуючи вираз (7), довести, що

$$\mathbf{P}(\tau(u) < \infty) = 1 - \left(1 - \frac{\lambda m}{c}\right) \sum_{n=0}^\infty \left(\frac{\lambda m}{c}\right)^n G^{*n}(u)),$$

де $G^{*n}(x)$ є n-кратною згорткою функції розподілу $G(u)$ самої з собою.

4. Нехай $c-\lambda m>0$, де $m=\mathbf{M}\xi_1$ і $F(x)=1-e^{-x/m}, x>0$. Довести, що

$$\mathbf{P}(\tau(u)<\infty)=\frac{\lambda m}{c}e^{-(\frac{1}{m}-\frac{\lambda}{c})u}.$$

Лекція 7. Процеси Маркова зі зліченною множиною станів 1

7.1 Визначення процесу Маркова. Функція ймовірності переходу

Нехай випадковий процес $\xi(t)$, $t \geq 0$ є визначеним на ймовірносному просторі $(\mathbf{\Omega}, \mathcal{B}, \mathbf{P})$. Будемо передбачати, що процес приймає значення в множині $E = \{0, \pm 1, ...\}$.

Визначення 1. *Випадковий процес $\xi(t)$, $t \geq 0$ називається процесом Маркова , якщо виконується рівність*

$$\mathbf{P}(\xi(t+s) = j/\xi(t_0) = i_0, \xi(t_1) = i_1, ..., \xi(t_{n-1}) = i_{n-1}, \xi(s) = i) = \\ = \mathbf{P}(\xi(t+s) = j/\xi(s) = i) \tag{1}$$

для довільних $n \geq 0, t, s \geq 0, i, j, i_0, ..., i_{n-1} \in E, 0 \leq t_1 < t_2 < ... < t_{n-1} < s$.

Рівність (1) має назву *властивість Маркова.* З цією властивістю ми вже зустрічались у лекції 4 (рівність(13)). З'ясуємо ще раз сенс цієї властивості. Якщо момент часу s трактувати як теперішній, тоді подія $\{\omega : \xi(t_0) = i_0, \xi(t_1) = i_1, ..., \xi(t_{n-1}) = i_{n-1}, \}$ відбулась у минулому стосовно цього моменту, тобто належить до історії $\xi(t)$. Подія $\{\omega : \xi(s) = i\}$ належить до теперішнього часу, а подія $\{\omega : \xi(t+s) = j\}$ – до майбутнього процесу $\xi(t)$. Рівність (1) означає, що поведінка процесу в майбутньому залежить тільки від його стану в теперішній момент часу і не залежить від історії цього процесу.

Визначення 2. *Марковський процес $\xi(t)$, $t \geq 0$ називається однорідним, якщо*

$$\mathbf{P}(\xi(t+s) = j/\xi(s) = i) = \mathbf{P}(\xi(t) = j/\xi(0) = i), \qquad \forall\ t, s \geq 0.$$

Властивість однорідності означає, що ймовірносні характеристики процесу на інтервалах $[s; t+s]$ і $[0; t]$ є однаковими. Іншими словами ці характеристики залежать тільки від довжини інтервалу і не залежать від його розташування на часовій вісі.

Зауваження 1. У подальшому будемо вивчати тільки однорідні процеси.

Зауваження 2. Також будемо припускати, що траєкторії процесу $\xi(t)$ належать до $D_{[0,\infty)}$, тобто є неперервними справа і мають границі зліва.

Визначення 3. *Функція*

$$P_{ij}(t) = \mathbf{P}(\xi(t) = j/\xi(0) = i)$$

називається перехідною ймовірністю, а матриця $P = (P_{ij}(t))$ називається матрицею перехідних ймовірностей процесу $\xi(t)$.

Перехідна ймовірність має наступні властивості:

1) $P_{ij}(t) \geq 0, \qquad \sum_j P_{ij}(t) = 1, \;\; \forall\, t \geq 0, \; i \in E;$
2) $P_{ij}(0) = \delta_{ij} \qquad \forall\, i, j \in E;$
3) $P_{ij}(t+s) = \sum_{k\in E} P_{ik}(t)P_{kj}(s) \qquad \forall\, t, s \geq 0, \; i, j \in E.$

Властивості 1) і 2) є очевидними. Властивість 3), яку називають *рівнянням Чепмена-Колмогорова*, випливає з властивості Маркова (завдання 1). Умови 1)–3) є необхідними умовами для ймовірностей переходу однорідного марковського процесу. Можна довести, що ці умови також є достатніми – якщо певна функція $P_{ij}(t), i, j \in E, t \geq 0$ має властивості 1)–3), тоді існує однорідний марковський процес $\xi(t)$ з множиною станів E, для якого функція $P_{ij}(t)$ є перехідною ймовірністю.

Нехай $P_i = \mathbf{P}(\xi(0) = i)$ є *початковим розподілом* процесу $\xi(t)$. Для довільних $n = 1, 2, ..., 0 \leq t_0 < t_1 < ... < t_n$, $i_0, i_1, ...i_n \in E$, з властивості Маркова, однорідності і рівняння Чепмена-Колмогорова випливає наступна формула для скінченно вимірного розподілу $\xi(t)$ (завдання 2)

$$\mathbf{P}(\xi(t_0){=}i_0, \xi(t_1){=}i_1, ..., \xi(t_n){=}i_n){=}P_{i_0}P_{i_0 i_1}(t_1{-}t_0)\cdots P_{i_{n-1}i_n}(t_n{-}t_{n-1}). \qquad (2)$$

Як вже зазначалось, властивості 1)-3) ймовірності переходів є наслідком визначення однорідного процесу Маркова. Наступна властивість не випливає з цього визначення, але в наших міркуваннях завжди будемо припускати, що вона має місце:

4) $\lim_{t\downarrow} P_{ij}(t) = \delta_{ij}$.

Властивість 4) накладає на ймовірність переходу $F_{ij}(t)$ умову, яка свідчить про те, що ця функція є неперервною в точці $t = 0$. Можна довести (завдання 3), що умова 4) є еквівалентна стохастичній непервності процесу $\xi(t)$. У подальшому стане зрозумілим, що ця умова забезпечує певні властивості регулярності як для самої ймовірності переходів, так і для траєкторії процесу $\xi(t)$. Виявляється, що при виконанні цієї умови ймовірність переходів є рівномірно неперервна стосовно t.

Теорема 1. *Функція $P_{ij}(t)$ є рівномірно неперервною стосовно t.*

Д о в е д е н н я. Нехай $h > 0$. Використовуючи властивість 3) ймовірностей переходів, отримаємо:

$$P_{ij}(t+h){-}P_{ij}(t) = \sum_{k\in E} P_{ik}(h)P_{kj}(t){-}P_{ij}(t){=}\sum_{k\neq i} P_{ik}(h)P_{kj}(t){+}P_{ii}(h)P_{ij}(t){-}$$
$$-\, P_{ij}(t) = \sum_{k\neq i} P_{ik}(h)P_{kj}(t) + (P_{ii}(h) - 1)P_{ij}(t).$$

Звідки

$$|P_{ij}(t+h) - P_{ij}(t)| \leq \sum_{k\neq i} P_{ik}(h)P_{kj}(t) + (1 - P_{ii}(h))P_{ij}(t) \leq \sum_{k\neq i} P_{ik}(h) +$$
$$+\, (1 - P_{ii}(h))P_{ij}(t) \leq 1 - P_{ii}(h) + 1 - P_{ii}(h) = 2(1 - P_{ii}(h)) \underset{h\to 0}{\to} 0.$$

Якщо $h{<}0$, тоді аналогічним чином будемо мати: $|P_{ij}(t+h) - P_{ij}(t)| \leq 2(1{-}P_{ii}(|h|)) \underset{h\to 0}{\to} 0.$ ◀

Наступний результат, який наведемо без доведення, показує, що умова 4) забезпечує певні глибші властивості ймовірності переходів, зокрема диференційованість у нулі.

Лема 1. *Існують наступні границі:*

$$\lim_{t\downarrow 0} \frac{P_{ii}(t)-1}{t} = a_{ii} \geq -\infty, \qquad \lim_{t\downarrow 0} \frac{P_{ij}(t)}{t} = a_{ij} < \infty. \tag{3}$$

Зауваження 3. Як випливає з цієї леми $a_{ij} < \infty$, якщо $i \neq j$, але значення a_{ii} може дорівнювати $-\infty$. Можна довести, що коли збіжність в умові 4) є рівномірною стосовно i, тоді $|a_{ij}| < \infty$ для всіх $i, j \in E$. Зокрема так відбувається, якщо множина станів процесу є скінченною.

7.2 Структура траєкторії процесу Маркова

Нехай $\tau_i = \inf\{t > 0 : \xi(t) \neq i/\xi(0) = i\}$ є першим моментом, коли процес, стартуючи в момент $t = 0$ зі стану i, залишить цей стан.

Теорема 2. *Для довільного $i \in E$ має місце:*

$$\mathbf{P}(\tau_i > t) = e^{a_{ii}t}, \qquad t \geq 0. \tag{4}$$

Д о в е д е н н я. Нехай t є фіксованим. Для кожного $n = 1, 2, ...$ визначимо подію

$$A_n = \{\xi(\frac{kt}{2^n}) = i, k = 0, 1, ..., 2^n\},$$

яка полягає в тому, що в точках $\frac{kt}{2^n}$, $k = 0, 1, ..., 2^n$ процес перебуває в стані i. Легко зрозуміти, що послідовність множин A_n, $n = 1, 2, ...$ є спадною, а множина точок $P(t) = \{\frac{kt}{2^n},\ n = 1, 2, ...,\ k = 0, 1, 2, ..., 2^n\}$ є щільною в інтервалі $[0, t]$. Позаяк траєкторії процесу $\xi(t)$ є неперервними справа і мають границі зліва, тоді $\{\tau_i > t\} = \{\xi(s) = i,\ 0 \leq s \leq t\} = \{\xi(s) = i,\ s \in P(t)\} = \cap_{n=1}^{\infty} A_n = \lim_{n\to\infty} A_n$. Після чого отримуємо:

$$\mathbf{P}(\tau_i > t) = \mathbf{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbf{P}(A_n) = \lim_{n\to\infty} \Big[P_{ii}(t/2^n)\Big]^{2^n}.$$

Остання рівність випливає з властивості Маркова й однорідності процесу.

З (3) можна зробити висновок, що $P_{ii}(h) = 1 + a_{ii}h + o(h)$, якщо $h \to 0$. Отже,

$$\mathbf{P}(\tau_i > t) = \lim_{n\to\infty} \Big[1 + a_{ii}t/2^n + o(t/2^n)\Big]^{2^n} = e^{a_{ii}t},$$

що й завершує доведення. ◀

З доведеного твердження випливає, що випадкова величина τ_i має показниковий розподіл з параметром $-a_{ii}$. Цей параметр можемо використати для класифікації станів процесу Маркова.

1) Якщо $a_{ii} = 0$, тоді $\mathbf{P}(\tau_i > t) = 1$, отже $\mathbf{P}(\tau_i \leq t) = 0$ для довільного $t \geq 0$. Це означає, що з ймовірністю 1 має місце $\tau_i = \infty$, тобто процес ніколи не залишить стан i. Тоді кажуть, що стан i є *поглинаючим.*
2) Якщо $a_{ii} = -\infty$, to $\mathbf{P}(\tau_i > t) = 0$, отже $\mathbf{P}(\tau_i \leq t) = 1$ для довільного $t \geq 0$. Це означає, що з ймовірністю 1 має місце $\tau_i = 0$, тобто процес після потрапляння до стану i миттєво його залишає. У такому випадку кажуть, що стан i є *миттєвим.*

3) Якщо $-\infty < a_{ii} < 0$, тоді стан i називають *станом зі затримкою.* Після потрапляння до такого стану процес перебуває в ньому певний скінченний час, а потім стається стрибок до іншого стану.

Нехай $|E| < \infty$, тобто число станів процесу є скінченним. Якщо рівняння $\sum_{j\in E\setminus\{i\}} P_{ij}(h) = 1 - P_{ii}(h)$ поділити на h і перейти до границі $h \to 0$, тоді отримаємо $\sum_{j\in E\setminus\{i\}} a_{ij} = - a_{ii}$. Позаяк $0 \le a_{ij} < \infty$, тоді з останньої рівності випливає, що $|a_{ii}| < \infty$. Отже, якщо множина станів процесу є скінченною, тоді він не має миттєвих станів і має місце:

$$\sum_{j\in E} a_{ij} = 0 \tag{5}$$

для кожного $i \in E$.

Якщо $|E| = \infty$, тоді, взагалі кажучи, рівність (5) не виконується, але можна довести, що $\sum_{j\in E\setminus\{i\}} a_{ij} \le -a_{ii}$ (або $\sum_{j\in E} a_{ij} \le 0$) (завдання 4). Тепер рівність $a_{ii} = -\infty$ для певних станів i може мати місце, тобто можуть бути миттєві стани. У такому випадку траєкторії процесу можуть бути досить екзотичними – теорія таких процесів є складною. У практичних застосуваннях, зазвичай, не мають справ з такими процесами. Ось чому дамо наступне визначення.

Визначення 4. *Стан i називається регулярним, якщо* $|a_{ii}| < \infty$ *і* $\sum_j a_{ij} = 0$. *Процес Маркова називається регулярним, якщо всі його стани є регулярними.*

Зауваження 4. Як вже було доведено перед тим, якщо множина станів процесу є скінченною, тоді він є регулярним.

Нехай, як це було й раніше, $\tau_i = \inf\{t > 0 : \xi(t) \ne i/\xi(0) = i\}$ є першим моментом, у який процес, стартуючи в момент $t = 0$ зі стану i, залишить цей стан.

Теорема 3. *Нехай стан i є регулярним і* $a_{ii} < 0$. *Тоді:*

1) *для довільного* $j \ne i$ *має місце рівність:*

$$\mathbf{P}(\tau_i > t, \xi(\tau_i) = j) = \frac{a_{ij}}{-a_{ii}} e^{a_{ii}t}, \qquad t \ge 0; \tag{6}$$

2) *випадкові величини* τ_i *і* $\xi(\tau_i)$ *є незалежними і мають розподіл:*

$$\mathbf{P}(\tau_i > t) = e^{a_{ii}t}, \qquad \mathbf{P}(\xi(\tau_i) = j) = -\frac{a_{ij}}{a_{ii}}. \tag{7}$$

Д о в е д е н н я. Перша формула в (7) була доведена в твердженні 2. Друга нескладно випливає з (6), якщо покласти $t = 0$. Тепер з формул (6), (7) отримаємо $\mathbf{P}(\tau_i > t, \xi(\tau_i) = j) = \mathbf{P}(\tau_i > t)\mathbf{P}(\xi(\tau_i) = j)$, що обумовлює незалежність випадкових величин $\tau_i,\ \xi(\tau_i)$.

Доведемо тепер перший пункт твердження. Нехай t є фіксованим. Для довільних $n = 1, 2, ...,\ k = 0, 1, 2, ...$ окреслимо подію

$$A_n(k) = \{\xi(\frac{lt}{2^n}) = i, l = 0, 1, ..., 2^n + k, \xi(\frac{(2^n + k + 1)t}{2^n}) = j\},$$

яка означає, що в точках $\frac{lt}{2^n}, l = 0, 1, ..., 2^n + k$ процес був у стані i, а в інтервалі $(\frac{(2^n+k)t}{2^n}), \frac{(2^n+k+1)t}{2^n}]$ настала зміна стану процесу і в момент $\frac{(2^n+k+1)t}{2^n}$ він є в стані j. Легко зрозуміти, що $A_n(k) \cap A_n(m) = \emptyset$ для $k \ne m$ і послідовність множин $A_n = \cup_{k=0}^{\infty} A_n(k)$ є спадною, якщо $n \to \infty$. Множина точок $P(t) = \{\frac{kt}{2^n}, n =$

$1, 2, ..., \; k = 0, 1, 2, ...\}$ є щільною на піввісі $x \geq 0$. Траєкторії процесу $\xi(t)$ є неперервними справа і мають границі зліва, отже неперервність або існування розривів цього процесу однозначно визначаються його поведінкою на множині $P(t)$, тому $\{\tau_i > t, \xi(\tau_i) = j\} = \lim_{n\to\infty} A_n$. Будемо мати:

$$\mathbf{P}(\tau_i > t, \xi(\tau_i) = j) = \mathbf{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbf{P}(A_n) = \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbf{P}(A_n(k)).$$

З властивості Маркова і однорідності робимо висновок, що

$$\mathbf{P}(A_n(k)) = \Big[P_{ii}(t/2^n)\Big]^{2^n+k} P_{ii}(t/2^n).$$

З (3) випливає, що $P_{ii}(h) = 1 + a_{ii}h + o(h)$ і $P_{ij}(h) = a_{ij}h + o(h)$, якщо $h \to 0$, звідки отримаємо:

$$\mathbf{P}(\tau_i > t, \xi(\tau_i){=}j){=}\lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbf{P}(A_n(k)){=}\lim_{n\to\infty} \sum_{k=0}^{\infty} \Big[P_{ii}(t/2^n)\Big]^{2^n+k} P_{ij}(t/2^n){=}$$

$$= \lim_{n\to\infty} \frac{\Big[P_{ii}(t/2^n)\Big]^{2^n} P_{ij}(t/2^n)}{1 - P_{ii}(t/2^n)} {=} \lim_{n\to\infty} \frac{\Big[1{+}\frac{a_{ii}t}{2^n}{+}o(\frac{t}{2^n})\Big]^{2^n} (\frac{a_{ij}t}{2^n}{+}o(\frac{t}{2^n}))}{-a_{ii}t/2^n{-}o(t/2^n)}{=}$$

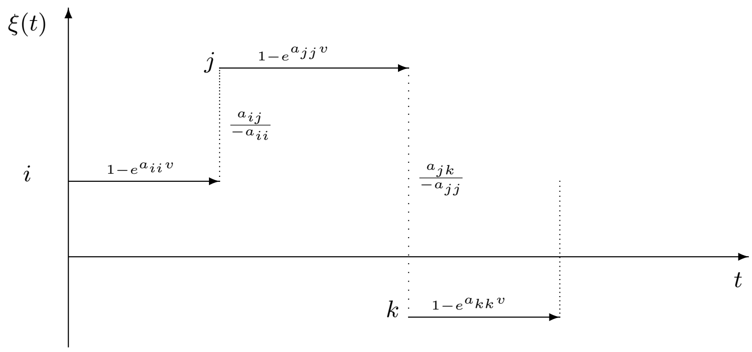
$$= -\lim_{n\to\infty} \frac{\Big[1 + a_{ii}t/2^n + o(2^{-n})\Big]^{2^n} \; a_{ij}t + 2^n o(2^{-n})}{a_{ii}t + 2^n o(2^{-n})} = -e^{a_{ii}t} \frac{a_{ij}}{a_{ii}},$$

що завершує доведення. ◄

Зауваження 5. Умова $a_{ii} \neq 0$ в доведеному твердженні є важливою, якби $a_{ii} = 0$, тоді стан i є поглинаючим і не відбудеться жодного стрибка з цього стану.

Зауваження 6. Якщо $-\infty < a_{ii} < 0$, але $\sum_{j\in E\setminus\{i\}} a_{ij} < -a_{ii}$, тобто стан i не є регулярним, тоді, як можна зауважити з доведення твердження 3, формула (6) залишається в силі. Але пункт 2) цього твердження буде мати інший вигляд (завдання 5).

Дамо тепер опис поведінки траєкторії регулярного процесу Маркова $\xi(t)$, заздалегідь припускаючи, що він не має поглинаючих станів. Нехай $\xi(0) = i$ (див. мал. 1). Процес $\xi(t)$ перебуває в стані i випадковий час з функцією розподілу $1 - e^{a_{ii}v}$. Потім відбувається стрибок до стану j з ймовірністю $\frac{-a_{ij}}{a_{ii}}$, причому момент стрибка і його величина є незалежними. З моменту потрапляння до стану j ситуація повторюється – процес $\xi(t)$ проводить у стані j випадковий час з функцією розподілу $1 - e^{a_{jj}v}$, потім відбувається стрибок, скажімо, до стану k з ймовірністю $\frac{-a_{kj}}{a_{jj}}$. Так само момент стрибка і його величина є незалежними і т.д.



Мал. 1

Нехай $\tau(0) = 0$, а $\tau(n)$ є n-тим стрибком процесу $\xi(t)$, $n = 1, 2, ...$ Наступні твердження легко отримати з твердження 3.

Теорема 4. *Якщо процес $\xi(t)$ є регулярним і не має поглинаючих станів, тоді послідовність $\xi_n = \xi(\tau(n)), n = 0, 1, 2,$ є однорідним ланцюгом Маркова з матрицею переходів*

$$P = \left(-\frac{a_{ij}}{a_{ii}}(1 - \delta_{ij}) \right)_{ij \in E}.$$

Визначення 5. *Ланцюг $\xi_n, n = 0, 1, 2, ...$ називається вкладеним ланцюгом Маркова.*

Зауваження 7. Якщо процес $\xi(t)$ не є регулярним, тоді, як випливає з завдання 5., послідовність $\tau(n),\ n \geq 0$ (а, отже і послідовність $\xi_n = \xi(\tau(n)), n \geq\ 0$) може бути скінченною. Відповідно в такій ситуації ланцюг $\xi_n = \xi(\tau(n)), n \geq 0$ може обірватись.

Зауваження 8. Якщо подія $\{\xi(\tau(i)) = \infty\}$ має місце, тоді процес $\xi(t)$ у момент $\tau(i)$ зникає з простору E. У цій ситуації будемо говорити, що процес обірвався. Іншими словами до моменту i-того стрибка процес $\xi(t)$ існував, а після цього моменту зникає.

ЗАВДАННЯ

1. Довести рівняння Чепмена-Колмогорова.
2. Довести формулу (2).
3. Довести, що умова 4) стосовно ймовірності переходів є еквівалентною стохастичній неперервності процесу $\xi(t)$.
4. Довести, що завжди $\sum_{j\in E} a_{ij} \le 0$.

   Вказівка. *Записати множину E у вигляді $E = \cup_{m=1}^{\infty} E_m$, де $E_m \subset E_{m+1}$, і $|E_m| = m$, $m = 1, 2, ...$*

5. Нехай $-\infty < a_{ii} < 0$ і $\sum_{j\in E\setminus\{i\}} a_{ij} < -a_{ii}$. Довести, що:

   a) випадкова величина $\xi(\tau_i)$ приймає значення $+\infty$ з додатньою ймовірністю (тобто має невласний розподіл [1]). Знайти $\mathbf{P}(\xi(\tau_i) = \infty)$.

   b) $\mathbf{P}(\tau_i > t, \xi(\tau_i) = \infty) = \mathbf{P}(\tau_i > t)\mathbf{P}(\xi(\tau_i) = \infty)$, тобто пункт 2) твердження 3 формально залишається в силі.

2.1. Описати вкладений ланцюг Маркова, якщо регулярний процес МАркова $\xi(t)$ має поглинаючі стани.

[1] Будемо говорити, що випадкова величина ξ має *невласний розподіл*, якщо $\mathbf{P}(|\xi| = \infty) > 0$. У цьому випадку також будемо казати, що випадкова величина ξ є *невласною випадковою величиною.*

Лекція 8. Процеси Маркова зі зліченною множиною станів 2

З результатів, представлених у попередній лекції, випливає, що коли виконується умова 4) (стор. 329), тоді ймовірність переходів $P_{ij}(t)$ є рівномірно неперервною стосовно t і диференційованою в точці $t = 0$. У даній лекції доведемо, що для регулярних процесів цей результат можна посилити, показавши, що ймовірність переходів є диференційованою для кожного t.

8.1 Системи рівнянь Колмогорова

Теорема 1. *Нехай $|E| < \infty$, тобто множина станів процесу $\xi(t)$ є скінченною. Тоді ймовірність переходів задовільняє наступним системам рівнянь:*

$$\frac{dP_{ij}(t)}{dt} = \sum_{k \in E} a_{ik} P_{kj}(t), \tag{1}$$

$$\frac{dP_{ij}(t)}{dt} = \sum_{k \in E} P_{ik}(t) a_{kj}. \tag{2}$$

Зауваження 1. Система рівнянь (1) називається *першою* (або *зворотньою*) системою рівнянь Колмогорова, система рівнянь (2) натомість називається *другою* (або *прямою*) системою рівнянь Колмогорова.

Д о в е д е н н я. Для $h > 0$ будемо мати:

$$P_{ij}(t+h) = \sum_{k \in E} P_{ik}(h) P_{kj}(t),$$

звідки

$$\frac{P_{ij}(t+h) - P_{ij}(t)}{h} = \sum_{k \in E, k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) + \frac{P_{ii}(h) - 1}{h} P_{ij}(t). \tag{3}$$

Позаяк існує границя правої сторони цієї рівності, якщо $h \to 0$, тоді існує також границя лівої сторони. Після переходу до границі $h \to 0$ в (3) отримаємо:

$$\frac{d^+ P_{ij}(t)}{dt} = \sum_{k \in E, k \neq i} a_{ik} P_{kj}(t) + a_{ii} P_{ij}(t) = \sum_{k \in E} a_{ik} P_{kj}(t), \tag{4}$$

де $\frac{d^+}{dt}$ означає правосторонню похідну. Так само для $h < 0$ можемо довести (завдання 1), що

$$\frac{d^- P_{ij}(t)}{dt} = \sum_{k \in E} a_{ik} P_{kj}(t), \tag{5}$$

де $\frac{d^-}{dt}$ означає лівосторонню похідну. Тепер з (4), (5) будемо мати (1).
Доведення (2) є практично аналогічним. ◂

Якщо множина станів є нескінченною, тоді в загальному випадку рівняння (1) і (2) не виконуються без додаткових припущень. Натомість має місце наступне твердження.

Теорема 2. *Нехай* $|E| = \infty$. *Тоді:*

1) *для регулярного процесу* $\xi(t)$ *ймовірність переходів* $P_{ij}(t)$ *задовільняє зворотній системі рівнянь Колмогорова;*
2) *якщо рівномірно стосовно* i*-го стану виконується* $\lim_{h\to 0} P_{ij}(h)/h = a_{ij}$, *тоді ймовірність переходів* $P_{ij}(t)$ *задовільняє прямій системі рівнянь Колмогорова.*

Доведення. Якщо взяти до уваги доведення попереднього твердження, тоді для доведення пункту 1) достатньо показати, що для регулярного процесу можна перейти до границі $h \to 0$ в (3) під знаком $\sum$. Отже, треба довести, що

$$\lim_{h\to 0} h^{-1} \sum_{k\in E, k\neq i} P_{ik}(h)P_{kj}(t) = \sum_{k\in E, k\neq i} a_{ik}P_{kj}(t). \tag{6}$$

Розглянемо лише випадок $h > 0$. Нехай E_m є підмножиною E, яка містить m елементів і такою, що $i \notin E_m$, $E_m \subset E_{m+1}, m = 1, 2, ...$ і $E \backslash \{i\} = \cup_{k=1}^{\infty} E_k = \lim_{m\to\infty} E_m$.

Зрозуміло, що

$$\varliminf_{h\to 0} h^{-1} \sum_{k\in E, k\neq i} P_{ik}(h)P_{kj}(t) \geq \varliminf_{h\to 0} h^{-1} \sum_{k\in E_m} P_{ik}(h)P_{kj}(t) = \sum_{k\in E_m} a_{ik}P_{kj}(t).$$

Звідки при $m \to \infty$ будемо мати:

$$\varliminf_{h\to 0} h^{-1} \sum_{k\in E, k\neq i} P_{ik}(h)P_{kj}(t) \geq \sum_{k\in E, k\neq i} a_{ik}P_{kj}(t). \tag{7}$$

З іншого боку

$$\sum_{k\in E, k\neq i} P_{ik}(h)P_{kj}(t) \leq \sum_{k\in E_m} P_{ik}(h)P_{kj}(t) + \sum_{k\in E\backslash E_m} P_{ik}(h) =$$
$$= \sum_{k\in E_m} P_{ik}(h)P_{kj}(t) + 1 - P_{ii}(h) - \sum_{k\in E_m} P_{ik}(h).$$

Поділивши на h і перейшовши до границі $\varlimsup\limits_{h\to 0}$, отримаємо

$$\varlimsup_{h\to 0} h^{-1} \sum_{k\in E, k\neq i} P_{ik}(h)P_{kj}(t) \leq \sum_{k\in E_m} a_{ik}P_{kj}(t) + a_{ii} - \sum_{k\in E_m} a_{ik}. \tag{8}$$

Якщо взяти до уваги припущення регулярності, тоді з (8) для $m \to \infty$ будемо мати:

$$\varlimsup_{h\to 0} h^{-1} \sum_{k\in E, k\neq i} P_{ik}(h)P_{kj}(t) \leq \sum_{k\in E, k\neq i} a_{ik}P_{kj}(t),$$

що разом з (7) дає (6).

Аби довести 2) зауважимо, що

$$\frac{P_{ij}(t+h)-P_{ij}(t)}{h}=\sum_{k\in E,k\neq j}P_{ik}(t)\frac{P_{kj}(t)}{h}+\frac{P_{jj}(h)-1}{h}P_{ij}(t).$$

Припущення для пункту 2) дає можливість перейти до границі під знаком $\sum$ у цій рівності, що й завершує доведення (2).

◄

8.2 Інфінітезимальна матриця переходів

Визначення 1. *Матриця $A = \|a_{ij}\|$ називається інфінітезимальною матрицею переходів або простіше інфінітезимальною матрицею.*

Вочевидь, інфінітезимальна матриця задовільняє умовам:

a) $a_{ij} \geq 0, i \neq j, a_{ii} \leq 0$ для всіх $i, j \in E$;
b) $\sum_{j\in E} a_{ij} = 0$ для всіх $i \in E$.

Зворотню і пряму систему рівнянь Колмогорова тепер можна записати в матричному вигляді

$$\frac{dP(t)}{dt}=AP(t), \qquad \frac{dP(t)}{dt}=P(t)A. \tag{9}$$

Якщо взяти до уваги, що $P(0) = I = \|\delta_{ij}\|$, тоді крім рівнянь (9) отримаємо наступні початкові умови:

$$\begin{cases}\frac{dP(t)}{dt} &= AP(t),\\ P(0) &= I,\end{cases} \qquad \begin{cases}\frac{dP(t)}{dt} &= P(t)A,\\ P(0) &= I.\end{cases} \tag{10}$$

Виникає наступне питання. Нехай задана матриця $A = \|a_{ij}\|$, для якої виконуються умови a) і b). Чи завжди існує регулярний марковський процес $\xi(t)$ з інфінітезимальною матрицею A? Якщо такий процес існує, тоді з твердження 2 випливає, що його ймовірність переходів повинна бути розв'язком першої системи рівнянь (10), тобто

$$\frac{dP(t)}{dt}=AP(t), \qquad P(0)=I. \tag{11}$$

Або це ж питання можна сформулювати наступним чином. Задана є матриця $A = \|a_{ij}\|$, для якої виконуються умови a) і b). Чи існує розв'язок задачі (11) такий, який одночасно є ймовірністю переходів певного регулярного процесу Маркова? Вочевидь, тоді матриця A буде інфінітезимальною матрицею цього процесу. Якщо цей розв'язок буде єдиним, тоді матриця A однозначно визначає процес Маркова. Виявляється, що відповідь на це питання не є легкою. Далі покажемо і доведемо для одного з прикладів, що існує позитивна відповідь на наше запитання. Але для цього нам буде потрібне наступне визначення.

Нехай задана иатриця $B = \|b_{ij}\|$, норма якої визначається як $||B|| = \sup_i \sum_j |b_{ij}|$.

Визначення 2. *Експоненту матриці B, норма якої $||B|| < \infty$, будемо визначати як*

$$e^B \stackrel{def}{=} \sum_{n=0}^{\infty}\frac{B^n}{n!}.$$

Якщо скористатись властивостями рядів, легко пересвідчитись, що операції з експонентою матриці мають ті самі властивості, що й операції зі звичайною числовою показниковою функцією, тобто $e^{A+B} = e^A e^B$, $\frac{d^n e^{Bt}}{dt^n} = B^n e^{Bt} = e^{Bt} B^n$. Аналогічно, як і для звичайної показникової функції, має місце:

$$e^B = \lim_{n\to\infty} \left(I + \frac{B}{n} \right)^n.$$

Теорема 3. *Нехай $E = \{0, \pm 1, ...\}$ і задана иатриця $A = (a_{ij})_{ij\in E}$ така, що для довільних $i, j \in E$ має місце $0 \le a_{ij} < \infty, -\infty < a_{ii} < 0$ і $\sum_j a_{ij} = 0$. Якщо $\sup_i |a_{ii}| < \infty$, тоді існує однорідний процес Маркова $\xi(t)$ з інфінітезимальними характеристиками a_{ij} і матрицею ймовірністей переходів $P(t) = e^{At}$.*

Д о в е д е н н я. Маємо $||A|| = \sup_i \sum_j |a_{ij}| = \sup_i \left(\sum_{j\ne i} a_{ij} + |a_{ii}| \right) = 2\sup_i |a_{ii}| < \infty$, отже матрична функція $\left(P_{ij}(t) \right) = P(t) = e^{At}$ є повністю визначеною. Вочевидь, вона є розв'язком (11). Доведемо, що для цієї функції виконуються умови 1)–4) ймовірностей переходів з попередньої лекції. Позаяк ці умови є достатніми для існування однорідного марковського процесу $\xi(t)$ з множиною станів E, у якого $P_{ij}(t)$ є ймовірностями переходів.

У матриці $I + At/n$ всі елементи, за вийнятком тих, які знаходяться на діагоналі, є невід'ємними. Для елементів на діагоналі будемо мати: $1 + a_{ii}t/n \underset{n\to\infty}{\longrightarrow} 1$. Завдяки умові $\sup_i |a_{ii}| < \infty$, ця збіжність є рівномірною стосовно i для кожного фіксованого t. Це в свою чергу означає, що для кожного фіксованого t всі елементи матриці $I + At/n$ є невід'ємні, починаючи з певного n. Тепер з рівності

$$P(t) = \lim_{n\to\infty} \left(I + \frac{At}{n} \right)^n$$

випливає, що всі елементи матриці $P(t)$ є невід'ємними. Отже, властивість 1) для ймовірностей переходів виконується.

Нехай $A^n = \left(a_{ij}^{(n)} \right)$. Позаяк $\sum_j a_{ij} = 0,\ i \in E$, тоді будемо мати:

$$\sum_j P_{ij}(t) = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_j a_{ij}^{(n)} = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_j \sum_k a_{ik}^{(n-1)} a_{kj} =$$
$$= 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_k a_{ik}^{(n-1)} \sum_j a_{kj} = 1,$$

що забезпечує виконання властивості 2) для ймовірностей переходів.

Властивість 3) для $P(t)$, тобто рівняння Чепмена-Колмогорова випливає з відношення

$$P(t+s) = e^{A(t+s)} = e^{At} e^{As} = P(t)P(s).$$

Властивість 4) є очевидною. ◄

На останок розглянемо питання про існування $\lim_{t\to\infty} P_{ij}(t)$ для регулярного процесу Маркова $\xi(t)$. Якщо ξ_n є вкладеним ланцюгом Маркова цього процесу з матрицею переходів

$$P = \left(-\frac{a_{ij}}{a_{ii}} (1 - \delta_{ij}) \right)_{ij\in E},$$

і $p_{ij}^{(n)}$ є ймовірністями переходів за n кроків для цього ланцюгп, тоді інтуїтивно зрозуміло, що границя $\lim_{t\to\infty} P_{ij}(t)$ може існувати, якщо існує $\lim_{n\to\infty} p_{ij}^{(n)}$, іншими словами, коли вкладений ланцюг є ергодичним.

Має місце наступне твердження.

Теорема 4. *Нехай ξ_n є вкладеним ланцюгом Маркова для процесу $\xi(t)$. Якщо існує ергодичний розподіл $\rho_k, k \in E$ цього ланцюга і $\sum_k \frac{\rho_k}{|a_{kk}|} < \infty$, тоді*

$$\pi_j \stackrel{def}{=} \lim_{t\to\infty} P_{ij}(t) = \frac{\rho_j/|a_{jj}|}{\sum_k \rho_k/|a_{kk}|}.$$

Якщо ергодичний розподіл $\pi_i, i \in E$ існує, тоді для його знаходження зазвичай використовують пряму систему рівнянь Колмогорова

$$\frac{dP_{ij}(t)}{dt} = \sum_{k\in E} P_{ik}(t)a_{kj}.$$

Після переходу до границі $t \to \infty$ отримаємо $\sum_k \pi_k a_{kj} = 0$, що разом з умовою нормування $\sum_k \pi_k = 1$ дасть нам систему рівнянь

$$\begin{cases} \sum_k \pi_k a_{kj} = 0, \quad j \in E, \\ \sum_k \pi_k = 1. \end{cases} \tag{12}$$

У матричному вигляді цю систему можемо записати як

$$\vec{\pi}A = 0, \qquad \sum_{k\in E} \pi_k = 1, \qquad \vec{\pi} = (\pi_i, i \in E).$$

Система рівнянь (12) дозволяє знайти ергодичний розподіл $\pi_i, i \in E$ без необхідності знаходження ергодичного розподілу вкладеного ланцюга.

ЗАВДАННЯ

1. Для процесу Маркова з $N < \infty$ станами довести рівність (5).
2. Нехай $\xi(t)$ є процесом Маркова з $N < \infty$ станами, A - інфінітезимальна матриця цього процесу. Довести, що ранг матриці A становить $N-1$.
3. Нехай $\xi(t)$ є процесом Маркова з інфінітезимальною матрицею

   $$A = \begin{matrix} -a & a \\ b & -b \end{matrix}, \qquad a, b > 0.$$

   Довести, що зворотню систему рівнянь Колмогорова можна записати у вигляді:

   $$P_{ij}^{''}(t) - (a+b)P_{ij}'(t) = 0.$$

   Знайти $P_{ij}(t)$.
4. Маємо процес Маркова з інфінітезимальною матрицею

   $$A = \begin{pmatrix} -4 & 2 & 2 \\ 1 & -2 & 1 \\ 0 & 3 & -3 \end{pmatrix}.$$

a) Записати системи рівнянь Колмогорова для цього процесу.
b) Знайти його ергодичний розподіл.
c) Знайти матрицю переходів вкладеного ланцюга Маркова.

5. Розглянемо проце Пуассона $\xi(t), \xi(0)=0$ з параметром λ, і нехай $\eta(t)=(-1)^{\xi(t)}$.

a) Знайти матрицю ймовірностей переходів $P(t)$ для процесу $\eta(t)$.
b) Знайти ергодичний розподіл для $\eta(t)$.

Лекція 9. Процеси народження та загибелі

У цьому параграфі визначимо та проведемо лослідження процесу народження та загибелі, який є частковим випадком процесу Маркова. Він відіграє важливу роль, зокрема в моделюванні роботи технічних пристроїв, у теорії масового обслуговування, а також при дослідженні числа осіб в популяціях. Конкретні приклади застосувань цього процесу будуть продемонстровані у подальшому, а тепер перейдемо до його визначення.

9.1 Визначення і ергодичний розподіл

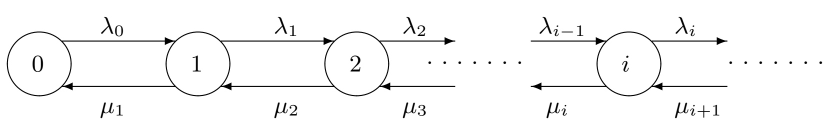
Визначення 1. *Марковський процес $\xi(t) \in E = \{0, 1, 2, ...\}$ будемо називати процесом народження та загибелі , якщо виконані умови:*

$$a_{ij} = \begin{cases} \lambda_i, & j = i+1, \\ \mu_i, & j = i-1, \\ -\lambda_i - \mu_i, & j = i, \\ 0, & |j-i| > 1, \end{cases} \qquad \text{для } i \geq 1,$$

$$a_{0j} = \begin{cases} \lambda_0, & j = 1, \\ -\lambda_0, & j = 0, \\ 0, & j \geq 2, \end{cases}$$

де $0 \leq \mu_i, \lambda_i < \infty$.

З цього визначення випливає, що процес народження та загибелі є процесом регулярним. Якщо в певний момент процес є в стані $i \geq 1$, тоді в цьому стані він перебуває випадковий проміжок часу, який має показниковий розподіл з параметром $\lambda_i + \mu_i$. Після закінчення часу перебування є дві можливості: стрибок з ймовірністю $\lambda_i/(\lambda_i + \mu_i)$ до стану $i+1$ (народження) або стрибок з ймовірністю $\mu_i/(\lambda_i + \mu_i)$ до стану $i-1$ (загибель). Якщо $i = 0$, тоді в цьому стані процес залишається випадковий проміжок часу, який має показниковий розподіл з параметром λ_0. Після чого з ймовірністю 1 відбувається стрибок до стану 1. Наочно цей процес млжна представити у вигляді направленого графа (мал. 1).



Мал. 1

Назва „процес народження та загибелі" походить від того, що стан процесу може бути проінтерпретований як кількість осіб певної популяції. Якщо в певний момент в популяції є i осіб, тоді зміна стану $i \longrightarrow i+1$ трактується як народження чергової особи, а зміна стану $i \longrightarrow i-1$ – як загибель (смерть) однієї з осіб популяції. Зміна стану $0 \longrightarrow 1$ інтерпретується як виникнення (зародження) популяції знову.

Якщо $\mu_i = 0$, $i \geq 1$, тобто особи ніколи не вмерають, тоді будемо говорити, що маємо справу з *чистим процесом народження.*

З визначення процесу народження та загибелі випливає, що його перехідна ймовірність $P_{ij}(t)$ має наступні властивості для $t \to 0$

$$P_{ij}(t) = \begin{cases} \lambda_i t + o(t), & j = i+1, \\ \mu_i t + o(t), & j = i-1, \\ 1-(\lambda_i+\mu_i)t + o(t), & j = i, \\ o(t), & |j-i| > 1, \end{cases} \qquad \text{для } i \geq 1, \tag{1}$$

$$P_{0j}(t) = \begin{cases} \lambda_0 t + o(t), & j = 1, \\ 1-\lambda_0 t + o(t), & j = 0, \\ o(t), & j \geq 2. \end{cases} \tag{2}$$

Зробимо дослідження умов існування і вигляд ергодичного розподілу процесу народження та загибелі. Якщо ергодичний розподіл існує, тоді згідно з результатами попередньої лекції він є розв'язком наступної системи рівнянь:

$$\vec{\pi}A = 0, \qquad \sum_{k=0}^{\infty} \pi_k = 1, \qquad \vec{\pi} = (\pi_0, \pi_1, ...),$$

де

$$A = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & ... \\ \mu_1 & -\lambda_1-\mu_1 & \lambda_1 & 0 & 0 & ... \\ 0 & \mu_2 & -\lambda_2-\mu_2 & \lambda_2 & 0 & ... \\ ... & ... & ... & ... & ... & ... \end{pmatrix}.$$

Якщо взяти до уваги вигляд матриці A, тоді рівняння $\vec{\pi}A = 0$ можемо переписати у наступний спосіб:

$$\begin{aligned} &-\lambda_0\pi_0 + \mu_1\pi_1 = 0, \\ &-(\lambda_i+\mu_i)\pi_i + \lambda_{i-1}\pi_{i-1} + \mu_{i+1}\pi_{i+1} = 0, \quad i = 1, 2, ... \end{aligned}$$

Нехай $\mu_i > 0$, $i = 1, 2, ...$ З першого рівняння отримаємо $\pi_1 = \pi_0\lambda_0/\mu_1$. Для $i = 1$ маємо рівняння $-(\lambda_1+\mu_1)\pi_1 + \lambda_0\pi_0 + \mu_2\pi_2 = 0$, звідки

$$\pi_2 = \frac{(\lambda_1+\mu_1)\pi_1 - \lambda_0\pi_0}{\mu_2} = \frac{\pi_0\frac{(\lambda_1+\mu_1)\lambda_0}{\mu_1} - \lambda_0\pi_0}{\mu_2} = \pi_0\frac{\lambda_0\lambda_1}{\mu_1\mu_2}.$$

Продовжуючи цю процедуру для $i = 2, 3, ...$, будемо мати

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}, \quad k = 1, 2, ... \tag{3}$$

Умова нормування $\sum_{k=0}^{\infty} \pi_k = 1$ дає нам

$$\pi_0 \;\; 1 + \sum_{k=1}^{\infty}\prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} \;\; = 1. \tag{4}$$

Якщо ряд $\sum_{k=1}^{\infty}\prod_{i=1}^{k}\lambda_{i-1}/\mu_i$ розбігається, тоді з (4) випливає, що $\pi_0=0$. Тоді з (3) робимо висновок, що $\pi_k=0$ для всіх $k=1,2,....$ Отже, у цьому випадку ергодичний розподіл не існує. Якщо ж ряд є збіжним, тоді з (3) і (4) отримаємо наступне твердження.

Теорема 1. *Процес народження та загибелі має ергодичний розподіл тоді і тільки тоді, коли* $\sum_{k=1}^{\infty}\prod_{i=1}^{k}\frac{\lambda_{i-1}}{\mu_i}<\infty$. *Якщо зазначена умова виконується, ергодичний розподіл має вигляд:*

$$\pi_0=\frac{1}{1+\sum_{k=1}^{\infty}\prod_{i=1}^{k}\frac{\lambda_{i-1}}{\mu_i}},\quad \pi_k=\pi_0\prod_{i=1}^{k}\frac{\lambda_{i-1}}{\mu_i},\qquad k\geq 1.$$

Зауваження 1. Якщо інтенсивністі народжень λ_i зростають при $i\to\infty$ занадто швидко в порівнянні з μ_i, тоді може так статись, що в певний скінченний (але випадковий!) момент τ_* відбудеться подія $\xi(\tau_*)=\infty$. Іншими словами, процес $\xi(t)$ зникне з E. Необхідною і достатньою умовою, аби цього не сталось є умова:

$$\sum_{k=1}^{\infty}\prod_{i=1}^{k}\frac{\mu_i}{\lambda_i}=\infty.$$

Зокрема, для чистого процесу народжень необхідною і достатньою умовою регулярності є умова (твердження Феллера) $\sum_{k=1}^{\infty}\lambda_k^{-1}=\infty$.

9.2 Процес народження та загибелі з поглинанням

Якщо $\lambda_0=0$, тоді процес народження та загибелі, потрапивши до стану 0, залишеться в ньому назавжди. У цьому випадку будемо говорити про процес з *поглинанням.* Нашою метою є знаходження ймовірності поглинання $P_n=\mathbf{P}(\text{ процес закінчиться}/\xi(0)=n)$.

Теорема 2. *Якщо* $\sum_{i=1}^{\infty}\prod_{k=1}^{i}\frac{\mu_k}{\lambda_k}=\infty$, *тоді* $P_n=1$ *для всіх* $n\geq 0$. *У протилежному випадку*

$$P_n=\frac{\sum_{i=n}^{\infty}\prod_{k=1}^{i}\frac{\mu_k}{\lambda_k}}{1+\sum_{i=1}^{\infty}\prod_{k=1}^{i}\frac{\mu_k}{\lambda_k}},\ n\geq 1. \tag{5}$$

Д о в е д е н н я. З властивості Маркова і формули повної ймовірності з урахуванням першого після моменту $t=0$ стрибка процесу отримаємо:

$$P_i=\frac{\lambda_i}{\mu_i+\lambda_i}P_{i+1}+\frac{\mu_i}{\mu_i+\lambda_i}P_{i-1},\quad P_0=1.$$

Звідки після нескладних перетворень будемо мати:

$$P_{i+1}-P_i=\frac{\mu_i}{\lambda_i}(P_i-P_{i-1}),$$

що дасть

$$P_{i+1}-P_i=(P_1-1)\prod_{k=1}^{i}\frac{\mu_k}{\lambda_k}.$$

Просумував в останній рівності по i від 1 до $n-1$, будемо мати:

$$P_n - P_1 = (P_1 - 1)\sum_{i=1}^{n-1}\prod_{k=1}^{i}\frac{\mu_k}{\lambda_k}. \tag{6}$$

Якщо $\sum_{i=1}^{\infty}\prod_{k=1}^{i}\frac{\mu_k}{\lambda_k}=\infty$, тоді з цього рівняння випливає (треба взяти $n\to\infty$) $P_1 = 1$, а тоді $P_n = 1$ для $n \geq 1$.

Отже, якщо ми хочемо, аби $0 < P_1 < 1$, тоді умова $\sum_{i=1}^{\infty}\prod_{k=1}^{i}\frac{\mu_k}{\lambda_k} < \infty$ є обов'язковою. Припустимо, що ця умова виконується. Позаяк P_n є незростаючою по n функцією (це випливає з (6)), тоді існує $\lim_{n\to\infty} P_n = \alpha$. З рівності $P_{2n} = P_n \cdot P_n$ (завдання 5) маємо, що $\alpha = \alpha^2$. Отже, $\alpha = 0$ або $\alpha = 1$.

Якщо $\alpha = 1$, тоді з (6) при $n\to\infty$ отримуємо протиріччя

$$-1 = \sum_{i=1}^{\infty}\prod_{k=1}^{i}\frac{\mu_k}{\lambda_k}.$$

Отже, $\alpha = 0$ і тепер з (6) при $n\to\infty$ будемо мати:

$$-P_1 = (P_1 - 1)\sum_{i=1}^{\infty}\prod_{k=1}^{i}\frac{\mu_k}{\lambda_k},$$

звідки

$$P_1 = \frac{\sum_{i=1}^{\infty}\prod_{k=1}^{i}\frac{\mu_k}{\lambda_k}}{1+\sum_{i=1}^{\infty}\prod_{k=1}^{i}\frac{\mu_k}{\lambda_k}}.$$

З останньої рівності і виразу (6) отримаємо (5). ◄

ЗАВДАННЯ

1. Нехай $\xi(t)$ є процесом народження та загибелі таким, що $\lambda_i, \mu_i>0$ за вийнятком $\mu_m = 0$ для певного m. Довести, що коли $\sum_{k=m+1}^{\infty}\prod_{i=m+1}^{k}\frac{\lambda_{i-1}}{\mu_i} < \infty$, тоді

$$\lim_{t\to\infty} P_{ij}(t) = \begin{cases} 0 & 0 \leq j \leq m-1, \\ \pi_m \prod_{i=m+1}^{k}\frac{\lambda_{i-1}}{\mu_i} & j \geq m, \end{cases}$$

де

$$\pi_m = \frac{1}{1+\sum_{k=m+1}^{\infty}\prod_{i=m+1}^{k}\frac{\lambda_{i-1}}{\mu_i}}.$$

2. Написати рівняння Колмогорова для чистого процесу народжень.

2.1. Нехай $\xi(t)$ є чистим процесом народжень і $P_{ij}(t)$ його ймовірність переходів. Позначимо через $\varphi_{ij}(s)=\int_0^{\infty} e^{-st}P_{ij}(t)dt$, $s > 0$ перетворення Лапласа для функції $P_{ij}(t)$. Використовуючи другу систему рівнянь Колмогорова, довести, що

$$\varphi_{ij}(s) = \frac{1}{\lambda_i + s}\prod_{k=i}^{j-1}\frac{\lambda_k}{\lambda_{k+1}+s}. \tag{7}$$

3. Нехай $\xi(t)$ є процесом з попереднього прикладу і таким, що $\lambda_i \neq \lambda_j$ для $i \neq j$. Використовуючи вираз (7), довести, що

$$P_{ij}(t) = \sum_{l=i}^{j} a_l e^{-\lambda_l t},$$

де

$$a_i = \prod_{k=i}^{j-1} \frac{\lambda_k}{\lambda_{k+1} - \lambda_i}, \quad a_l = \frac{\lambda_{l-1}}{\lambda_i - \lambda_l} \prod_{k=i,\neq l-1}^{j-1} \frac{\lambda_k}{\lambda_{k+1} - \lambda_l}, \quad l = i+1, ..., j.$$

4. Нехай $\xi(t)$ є процесом з прикладу 2 і таким, що $\lambda_i = i\lambda$. (Такий процес називається *процесом лінійного зростання*). Довести, що

$$P_{ij}(t) = C_{j-1}^{j-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{j-i}, \ j \geq i.$$

2.2. Нехай $\xi(t)$ є процесом лінійного зростання з попереднього прикладу, для якого маємо $\xi(0) = i$. Довести, що

$$\mathbf{M}\xi(t) = ie^{\lambda t}, \quad \mathbf{D^2}\xi(t) = ie^{\lambda t}(e^{\lambda t} - 1).$$

5. Для ймовірності поглинання P_n процесу з параграфа 9.2 довести, що $P_{2n} = P_n{}^2$.

Лекція 10. Застосування марковських процесів 1

10.1 Моделі обслуговування

Для того, щоб визначити модель обслуговування, треба описати її складові елементи: a) відний потік вимог, b)обслуговуючий пристрій (або пристрої), c) дисципліну обслуговування.

Для зручності і бульш кращого розуміння, зазвичай, вживають позначення систем у вигляді $A/B/n/m$, який був запропонований Кендаллом. У цьому позначенні літера A використовується для опису вхідного потоку вимог, літера B характеризує процес обслуговування вимог, на позиціях n і m у свою чергу записують натуральні числа. Перше з них n фіксує кількість обслуговуючих приладів, а друге m – максимально можливу довжину черги.

Розглянемо тепер детально головні складові системи обслуговування.

Вхідний потік вимог. Найчастіше мають справу з ситуацією, коли вимоги прибувають по одній. Якщо α_n є моментом прибуття n-тої вимоги, тоді зазвичай припускають, що випадкові величини $\alpha_n - \alpha_{n-1}$, $n \geq 1$, $\alpha_0 = 0$, тобто інтервали часу між моментами прибуття вимог є незалежними з одним і тим самим розподілом. У випадку показникового розподілу на позиції A у позначенні $A/B/n/m$ пишуть M. Якщо цей розподіл є розподілом Ерланга порядку n, тоді замість A використовують E_n. Літера G на позиції A у свою чергу означає, що інтервали поміж моментами приходів вимог є незалежними і мають довільний розподіл (анг. *general*). Крім показникового або Ерланга можна використовувати також різні інші розподіли. Кожний з них має своє абревіатурне позначення, наприклад, рівномірний розподіл має позначення – U (анг. *uniform*), детермінована величина (одноточковий розподіл) – D (анг. *deterministic*) і т. д. Вимоги можуть також прибувати групами, а не по одній. Це знаходить своє відображення в позначеннях Кендалла, проте ми не будемо розглядати такі випадки.

Обслуговуючі прилади. Символ n у позначенні Кендалла означає число обслуговуючих приладів. Зазвичай кожний прилад обслуговує вимоги по одній. Якщо час обслуговування n-тої вимоги дорівнює β_n, тоді як правило припускають, що випадкові величини β_n є незалежними і однаково розподіленими. Подібно до вхідного потоку літера B визначає тип цього розподілу, наприклад, M, E, G, U, D. Якщо приладів є декілька, тоді у більшості випадків вважають, що вони працюють незалежно, але час обслуговування кожним приладом може мати свій розподіл.

Дисципліна обслуговування. Дисципліна обслуговування визначає, в якій послідовності вимоги обслуговуються, а також що відбувається з вимогою, якщо в момент її надходження прилади зайняті. Типовим припущенням у випадку одного приладу є наступне: прилад починає працювати, коли надходить перша вимога. Якщо наступна вимога з'являється, коли прилад зайнятий, вона стає у чергу. Максимальне число місць у черзі визначається літерою m у позначенні Кендалла. Черга зазвичай функціонує згідно правилу: вимога, яка надійшла першою, є першою у черзі. Після закінчення обслуговування вимога залишаю систему, а прилад миттєво починає обслуговування вимоги, яка є наступною у черзі. Така дисципліна обслуговування називається FIFO (анг. *first in–first out*).

Вочевидь, всі можливі варіанти вхідного потоку, функціонування приладів, дисципліни обслуговування і т.д. важко відобразити в простій символіці Кендалла. Через те, її часто доповнюють певними записами, які несуть інформацію про особливості систем, що розглядаються.

Зауваження 1. Зазвичай за початок функціонування приймають момент появи першої вимоги в порожній системі. Тоді $\xi(0) = 1$, але можлива ситуація, коли в початковий момент часу в системі перебуває більше ніж $n > 1$ вимог.

10.1.1 Марковські моделі обслуговування

Система $M/M/1/\infty$**. Опис моделі**. Один прилад обслуговує вимоги по одній, час обслуговування n-тої вимоги β_n має показниковий розподіл з параметром μ. Прилад починає працювати, коли надходить перша вимога. Вимога стає до черги, якщо в момент її приходу прилад є зайнятим. Черга не обмежена. Після закінчення обслуговування вимога залишає систему, а прилад миттєво починає обслуговування наступної вимоги з черги. Вимоги надхолять по одній незалежно від роботи приладу в моменти α_n, $\alpha_0 = 0$. Інтервали часу між моментами надходжень вимог $\alpha_n - \alpha_{n-1},\ n \geq 1$ є незалежними і мають показниковий розподіл з параметром λ. Згідно Кендаллу ця система позначається як $M/M/1/\infty$.

Нехай $\xi(t)$ є числом вимог у системі в момент часу t. Значення $\xi(t)$ складається з числа вимог у черзі і вимоги, яка зараз обслуговується за умови, що система не є порожньою.

Теорема 1. *$\xi(t)$ є процесом народження та загибелі з інтенсивністю народження λ та інтенсивністю смерті μ.*

Д о в е д е н н я. Процес $\xi(t)$ буде мати марковську властивість, якщо інформації про те, що в довільний момент часу t_0 процес $\xi(t)$ мав значення $\xi(t_0) = n \geq 0$, достатньо для опису ймовірносних характеристик процесу в момент $t > t_0$. З опису системи випливає, що траєкторії процесу $\xi(t)$ є сходинкові. Для того, щоб переконатись у властивості Маркова покажемо, що інформація $\xi(t_0) = n$ є вичерпною для знаходження розподілу часу до чергового стрибка і ймовірності переходу до нового стану в момент першого після t_0 стрибка процесу. З властивості відсутності пам'яті показникового розподілу випливає, що перша вимога після моменту t_0 прийде через інтервал часу, який має показниковий розподіл з параметром λ. У свою чергу перша вимога з черги після моменту t_0 отримає обслуговування через інтервал часу, який має показниковий розподіл з параметром μ (якщо $n = 0$, тоді після момента t_0 буде тільки прихід вимоги). З цього випливає, що ми можемо знайти ймовірносні характеристики першого після момента t_0 стрибка процесу. Нехай $\xi(t_0) = n \geq 1$ і випадкова величина α описує час від момента t_0 до першого прибуття вимоги з потоку, а випадкова величина β – час після момента t_0 , через який перша обслужена вимога залишить систему. Згідно припущень α і β є незалежними випадковими величинами з показниковим розподілом з параметрами λ і μ відповідно. Отже, інтервал часу від моменту t_0 до першого після цього моменту стрибка процесу буде мати розподіл:

$$\mathbf{P}(\min\{\alpha,\beta\} < x) = 1 - \mathbf{P}(\min\{\alpha,\beta\} \geq x) = 1 - \mathbf{P}(\alpha \geq x, \beta \geq x) =$$
$$= 1 - \mathbf{P}(\alpha \geq x)\mathbf{P}(\beta \geq x) = 1 - e^{-(\lambda+\mu)x}. \quad (1)$$

Перехід до стану $n+1$ в момент стрибка станеться з ймовірністю

$$\mathbf{P}(\alpha < \beta) = \mu \int_0^\infty \mathbf{P}(\alpha < x) e^{-\mu x} dx = \mu \int_0^\infty (1 - e^{-\lambda x} e^{-\mu x}) dx = \frac{\lambda}{\lambda + \mu}, \quad (2)$$

а перехід до стану $n-1$ – з ймовірністю

$$\mathbf{P}(\alpha > \beta) = 1 - \mathbf{P}(\alpha \leq \beta) = 1 - \frac{\lambda}{\lambda+\mu} = \frac{\mu}{\lambda+\mu}. \quad (3)$$

Якщо $\xi(t_0) = 0$, тоді в момент стрибка відбудеться перехід до стану 1 з ймовірністю 1, розподіл інтервала часу від моменту t_0 до цього стрибка буде мати показниковий розподіл з параметром λ.

З виразу (1) випливає, що інтенсивність виходу зі стану процесу $\xi(t)$ має вигляд $a_{nn} = -(\lambda+\mu)$, $n \geq 1$ і $a_{00} = -\lambda$. Позаяк $a_{n,n\pm1}/a_{nn}$ є ймовірністю переходу $n \to n\pm1$ в момент стрибка, тоді з виразів (2), (3) отримаємо:

$$\frac{a_{n,n+1}}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}, \quad \frac{a_{n,n-1}}{\lambda+\mu} = \frac{\mu}{\lambda+\mu},$$

отже $a_{n,n+1} = \lambda$ і $a_{n,n-1} = \mu$, якщо $n \geq 1$. Решта інтенсивностей переходів дорівнює нулеві. Наше твердження є доведеним. ◀

Існування ернодичного розподілу процесу $\xi(t)$ означає, що робота системи обслуговування $M/M/1/\infty$ з плином часу стабілізується і черга з вимог не зростає до нескінченності. Для дослідження умов існування цього розподілу можна застосувати твердження 1 (стор. 343). Позаяк тепер

$$\sum_{k=1}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} = \sum_{k=1}^{\infty} \frac{\lambda}{\mu}^{k} = \begin{cases} \infty, & \text{якщо } \lambda/\mu \geq 1, \\ \frac{\lambda}{\mu-\lambda} < \infty, & \text{якщо } \lambda/\mu < 1, \end{cases}$$

тоді отримаємо наступне твердження.

Теорема 2. *Нехай $\rho = \lambda/\mu$. Система $M/M/1/\infty$ має ергодичний розподіл тоді і тільки тоді, коли $\rho < 1$. Тоді $\pi_k = \rho^k(1-\rho)$, $k = 0, 1, ...$*

Прокоментуємо ще умову $\rho = \lambda/\mu < 1$, яка забеспечує існування ергодичного розподілу системи. Нехай α_i, $i \geq 1$ є інтервалами часу між приходами послідовних вимог, а β_i, $i \geq 1$, інтервалами часу, протягом яких відбувається обслуговування цих вимог. Позаяк випадкові величини α_i і β_i мають показникові розподіли відповідно з параметрами λ, μ, тоді $\mathbf{M}\alpha_i = \lambda^{-1}$, $\mathbf{M}\beta_i = \mu^{-1}$. Отже, умову $\lambda/\mu < 1$ можемо записати в наступному вигляді: $\mathbf{M}\beta_i < \mathbf{M}\alpha_i$. Іншими словами, для існування ергодичного розподілу, середній час обслуговування має бути **строго** меньшим за середній час між прибуттям вимог. З'ясувалось, що така умова існування ергодичного розподілу є обов'язковою також і для більш складних систем типу $G/G/1/\infty$, які не є марковськими.

10.1.2 Система $M/M/1/m$

Система $M/M/1/m$ відрізняється від попередньої тим, що в черзі є тільки m місць для очікування обслуговування. Вимога не обслуговується і втрачається, якщо в момент її приходу прилад є зайнятим, а в черзі перебуває m інших вимог.

Нехай $\xi(t)$ є числом вимог у системі в момент часу t,. Нескладно зрозуміти, що $\xi(t) \in \{0, 1, ..., m+1\}$, $\xi(t)$ є процесом народження та загибелі з інтенсивностями переходів

$$\lambda_i = \begin{matrix} \lambda, & \text{якщо} & 0 \le i \le m, \\ 0, & \text{якщо} & i \ge m+1, \end{matrix} \qquad \mu_i = \begin{matrix} \mu, & \text{якщо} & 1 \le i \le m+1, \\ 0, & \text{якщо} & i \ge m+2. \end{matrix}$$

Застосовуючи твердження 1 на стор. 343, отримаємо вираз ергодичного розподілу для цієї системи

$$\pi_k = \frac{(1-\rho)\rho^k}{1-\rho^{m+2}}, \qquad 0 \le k \le m+1,$$

де $\rho = \lambda/\mu$.

10.1.3 Система $M/M/s/\infty, s \ge 2$

Система $M/M/s/\infty, s \ge 2$ є подібною до системи $M/M/1/\infty$. Відмінність полягає в тому, що тепер ми маємо s однакових приладів, які працюють незалежним чином. Якщо в момент приходу вимоги є декілька вільних приладів, тоді вона обирає будь-який для свого обслуговування. Якщо ж всі прилади є зайнятими, тоді вимога потрапляє до черги, яка є необмеженою.

Нехай $\xi(t)$ є числом вимог у системі в момент часу t, тоді $\xi(t)$ є процесом народження та загибелі з інтенсивностями переходів

$$\lambda_i = \lambda, \qquad i \ge 0, \qquad \mu_i = \begin{matrix} i\mu, & \text{якщо} & 1 \le i \le s, \\ s\mu, & \text{якщо} & i \ge s+1. \end{matrix}$$

Застосовуючи твердження 1 (стор. 343), отримаємо:

$$\pi_0 = \frac{1}{1 + \sum\limits_{k=1}^{\infty} \prod\limits_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}} = \frac{1}{1 + \sum\limits_{k=1}^{s} \prod\limits_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} + \sum\limits_{k=s+1}^{\infty} \prod\limits_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}} =$$

$$= \frac{1}{1 + \sum\limits_{k=1}^{s} \prod\limits_{i=1}^{k} \frac{\lambda^k}{i\mu} + \sum\limits_{k=s+1}^{\infty} \frac{\lambda^k}{\prod\limits_{i=1}^{s} i\mu \prod\limits_{i=s+1}^{k} s\mu}} = \frac{1}{1 + \sum\limits_{k=1}^{s} \frac{\rho^k}{k!} + \frac{\rho^{s+1}}{s!(s-\rho)}},$$

що дасть наступні вирази для ергодичного розподілу системи.

Теорема 3. *Система $M/M/s/\infty$, $s \ge 2$ має ергодичний розподіл тоді і тільки тоді, коли $\lambda/s\mu = \rho/s < 1$. За цієї умови будемо мати:*

$$\pi_k = \begin{cases} \left(1 + \sum\limits_{k=1}^{s} \frac{\rho^k}{i!} + \frac{\rho^{s+1}}{s!(s-\rho)}\right)^{-1}, & \textit{якщо} \quad k = 0, \\ \pi_0 \frac{\rho^k}{k!}, & \textit{якщо} \quad 1 \le k \le s, \\ \pi_0 \frac{\rho^k}{s! s^{k-s}}, & \textit{якщо} \quad k \ge s+1. \end{cases}$$

Користуючись цим твердженням, нескладно знайти ймовірність того, що всі прилади є зайнятими в стаціонарному режимі. Іншими словами ймовірність того, що вимога буде вимушена стати до черги, коли система перебуває в стаціонарному режимі.

$$\mathbf{P}(\text{ всі прилади є зайнятими }) = \sum_{k=s}^{\infty} \pi_k = \frac{\rho^s}{(s-1)!(s-\rho)\ \sum\limits_{k=0}^{s} \frac{\rho^k}{i!} + \frac{\rho^{s+1}}{s!(s-\rho)}}.$$

Останній вираз називають *формулою Ерланга.*

10.1.4 Нестаціонарні характеристики. Період зайнятості

Марковські процеси можна також застосовувати для дослідження характеристик розглянутих систем, коли вони перебувають у нестаціонарному режимі. Вивчемо це питання на прикладі однієї з найважливіших характеристик систем обслуговування – *періоді зайнятості.* Спершу дамо визначення періоду зайнятості. Нехай $\xi(t)$ є числом вимог, які знаходяться в системі в момент часу t. Тоді $\tau(1) = \inf\{t : \xi(t) = 0/\xi(0) = 1\}$ будемо називати періодом зайнятості системи. Зазвичай початком роботи системи вважають момент приходу першої вимоги (див. зауваження 1), тобто $\xi(0) = 1$. Саме для такої ситуації сформульоване подане вище визначення періоду зайнятості. Але є можливим, що на початку в системі знаходиться n вимог (див. увага 1). Тоді $\tau(n) = \inf\{t : \xi(t) = 0/\xi(0) = n\}$ є періодом зайнятості за умови, що на початку в системі n вимог.

Розглянемо тепер систему $M/M/1/\infty$.

Теорема 4. *Для всіх $s > 0$ має місце*

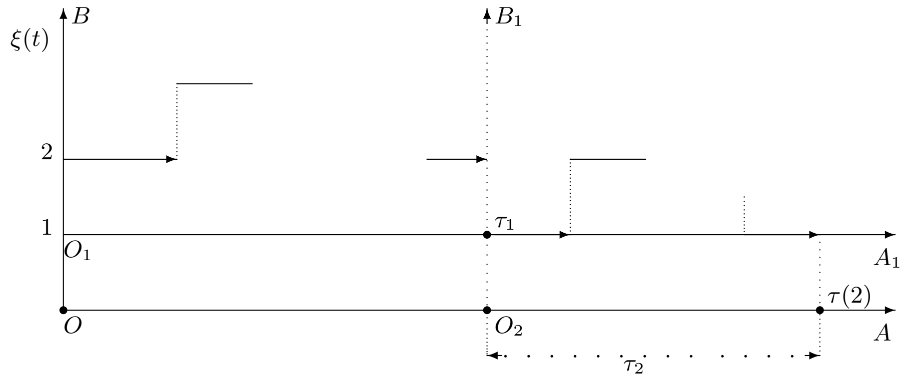
$$\mathbf{M}e^{-s\tau(n)} = \ \mathbf{M}e^{-s\tau(1)}\ {}^{n}, \qquad n \geq 1, \tag{4}$$

$$\mathbf{M}e^{-s\tau(1)} = \frac{\mu+\lambda+s-\sqrt{(\mu+\lambda+s)^2-4\lambda\mu}}{2\lambda}. \tag{5}$$

Д о в е д е н н я. Нехай $\varphi_n(s) = \mathbf{M}e^{-s\tau(n)},\ s > 0$. Спершу доведемо, що має місце $\varphi_n(s) = \ \varphi_1(s)\ {}^{n}$. Доведення проведемо тільки для $n = 2$, залишаючи читачеві загальний випадок (завдання 2).

Нехай $\xi(0) = 2$. Процес досягнення нульового рівня можемо поділити на два етапи (див. мал. 1). На першому з них процес $\xi(t)$ досягає рівня 1 вперше в певний момент τ_1. На другому етапі через інтервал часу τ_2 процес досягає вперше рівня нуль, стартуючи в момент τ_1 зі значення $\xi(\tau_1) = 1$. У зв'язку з чим будемо мати $\tau(2) = \tau_1 + \tau_2$. З властивості Маркова випливає, що випадкові величини τ_1, τ_2 є незалежними. Крім того, ці величини мають такий самий розподіл як $\tau(1)$. Дійсно, якщо розглядати еволюцію процесу $\xi(t)$ в системах координат A_1O_1B і AO_2B_1, тоді в кожній системі процес стартує зі стану 1 і τ_1 є його першим досягнення нуля в першій системі координат, а τ_2 є його першим моментом досягнення нуля в другій системі координат. Позаяк процес $\xi(t)$ є однорідним (тобто його ймовірносні властивості не залежать від зсуву), тоді випадкові величини τ_1 і τ_2 мають однакові розподіли. Випадкова величина τ_1 в системі координат A_1O_1B має такий самий сенс, що й $\tau(1)$ в системі координат AOB, отже, випадкова величина τ_1 має такий самий розподіл як $\tau(1)$. Звідки $\tau(2) = \tau_1 + \tau_2$, де τ_1, τ_2 є незалежними і мають такий самий розподіл як $\tau(1)$. Отже, будемо мати:

$$\varphi_2(s) = \mathbf{M}e^{-s(\tau_1+\tau_2)} = \ \mathbf{M}e^{-s\tau_1}\ {}^{2} = \ \mathbf{M}e^{-s\tau(1)}\ {}^{2} = \ \varphi_1(s)\ {}^{2}. \tag{6}$$



Мал. 1

Нехай $\xi(0) = 1$, випадкова величина α є часом до появи першої вимоги, а випадкова величина β – інтервалом часу, через який перша обслужена вимога залишить систему. Випадкові величини α, β є незалежними і мають показникові розподіли відповідно з параметрами λ, μ.

Подія {*перший стрибок процесу станеться догори і до моменту часу* t }= $\{\alpha < \beta, \alpha < t\}$ має ймовірність

$$\mathbf{P}(\alpha < \beta, \alpha < t) = \lambda \int_0^t \mathbf{P}(u < \beta) e^{-\lambda u} du = \lambda \int_0^t e^{-(\mu+\lambda)u} du.$$

Тоді щільність цієї ймовірності є

$$f_+(t) = \frac{d}{dt}\mathbf{P}(\alpha < \beta, \alpha < t) = \lambda e^{-(\mu+\lambda)t}. \tag{7}$$

Аналогічно, щільність ймовірності настання події {*перший стрибок процесу станеться додолу і до моменту часу* t }=$\{\alpha \geq \beta, \beta < t\}$ дорівнює

$$f_-(t) = \frac{d}{dt}\mathbf{P}(\alpha \geq \beta, \beta < t) = \mu e^{-(\mu+\lambda)t}. \tag{8}$$

Застосовуючи формулу повної ймовірності, отримаємо:

$$\mathbf{M}e^{-s\tau(1)} = \int_0^\infty \mathbf{M}\{e^{-s\tau(1)}/\alpha < \beta, \alpha = u\} f_+(u)du +$$

$$+ \int_0^\infty \mathbf{M}\{e^{-s\tau(1)}/\alpha \geq \beta, \beta = u\} f_-(u)du = \lambda \int_0^\infty \mathbf{M}\{e^{-s\tau(1)}/\alpha < \beta, \alpha = u\} e^{-(\mu+\lambda)u} du +$$

$$+ \mu \int_0^\infty \mathbf{M}\{e^{-s\tau(1)}/\alpha \geq \beta, \beta = u\} e^{-(\mu+\lambda)u} du. \tag{9}$$

Якщо відомо, що сталась подія $\{\alpha < \beta, \alpha = u\}$, тоді в момент часу $t = u$ будемо мати $\xi(u) = 2$. Якщо $\widehat{\tau}$, є інтервалом часу, після якого процес $\xi(t)$, стартуючи в момент $t = u$ зі стану 2, досягне вперше нульового рівня, тоді, вочевидь, $\tau(n) = u + \widehat{\tau}$. Випадкова величина $\widehat{\tau}$ залежить виключно від поведінки процесу $\xi(t)$ для $t \geq u$. Завдяки властивості Маркова, будемо мати: $\mathbf{P}(\widehat{\tau} < x/\alpha{<}\beta, \alpha{=}u){=}\mathbf{P}(\widehat{\tau}{<}x/\xi(u){=}2) = \mathbf{P}(\tau(2) < x)$. Остання рівність випливає з однорідності процесу $\xi(t)$. Ймовірність досягнення нульового рівня за умови $\{\xi(u) = 2\}$ є такою самою як і ймовірність досягнення цього рівня за умови $\{\xi(0) = 2\}$. Іншими словами розподіл випадкової величини $\widehat{\tau}$ за умови $\xi(u) = 2$ є аналогічним розподілу випадкової величини $\tau(2)$. Підсумовуючи ці міркування, отримаємо:

$$\mathbf{M}\{e^{-s\tau(1)}/\alpha < \beta, \alpha = u\} = \mathbf{M}\{e^{-s(u+\widehat{\tau})}/\alpha < \beta, \alpha = u\} = \\ = e^{-su}\mathbf{M}\{e^{-s\widehat{\tau}}/\xi(u) = 2\} = e^{-su}\mathbf{M}e^{-s\tau(2)}.$$

У випадку, коли перший стрибок процесу буде додолу, тобто станеться подія $\{\alpha{\geq}\beta, \beta{=}\,u\}$, у момент часу u настає кінець періоду зайнятості, тоді $\tau(1) = u$, що дасть

$$\mathbf{M}\{e^{-s\tau(1)}/\alpha \geq \beta, \beta = u\} = e^{-su}.$$

Використовуючи це в (9), отримаємо:

$$\mathbf{M}e^{-s\tau(1)} = \lambda \mathbf{M}e^{-s\tau(2)} \int_0^\infty e^{-(\mu+\lambda+s)u}du + \mu \int_0^\infty e^{-(\mu+\lambda+s)u}du = \frac{\lambda \mathbf{M}e^{-s\tau(2)} + \mu}{\mu + \lambda + s}.$$

Якщо взяти до уваги співвідношення (6), тоді для функції $\varphi_1(s) = \mathbf{M}e^{-s\tau(1)}$ отримаємо рівняння

$$\varphi_1(s) = \frac{\lambda\ \varphi_1(s)\ ^2 + \mu}{\mu + \lambda + s},$$

або

$$\lambda\ \varphi_1(s)\ ^2 - (\mu + \lambda + s)\varphi_1(s) + \mu = 0, \tag{10}$$

звідки

$$\varphi_1(s) = \frac{\mu + \lambda + s \pm \sqrt{(\mu + \lambda + s)^2 - 4\lambda\mu}}{2\lambda}.$$

Нескладно довести (завдання 3), що для $s > 0$ корінь рівняння (10) при $+\sqrt{\ }$ є більшим за 1, а при $-\sqrt{\ }$ –менший за 1, тому

$$\varphi_1(s) = \frac{\mu + \lambda + s - \sqrt{(\mu + \lambda + s)^2 - 4\lambda\mu}}{2\lambda}$$

і твердження є доведеним. ◄

Вираз (5) можна обернути і отримати розподіл періоду зайнятості системи. Наведемо лише щільність цього розподілу (див. [?], розділ 14, формула (6.16)).

$$\frac{d}{dt}\mathbf{P}(\tau(1) < t) = \sqrt{\frac{\mu}{\lambda}} t^{-1} e^{-t(\lambda+\mu)} I_1(2\sqrt{2\lambda\mu t}),$$

де $I_1(x)$ є фукцією Бесселя порядку 1, яка може бути записана у вигляді:

$$I_1(x) = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{k!(k+1)!2^{2k+1}}.$$

ЗАВДАННЯ

1. Знайти середню довжину черги в стаціонарному режимі для систем типу $M/M/1/\infty$, $M/M/1/m$ і $M/M/2/\infty$.
2. Довести вираз (4).
3. Довести, що коли $z_\pm(s)$ є розв'язками рівняння

   $$\lambda z^2 - (\mu + \lambda + s)z + \mu = 0,$$

   тоді $z_-(s) < 1 < z_+(s)$ для $s > 0$.
4. Для системи $M/M/1/\infty$ знайти умови, за яких $\mathbf{P}(\tau(1)<\infty) = 1$.

   Вказівка. *Скористатися тим, що* $\mathbf{P}(\tau(1) < \infty) = \lim_{s\to 0} \mathbf{M}e^{-s\tau(1)}$, *і з виразу* (5).
5. Знайти середню довжину періоду зайнятості для системи типу $M/M/1/\infty$.

Лекція 11. Застосування марковських процесів 2

11.1 Моделі з двома обслуговуючими приладами

Розлянемо систему з двома обслуговуючими приладами. Ця система є подібною до вивченої раніше системи $M/M/s/\infty$, але з важливою засадничою різницею.

Опис моделі. Два прилади працюють незалежно і обслуговують вимоги по одній. Час обслуговування вимоги має показниковий розподіл з параметром μ_1 для першого приладу і μ_2 – для другого. Якщо вимога надійде, коли прилади є зайнятими, тоді вона стає до черги, в якій можуть перебувати максимум m вимог. Після закінчення обслуговування вимога залишає систему, а звільнений прилад миттєво починає обслуговувати наступну вимогу з черги. Якщо вимога прибуває, коли обидва прилади є вільними, вона потрапляє на обслуговування до першого приладу. Вимога, яка приходить, коли обидва прилади є зайнятими, і у черзі вже є m вимог, втрачається. Вимоги прибувають по одній і незалежно від приладів. Інтервали часу між моментами приходу вимог мають показниковий розподіл з параметром λ.

Нехай, як і раніше, $\xi(t)$ є числом вимог у системі в момент часу t. Якщо $\mu_1 \neq \mu_2$, тоді процес $\xi(t)$ не є марковським. Якщо наприклад відомо, що $\xi(t_0) = 1$, тоді розподіл наступного стрибка процесу $\xi(t)$ буде залежати від того, який прилад обслуговує цю єдину вимогу в момент t_0. Насправді, якщо відомо, що в момент t_0 працює перший прилад, тоді з формули (2) (стор. 348) випливає, що наступний після t_0 стрибок догори процесу $\xi(t)$ відбудеться з ймовірністю $\lambda/(\lambda + \mu_1)$. У випадку, коли в момент t_0 вимога обслуговується другим приладом, тоді наступний після момента t_0 стрибок догори процесу $\xi(t)$ станеться ймовірністю $\lambda/(\lambda+\mu_2)$. З цього випливає, що для опису еволюції системи в майбутньому є потрібною інформація про стан приладів у момент часу t_0. Розглянуту систему можна було б описати за допомогою трьохвимірного марковського процесу $(\xi(t), \eta_1(t), \eta_2(t))$, в якому складові $\eta_1(t), \eta_2(t)$ описують стани обох приладів. Але ми змоделюємо нашу систему за допомогою іншого марковського процесу більш простим способом.

Спочатку позначимо стани нашої системи наступним чином:

$(0, 0)$ – обидва прилади є вільними,

$(1, 0)$ – перший прилад є зайнятим, а другий вільний,

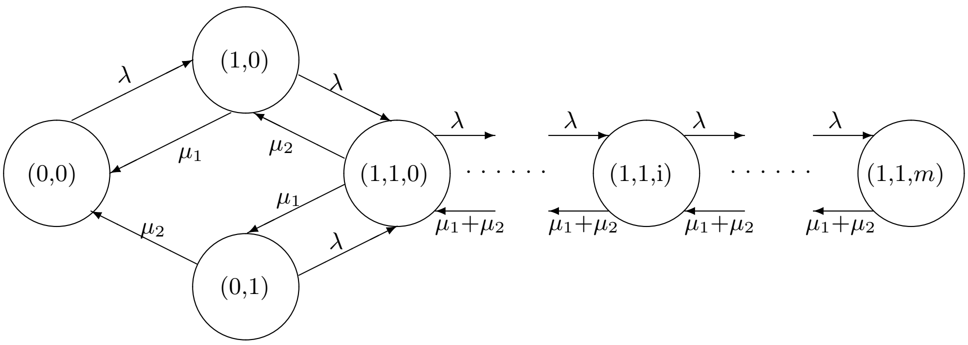
$(0, 1)$ – другий прилад є зайнятим, а перший вільний,

$(1, 1, i)$ – обидва прилади є зайнятими і в черзі є i вимог, i=$0, 1, ..., m$.

Визначимо тепер процес $\xi(t)$ з множиною станів

$$E = \{(0, 0), (1, 0), (0, 1), (1, 1, i), i = 0, 1, 2, ..., m\} \tag{1}$$

таким чином, що коли в момент t система є в стані $\alpha \in E$, тоді $\xi(t) = \alpha$. Процес $\xi(t)$ буде тепер процесом Маркова (завдання 1). Наступна схема представляє граф інтенсивностей переходів системи.



Нехай

$$\pi_{00} = \mathbf{P}((0,0)),\ \pi_{10} = \mathbf{P}((1,0)),\ \pi_{01} = \mathbf{P}((0,1)),\ \pi_{11}(i) = \mathbf{P}((1,1,i))$$

будуть ймовірностями балансу системи (ергодичним розподілом).

Теорема 1. *Для ергодичного розподілу мають місце*

$$\pi_{00} = \frac{\mu_1\mu_2(1+2\rho)}{\lambda\rho(\lambda+\mu_2)}C, \quad \pi_{10} = \frac{\mu_2(1+\rho)}{\rho(\lambda+\mu_2)}C, \quad \pi_{01} = \frac{\mu_1}{\lambda+\mu_2}C, \tag{2}$$

$$\pi_{11}(i) = C\rho^i, \quad 0 \le i \le m, \tag{3}$$

де $\rho = \lambda(\mu_1+\mu_2)^{-1}$ *і*

$$C = \left(1 + \frac{\mu_1\mu_2(1+2\rho)}{\lambda\rho(\lambda+\mu_2)} + \frac{\mu_2(1+\rho)}{\rho(\lambda+\mu_2)} + \frac{\mu_1}{\lambda+\mu_2} + \frac{1-\rho^{m+1}}{1-\rho}\right)^{-1}. \tag{4}$$

Зауваження 1. Якщо $\rho = 1$, тоді в наведених виразах треба перейти до границі $\rho \to 1$.

Д о в е д е н н я. Позаяк множина станів процесу $\xi(t)$ є складною, записати рівняння балансу цього процесу за допомогою рівнянь Колмогорова є непростою проблемою. Продемонструємо підхід, який полегшує це завдання.

Нехай є заданим процес Маркова $\eta(t) \in E = \{0, \pm1, ...\}$ з інтенсивностями переходів a_{ij}. Для фіксованого стану k визначимо через $S_-(k)$ сукупність станів, до яких є можливим перехід за один крок зі стану k, а через $S_+(k)$ множину станів, з яких є можливим перехід за один крок до стану k. Нехай π_i, $i \in E$ є ергодичним розподілом процесу $\eta(t)$. Суму $\pi_k \sum_{j\in S_-(k)} a_{kj}$ називають потоком ймовірностей зі стану k, а суму $\sum_{j\in S_+(k)} \pi_j a_{jk}$ – потоком ймовірностей до стану k. В умовах рівноваги потік ймовірностей зі стану k мусить дорівнювати потоку ймовірностей до цього стану, тобто

$$\pi_k \sum_{j\in S_-(k)} a_{kj} = \sum_{j\in S_+(k)} \pi_j a_{jk}. \tag{5}$$

Тепер нескладно написати рівняння балансу для кожного стану нашого процесу:

$$\begin{cases} \pi_{00}\lambda &= \pi_{10}\mu_1 + \pi_{01}\mu_2, \\ \pi_{10}(\lambda+\mu_1) &= \pi_{00}\lambda + \pi_{11}(0)\mu_2, \\ \pi_{01}(\lambda+\mu_2) &= \pi_{11}(0)\mu_1, \\ \pi_{11}(0)(\lambda+\mu_1+\mu_2) &= \pi_{10}\lambda + \pi_{01}\lambda + \pi_{11}(1)(\mu_1+\mu_2), \\ \pi_{11}(i)(\lambda+\mu_1+\mu_2) &= \pi_{11}(i-1)\lambda + \pi_{11}(i+1)(\mu_1+\mu_2), \quad 1 \le i \le m-1, \\ \pi_{11}(m)(\mu_1+\mu_2) &= \pi_{11}(m-1)\lambda. \end{cases} \tag{6}$$

Нехай $\rho \neq 1$. Загальний розв'язок рівняння

$$\pi_{11}(i)(\lambda+\mu_1+\mu_2) = \pi_{11}(i-1)\lambda + \pi_{11}(i+1)(\mu_1+\mu_2), \quad i \ge 1 \tag{7}$$

може бути записаний у вигляді (завдання 2)

$$\pi_{11}(i) = A + B\left(\frac{\lambda}{\mu_1+\mu_2}\right)^i = A + B\rho^i, \ i \ge 1. \tag{8}$$

Умова $\pi_{11}(m)(\mu_1+\mu_2) = \pi_{11}(m-1)\lambda$ або $\pi_{11}(m) = \pi_{11}(m-1)\rho$ дає $A = 0$. За B можна взяти $\pi_{11}(0) \equiv C = const$, отже маємо

$$\pi_{11}(i) = C\rho^i, \quad 0 \le i \le m.$$

З перших трьох рівнянь (6) після заміни $\pi_{11}(0)$ на C будемо мати:

$$\pi_{00} = \frac{\mu_1\mu_2(1+2\rho)}{\lambda\rho(\lambda+\mu_2)}C, \quad \pi_{10} = \frac{\mu_2(1+\rho)}{\rho(\lambda+\mu_2)}C, \quad \pi_{01} = \frac{\mu_1}{\lambda+\mu_2}C. \tag{9}$$

Умова нормування дасть наступним кроком (4). Легко пересвідчитись, що четверте рівняння в (6) теж виконується.

Якщо $\rho = 1$, тоді в формулах (2)-(4) треба перейти до границі $\rho \to 1$ (завдання 3).

◄

Система $M/M/2/\infty$

Різниця між зазначеною системою і системою, яка була розглянуто попередньо, полягає в тому, що тепер черга буде необмеженою. Дослідження ергодичного розподілу цієї системи можна провести так само як і у випадку обмеженої черги або перейти до границі $m \to \infty$ у виразах (2)-(4). Як результат отримаємо:

Теорема 2. *Ергодичний розподіл системи $M/M/2/\infty$ існує тоді і тільки тоді, коли $\rho = \frac{\lambda}{\mu_1+\mu_2} < 1$. Тоді*

$$\pi_{00} = \frac{\mu_1\mu_2(1+2\rho)}{\lambda\rho(\lambda+\mu_2)}C, \quad \pi_{10} = \frac{\mu_2(1+\rho)}{\rho(\lambda+\mu_2)}C, \quad \pi_{01} = \frac{\mu_1}{\lambda+\mu_2}C,$$
$$\pi_{11}(i) = C\rho^i, \quad i \ge 0,$$

де

$$C = \left(1 + \frac{\mu_1\mu_2(1+2\rho)}{\lambda\rho(\lambda+\mu_2)} + \frac{\mu_2(1+\rho)}{\rho(\lambda+\mu_2)} + \frac{\mu_1}{\lambda+\mu_2} + \frac{1}{1-\rho}\right)^{-1}. \tag{10}$$

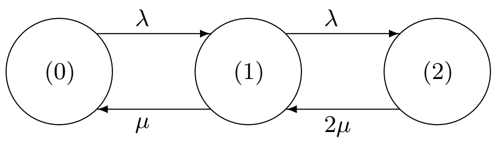
11.2 Система з резервним приладом

Опис моделі. Система має два ідентичні прилади. Один з них працює, а другий є резервним. Час функціонування активного прилада до момента його аварії є показниковим з параметром λ. Якщо відбувається аварія працюючого прилада, тоді він миттєво замінюється на резервний. Аварійний прилад направляється на ремонт, який триває показниковий час з параметром μ. Відмова системи настає тоді, коли обидва прилади виходять з ладу. Після ремонту хоча б одного з них система відновлює своє функціонування.

Введемо позначення ймовірностей стану рівноваги

$$\pi_0 = \mathbf{P}(\text{обидва прилади є справними}),$$
$$\pi_1 = \mathbf{P}(\text{один прилад працює, а другий є на ремонті}),$$
$$\pi_2 = \mathbf{P}(\text{обидва прилади є на ремонті}).$$

Наступна схема представляє граф інтенсивностей переходів нашої системи.



На цій схемі стани $(0),(1),(2)$ означають відповідно: „обидва прилади є справними", „один прилад працює, а другий є на ремонті", „обидва прилади ремонтуються".

Тепер нескладно за допомогою цього графу написати рівняння стану рівноваги:

$$\lambda\pi_0 = \mu\pi_1, \quad (\lambda+\mu)\pi_1 = 2\mu\pi_2 + \lambda\pi_0, \quad 2\mu\pi_2 = \lambda\pi_1.$$

Звідки маємо наступний результат:

Теорема 3.

$$\pi_0 = \frac{2\mu^2}{2\mu^2+\lambda^2+2\mu\lambda}, \quad \pi_1 = \frac{2\lambda\mu}{2\mu^2+\lambda^2+2\mu\lambda}, \quad \pi_2 = \frac{2\lambda^2}{2\mu^2+\lambda^2+2\mu\lambda}.$$

Визначимо тепер для нашої моделі розподіл часу до відмови. Аби звести цю задачу до дослідження ймовірності переходу процесу Маркова, змодифікуєми попередний опис моделі наступним чином: якщо система потрапляє до стану $\{2\}$ (обидва прилади є аварійними), тоді вона залишається в цьому стані назавжди. Якщо $\xi(t)$ описує тепер стан модифікованої системи, тоді нам треба знайти $P_{02}(t)$ і $P_{12}(t)$ для процесу Маркова $\xi(t)$ зі станами $\{0\}$,$\{1\}$,$\{2\}$ і інтенсивностями переходів $a_{01}=\lambda$, $a_{11}=-\lambda$, $a_{10}=\mu$, $a_{12}=\lambda$, $a_{11}=-\mu-\lambda$. Всі інші інтенсивності дорівнюють нулеві.

Теорема 4.

$$P_{12}(t) = \frac{\mu-\sqrt{4\lambda\mu+\mu^2}}{2\sqrt{4\lambda\mu+\mu^2}}e^{\alpha_- t} - \frac{\mu+\sqrt{4\lambda\mu+\mu^2}}{2\sqrt{4\lambda\mu+\mu^2}}e^{\alpha_+ t} + 1,$$
$$P_{02}(t) = \frac{\mu+\sqrt{4\lambda\mu+\mu^2}^{\;2}}{4\mu\sqrt{4\lambda\mu+\mu^2}}e^{\alpha_- t} + \frac{\mu-\sqrt{4\lambda\mu+\mu^2}^{\;2}}{4\mu\sqrt{4\lambda\mu+\mu^2}}e^{\alpha_+ t} + 1,$$

де

$$\alpha_{\pm} = \frac{-2\lambda - \mu \pm \sqrt{4\lambda\mu + \mu^2}}{2}.$$

Д о в е д е н н я. З першої системи рівнянь Колмогорова отримаємо:

$$\begin{cases} \frac{dP_{02}(t)}{dt} & = \lambda P_{12}(t) - \lambda P_{02}(t), \\ \frac{dP_{12}(t)}{dt} & = \mu P_{02}(t) + \lambda - (\lambda + \mu) P_{12}(t). \end{cases} \tag{11}$$

Звідки

$$\frac{d^2 P_{12}(t)}{dt^2} + (2\lambda + \mu)\frac{dP_{12}(t)}{dt} + \lambda^2 P_{12}(t) = \lambda^2.$$

З другого рівняння в (11) випливає $\left.\frac{dP_{12}(t)}{dt}\right|_{t=0} = \lambda$, що дасть наступну задачу Коші

$$\begin{cases} \frac{d^2 P_{12}(t)}{dt^2} + (2\lambda + \mu)\frac{dP_{12}(t)}{dt} + \lambda^2 P_{12}(t) & = \lambda^2, \\ \left.\frac{dP_{12}(t)}{dt}\right|_{t=0} = \lambda, \qquad P_{12}(0) = 0. \end{cases} \tag{12}$$

Загальний розв'язок першого рівняння в (12) має вигляд

$$P_{12}(t) = C_+ e^{\alpha_+ t} + C_- e^{\alpha_- t} + 1,$$

де

$$\alpha_{\pm} = \frac{-2\lambda - \mu \pm \sqrt{4\lambda\mu + \mu^2}}{2}.$$

З початкових умов в (12) випливає, що

$$C_+ = -\frac{\mu + \sqrt{4\lambda\mu + \mu^2}}{2\sqrt{4\lambda\mu + \mu^2}}, \qquad C_- = \frac{\mu - \sqrt{4\lambda\mu + \mu^2}}{2\sqrt{4\lambda\mu + \mu^2}}.$$

Тепер з другого рівняння в (11) булемо мати:

$$P_{02}(t) = e^{\alpha_+ t}\frac{C_+(\alpha_+ + \lambda + \mu)}{\mu} + e^{\alpha_- t}\frac{C_-(\alpha_- + \lambda + \mu)}{\mu} + 1.$$

◄

ЗАВДАННЯ

1. Довести, що процес $\xi(t)$ з множиною станів (1) є марковським.
2. Довести, що загальний розв'язок рівняння

$$u(i+1)a + u(i)b + u(i-1)c = 0, \qquad i \geq 1$$

має вигляд

$$u(i) = At_1^i + Bt_2^i, \qquad i \geq 1,$$

якщо $t_1 \neq t_2$, де t_1, t_2 є коренями оівняння $at^2 + bt + c = 0$, і

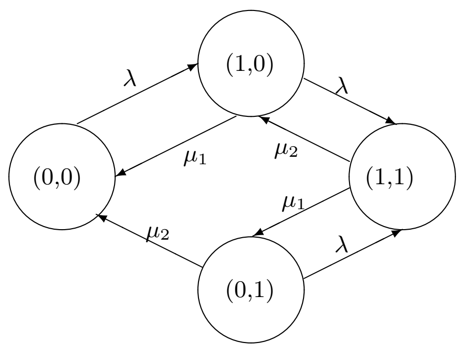
$$u(i) = t^i(Ai + B), \qquad i \geq 1,$$

якщо $t_1 = t_2 = t$.

3. Довести твердження 1 для $\rho = 1$.
4. Знайти ергодичний розподіл для наступної системи.

Опис моделі. Два прилади працюють незалежно і починають функціонування по черзі. Вимоги обслуговуються по одній, а час обслуговування вимоги має показниковий розподіл з параметром μ_1 для першого прилада і μ_2 для другого. Вимоги спочатку поступають на перший прилад. Якщо перший прилад є зайнятим, тоді вимога йде на другий прилад. Якщо обидва прилади є зайнятими, тоді вимога залишає систему і втрачається. Вимоги надходять по одній і незалежно від роботи приладів, час між моментами надходжень має показниковий розподіл з параметром λ.

Вказівка. *Використати граф переходів системи.*



Відповідь.

$$\pi_{00} = \frac{\mu_1\mu_2(\mu_1 + \mu_2 + 2\lambda)}{\lambda^2(\lambda + \mu_2)}\pi_{11}, \quad \pi_{10} = \frac{\mu_2(\mu_1 + \mu_2 + \lambda)}{\lambda(\lambda + \mu_2)}\pi_{11},$$

$$\pi_{01} = \frac{\mu_1}{\lambda + \mu_2}\pi_{11}, \quad \pi_{11} = \lambda^2\Big[\lambda^2 + \lambda(\mu_1 + \mu_2) + \frac{\mu_1\mu_2(\mu_1 + \mu_2 + 2\lambda)}{\lambda + \mu_2}\Big]^{-1},$$

де

$\pi_{00} = \mathbf{P}$(система є вільною), π_{10}=$\mathbf{P}$(перший прилад є зайнятим, а другий є вільний),
$\pi_{01} = \mathbf{P}$(другий прилад є зайнятим, а перший є вільним),
$\pi_{11} = \mathbf{P}$(обидва прилади є зайнятими).

5. (Система з нетерплячими вимогами). До системи обслуговування надходять вимоги двох типів: звичайні і нетерплячі. Якщо в момент прибуття вимоги система є вільною, тоді відразу розпочинається її обслуговування. Якщо вимога, яка прибула, є нетерплячою, а система є зайнятою обслуговуванням звичайної вимоги, тоді вимога, яка обслуговується в цей момент, втрачається і розпочинається обслуговування нетерплячої вимоги. Якщо система вже зайнята обслуговуванням нетерплячої вимоги, тоді нетерпляча вимога, яка щойно надійшла, втрачається. У випадку, коли приходить звичайна вимога, а система є зайнятою обслуговуванням будь-якої вимоги, тоді вимога, яка прийшла втрачається. Моменти приходів звичайних і нетерплячих вимог утворюють незалежні процеси Пуассона з параметрами λ_1, λ_2 відповідно. Час обслуговування не залежить від типу вимоги і має показниковий розподіл з параметром μ. Знайти ймовірності стану рівноваги цієї системи.
6. Знайти ймовірність того, що система $M/M/1/0$ є вільною в момент t, якщо інтенсивність вхідного потоку дорівнює λ, параметр часу обслуговування дорівнює μ, а також:

   a) в момент $t = 0$ система була вільною;

   b) в момент $t = 0$ система була зайнятою.

Лекція 12. Гіллясті процеси

Нехай в момент $t = 0$ є i ідентичних частинок. Після випадкового інтерваду часу, який має показниковий розподіл з параметром $\lambda > 0$, кожна частинка ділеться незалежно від інших частинок на випадкову кількість $\zeta \in \{0, 1, 2, ...\}$ нових частинок. (Випадок $\zeta = 0$ означає, що частинка зникла). Нові частинки, які виникли таким чином, будемо називати *першим поколінням.* Еволюція кожної з частинок першого покоління відбувається точно в той самий спосіб як і еволюція частинок, які були в момент $t = 0$. Частинки, які з'явились після поділу частинок першого покоління, називають *другим поколінням.* Частинки другого покоління діляться і утворюють частинки третього покоління і т.д. Якщо $\xi(t)$ є числом частинок у момент t, тоді $\xi(t)$ є прикладом *гіллястого процесу.* З опису моделі випливає, що $\xi(t)$ є однорідним марковським процесом з множиною станів $E = \{0, 1, 2, ...\}$.

12.1 Визначення і властивості гіллястого процесу

Перейдемо від представленого вище конструктивного опису до формального визначення. З цією метою введемо наступні позначення. Нехай τ є часом життя частинки до поділу і $\mathbf{P}(\tau < t) = 1 - e^{-\lambda t}$, $\lambda > 0$. Будемо також вважати, що $\mathbf{P}(\zeta = k) = p_k$, $k = 0, 1, 2, ...$ є розподілом числа нових частинок, які виникають в результаті поділу однієї частинки.

Визначення 1. *Марковський процес $\xi(t) \in E = \{0, 1, 2, ...\}$ будемо називати гіллястим процесом, якщо його інтенсивності мають вигляд*

$$a_{ij} = \begin{cases} i\lambda p_{j-i+1}, & j = i-1 \textit{ albo } j \geq i+1, \\ -i\lambda(1-p_1), & j = i, \\ 0, & j \leq i-2. \end{cases} \tag{1}$$

Якщо $P_{ij}(t)$ є ймовірністю переходу гіллястого процесу $\xi(t)$, тоді (1) є еквівалентним

$$P_{ij}(t) = \begin{cases} i\lambda p_{j-i+1} t + o(t), & j = i-1 \text{ albo } j \geq i+1, \\ 1 - i\lambda(1-p_1)t + o(t), & j = i, \\ o(t), & j \leq i-2, \end{cases} \tag{2}$$

якщо $t \to 0$.

Покажемо тепер, що поданий вище конструктивний опис гіллястого процесу відповідає наведеному визначенню, тобто виразам (1), (2). Припустимо, що в момент $t = 0$ маємо i ідентичних частинок. Позначимо тривалість життя кожної з частинок через $\tau(k), k = 1, 2, ..., i$. Згідно з конструктивним описом випадкові величини $\tau(k), k = 1, 2, ..., i$ є незалежними і мають однаковий розподіл $\mathbf{P}(\tau(k) < t) = 1 - e^{-\lambda t}$, $k = 1, 2, ..., i$. Тому для першого після момента часу $t = 0$ стрибка процесу $\xi(t)$, тобто для $\min_{1 \leq k \leq i} \tau(k)$ має місце

$$\mathbf{P}(\min_{1 \leq k \leq i} \tau(k) > t) = \mathbf{P}(\cap_{k=1}^{i} \{\tau(k) > t\}) = \prod_{k=1}^{i} \mathbf{P}(\tau(k) > t) = e^{-i\lambda t}. \tag{3}$$

Ймовірність того, що одна частинка на інтервалі $[0, t]$ згенерує принаймні два покоління, є не більшою ніж

$$\lambda^2 \int_0^t e^{-\lambda v} \int_0^{t-v} e^{-\lambda u} du dv = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} = o(t), \quad \text{якщо } t \to 0. \tag{4}$$

Ймовірність того, що принаймні дві частинки будуть мати потомство в інтервалі $[0, t]$, також становить $o(t)$, якщо $t \to 0$ (завдання 1). Отже, для того, щоб дослідити поведінку ймовірності переходу $P_{ij}(t)$ для малих t, вистачить розглянути ситуацію, за якої не більше ніж одна частинка буде мати потомство в інтервалі $[0, t]$, і то не більше ніж один раз.

Нехай $j = i - 1$ або $j \geq i + 1$. Перехід $i \to j$ в інтервалі $[0, t]$ є можливим, коли якась частинка з i частинок буде мати потомство з $j - i + 1$ частинок в інтервалі $[0, t]$. Ймовірність такої події є $(1 - e^{-\lambda t})p_{j-i+1}$. Позаяк маємо i частинок, тоді $P_{ij}(t) = i(1 - e^{-\lambda t})p_{j-i+1} + o(t) = i\lambda p_{j-i+1} t + o(t)$, якщо $t \to 0$. Таким чином отримаємо першу рівність в (2).

Нехай тепер $j = i$. Перехід $i \to i$ в інтервалі $[0, t]$ є можливим тільки тоді, коли якась частинка буде мати в цьому інтервалі одиничне потомство, або жодна частинка не буде мати потомства. Ймовірність цих подій становить відповідно $i(1 - e^{-\lambda t})p_1$ і (див. вираз (3)) $e^{-i\lambda t}$. Ось чому $P_{ii}(t) = i(1 - e^{-\lambda t})p_1 + e^{-i\lambda t} + o(t) = 1 - i\lambda(1 - p_1)t + o(t)$, якщо $t \to 0$, що доводить другу рівність в (2). Третя рівність в (2) випливає з того, що ймовірність зникнення принаймні двох частинок в інтервалі $[0; t]$ є не більшою за (див. вираз (4))

$$[1 - e^{-\lambda t} - \lambda t e^{-\lambda t}]p_0^2 = o(t), \quad \text{jeњli } t \to 0.$$

Наступним нашим завданням буде дослідження ймовірності переходу $P_{ij}(t) = \mathbf{P}(\xi(t) = j/\xi(0) = i)$. Нехай в момент часу $t = 0$ є i частинок. Позначимо через $\xi_k(t)$, $k = 1, 2, ..., i$ число потомків k-тої частинки в момент t. Зрозуміло, що тоді $\xi_k(t)$, $k = 1, 2, ..., i$ є незалежними гіллястими процесами з функціями переходів $P_{1j}(t) = \mathbf{P}(\xi_k(t) = j/\xi_k(0) = 1), k = 1, 2, ..., i$ і $\xi(t) = \sum_{k=1}^{i} \xi_k(t)$. Для $0 < z < 1$ введемо генератриси випадкових величин $\xi_k(t)$ і $\xi(t)$

$$P_i(z, t) = \mathbf{M}\{z^{\xi(t)}/\xi(0) = i\} = \sum_{j=0}^{\infty} z^j P_{ij}(t),$$

$$f(z, t) = \mathbf{M}\{z^{\xi_k(t)}/\xi_k(0) = 1\} = \sum_{j=0}^{\infty} z^j P_{1j}(t).$$

Тоді

$$P_i(z, t) = \mathbf{M}\{z^{\xi(t)}/\xi(0) = i\} = \mathbf{M}\{z^{\sum_{k=1}^{i} \xi_k(t)}/\xi(0) = i\} =$$

$$= \prod_{k=1}^{i} \mathbf{M}\{z^{\xi_k(t)}/\xi_k(0) = 1\} = \ f(z, t)\ ^i. \tag{5}$$

Звідки випливає, що досить дослідити ймовірність переходу для $\xi(0) = 1$, тобто, $P_{1j}(t)$.

З першої системи рівнянь Колмогорова отримаємо:

$$\frac{dP_{1j}(t)}{dt} = -\lambda(1 - p_1)P_{1j}(t) + \lambda \sum_{k=0,\neq 1}^{\infty} p_k P_{kj}(t), \qquad j \geq 0. \tag{6}$$

Якщо помножити рівняння (6) на z^j, а потім просумувати від 0 до ∞ по j, тоді, беручи до уваги (5), будемо мати:

$$\frac{df(z,t)}{dt} = -\lambda(1-p_1)f(z,t) + \lambda \sum_{k=0,\neq 1}^{\infty} p_k P_k(z,t) =$$

$$= -\lambda(1-p_1)f(z,t) + \lambda \sum_{k=0,\neq 1}^{\infty} p_k \ \ f(z,t) \ \ ^k = -\lambda f(z,t) + \lambda \sum_{k=0}^{\infty} p_k \ \ f(z,t) \ \ ^k.$$

Після чого для фіксованого $z \in (0,1)$ отримаємо наступну крайову задачу на піввісі $t \geq 0$

$$\begin{cases} \frac{df(z,t)}{dt} = -\lambda f(z,t) + \lambda \sum_{k=0}^{\infty} p_k \ \ f(z,t) \ \ ^k, \quad t > 0, \\ f(z,0) = z. \end{cases} \tag{7}$$

Нехай функція $U(v), 0 \leq v \leq 1$ є визначеною наступним чином:

$$U(v) = \lambda p_0 - \lambda(1-p_1)v + \lambda \sum_{k=2}^{\infty} p_k v^k. \tag{8}$$

Тепер з (7) отримаємо наступне твердження.

Теорема 1. *Функція* $f(z,t) = \sum_{j=0}^{\infty} z^j P_{1j}(t)$ *є розв'язком крайової задачі*

$$\frac{df(z,t)}{dt} = U \ \ f(z,t) \ \ , \qquad f(z,0) = z, \tag{9}$$

де $U(v) = \lambda p_0 - \lambda(1-p_1)v + \lambda \sum_{k=2}^{\infty} p_k v^k$ *і тоді*

$$P_i(z,t) = \mathbf{M} z^{\xi(t)} = \ \ f(z,t) \ \ ^i. \tag{10}$$

Визначимо тепер функцію

$$\Psi(z) = \int_0^z \frac{dv}{U(v)}, \qquad 0 \leq z \leq 1. \tag{11}$$

Зауваження 1. Якщо взяти до уваги властивості функції $U(v)$ (див. нижче графік 1 цієї функції), тоді нескладно довести, що інтеграл у виразі (11) є скінченним, принаймні для достатньо малих z, а, отже для цих самих z функція $\Psi(z)$ є коректно визначеною.

Функція f, яка є розв'язком наступного функціонального рівняння

$$\Psi(f) - \Psi(z) = t, \tag{12}$$

є також розв'язком крайової задачі (9) (завдання 2).

Якщо існує такий момент t_0, що $\xi(t_0) = 0$, тоді, вочевидь, $\xi(t_0) = 0$ для всіх $t \geq t_0$. У зв'язку з чим випадковий момент $\zeta = \inf\{t > 0 : \xi(t) = 0\}$ будемо називати *моментом виродження* гіллястого процесу $\xi(t)$. Позначимо $\alpha(i) = \mathbf{P}(\zeta < \infty/\xi(0) = i)$. Нехай $\xi(0) = i \geq 1$ і, як раніше, $\xi_k(t)$, $k = 1, 2, ..., i$ є числом потомків k-тої частинки в момент t. Якщо $\zeta_k = \inf\{t > 0 : \xi_k(t) = 0\}$, тоді зрозуміло, що ζ_k, $k = 1, 2, ..., i$ є незалежними однаково розподіленими випадковими величинами і $\zeta = \max_{1\le k\le i} \zeta_k$, а, отже,

$$\alpha(i) = \mathbf{P}(\zeta < \infty / \xi(0) = i) = \mathbf{P}(\max_{1\le k\le i} \zeta_k < \infty) = \mathbf{P}(\zeta_1 < \infty, ..., \zeta_i < \infty) =$$

$$= \prod_{k=1}^{i} \mathbf{P}(\zeta_k < \infty) = \alpha^i,$$

де $\alpha = \alpha(1) = \mathbf{P}(\zeta < \infty / \xi(0) = 1)$. З цього випливає, що достатньо дослідити випадок $\xi(0) = 1$.

Легко зрозуміти, що $\{\xi(t_1) = 0\} \subset \{\xi(t_2) = 0\}$, $t_1 < t_2$, а, отже послідовність подій $\{\xi(t_2) = 0\}$ є неспадною по t. Тому

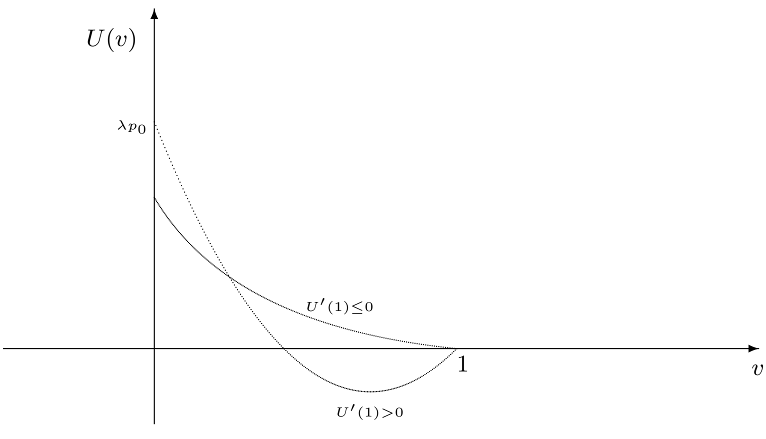
$$\alpha = \mathbf{P}(\zeta < \infty / \xi(0) = 1) = \mathbf{P}(\lim_{t\to\infty} \xi(t) = 0 / \xi(0) = 1) =$$
$$= \lim_{t\to\infty} \mathbf{P}(\xi(t) = 0 / \xi(0) = 1) = \lim_{t\to\infty} P_{10}(t).$$

Наступне твердження визначає, коли $\alpha = 1$, а коли $\alpha < 1$.

Теорема 2. *Нехай v_0 є найменшим коренем рівняння $U(v) = 0$ в інтервалі* $[0, 1]$.

$$\alpha = \begin{cases} v_0 < 1, & \text{якщо} \quad U'(1) = -\lambda(1 - p_1) + \lambda \sum_{k=2}^{\infty} kp_k > 0, \\ 1, & \text{якщо} \quad U'(1) = -\lambda(1 - p_1) + \lambda \sum_{k=2}^{\infty} kp_k \le 0. \end{cases}$$

Д о в е д е н н я. Насамперед зауважимо, що для функції $U(v)$, визначеної в (8), будемо мати $U(0) = \lambda p_0 \ge 0$, $U(1) = 0$ і $U''(v) \ge 0$, $0 \le v \le 1$, а, отже графік функції $U(v)$ має вигляд



Мал. 1

Як видно, рівняння $U(v) = 0$ завше має в інтервалі $[0, 1]$ найменший розв'язок v_0, такий, що $v_0 < 1$, коли $U'(1) > 0$, і $v_0 = 1$, коли $U'(1) \leq 0$.

Якщо $p_0 = 0$, тоді частинки мають ≥ 1 потомків. Тоді $\zeta = \infty$ з ймовірністю 1, а, отже, процес ніколи не вироджується. Легко зрозуміти, що $U'(1) = \lambda \sum_{k=2}^{\infty}(k-1)p_k > 0$ і $v_0 = 0$, тому твердження є справедливим.

Нехай тепер $p_0 > 0$. Графіки на мал. 1 демонструють власне таку ситуацію. Якщо в (9) покласти $z = 0$, тоді, позаяк $P_{10}(t) = f(0, t)$, ми отримаємо наступну крайову задачу:

$$\frac{dP_{10}(t)}{dt} = U\ \ P_{10}(t)\ \ , \qquad P_{10}(t) = 0. \tag{13}$$

Позаяк $\lim_{t\to\infty} P_{10}(t) = \alpha \leq 1$, тоді з (13) отримаємо $\lim_{t\to\infty} P'_{10}(t) = U(\alpha)$. Якщо має місце $U(\alpha) > 0$ або $U(\alpha) < 0$, тоді функція

$$P_{10}(t) = \int_0^t P'_{10}(v)dv$$

мала би зростати до ∞ (при $U(\alpha) > 0$) або спадати до $-\infty$ (при $U(\alpha) < 0$). А це є неможливим, отже, $U(\alpha) = 0$. Тепер доведення легко завершити, використовуючи мал. 1. ◄

12.2 Приклади гіллястих процесів

Розглянемо гіллястий процес, для якого $\mathbf{P}(\zeta{=}k){=}p_k$, $k{=}0, 1, 2$, а час життя частинки має показниковий розподіл з параметром $\lambda > 0$. Іншими словами в момент поділу частинка зникає з ймовірністю p_0, залишається одна з ймовірністю p_1, або ділеться на дві частинки з ймовірністю p_2.

Інтенсивності переходів цього процесу мають вигляд:

$$a_{ij} = \begin{cases} i\lambda p_0, & j = i-1, \\ i\lambda p_2, & j = i+1, \\ -i\lambda(1-p_1), & j = i, \\ 0, & \text{для решти } j. \end{cases} \tag{14}$$

Функція $U(v)$ у виразі (8) набуває вигляду

$$U(v) = \lambda p_0 - \lambda(1-p_1)v + \lambda p_2 v^2. \tag{15}$$

Згідно з твердженням 1 для функції $f(z, t) = \sum_{j=0}^{\infty} z^j P_{1j}(t)$ отримаємо наступну крайову залачу:

$$\frac{df(z,t)}{dt} = \lambda p_0 - \lambda(1-p_1)f(z,t) + \lambda p_2 f^2(z,t), \qquad f(z, 0) = z. \tag{16}$$

Функцію $\Psi(z)$ можемо записати у вигляді (припускаємо, що $p_2 \neq p_0$)

$$\Psi(z) = \lambda^{-1}\int_0^z \frac{dv}{p_0 - (1-p_1)v + p_2 v^2} = \frac{1}{\lambda(p_2 - p_0)} \ln \frac{1-z}{1-\alpha z}, \qquad \alpha = \frac{p_2}{p_0}.$$

Для розв'язування крайової задачі (16) використовуємо функціональне рівняння

$$\frac{1}{\lambda(p_2-p_0)}\ln\frac{1-f(z,t)}{1-\alpha f(z,t)}-\frac{1}{\lambda(p_2-p_0)}\ln\frac{1-z}{1-\alpha z}=t.$$

З цього випливає, що

$$\sum_{n=0}^{\infty} z^n P_{1n}(t)=f(z,t)=\frac{1-\alpha z-(1-z)e^{\lambda(p_2-p_0)t}}{1-\alpha z-\alpha(1-z)e^{\lambda(p_2-p_0)t}}. \tag{17}$$

Якщо скористатись розкладенням

$$\frac{1-az-b(1-z)}{1-az-ab(1-z)}=\frac{1-b}{1-ab}+\frac{b(1-a)^2}{(1-ab)^2}\sum_{n=1}^{\infty}\frac{a(1-b)}{1-ab}\;^{n-1}z^n,$$

яке має місце принаймні для малих z, тоді з (17) отримаємо:

$$\begin{cases} P_{10}(t)=\dfrac{1-e^{\lambda(p_2-p_0)t}}{1-\alpha e^{\lambda(p_2-p_0)t}}, \\ P_{1n}(t)=\dfrac{(1-\alpha)^2\alpha^{n-1}(1-e^{\lambda(p_2-p_0)t})^{n-1}}{(1-\alpha e^{\lambda(p_2-p_0)t})^{n+1}}e^{\lambda(p_2-p_0)t}, \quad n=1,2,... \end{cases} \tag{18}$$

У виразах (18) ми припускали, що $p_2 \neq p_0$. Випадок $p_2 = p_0$ є легшим, і ми залишаємо його читачеві (завдання 3).

Якщо $p_2 < p_0$, тоді $\alpha < 1$ і з виразів (18) випливає, що $P_{10}(t) \to 1$, а $P_{1n}(t) \to 0$ $n \geq 1$ при $t \to \infty$. Тому наш процес з ймовірністю 1 зникне після певного часу.

Якщо $p_2 > p_0$, тоді $\alpha > 1$ і з виразу (18) випливає, що $P_{10}(t) \to \alpha^{-1} = p_0/p_2 < 1$, а $P_{1n}(t) \to 0$, $n \geq 1$, при $t \to \infty$. У цьому випадку ймовірність того, що процес зникне, є $p_0/p_2 < 1$.

Якщо частинки не зникають, тоді їх число росте до нескінченності. Дійсно,

$$\begin{aligned} &\mathbf{P}(\xi(t) > N/\xi(t) > 0) = 1-\mathbf{P}(1 \leq \xi(t) \leq N/\xi(t) > 0) = \\ &= 1-\frac{\mathbf{P}(1 \leq \xi(t) \leq N, \xi(t) > 0)}{\mathbf{P}(\xi(t) > 0)} = 1-\frac{\mathbf{P}(1 \leq \xi(t) \leq N)}{\mathbf{P}(\xi(t) > 0)} = \\ &= 1-\frac{1}{1-P_{10}(t)}\sum_{k=1}^{N}P_{1k}(t) \underset{t\to\infty}{\to} 1 \end{aligned}$$

для довільного $N > 1$.

Якщо в наведеному прикладі гіллястого процесу має місце $p_0 = p_1 = 0$ і $p_2 = 1$, тоді такий процес називається *процесом Ула.* Тоді

$$\begin{cases} P_{10}(t)=0, \\ P_{1n}(t)=\dfrac{1-e^{\lambda(n-1)t}}{e^{\lambda(n+1)t}}, \quad n=1,2,... \end{cases} \tag{19}$$

Доведення формул (19) залишаємо читачеві (завдання 4).

ЗАВДАННЯ

1. Довести, що ймовірність того, що принаймні дві частинки мають потомків в інтервалі $[0,t]$, є $o(t)$, якщо $t \to 0$.

2. Довести, що для фіксованого $\delta > 0$ існує розв'язок $f(z,t)$ в інтервалі $0 \le z \le \delta$ і функція $f(z,t)$ є розв'язком крайової задачі (9) на цьому інтервалі.
3. Довести, що коли у виразах (18) $p_2 = p_0 = p$, тоді

$$P_{10}(t) = \frac{\lambda pt}{1+\lambda pt}, \qquad P_{1n}(t) = \frac{(\lambda pt)^{n-1}}{(1+\lambda pt)^{n+1}}, \quad n = 1, 2, ...$$

4. Довести вирази (19).

Лекція 13. Процес Вінера

У 1827 році шотландський біолог Роберт Броун, спостерігаючи квітковий пилок у водяній суспензії, виявив, що він перебуває у безперервному хаотичному русі. Частинки рухались без зупинки, крім того, їх рух не зменшувався з часом. Швидкість руху була тим більша, чим меньшою була частинка і вищою температура води. Броун описав цей рух як певне фізичне явище. Воно було згодом досліджене в працях Енштейна (1905), Смолуховського (1906) і інших відомих фізиків. Математичну модель цього явища, яке дістало назву *броунівський рух*, запропонував Норберт Вінер [1] у 1918 році. Математична модель броунівського руху називається *вінерівським процесом.*

Вінерівський процес має чисельні застосування при моделюванні реальних процесів, які зустрічаються в різних галузях, наприклад економіці, страхуванні та інших.

13.1 Визначення та властивості і вінерівського процесу

Визначення 1. *Випадковий процес $W(t)$ будемо називати вінерівським процесом (або броунівсьвим рухом), Якщо:*

1) *для всіх $0 \le t_0 < t_1 < ... < t_n$ випадкові величини $W(t_i) - W(t_{i-1})$, $i = 1, ..., n$ є незалежними (тобто, процес $W(t)$ є процесом з незалежними приростами);*
2) *випадкова величина $W(t) - W(s)$ має розподіл $N(0, \sigma^2(t-s))$ для всіх $s < t$, тобто,*

$$\mathbf{P}(W(t) - W(s) < x) = \frac{1}{\sigma\sqrt{2\pi(t-s)}} \int\limits_{-\infty}^{x} e^{-\frac{v^2}{2\sigma^2(t-s)}} dv. \tag{1}$$

З пункту 1) випливає, що процес Вінера є процесом марковським, а з пункту 2), що він також є однорідним процесом.

Якщо $\sigma = 1$, тоді такий процес будемо називати *стандартним вінерівським процесом.* У подальшому будемо розглядати тільки такий процес, і тоді замість (1) буде мати місце

$$\mathbf{P}(W(t) - W(s) < x) = \frac{1}{\sqrt{2\pi(t-s)}} \int\limits_{-\infty}^{x} e^{-\frac{v^2}{2(t-s)}} dv, \quad s < t. \tag{2}$$

[1] *Норберт Вінер* (1894-1964) – американський математик, засновник кібернетики. Праці Вінера стосуються фундаментальних засад математики, теорії ймовірностей і функціонального аналізу.

Зауваження 1. У визначенні процесу Вінера часто додають умову, що $W(0) = 0$. Ця умова не є обов'язковою, може бути так, що $W(0) = u \neq 0$, де u може навіть бути випадковою величиною. Ми ж зазвичай будемо передбачати, що $W(0) = 0$.

Позаяк випадкова величина $W(t)$ має розподіл $N(0, t)$, тоді характеристична функція $W(t)$ буде мати вигляд

$$\mathbf{M}e^{isW(t)} = e^{-s^2 t/2}, \quad s \in R^1, \quad t \geq 0. \tag{3}$$

Наступні два твердження містять у собі найважніші властивості вінерівського процесу.

Теорема 1. *З ймовірністю 1 траєкторії процесу Вінера є неперервні.*

Д о в е д е н н я. Знайдемо $\mathbf{M}|W(t) - W(s)|^3$. Будемо мати

$$\mathbf{M}|W(t) - W(s)|^3 = \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{\infty} |x|^3 e^{-\frac{x^2}{2(t-s)}} dx = \frac{(t-s)^{3/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |v|^3 e^{-\frac{v^2}{2}} dv.$$

Звідки

$$\mathbf{M}|W(t) - W(s)|^3 = C(t-s)^{3/2},$$

де $C = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |v|^3 e^{-\frac{v^2}{2}} dv$. Тепер результат твердження випливає з твердження 2, стор. 288. ◄

Теорема 2. *З ймовірністю 1 трєкторії процесу Вінера не є диференційованими в жодній точці,тобто*

$$\mathbf{P}(\lim_{h\to 0} \frac{W(t+h) - W(t)}{h} < \infty, \ t \geq 0) = 0. \tag{4}$$

Доведення цього твердження можна знайти, наприклад, в [?].

Теорема 3. *Нехай* $a = t_0 < t_1 < \ldots < t_n = b$ *і* $\Delta_n = \max_{1\leq i\leq n}(t_i - t_{i-1})$. *Тоді*

$$\sum_{i=1}^{n} \left(W(t_i) - W(t_{i-1}) \right)^2 \xrightarrow[\Delta_n \to 0]{\text{ъ}.k} b - a,$$

тобто

$$\lim_{\Delta_n \to 0} \mathbf{M} \left(\sum_{i=1}^{n} \left(W(t_i) - W(t_{i-1}) \right)^2 - (b-a) \right)^2 = 0. \tag{5}$$

Д о в е д е н н я. Маємо

$$\mathbf{M}\sum_{i=1}^{n} \left(W(t_i) - W(t_{i-1}) \right)^2 = \sum_{i=1}^{n} \mathbf{M} \left(W(t_i) - W(t_{i-1}) \right)^2 = \sum_{i=1}^{n} (t_i - t_{i-1}) = b - a.$$

У подальших міркуваннях скористаємось фактом, що коли $\xi \in N(0,\sigma^2)$, тоді $\mathbf{M}\xi^4 = 3\sigma^2$. Маємо

$$\mathbf{M}\Big(\sum_{i=1}^{n}\big(W(t_i)-W(t_{i-1})\big)^2-(b-a)\Big)^2=$$
$$=\mathbf{M}\Big(\sum_{i=1}^{n}\big(W(t_i)-W(t_{i-1})\big)^2-\mathbf{M}\sum_{i=1}^{n}\big(W(t_i)-W(t_{i-1})\big)^2\Big)^2$$
$$=\mathbf{D^2}\sum_{i=1}^{n}\big(W(t_i)-W(t_{i-1})\big)^2=\sum_{i=1}^{n}\mathbf{D^2}\big(W(t_i)-W(t_{i-1})\big)^2=$$
$$=\sum_{i=1}^{n}\Big[\mathbf{M}\big(W(t_i)-W(t_{i-1})\big)^4-\Big(\mathbf{M}\big(W(t_i)-W(t_{i-1})\big)^2\Big)^2\Big]=$$
$$=\sum_{i=1}^{n}\Big[3(t_i-t_{i-1})^2-(t_i-t_{i-1})^2\Big]=2\sum_{i=1}^{n}(t_i-t_{i-1})^2\le\Delta_n\sum_{i=1}^{n}(t_i-t_{i-1})=$$
$$=\Delta_n(b-a)\underset{\Delta_n\to 0}{\longrightarrow}0,$$

що завершує доведення. ◂

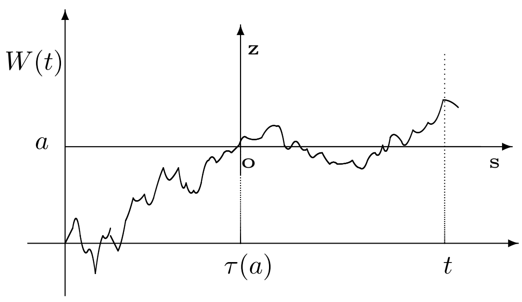
13.2 Розподіл $\inf_{0\le u\le t} W(u)$

Визначимо тепер розподіли деякмх важливих функціоналів вінерівського процесу. Для цього скористаємось властивістю Маркова і неперервністю траєкторій процесу.

Теорема 4. *Для $a \ge 0$ має місце:*

$$\mathbf{P}(\sup_{0\le u\le t} W(u)>a)=\mathbf{P}(\inf_{0\le u\le t} W(u)<-a)=\frac{2}{\sqrt{2\pi t}}\int_a^{\infty}e^{-\frac{v^2}{2t}}\,dv. \tag{6}$$

Д о в е д е н н я. Для фіксованого t припустимо, що подія $\sup_{0\le u\le t} W(u)>a$ сталась. Позаяк траєкторія $W(u)$ є неперервною, тоді можемо стверджувати, що існує $\tau(a)=\inf\{u\ge 0; W(u)=a\}<t$ і, вочевидь, $W(\tau(a))=a$ (див мал. 1).



Мал. 1

Позаяк $\{W(t) > a\} \subset \{\tau(a) < t\}$, тоді

$$\mathbf{P}(W(t) > a) = \mathbf{P}(\tau(a) < t, W(t) > a) =$$
$$= \int_0^t \mathbf{P}(W(t) > a/\tau(a) = u) d\mathbf{P}(\tau(a) < u). \quad (7)$$

Припустимо, що сталась подія $\{\tau(a) = u\}$. Поведінка процесу $W(t)$ після момента u не залежить від історії цього процесу до момента $\tau(a) = u$ (властивість Маркова!), а позаяк $W(u) = a$, тоді в системі координат ZOS (див. мал. 1) отримаємо новий процес Вінера $\widehat{W}(s)$ і, вочевидь,

$$\mathbf{P}(W(t) > a/\tau(a) = u) = \mathbf{P}(\widehat{W}(t-u) > 0),$$

але також має місце $\mathbf{P}(\widehat{W}(t-u) > 0) = 1/2$ (див. завдання 2). З (7) робимо висновок

$$\mathbf{P}(W(t) > a) = \frac{1}{2}\int_0^t d\mathbf{P}(\tau(a) < u) = \frac{1}{2}\mathbf{P}(\tau(a) < t),$$

отже

$$\mathbf{P}(\sup_{0\le u\le t} W(u) > a) = \mathbf{P}(\tau(a) < t) = 2\mathbf{P}(W(t) > a) = \frac{2}{\sqrt{2\pi t}}\int_a^\infty e^{-\frac{v^2}{2t}} dv.$$

Тепер

$$\mathbf{P}(\inf_{0\le u\le t} W(u) < -a) = \mathbf{P}(-\sup_{0\le u\le t}(-W(u)) < -a) = \mathbf{P}(\sup_{0\le u\le t}(-W(u)) > a) =$$
$$= \mathbf{P}(\sup_{0\le u\le t} W(u) > a) = \frac{2}{\sqrt{2\pi t}}\int_a^\infty e^{-\frac{v^2}{2t}} dv.$$

Ми використали факт, що $-W(t)$ також є вінерівським процесом (завдання 1). ◄

13.3 Процеси з незалежними приростами

У п'ятій лекції в твердженні 1 про властивості узагальненого процесу Пуассона, було зазначено, що узагальнений процес Пуассона $\xi(t)$ є процесом з незалежними приростами, тобто для довільних $0 \le t_1 < t_2 < ... < t_n$ випадкові величини $\xi(t_i) - \xi(t_{i-1})$, $i = 1, ..., n$ є незалежними. Ця властивість може бути використана для визначення цілого класу процесів.

Визначення 2. *Випадковий процес $\xi(t)$ будемо називати процесом з незалежними приростами, якщо для довільних $0 \le t_1 < t_2 < ... < t_n$ випадкові величини $\xi(t_i) - \xi(t_{i-1})$, $i = 1, ..., n$ є незалежними.*

Процеси з незалежними приросами є марковськими процесами. Теорія таких процесів є дуже добре розвинутою. (див.напр.[**?**].) Особливо глибоко досліджені ситуації, коли процес з незалежними приростами є однорідним, тобто, якщо $\mathbf{P}(\xi(t+s) < x/\xi(s) = u) = \mathbf{P}(\xi(t) < x/\xi(0) = u)$ для довільних $x, u \in R^1$, $t, s \ge 0$. Нехай $\xi(t)$

є однорідним процесом з незалежними приростами, припустимо, що його траєкторії належать до $D_{[0,\infty)}$, тобто траєкторії процесу є неперервними справа і мають границі зліва. Нехай також $\nu(t)$ є числом стрибків процесу $\xi(t)$ в інтервалі $[0,t]$. Має місце наступне твердження.

Теорема 5. *Кожний однорідний стохастично неперервний процес з незалежними приростами, такий що* $\mathbf{P}(\nu(t) < \infty) = 1$ *для довільного* $0 \le t < \infty$, *може бути представлений у вигляді:*

$$\xi(t) = \xi(0) + \sigma W(t) + \eta(t), \tag{8}$$

де $\sigma \ge 0$, $W(t)$ *є вінерівським процесом,* $\eta(t)$ *є узагальненим процесом Пуассона і процеси* $W(t), \eta(t)$ *є незалежними.*

Умова $\mathbf{P}(\nu(t) < \infty) = 1$ означає, що процес має скінченне (але випадкове!) число стрибків на кожному скінченному інтервалі. Тільки такі процеси будемо вивчати в цій лекції.

З виразів (3) і формули (2) (стор. 317) отримаємо вираз характеристичної функції процесу з незалежними приростами

$$\mathbf{M}e^{is(\xi(t)-\xi(0))} = e^{tk(s)}, \quad s \in R^1, t \ge 0, \tag{9}$$

де

$$k(\mu) = -s^2\sigma^2/2 + isc + \lambda \int\limits_{-\infty}^{\infty} (e^{isx} - 1)dF(x). \tag{10}$$

В останньому виразі доданок $-s^2\sigma^2/2$ відповідає процесу $\sigma W(t)$, доданок $isc + \lambda\int_{-\infty}^{\infty}(e^{isx} - 1)dF(x)$ відповідає узагальненому процесу Пуассона $\eta(t)$ з (8). У свою чергу $c, \lambda, F(x)$ є параметрами цього процесу. А саме, c є коефіцієнтом зсуву, λ – параметром процесу Пуассона, який описує стрибки процесу $\eta(t)$, а $F(x)$ є розподілом величини стрибка цього процесу. Функція $k(s)$, так само як і для узагальненого процесу Пуассона, називається *кумулянтою* процесу $\xi(t)$.

13.4 Процес ризику зі збуреннями

Класичний процес ризику, про який шла мова в лекції 5, є досить наближеною моделлю реальної еволюції капіталу страхової компанії. Він не враховує випадкових коливань капіталу, спричинених впливом короткотермінових чинників (наприклад, випадкове коливання курсів валют, темпу інфляції та інших). У зв'язку з чим є різні узагальнення класичної моделі ризику. Далі розглянемо одне з них.

Нехай $\xi(t)$ є класичним процесом ризику згідно визначення 3 (стор. 318), тобто

$$\xi(t) = u + ct - \sum_{i=1}^{N(t)} \xi_i, \tag{11}$$

де u є початковим капіталом, а c – інтенсивністю надходжень внесків. Процес Пуассона $N(t)$ з параметром λ представляє число виплат, які були до момента часу t, а невід'ємна випадкова величина ξ_i є величиною виплати відшкодування i-ому клієнту.

Визначення 3. *Процес*

$$\xi(t) = u + \sigma W(t) + ct - \sum_{i=1}^{N(t)} \xi_i, \tag{12}$$

де $\sigma > 0$, а $W(t)$ є процесом Вінера, незалежним від $N(t)$ і випадкових величин ξ_i називається процесом ризику зі збуреннями.

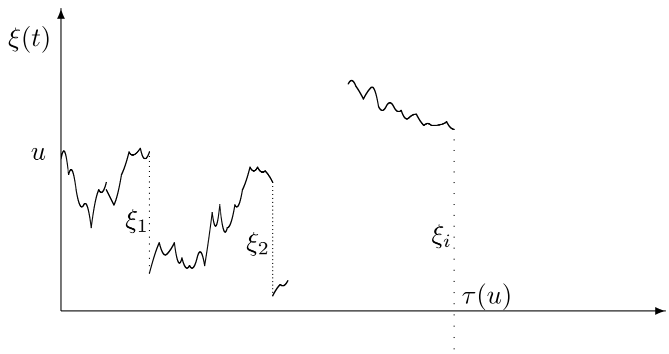
З виразів (9) і (10) для процесу (12) отримаємо

$$\mathbf{M}e^{s(\xi(t)-\xi(0))} = \exp\{tk(s)\}, \quad s \geq 0, \tag{13}$$

де функція $k(s)$, яку також називають кумулянтою, може бути записана у вигляді

$$k(s) = \frac{\sigma^2 s^2}{2} + sc + \lambda \int_0^\infty (e^{-sx} - 1)dF(x). \tag{14}$$

Наступний графік представляє траєкторію цього процесу.



Мал. 2

Нехай $\tau(u) = \inf\{t : \xi(t) < 0/\xi(0) = u\}$ є ймовірністю банкрутства для цього процесу.

У шостій лекції ми вивчили ймовірність банкрутства за припущення, що має місце умова **A)**, тобто

A) $s_0 = \inf\{s < 0 : \int_0^\infty e^{-sx} dF(x) < \infty\} < 0$ і $k(s_0) > 0$.

Подібно до того, як це було в шостій лекції, з цієї умови випливає, що рівняння $k(s) = \mu$, $0 < \mu \leq \delta$ має для достатньо малих $\delta > 0$ два корені $s_\pm(\mu)$ такі, що $s_-(\mu) < 0 < s_+(\mu)$. Нескладно зрозуміти, що лема 1 (стор. 324) залишається в силі. Тепер доведемо, що твердження 1 і 2, подані відповідно на сторінках 324 і 325, теж залишаються в силі для процесу ризику зі збуреннями.

Теорема 6. *Для процесу ризику зі збуреннями вигляду* (12) *має місце:*

$$\int_0^\infty e^{-su} \mathbf{M}e^{-\mu\tau(u)} du = \frac{1}{s} + \frac{\mu}{s_+(\mu)} \frac{s_+(\mu) - s}{s(k(s) - \mu)}, \qquad s > 0. \tag{15}$$

Д о в е д е н н я. Можна показати (завдання 8), що для кожного процесу вигляду (12) з кумулянтою $k(s)$ існує послідовність $\xi_n(t)$, $n \geq 1$ процесів вигляду (11) з кумулянтами $k_n(s)$ і така, що $k_n(s) \to k(s)$, якщо $n \to \infty$ для довільного фіксованого $s > 0$. Але тоді $\mathbf{M}e^{s(\xi_n(t)-\xi_n(0))} \to \mathbf{M}e^{s(\xi(t)-\xi(0))}$, якщо $n \to \infty$ для кожного фіксованого $s > 0$. З цього випливає (див. наприклад [?]) наступне, якщо $\tau_n(u)$ є моментом банкрутства для процесу $\xi_n(t)$, тоді $\mathbf{M}e^{-s\tau_n(u)} \to \mathbf{M}e^{-s\tau(u)}$ при $n \to \infty$. Для процесів $k_n(t)$ з твердження 1 (стор. 324) має місце

$$\int_0^\infty e^{-su}\mathbf{M}e^{-\mu\tau_n(u)}du = \frac{1}{s} + \frac{\mu}{s_+(n,\mu)}\frac{s_+(n,\mu)-s}{s(k_n(s)-\mu)}, \qquad s > 0, \tag{16}$$

де $s_+(n,\mu) > 0$ є додатним розв'язком рівняння $k_n(s) = \mu$. Зрозуміло, що $s_+(n,\mu) \to s_+(\mu)$, якщо $n \to \infty$. Перехід до границі $n \to \infty$ в (16) завершує доведення твердження. ◀

Доведення наступного твердження є практично аналогічним доведенню твердження 2 на сторінці 325.

Теорема 7. *Нехай $m = \mathbf{M}\xi_1$. Тоді*

$$\mathbf{P}(\tau(u) < \infty) = \begin{cases} 1, & \text{якщо} \quad c - \lambda m \leq 0, \\ \frac{e^{s_-u}(\lambda m - c)}{k'(s_-)} + o\left(e^{s_-u}\right), & \text{якщо} \quad c - \lambda m > 0. \end{cases} \tag{17}$$

ЗАВДАННЯ

1. Нехай $W(t), t \geq 0$ є вінерівським процесом. Показати, що наступні процеси також є вінерівськими процесами:
   a) $-W(t)$;
   b) $c^{-1}W(c^2t),\ c > 0$;
   c) $tW(1/t)$.
2. Нехай $W(t), t \geq 0$ є процесом Вінера. Використовуючи факт, що $-W(t)$ теж є процесом Вінера, довести, що $\mathbf{P}(W(t) > 0) = 1/2$ для всіх $t > 0$.
3. Нехай $W(t),\ t \geq 0$ є вінерівським процесом. Довести, що $\widehat{W}(t) = (1+t)W(\frac{1}{1+t}) - W(1),\ t \geq 0$ також є вінерівським процесом.
4. Довести, що $\mathbf{M}W^4(t) = 3t$.
5. Довести, що $\mathbf{M}W(t)W(s) = \min(t,s)$.

   Вказівка. *Нехай $t > s$. Треба використати тотожність:*

   $$W(t)W(s) = [W(t) - W(s)]W(s) + W^2(s).$$

6. Довести, що $\lim_{t\to\infty} \frac{W_t}{t} = 0$ за ймовірністю.
7. Довести, що функція $p(t,x) = \frac{d\mathbf{P}(W(t)<x)}{dx} = \frac{1}{\sqrt{2\pi t}}exp[-x^2/2t]$ задовільняє рівнянню
   $$\frac{\partial p(t,x)}{\partial t} = \frac{1}{2}\frac{\partial^2 p(t,x)}{\partial x^2}.$$

8. Довести наступне, якщо

$$k(s) = \frac{\sigma^2 s^2}{2} + sc + \lambda \int_0^\infty (e^{-sx} - 1) dF(x),$$

а

$$k_n(s) = sc_n + \lambda_n \int_0^\infty (e^{-sx} - 1) dF_n(x),$$

де

$$c_n = c + \frac{3n\sigma^2}{2}, \quad \lambda_n = \lambda + 3n^2\sigma^2,$$

$$F_n(x) = \frac{\lambda F(x) + 3n^3\sigma^2 \quad xI\{0 \le x \le n^{-1}\} + n^{-1}I\{x > n^{-1}\}}{\lambda + 3n^2\sigma^2},$$

тоді $k_n(s) \to k(s)$, якщо $n \to \infty$ для кожного фіксованого $s > 0$.

Лекція 14. Елементи теорії відновлення 1

14.1 Процес відновлення

Нехай $\xi_1, \xi_2, ...$ є незалежними однаково розподіленими невід'ємними випадковими величинами з функцією розподілу $F(x)$, і $S_n = \sum_{i=1}^{n} \xi_i$, $n \geq 1$, $S_0 = 0$. Будемо припускати, що $F(0) < 1$, тому, що інакше $\mathbf{P}(S_n = 0) = 1$ для всіх $n = 1, 2, ...$

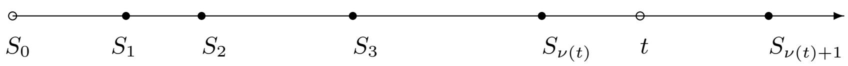
Визначення 1. *Випадковий процес*

$$\nu(t) = \max\{n \geq 0 : S_n \leq t\} \tag{1}$$

називається процесом відновлення.

Зауваження 1. Якщо $F(x) = 1 - e^{-\lambda x}, \lambda > 0,\ x \geq 0$, а, значить, випадкові величини $\xi_1, \xi_2, ...$ мають показниковий розподіл, тоді $\nu(t)$ буде рахуючим процесом $N(t)$ з лекції 4, тобто процесом Пуассона.

Подібно тому, як це було в лекції 4, суми S_n, $n = 1, 2, ...$ можна представити як точки на чисельній вісі. Тоді $\nu(t)$ є числом сум S_n, які помістились в інтервалі $[0, t]$.



Мал. 1

Назву „процес відновлення" можна витлумачити наступним чином. Представимо собі, що певна система складається з одного приладу і час його безаварійної роботи триває ξ_1. Система починає працювати в момент часу $t = 0$. Потім в момент часу $S_1 = \xi_1$ настає відмова приладу, тобто перша аварія системи. У ту саму мить відбувається миттєва заміна зіпсутого приладу на ідентичний, і система продовжує функціонування спочатку. Нехай ξ_2 є часом безаварійної роботи другого приладу. Позаяк зіпсутий прилад замінюється на ідентичний, тоді природним є припущення, що випадкові величини ξ_1, ξ_2 є незалежними і мають однаковий розподіл. Іншими словами від моменту S_1 система функціонує точнісінько так само, як і від моменту $t = 0$. Моменти S_n, $n = 1, 2, ...$ називаються *моментами відновлення.* Зазвичай момент $t = 0$ долучається до множини моментів відновленя, отже S_n, $n = 0, 1, 2, .., S_0 = 0$ складають всі моменти відновлення процесу $\nu(t)$.

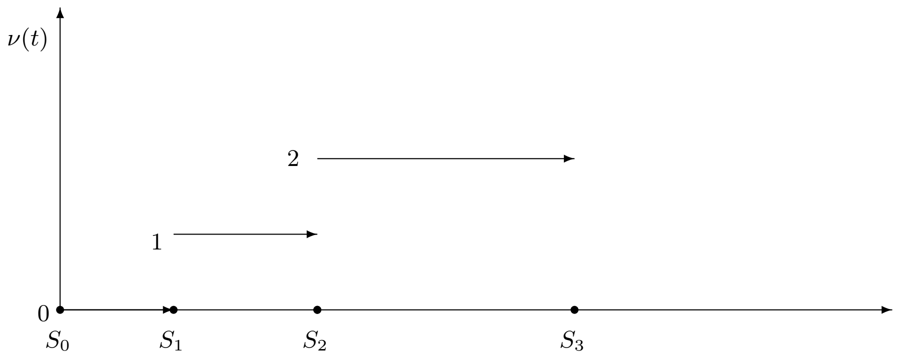
Звернемо тут увагу на один факт.

Зауваження 2. Якщо $F(0) = p > 0$, тоді ймовірність того, що станеться подія $\{S_n = S_{n+1}\}$, не дорівнює нулеві (завдання 1.), отже моменти відновлення S_n, S_{n+1} могли би накладатись. Проте нам буде зручніше уникати такої ситуації, ось чому будемо припускати, що $F(0) = 0$. Як випливає з завдань 2. і 3., найважливіші результати теорії відновлення залишаються в силі й без такого припущення.

Якщо $I\{A\}$ дорівнює 1 або 0 в залежності від того, чи сталась подія A чи ні, тоді нескладно зрозуміти, що

$$\nu(t) = \sum_{n=1}^{\infty} I\{S_n \leq t\} = \sum_{n=1}^{\infty} I\{\sum_{k=1}^{n} \xi_k \leq t\}. \tag{2}$$

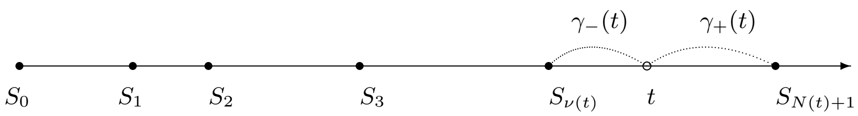
Наступний графік представляє приклад траєкторії процесу відновлення



Мал. 2

З цього графіку випливає, що послідовність сум S_n, $n = 0, 1, 2, ...$ однозначно визначає процес $\nu(t)$ і навпаки. (Зауважимо, що у випадку $F(0) > 0$ так би не було). Така інтерпретація є зручною, тому часто будемо нею користуватися.

Якщо розподіл випадкових величин ξ_i, $i = 1, 2, ...$ є показниковим, тоді $\nu(t)$ є процесом Пуассона (див. лекцію 4), але в загальному випадку процес $\nu(t)$ не буде марковським. Це випливає з того, що коли позначити $\gamma_+(t) = S_{\nu(t)+1} - t$, $\gamma_-(t) = t - S_{\nu(t)+1}$ (перескок моменту t і недоскок до моменту t послідовності S_n, $n = 0, 1, 2, ..$, див. мал. 3), тоді у випадку, коли розподіл випадкових величин ξ_i, $i = 1, 2, ...$ не є показниковим, розподіл випадкової величини $\gamma_+(t)$ залежить від випадкової величини $\gamma_-(t)$. Таким чином, починаючи з моменту t, ймовірносні характеристики процесу $\nu(t)$ залежать від історії цього процесу до моменту t, що означає відсутність властивості Маркова у процесу $\nu(t)$.



Мал. 3

З іншого боку з конструкції процесу відновлення легко випливає (завдання 4.), що для довільних натуральних $n < m$ і $s_1 < s_2 < ... < s_{n-1} < s < t$ має місце

$$\mathbf{P}(\nu(t){=}m/S_1{=}s_1, S_2{=}s_2, ..., S_{n-1}{=}s_{n-1}, S_n{=}s){=}\mathbf{P}(\nu(t-s){=}m-n). \tag{3}$$

Це означає, що хоча процес $\nu(t)$ не є марковським, проте він у моменти S_n забуває свою історію до цих моментів. У такому випадку будемо говорити, що S_n є *марковськими моментами* процесу $\nu(t)$. Опріч того з (3) випливає однорідність процесу $\nu(t)$, але тільки стосовно моменту S_n, тобто

$$\mathbf{P}(\nu(t+s)=m/S_n=s)=\mathbf{P}(\nu(t)=m-n), \qquad t,s,\geq 0, \quad n\leq m.$$

14.2 Функція відновлення

Визначення 2. *Функція*

$$H(t)=1+\mathbf{M}\nu(t) \tag{4}$$

називаєтьcя функцією відновлення.

Як вже зазначалося раніше, момент $t=0$ долучаємо до множини моментів відновлення процесу $\nu(t)$. Це обумовлює появу одиниці у визначенні (4) і разом з тим дозволяє надати дуже простий сенс функції $H(t)$, яким будемо користуватися у подальшому: $H(t)$ є середнім числом моментів відновлення процесу $\nu(t)$ в інтервалі $[0;t]$.

Позаяк для $n\geq 1$ маємо $\mathbf{M}I\{\sum_{k=1}^{n}\xi_k\leq t\}=\mathbf{P}(\sum_{k=1}^{n}\xi_k\leq t)=F^{*n}(t)$ і $\mathbf{M}I\{\sum_{k=1}^{0}\xi_k\leq t\}=1$, отже з виразів (2) і (4) випливає

$$H(t)=1+\sum_{n=1}^{\infty}F^{*n}(t)=\sum_{n=0}^{\infty}F^{*n}(t), \tag{5}$$

де з визначення $F^{*0}(t)=1,\ t\geq 0$ і $F^{*0}(t)=0,\ t<0$.

Теорема 1. *Функція відновлення $H(t)$ має наступні властивості:*

1) $H(0)=1$ *і* $H(t)<\infty$ *для довільного* $t>0$.
2) *Функція* $H(t)$ *є неспадною і* $H(t+s)\leq H(t)+H(s)$.
3) $H(t)=\frac{t}{m}+o(t)$ *для* $t\to\infty$, *де* $m=\mathbf{M}\xi_1=\int_0^{\infty}xdF(x)$.
4) *Має місце*

$$H(t)=1+\int\limits_0^t H(t-x)dF(x), \qquad t\geq 0. \tag{6}$$

Зауваження 3. Згортку $\int_0^t g(t-x)d\varphi(x)$ часто є зручним записувати у символічному вигляді як $g*\varphi(t)$. Тоді (6) можемо записати у наступному символічному вигляді: $H(t)=1+H*F(t),\ t\geq 0$.

Д о в е д е н н я. 1) Той факт, що $H(0)=1$, є очевидним. Доведемо тепер, що $H(t)<\infty$ для кожного $t>0$. Нехай $t>0$ є фіксованим. Якщо $F(t)=q<1$, тоді

$$F^{*n}(t)=\mathbf{P}(\sum_{k=1}^{n}\xi_k\leq t)\leq\mathbf{P}(\xi_1\leq t,\xi_2\leq t,....,\xi_n\leq t)=\prod_{k=1}^{n}\mathbf{P}(\xi_k\leq t)=F^n(t)=q^n,$$

отже

$$H(t)=1+\sum_{n=1}^{\infty}F^{*n}(t)\leq 1+\sum_{n=1}^{\infty}q^n=\frac{1}{1-q}<\infty.$$

нехай тепер $F(t) = 1$. Згідно Закону великих чисел будемо мати: $n^{-1}\sum_{k=1}^{n}\xi_k \xrightarrow[n\to\infty]{p.n.}$ $\mathbf{M}\xi_1 > 0$, тому

$$F^{*n}(t) = \mathbf{P}(\sum_{k=1}^{n}\xi_k \le t) = \mathbf{P}(n^{-1}\sum_{k=1}^{n}\xi_k \le t/n) \underset{n\to\infty}{\longrightarrow} \mathbf{P}(\mathbf{M}\xi_1 \le 0) = 0. \qquad (7)$$

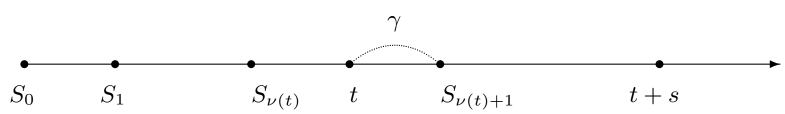
У зв'язку з чим для певного n_0 маємо $F^{*n_0}(t) = q < 1$. Нехай $U(t)$ є функцією відновлення, яка відповідає функції розподілу $G(x) = F^{*n_0}(x)$. Позаяк $G(t) = F^{*n_0}(t) = q < 1$, тоді $U(t) < \infty$. Тепер будемо мати:

$$H(t) = \sum_{n=0}^{\infty}F^{*n}(t) = \sum_{n=0}^{n_0-1}F^{*n}(t) + \sum_{n=n_0}^{2n_0-1}F^{*n}(t) + \sum_{n=2n_0}^{3n_0-1}F^{*n}(t) + ... =$$

$$= \sum_{n=0}^{n_0-1}F^{*n}(t) + \sum_{n=0}^{n_0-1}F^{*n}*G(t) + \sum_{n=0}^{n_0-1}F^{*n}*G^{*2}(t) + ... = \sum_{n=0}^{n_0-1}F^{*n}*\sum_{n=0}^{\infty}G^{*n}(t) =$$

$$= \sum_{n=0}^{n_0-1}F^{*n}(t)*U(t) = \sum_{n=0}^{n_0-1}\int_0^t F^{*n}(t-v)dU(v) \le \sum_{n=0}^{n_0-1}\int_0^t dU(v) = n_0U(t) < \infty.$$

2) У свою чергу факт, що $H(s) \le H(t)$ для $s < t$, теж є очевидним. Нехай $\nu(A)$ означає число відновлень у множині $A \subset [0,\infty]$. Зрозуміло, що $\nu(A) \le \nu(B)$, якщо $A \subset B$ і $H(t) = \mathbf{M}\nu([0,t])$.



Мал. 4

Нехай за визначенням $\nu((a,b]) = 0$, якщоі $a > b$. Наступна рівність є очевидною (див. мал. 4)

$$\nu([0,t+s]) = \nu([0,t]) + \nu([t+\gamma,t+s]). \qquad (8)$$

Будемо мати

$$\mathbf{M}\nu([t+\gamma,t+s]) = \int_0^s \mathbf{M}\{\nu([t+\gamma,t+s])/\gamma = v\}d\mathbf{P}(\gamma \le v) =$$

$$= \int_0^s \mathbf{M}\{\nu([t+v,t+s])/\gamma = v\}d\mathbf{P}(\gamma \le v) = \int_0^s H(s-v)d\mathbf{P}(\gamma \le v) \le$$

$$\le H(s)\int_0^s d\mathbf{P}(\gamma \le v) = H(s)\mathbf{P}(\gamma \le s) \le H(s). \qquad (9)$$

У (9) ми використали той факт, що за умови $\gamma = v \le s$ момент $t + v = S_{\nu(t)}$ є марковським моментом процесу відновлення. Тоді

$$\mathbf{M}\{\nu([t+v,t+s])/\gamma = v\} = \mathbf{M}\nu([0,s-v]) = H(s-v) \le H(s).$$

Тепер з (8) отримаємо:

$$H(t) = \mathbf{M}\nu([0, t+s]) = \mathbf{M}\nu([0, t]) + \mathbf{M}\nu([t+\gamma, t+s]) \le H(t) + H(s).$$

3) Нехай $c > 0$ є фіксованою і $t = cn + h$, де $0 \le h < c$. Вочевидь, $n \to \infty$, якщо $t \to \infty$. Застосовуючи наслідки пункту 2), отримаємо:

$$H(t) = H(cn+h) \le H(cn) + H(h) \le nH(c) + H(h),$$

звідки

$$\overline{\lim_{t\to\infty}} \frac{H(t)}{t} = \overline{\lim_{n\to\infty}} \frac{H(cn+h)}{cn+h} \le \overline{\lim_{n\to\infty}} \frac{nH(c)+H(h)}{cn+h} = \frac{H(c)}{c}.$$

Отже, для довільного $c > 0$ отримаємо:

$$\overline{\lim_{t\to\infty}} \frac{H(t)}{t} \le \frac{H(c)}{c}.$$

Якщо у правій частині цієї нерівності перейти до границі $\underline{\lim}_{c\to\infty}$, тоді будемо мати:

$$\overline{\lim_{t\to\infty}} \frac{H(t)}{t} \le \underline{\lim_{c\to\infty}} \frac{H(c)}{c}. \tag{10}$$

Позаяк завжди $\overline{\lim} \ge \underline{\lim}$, тоді в (10) має місце рівність. А, отже існує границя $\lim_{t\to\infty} t^{-1}H(t) = \alpha$. Знадемо тепер α. Легко перевірити, що для $s > 0$ настає рівність (завдання 4.)

$$\int_0^\infty e^{-st} H(t)dt = \frac{1}{s(1-\varphi(s))}, \tag{11}$$

де $\varphi(s) = \mathbf{M}e^{-s\xi_1} = \int_0^\infty e^{-sx} dF(x)$. Заміна змінної $st = v$ дозволяє переписати рівність (11) у вигляді:

$$\int_0^\infty e^{-v} v \frac{H(v/s)}{v/s} dv = \frac{s}{1-\varphi(s)}. \tag{12}$$

Позаяк для кожного фіксованого v має місце

$$\lim_{s\to 0} \frac{H(v/s)}{v/s} = \alpha,$$

а $\int_0^\infty e^{-v} v dv = 1$, тоді перехід до границі $\lim_{s\to 0}$ в (12) дає

$$\alpha = \frac{1}{-\varphi'(0)} = \frac{1}{\mathbf{M}\xi_1}.$$

Доведення пункту 4) є нескладним і ми залишаємо його читачеві. ◀

ЗАВДАННЯ

1. Нехай $F(0) = p \in (0,1)$ і $\zeta = \min\{k : S_0 = S_k\}$. Довести, що $\mathbf{P}(\zeta = k) = p^k(1-p)$.

2. Нехай $F(0) = p \in (0,1)$ і $H(t)$ є функцією відновлення, яка відповідає функції розподілу $F(x)$. Довести, що $H(0) = (1-p)^{-1}$, і рівняння для $H(t)$ набуває вигляду [1]

$$H(t) = 1 + \int\limits_{-0}^{t} H(t-x)dF(x), \qquad t \geq 0. \tag{13}$$

3. Нехай $F(0) = p \in (0,1)$ і $H(t)$ є функцією відновлення, яка відповідає функції розподілу $F(x)$. Довести, що коли $\widehat{H}(t)$ є функцією відновлення, яка відповідає функції розподілу $\widehat{F}(t) = (F(t)-p)/(1-p)$ (а, отже $\widehat{F}(0) = 0$), тоді $\widehat{H}(t) = (1-p)H(t)$.

   Вказівка. *Насамперед скоритатися з того, що коли функція розподілу $F(x)$ має атом у точці $t = 0$, тоді рівняння для $H(t)$ набуває вигляду* (13). *Далі переписати це рівняння для функції розподілу $\widehat{F}(t)$.*

4. Довести рівність (11).

5. Показати, що для показникового розподілу з функцією розподілу $F(t) = 1 - e^{-\lambda t}, \quad \lambda > 0,\ t \geq 0$ функція відновлення набуває вигляду: $H(t) = 1 + \lambda t$.

6. Нехай $F(x)$ відповідає рівномірному розподілу на інтервалі $[0,1]$. Довести, що

$$H(t) = \sum_{k=0}^{n} (-1)^k e^{t-k} \frac{(t-k)^k}{k!}, \qquad n \leq t \leq n+1.$$

7. Для яких a функція $1 + t + a(1 - e^{-t})$ може бути функцією відновлення?

8. Нехай $p(t,z) = \mathbf{M} z^{\nu(t)}$. Довести, що

$$\int\limits_{0}^{\infty} e^{-st} p(t,z)dt = \frac{1-\varphi(s)}{s(1-z\varphi(s))},$$

   де $\varphi(s) = \mathbf{M}e^{-s\xi_1}$.

9. Випадкова величина ξ_i має розподіл Ерланга з параметрами k, λ, тобто щільність розподілу має вигляд $\frac{\lambda(\lambda x)^{k-1}e^{-\lambda x}}{(k-1)!}$, $x \geq 0$. Знайти розподіл процесу відновлення $\nu(t)$.

   Вказівка. *Викоритати той факт, що випадкова величина, яка має розподіл Ерланга з параметрами k, λ може бути записана у вигляді суми k незалежних показниково розподілених з параметром λ випадкових величин.*

---

[1] Інтеграл $\int_{-0}^{t} g(x)dF(x)$ означає, що можливий стрибок функції розподілу $F(x)$ в точці $t = 0$ береться до уваги, у той же час інтеграл $\int_0^t g(x)dF(x) = \int_{0+}^t g(x)dF(x)$ цього стрибка не враховує. Іншими словами, $\int_{-0}^t g(x)dF(x) = \int_0^t g(x)dF(x) + g(0)F(0)$.

Лекція 15. Елементи теорії відновлення 2

15.1 Рівняння відновлення і вузлова теорема відновлення

Визначення 1. *Рівняння*

$$W(t) = z(t) + \int_0^t W(t-x)dF(x), \tag{1}$$

де функція $z(t)$ є обмеженою на кожному скінченному інтервалі, називаеться рівнянням відновлення.

Роз'язок рівняння відновлення шукаємо в класі функцій, обмежених на кожному скінченному інтервалі. Нехай $H(t)$ є функцією відновлення, яка відповідає функції розподілу $F(x)$.

Теорема 1. *Рівняння відновлення має розв'язок вигляду:*

$$W(t) = \int_0^t z(t-x)dH(x), \qquad t \geq 0. \tag{2}$$

Цей розв'язок є єдиним в класі функцій, обмежених на кожному скінченному інтервалі.

Д о в е д е н н я. Доведемо, що функція (2), яка, вочевидь, є обмеженою на кожному скінченному інтервалі, є розв'язком рівняння відновлення. З цією метою запишемо рівняння відновлення і вираз для функції $W(t)$ в символічному вигляді, тобто, $W(t) = z(t) + W * F(t)$ і $W(t) = z * H(t)$. Тепер, якщо взяти до уваги рівняння (6) зі стор. 378 і зауваження 3, отримаємо:

$$z(t){+}W * F(t){=}z(t) + z * H * F(t){=}z * \ \ 1(t) + H * F(t) \ \ {=}z * H(t){=}W(t),$$

де $1(t)$ означає функцію таку, що $1(t) = 1,\ t \geq 0$ і $1(t) = 0,\ t \leq 0$.

Якщо рівняння (1) має два розв'язки $W_1(t)$ і $W_2(t)$, тоді функція $U(t) = W_1(t) - W_2(t)$ є розв'язком рівняння

$$U(t) = \int_0^t U(t-x)dF(x),$$

звідки отримаємо

$$U(t) = U * F(t) = U * F^{*2}(t) = ... = U * F^{*n}(t) \tag{3}$$

для довільного натурального n. Нехай $T > 0$ є фіксованим і $C = \sup_{0\le x\le T} |U(x)| < \infty$. Z (3) для $0 \le t \le T$ будемо мати

$$|U(t)| = \int\limits_0^t U(t-x)dF^{*n}(x) \le \int\limits_0^t |U(t-x)|dF^{*n}(x) \le CF^{*n}(T) \underset{n\to\infty}{\longrightarrow} 0.$$

Останнє відношення випливає з (7), стор. 379, тому $U(t) \equiv 0$ для $0 \ge t \ge T$ і довільного $T \ge 0$, а, отже $W_1(t) \equiv W_2(t),\ t \ge 0$. ◀

Однією з найважливіших проблем теорії відновлення є дослідження поведінки функції відновлення $H(t)$ і поведінки розв'язку рівняння відновлення $W(t)$ при $t \to \infty$. Що до функції $H(t)$, то пункт 3) твердження 1 (стор. 378) дає частково на це відповідь. Позаяк згідно з цим пунктом $H(t) \sim t/m$, якщо $t \to \infty$[1], де $m = \mathbf{M}\xi_1$, тому формальний розв'язок дає

$$W(t)=\int\limits_0^t z(t-x)dH(x)\sim\frac{1}{m}\int\limits_0^t z(t-x)dx=\frac{1}{m}\int\limits_0^t z(x)dx \underset{t\to\infty}{\longrightarrow} \frac{1}{m}\int\limits_0^\infty z(x)dx. \tag{4}$$

З цього випливає, що для існування границі $\lim_{t\to\infty} W(t)$ необхідним є інтегрованість функції $z(t)$ на піввісі $[0,\infty)$, і що границя $\lim_{t\to\infty} W(t)$ правдоподібно має такий самий вигляд як в (4). Зараз нам знадобиться інше визначення інтеграла, подібне до класисного визначення інтеграла Рімана.

Нехай функція $f(x)$ є визначеною для $x \ge 0$ і обмеженою на кожному скінченному інтервалі $[0,b]$, $0 \le b < \infty$, і нехай $0 = a_0(n) < a_1(n) < ... < a_n(n) < a_i(n) < ....$, $n = 1, 2, ...$ буде розбиттям піввісі $[0,\infty)$. Позначимо

$$\Delta_i(n) = a_{i+1}(n) - a_i(n), \quad \Delta(n) = \sup_i \Delta_i(n),$$
$$m_i^-(n) = \inf\{f(x) : a_i(n) \le x < a_{i+1}(n))\},$$
$$m_i^+(n) = \sup\{f(x) : a_i(n) \le x < a_{i+1}(n))\}.$$

Нехай $S_n^\pm(f) = \sum_{i=0}^\infty m_i^\pm(n)\Delta_i(n)$.

Зробимо припущення, що $\Delta(n) \to 0$, якщо $n \to \infty$.

Визначення 2. Якщо

$$\lim_{n\to\infty} S_n^+(f) = \lim_{n\to\infty} S_n^-(f) \in (-\infty,\infty),$$

тоді будемо говорити, що функція $f(x)$ є *безпосередньо інтегрована за Ріманом* на піввісі $[0,\infty)$. Цю спільну границю будемо позначати так само, як класичний інтеграл Рімана, тобто $\int\limits_0^\infty f(x)dx$.

Зауваження 1. Класичний інтеграл Рімана на *скінченному* інтервалі визначається так само, як ми щойно зробили, але це визначення стосується тільки скінченного інтервалу $[0,b]$. У свою чергу класичний інтеграл Рімана на піввісі $[0,\infty)$ визначається як границя

$$\int\limits_0^\infty f(x)dx = \lim_{b\to\infty}\int\limits_0^b f(x)dx. \tag{5}$$

[1] Як завжди $f(t) \sim g(t), t \to \infty$, якщо $f(t)/g(t) \underset{t\to\infty}{\longrightarrow} 1$.

Якщо ця границя є скінченною, тоді кажуть, що функція $f(x)$ є інтегрованою в сенсі Рімана на піввісі $[0, \infty)$. Можна довести, що коли функція $f(x)$ є безпосередньо інтегрованою в сенсі Рімана на піввісі $[0, \infty)$, тоді вона також є інтегрованою в сенсі Рімана на цій піввісі. Зворотне твердження не є правдивим (завдання 1). Проте легко довести, що коли функція $f(x)$ є інтегрованою в сенсі Рімана на піввісі $[0, \infty)$ і є монотонною з певного x, тоді вона є безпосередньо інтегрованою в сенсі Рімана на цій піввісі (завдання 2).

Наступне твердження, відоме як *вузлова теорема теорії відновлення* або *головна теорема теорії відновлення*, наведемо без доведення.

Теорема 2. *Нехай функція* $z(t)$ *є безпосередньо інтегрованою в сенсі Рімана на піввісі* $[0, \infty)$, *а розподіл* $F(x)$ *не є гратчастим* [1]. *Тоді*

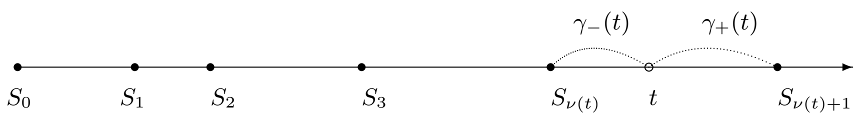
$$\lim_{t\to\infty} \int_0^t z(t-x)dH(x) = \frac{1}{m}\int_0^\infty z(t)dt.$$

15.2 Застосування теорії відновлення

Для процесу відновлення $\nu(t) = \max\{n \geq 0 : S_n = \sum_{i=1}^n \xi_i \leq t\}$ позначимо

$$\gamma_+(t){=}S_{\nu(t)+1} - t, \quad \gamma_-(t){=}t - S_{\nu(t)}, \quad \gamma(t) = \gamma_-(t) + \gamma_+(t){=}S_{\nu(t)+1} - S_{\nu(t)}.$$

Вочевидь, $S_{\nu(t)}$ є останнім відновленням перед моментом t, а $S_{\nu(t)+1}$ є першим відновленням після моменту t. Подібно до того, як це було в лекції 4, коли ми досліджували процес Пуассона, випадкову величину $\gamma_+(t) = S_{\nu(t)+1} - t$ будемо називати перескоком моменту t послідовністю $S_n, n = 0, 1, 2, ..$, а випадкову величину $\gamma_-(t) = t - S_{\nu(t)+1}$ – недоскоком до моменту t (див. мал. 1). Випадкова величина $\gamma(t) = \gamma_-(t) + \gamma_+(t) = S_{\nu(t)+1} - S_{\nu(t)}$ є довжиною інтервалу, який містить точку t.



Мал. 1

Випадкові величини $\gamma_\pm(t), \gamma(t)$ мають прості інтерпретації на практиці. Якщо процес відновлення моделює заміну зіпсутого приладу на справний, тоді випадкова величина $\gamma_+(t)$ є часом від моменту t до наступної аварії приладу, а $\gamma_-(t)$ є часом від останньої аврії до моменту t. Нашим завданням тепер буде дослідження розподілів цих випадкових величин.

Нехай $F(x) = \mathbf{P}(\xi \leq x)$ і $H(t)$ є функцією відновлення, яка відповідає розподілу $F(x)$. Позначимо

$$\varphi_\pm(t, x) = \mathbf{P}(\gamma_\pm(t) > x), \qquad \varphi(t, x) = \mathbf{P}(\gamma(t) > x).$$

[1] Див. визначення 3 далі.

Лема 1. *Для* $t, x \geq 0$ *виконуються рівняння:*

$$\varphi_+(t,x) = 1 - F(t+x) + \int_0^t \varphi_+(t-u,x)dF(u), \tag{6}$$

$$\varphi_-(t,x) = \ 1 - F(t)\ I\{t \geq x\} + \int_0^t \varphi_-(t-u,x)dF(u), \tag{7}$$

$$\varphi(t,x) = 1 - F(t \vee x) + \int_0^t \varphi(t-u,x)dF(u), \tag{8}$$

де $a \vee b = \max\{a, b\}$.

Д о в е д е н н я. Доведемо тільки (6), доведення (7), (8) залишаємо читачеві (завдання 3).

Для $\mathbf{P}(\gamma_+ > x)$ запишемо формулу повної ймовірності за умови першого стрибка процесу відновлення $\nu(t)$, тобто стосовно події $\{S_1 = u\} = \{\xi_1 = u\}$, $u \geq 0$. Тоді будемо мати

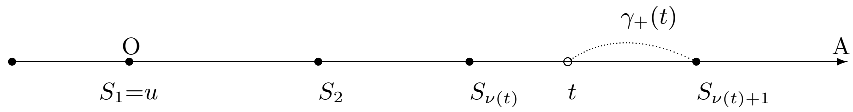
$$\mathbf{P}(\gamma_+(t) > x) = \int_0^\infty \mathbf{P}(\gamma_+(t) > x/\xi_1 = u)dF(u) =$$

$$= \int_0^t \mathbf{P}(\gamma_+(t) > x/\xi_1 = u)dF(u) + \int_t^\infty \mathbf{P}(\gamma_+(t) > x/\xi_1 = u)dF(u). \tag{9}$$

Якщо $u \geq t$, тоді реалізація події $\{\xi_1 = u\}$ означає, що стрибок стався після t і $\gamma_+(t) = u - t$. Отже, $\mathbf{P}(\gamma_+(t) > x/\xi_1 = u) = \mathbf{P}(u - t > x/\xi_1 = u) = I\{u > t + x\}$ і

$$\int_t^\infty \mathbf{P}(\gamma_+(t){>}x/\xi_1{=}u)dF(u) = \int_t^\infty I\{u{>}t{+}x\}dF(u){=}\int_{t+x}^\infty dF(u) = 1{-}F(t{+}x). \tag{10}$$

Якщо $u < t$, тоді реалізація події $\{\xi_1 = u\}$ означає, що перший стрибок процесу відновлення $\nu(t)$ був до моменту t. Від моменту u процес розпочинається знову (завдяки тому, що цей момент є моментом відновлення) і реалізується вже на півпрямій OA (див. мал. 2).



Мал. 2

Якщо позначити цей новий процес через $\widehat{\nu}(t)$, тоді стрибок після моменту t буде вже реалізований $\widehat{\nu}(t)$, але для процесу $\widehat{\nu}(t)$ момент t знаходиться на відстані $t - u$ від моменту O, отже $\gamma_+(t) = \widehat{\gamma}_+(t-u)$, де $\widehat{\gamma}_+(t-u)$ означає стрибок після моменту $t - u$ процесом $\widehat{\nu}(t)$, який розглядається на півпрямій OA. Як ми вже зазначали, момент $t = u$ є моментом відновлення, тоді ймовірносні характеристики процесів $\nu(t)$ і $\widehat{\nu}(t)$ є ідентичними, тому $\mathbf{P}(\widehat{\gamma}_+(t-u) > x) = \mathbf{P}(\gamma_+(t-u) > x)$. Будемо мати

$$\int_0^t \mathbf{P}(\gamma_+(t) > x/\xi_1{=}u)dF(u) = \int_0^t \mathbf{P}(\widehat{\gamma}_+(t-u) > x)dF(u) =$$

$$= \int_0^t \mathbf{P}(\gamma_+(t-u) > x)dF(u) = \int_0^t \varphi(t-u, x)dF(u). \tag{11}$$

Врешті з (9), (10) і (11) отримаємо (6). ◀

Наступне твердження випливає з леми 1 і твердження 1 в очевидний спосіб.

Теорема 3. *Мають місце відношення*

$$\mathbf{P}(\gamma_+(t) > x) = \int_0^t \left(1 - F(t-v+x)\right) dH(v), \tag{12}$$

$$\mathbf{P}(\gamma_-(t) > x) = \int_0^t \left(1 - F(t-v)\right) I\{t-v \geq x\}dH(v), \tag{13}$$

$$\mathbf{P}(\gamma(t) > x) = \int_0^t (1 - F((t-v) \vee x)dH(v). \tag{14}$$

Нехай $F(x)$ є функцією розподілу показникового розподілу з параметром λ. Тоді $H(v) = 1 + \lambda v$ (див. завдання 5, стор. 381) і з виразу (12) отримаємо:

$$\mathbf{P}(\gamma_+(t){>}x) = \int_0^t e^{-\lambda(t-v+x)}d(1+\lambda v){=}e^{-\lambda(t+x)}{+}\lambda e^{-\lambda(t+x)}\int_0^t e^{-\lambda v}dv{=}e^{-\lambda x}.$$

Для випадкової величини $\gamma_-(t)$ будемо мати:

$$\mathbf{P}(\gamma_-(t){>}x) = \int_0^t e^{-\lambda(t-v)}I\{t-v \geq x\}d(1+\lambda v){=}e^{-\lambda t}I\{t \geq x\}+$$

$$+\lambda\int_0^t e^{-\lambda v}I\{v \geq x\}dv = \begin{cases} e^{-\lambda x}, & \text{якщо} \quad 0 \leq x < t, \\ 0, & \text{якщо} \quad x \geq t. \end{cases}$$

Аналогічно для випадкової величини $\gamma(t) = \gamma_-(t) + \gamma_+(t)$

$$\mathbf{P}(\gamma(t){>}x) = \int_0^t e^{-\lambda\left((t-v)\vee x\right)}d(1+\lambda v){=}e^{-\lambda(t\vee x)} + \lambda\int_0^t e^{-\lambda(v\vee x)}dv =$$

$$= \begin{cases} e^{-\lambda x}(x\lambda + 1), & \text{якщо} \quad 0 \leq x < t \\ e^{-\lambda x}(t\lambda + 1), & \text{якщо} \quad x \geq t. \end{cases}$$

Отриманий розподіл випадкових величин $\gamma_\pm(t), \gamma(t)$ у випадку показникового розподілу співпадає з результатами, отриманими в твердженні 1 (стор. 310), застосовуючи інший метод.

З виразів (12)-(14) і твердження 2 після нескладних перетворень отримаємо:

Теорема 4.

$$\lim_{t\to\infty} \mathbf{P}(\gamma_+(t) > x) = \lim_{t\to\infty} \mathbf{P}(\gamma_-(t) > x) = \frac{1}{m}\int_x^\infty 1 - F(v)\ dv,$$

$$\lim_{t\to\infty} \mathbf{P}(\gamma(t) > x) = \frac{x\ 1-F(x)}{m} + \frac{1}{m}\int_x^\infty 1-F(v)\ dv.$$

Нехай $\gamma_\pm(\infty) = \lim_{t\to\infty}\gamma_\pm(t), \gamma(\infty) = \lim_{t\to\infty}\gamma(t)$. Легко підрахувати (завдання 3), що

$$\mathbf{M}\gamma_\pm(\infty) = \frac{\mathbf{M}\xi_1^2}{2m}, \qquad \mathbf{M}\gamma(\infty) = \frac{\mathbf{M}\xi_1^2}{m}. \tag{15}$$

З виразами (15) пов'язаний так званий *парадокс часу очікування*, про який піде мова в наступному параграфі.

15.3 Парадокс часу очікування

Розглянемо наступне завдання. На автобусну зупинку приїжджають автобуси. Інтервали часу між прибуттями є назалежними однаково розподіленими випадковими величинами $\xi_i,\ i = 1, 2, ...$ Припустимо, що людина приходить на цю зупинку у випадковий момент часу. Яким є математичне сподівання її часу очікування на автобус?

Можливими є два розв'язки цього завдання. Назвемо їх „науковим" та „інтуїтивним".

Інтуїтивний розв'язок. Прибуття автобусів будемо трактувати як процес відновлення – моменти їх приїзду на зупинку є моментами відновлення. Нехай $[a, b]$ є інтервалом часу між довільними послідовними відновленнями. Його середня довжина, вочевидь, дорівнює $\mathbf{M}\xi_1$. Позаяк момент приходу людини на зупинку є незалежним від руху автобусів, отже завдання можна переформулювати наступним чином: у випадковий спосіб кидаємо точку на інтервал $[a, b]$. Якою є середня відстань від цієї точки до правого кінця інтервалу $[a, b]$? Позаяк точка кидається рівномірно випадковим чином, тобто вона може потрапити в підінтервали $[a, b]$ однокової довжини з однаковою ймовірністю, середня відстань від місця потрапляння до точки b буде дорівнювати половині довжини інтервалу $[a, b]$ або $\mathbf{M}\xi_1/2$.

Науковий роз'язок. Аналогічно інтуїтивному розв'язку прибуття автобусів моделюємо процесом відновлення. Якщо t є моментом приходу людини на зупинку, тоді $\gamma_+(t)$ буде часом до приїзду чергового автобуса. Позаяк ми припускаємо, що рух автобусів триває вже довго, тоді час очікування людини дорівнює $\gamma_+(\infty)$ і згідно з (15) отримаємо $\mathbf{M}\gamma_+(\infty) = \mathbf{M}\xi_1^2/2m$.

Ми отримали зовсім інший результат. Слушним розв'язком є тільки розв'язок науковий. Інтуїтивний розв'язок є правильним до того місця, де ми приймаємо, що розподіл довжини інтервалу, який накриває момент приходу людини на зупинку, є такий самий як розподіл будь-якого іншого інтервалу між прибуттями автобусів. Але насправді розподіл довжини інтервалу, який накриває момент приходу людини є іншим, бо накриття моменту приходу людини інтервалами більшої довжини є більш ймовірним ніж накриття інтервалами меньшої довжини. Якщо взяти до уваги, як це зроблено в науковому розв'язку, що рух автобусів триває вже довго, тоді ця довжина складає $\gamma(\infty)$. Згідно з (15) маємо $\mathbf{M}\gamma(\infty) = \mathbf{M}\xi_1^2/m$. Решта інтуїтивного розв'язку є слушною. Згідно з ним час очікування автобуса буде дорівнювати половині довжини інтервалу, на який потрапив момент приходу люжини на зупинку або $2^{-1}\mathbf{M}\xi_1^2/m = \mathbf{M}\xi_1^2/2m$, що співпадає з науковим розв'язком.

15.4 Теорія відновлення для гратчастих випадкових величин

Визначення 3. *Розподіл випадкової величини ξ називається гратчастим (арифметичним) з кроком $h > 0$, якщо $\sum_{k=-\infty}^{\infty} \mathbf{P}(\xi = a + kh) = 1$ для певного $a \in R^1$. У цьому випадку будемо казати, що випадкова величина ξ є гратчастою з кроком h.*

Іншими словами гратчаста випадкова величина з кроком $h > 0$ приймає тільки значення з множини $\{a + kh,\ k = 0, \pm 1, \pm 2, ...\}$. Вочевидь, гратчаста випадкова величина є випадковою величиною дискретного типу.

Теорія відновлення для гратчастих розподілів з формальної точки зору є частинним випадком розглянутої перед цим теорії. Насправді, вистачить в якості функції розподілу $F(x)$ взяти функцію розподілу, яка відповідає гратчастій випадковій величині. Одначе у випадку використання дискретних випадкових величин ξ_i ми зустрінемося з пеними особливостями. Це насамперед пов'язано з тим, що в роботі з дискретними випадковими величинами ми не використовуємо функції розподілу $\mathbf{P}(\xi_i \le x)$, а радше самі розподіли $\mathbf{P}(\xi_i = x_k)$. Ось чому, зазвичай, випадок гратчастих випадкових величин досліджується окремо.

Позаяк, використовуючи перетворення $\eta_i = \frac{\xi_i - a}{h}$, гратчаста випадкова величина трансформується у цілочисельну η_i, тоді від самого початку будемо припускати, що випадкові величини ξ_n є незалежними, приймають значення з множини $E = \{0, 1, 2, \}$ і мають спільний розподіл $\mathbf{P}(\xi_i = k) = p_k$. Крім того, будемо передбачати, що НСД$\{n : p_n > 0\} = 1$, тобто розподіл випадкових величин не є зосередженим на множині типу $kd, k = 0, 1, ...$ з $d \ge 2$. Нехай, як раніше, $S_n = \sum_{i=1}^{n} \xi_i,\ n \ge 0$ і $S_0 = 0$.

Визначення процесу відновлення $\nu(t)$ і функції відновлення $H(t)$ залишається таким самим, як у попередній лекції:

$$H(t) = \sum_{n=0}^{\infty} \mathbf{P}(S_n \le t).$$

Позаяк тепер випадкова величина $S_n, n \ge 1$ може приймати тільки значення $0, 1, 2, ...$, тоді зрозуміло, що функція $H(t)$ є сходинковою і $H(n) = H(n + s)$ для $0 \le s < 1$. Отже, фактично достатньо розглянути тільки послідовність $H(n)$, $n = 0, 1, ...$ Тоді

$$H(k) = \sum_{n=0}^{\infty} \mathbf{P}(S_n \le k) = \sum_{n=0}^{\infty} \sum_{i=0}^{k} \mathbf{P}(S_n = i) = \sum_{i=0}^{k} \sum_{n=0}^{\infty} \mathbf{P}(S_n = i) = \sum_{i=0}^{k} h_i, \tag{16}$$

де

$$h_i = \sum_{n=0}^{\infty} \mathbf{P}(S_n = i). \tag{17}$$

Легко показати (завдання 6), що $h_0 = (1 - p_0)^{-1}$. Звідки $H(0) = h_0 = (1 - p_0)^{-1}$. Після чого, якщо $p_0 > 0$, тоді $H(0) = (1 - p_0)^{-1} \neq 1$ на відміну від неперервного випадку.

Можна навести твердження типу твердження 1 з попередньої лекції також для послідовності $H(n)$. Однак у випадку дискретних випадкових величин зазвичай працюємо не $H(n)$, а з послідовністю h_k, яка є визначеною в(17).

Запишемо без доведення властивості послідовності h_k, які є аналогічними відповідним властивостям функції відновлення.

1. Аналогом властивості 3) з твердження 1, стор. 378, є

$$h_k = \frac{1}{\mathbf{M}\xi_1} + o(1), \text{ dla } k \to \infty.$$

2. Аналогом рівняння (6), стор. 378, є

$$h_n = 1 + \sum_{k=0}^{n} h_{n-k} p_k, \qquad n = 0, 1, 2, ...$$

3. Аналогом рівняння відновлення (1) є рівняння

$$w_n = z_n + \sum_{k=0}^{n} w_{n-k} p_k, \qquad n = 0, 1, 2, ..., \tag{18}$$

   де послідовність z_n є відомою і такою, що $|z_n| < \infty$ для кожного $n = 0, 1, 2, ...$

4. Розв'язок рівняння (18) може бути представлений у вигляді:

$$w_n = \sum_{k=0}^{n} z_{n-k} h_k, \qquad n = 0, 1, 2, ... \tag{19}$$

5. Якщо $\sum_{k=0}^{\infty} |z_k| < \infty$, тоді

$$\lim_{n\to\infty} w_n = \frac{1}{\mathbf{M}\xi_1} \sum_{k=0}^{\infty} z_k. \tag{20}$$

Тепер є можливість довести твердження 3 зі сторінки 303. Нехай p_k, $k = 0, 1, 2...$ є ймовірносним розподілом, а b_k, $k = 0, 1, 2...$ – числовою послідовністю такою, що $\sum_{k=0}^{\infty} |b_k| < \infty$.

Тоді, згідно з розглянутим вище пунктом 4), рівняння

$$u_n - \sum_{k=0}^{n} p_k u_{n-k} = b_n, \qquad n \geq 0$$

має єдиний обмежений розв'язок, який може бути записаний у вигляді:

$$u_n = \sum_{k=0}^{n} b_{n-k} h_k, \qquad n = 0, 1, 2, ...$$

Якщо розподіл p_k, $k = 0, 1, 2...$ не є гратчастим, тоді згідно з (20) маємо:

$$\lim_{n\to\infty} u_n = \frac{\sum_{k=0}^{\infty} b_k}{\sum_{k=0}^{\infty} k p_k}.$$

Так само, як в параграфі 2 можна дослідити випадкові величини $\gamma_\pm(n)$, тобто величину перескоку рівня $n \geq 1$ і величину недоскоку цього ж рівня. Подамо лише кінцевий результат, залишаючи доведення читачеві (завдання 9).

$$\lim_{n\to\infty} \mathbf{P}(\gamma_+(n) = k, \gamma_-(n) = l) = \frac{p_{k+l}}{\mathbf{M}\xi_1} = \frac{\mathbf{P}(\xi_1 = k + l)}{\mathbf{M}\xi_1}. \tag{21}$$

ЗАВДАННЯ

1. Навести приклад функції, яка є інтегрованою в сенсі Рімана на пiввiсi $[0, \infty)$, але не є інтегрованою безпосередньо в сенсі Рімана на цій півісі.
2. Довести, що коли функція $f(x)$ є інтегрованою в сенсі Рімана на пiввiсi $[0, \infty)$ і є монотонною з певного x, тоді вона є безпосередньо інтегрованою в сенсі Рімана на цій пiввiсi.
3. Довести вирази (7), (8), (15).
4. Довести, що
$$\lim_{t\to\infty} \mathbf{P}(\gamma_+(t) > x, \gamma_-(t) > y) = \frac{1}{m} \int\limits_{x+y}^{\infty} 1 - F(v)\ dv.$$
5. Довести, що
$$H(t + s) - H(t) = \frac{s}{\mathbf{M}\xi_1} + o(1),\ \text{dla}\ \ s > 0\ \text{jeњli}\ \ t \to \infty.$$
6. Довести, що $h_0 = (1 - p_0)^{-1}$.
7. Нехай $p(z) = \mathbf{M}z^{\xi_1}$, $|z| < 1$. Довести, що
$$\sum_{n=0}^{\infty} z^n h_n = \frac{1}{1 - p(z)}, \qquad \sum_{n=0}^{\infty} z^n H(n) = \frac{1}{(1 - z)(1 - p(z))}.$$
8. Знайти функцію відновлення $H(n)$ і h_n, якщо $p_k = \mathbf{P}(\xi_i = k) = p^k(1 - p)$, $k = 0, 1, 2, ...$

   Вказівка. *Скористатися результатами попереднього завдання.*
9. Довести вираз (21) і знайти $\lim_{n\to\infty} \mathbf{P}(\gamma_\pm(n) = k)$.
10. Знайти $\lim_{n\to\infty} \mathbf{P}(\gamma(n) = \gamma_+(n) + \gamma_-(n) = k)$.

Бібліографія

[1] Боровков А.А. Математическая статистика М.: Наука 1984.- 472 с.

[2] Гантмахер Ф.Р. Теория матриц М.: Наука, 1966 с.

[3] Гаек Я., Шидак З., Теория ранговых критериев.-Наука,1971.-376 с.

[4] И.И. Гихман, А.В. Скороход, М.И.Ядренко Теория вероятностей и математическая статичтика.- К.: Вища школа, 1979.- 408 с.

[5] Б.В. Гнеденко Курс теории вероятностей.- М., Наука, 1965.-400с.

[6] Де Гроот М. Оптимальные статистические решения.-М: Мир. 1974.-493 с.

[7] Дейвид Г. Порядковые статистики.-М.: Наука. 1979.-336 с.

[8] Закс Ш. Теория статистичеких выводов.-М: Мир. 1975.-776 с.

[9] Кендал М. Дж.,Стьюарт А. Статистичеккие выводы и связи.-М.: Наука, 1973.-900 с.

[10] Крамер Г. Математиические методы статистки. М.: Мир, 1975, 648 с.

[11] Королюк В.С., Боровских Ю.В. Аналитические проблемы асимптотики вероятностных распределений.–Киев: Наук. думка.– 240с.

[12] Ивченко Г.И., Медведев Ю.И. Математическая статистика.- М: Высшая школа, 1984.-248с.

[13] Линьков Ю.Н. Лекции по математической статистике. Часть 1, Донецк: Истоки, 1999

[14] Линьков Ю.Н. Лекции по математической статистике. Часть 2, Донецк: Истоки, 2001.-276 с.

[15] Смирнов Н.В. Приближение законов распределения случайнiх величин по ємпирическим данным. Успехи мтатематических наук, 10, 1944.

[16] Смирнов Н.В. Теория вероятностей и математическая статистика. Избранные труды. М.: Наука, 1970,-270 с.

[17] Хеннекен П.Л., Тортра А. Теория вероятностей и некотррiе ее приложения.-М.: Наука, 1974.-472 с.

[18] Ширяев А.Н. Статистический последовательный анализ .-М: Наука. 1976.-272 с.

[19] Bartoszewicz J. Wykłady ze statystyki matematycznej. PWN, Warszawa, 1996.- 326 c.

[20] Bartlett M.S. Properties of sufficiency of statistical tests, Proc. Roy. Soc., 1937, A, 160.

[21] Bratijczuk M. Chydziński A. Rachunek prawdopodobieństwa. Wydawnictwo Politechniki Słaskiej. Gliwice, 2000. 294 str..

[22] Czesław Domański, Statystyczne testy nieparametryczne, Państwowe Wydawnictwo Ekonomiczne, Warszawa, 1979.

[23] Domański C., Pruskas K. Niekłasyczne metody statystyczne. PWE Warszawa 2000, 368 str.

[24] Durbin J. Kolmogorow -Smirnow Tests when Parametric are Estimated with Apllication to Test of Exponeniality and Tests on Spacings, Biometryka. 1975,62

[25] Jóżwiak J., Podgórski J. Statystyka od podstaw. Polskie Wydawnictwo Ekonomiczne, Warszawa, 1998.

[26] Fix E., Hodges J.L. Significance probabilities of the Wilcoxon Test. Annals of Mathematical Statistics 1955, 26.

[27] Fisz M. Rahunek prawdopodobieństwa i statystyka matematyczna, PWN Warszawa 1969, 688 c.

[28] Kolmogorov A.N. Sulla determinazione empirica di una legge di distribuzione, Giorn. Ist. Ital. Attuari, 4, 1933.

[29] Krysicki W., Bartos J., Dyczka W. Królikowska K., Wasiliewski M. Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, Сзеньж II Warszawa 1995.

[30] Mann, H. B., Whitney, D. R. "On a test of whether one of two random variables is stochastically larger than the other". Annals of Mathematical Statistics,1947, 18, 50–60.

[31] Mostowski A., Marceli Stark M.: Algebra liniowa. PWN, 1975

[32] Нилкс С. Математическая статистика. -М.: Наукаб 1967.- 632с.

[33] Saks S. Zymund A. Funkcje Analitycznet PWN Warszawa 1959 431 p.

[34] Schmid P. On the Koimogorov and Smirnov limit theorems for discontinuous distribution functions, AMS 29, 1958.

[35] Shapiro, S. S. and Wilk, M. B. "An analysis of variance test for normality (complete samples)", Biometrika, (1965), 52, 3 and 4, pages 591–611.

[36] Shorack R.A. Recursive Generation of the Distribution of the Mann-Whitney-Wilcoxon under Generalized Lehmann Alternatives. Annals of Mathematical Statistics 1966, 37.

[37] Wald, A. Sequential Tests of Statistical Hypotheses. The Annals of Mathematical Statistics, 1945, 16 (2): 117–186.

[38] Wilcoxon, F. "Individual comparisons by ranking methods". Biometrics Bulletin, 1945, 1, 80–83.

[39] Sobczyk M. Statystyka. PWN, Warszawa, 1999.

[40] Wald, A Sequential Analysis. New York: John Wiley and Sons. 1947.

[41] Zieliński R.,Tablice Statystyczne. 1972 , Wyd. PWN Warszawa

[42] Zieliński R., Przedziałufności dla frakcj Matematyka stosowana 2009, tom 10/50, c.51- 67.

[43] Хеннекен П.Л., Тортра А. Теория вероятностей и некотррiе ее приложения.-М.: Наука, 1974.-472 с.

[44] Ширяев А.Н. Статистический последовательный анализ .-М: Наука. 1976.-272 с.