Оригінали с.1-2 наведені у файлі VSP_HDR.DOC

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

М.В. КАРТАШОВ

ІМОВІРНІСТЬ, ПРОЦЕСИ, СТАТИСТИКА

Посібник

Затверджено
Міністерством освіти і науки України
як посібник для студентів університетів,
що навчаються за спеціальностями
"Математика" та "Статистика"

Київ
Видавничо-поліграфічний центр
'Київський університет'
2008

УДК 519.2 (075.8)
ББК 22.17Укр.я73
Л31

Рецензенти:
д-р фіз.-мат. наук, проф. О.І. Клесов,
д-р фіз.-мат. наук, проф. Є.О. Лєбєдєв,
д-р фіз.-мат. наук, проф. М.І. Портенко

Рекомендовано до друку вченою радою механіко-математичного факультету, протокол N 8 від 10 квітня 2006 року

Карташов М.В.

Л31 Імовірність, процеси, статистика : Посібник. – К.: Видавничо-поліграфічний центр 'Київський університет', 2008.– 494 с.
ISBN

Посібник містить матеріал курсів "Теорія ймовірностей", "Математична статистика" та "Додаткові розділи теорії ймовірностей" і призначений для студентів університетів, математичних та статистичних спеціальностей. Виклад ґрунтується на понятті інтеграла Лебега, однак всі необхідні властивості останнього визначаються та виводяться, що сприяє доступності курсу.

У курсах статистики та додаткових розділів теорії ймовірностей наводяться прямі посилання на відповідні теореми курсу теорії ймовірностей за їх назвами, що містяться у предметному покажчику.

Основні поняття та теореми ілюстровано на прикладах. Для повнішого засвоєння курсів наведені вправи для самостійного розв'язання.

Формулювання термінів і назв теорем та подальші посилання на них виділяються шрифтом. У додатку вміщено їх алфавітний перелік.

Для студентів університетів.

УДК 519.2 (075.8)
ББК 22.17Укр.я73

Гриф надано Міністерством освіти і науки України, лист N 14/18-Г-1029 від 07.11.2006 року

ISBN

Передмова

Посібник містить матеріал двосеместрових лекцій з теорії ймовірностей, математичної статистики, та семестрового курсу додаткових розділів теорії ймовірностей і випадкових процесів, які викладалися для студентів спеціальностей "Математика", "Статистика"третього курсу механіко-математичного факультету Київського національного університету імені Тараса Шевченка протягом 1988-2006 рр. Обсяг матеріалу розрахований на 137 лекційних годин, по 3 та 5 годин на тиждень відповідно у першому та другому семестрах.

На відміну від попередньої традиції [5], виклад спирається на поняття інтеграла Лебега (а не Фреше), як це зроблено в [4, 6, 7]. Необхідні властивості інтеграла доводяться для подальших посилань. Без доведення використовуються початкові результати загальної теорії міри – теореми Каратеодорі про продовження міри, Фубіні про кратний та повторний інтеграли, про заміну змінної та Радона – Нікодима про абсолютно неперервні міри. Відповідний курс викладається на факультеті одночасно з курсом теорії ймовірностей [1], однак для повного засвоєння предмету володіння основами теорії міри не обов'язкове. Водночас необхідними є знання фундаментальних понять лінійної алгебри та аналізу [2] – таких, як базис, ортонормованість, процедура Грама – Шмідта, границя, ряди, неперервність, похідна, інтеграл Рімана, комплекснозначні функції. Для повнішого засвоєння курсу та при розв'язанні задач корисним є володіння основами дискретної математики [3].

У курсах статистики та додаткових розділів теорії ймовірностей наводяться прямі посилання на відповідні теореми курсу теорії ймовірностей за їх назвами, що містяться у предметному покажчику.

У посібнику прийнято уніфіковану систему термінів та назв теорем. У додатку вміщено повний їх перелік зі вказівкою на сторінку. **Нові формулювання** та подальші *посилання на них* виділяються шрифтом. У електронному варіанті такі посилання є інтерактивними.

При створенні посібника використані підручники [4 – 10].

Укладач щиро вдячний своїм вчителям – професорам Анатолію Яковичу Дороговцеву, Михайлу Йосиповичу Ядренку, Анатолію Володимировичу Скороходу, Володимиру Семеновичу Королюку, Юрію Макаровичу Березанському – за їх блискучі лекції з математичного аналізу, теорії ймовірностей, математичної статистики та функціонального аналізу.

Особлива подяка рецензентам посібника – професорам О.І. Клесову, Є.О. Лєбєдєву та М.І. Портенку, за цінні зауваження й поради.

Розділ 1

Теорія ймовірностей

Вступ

Створення теорії ймовірностей пов'язане з іменами Г. Галілея, Б. Паскаля, П. Ферма, Х. Гюйгенса, Я. Бернуллі, А. Муавра (XVI – XVIII ст). Ці вчені вивчали випадкові явища, що спостерігаються в азартних іграх.

Імовірнісні проблеми теорії похибок вимірювань і стрільби, статистики народонаселення розв'язувалися в працях П. Лапласа, С. Пуассона, К. Гаусса, Т. Байеса, Ж. Бертрана в XVIII – XIX століттях.

У XIX – XX ст. значні досягнення в теорії ймовірностей та статистиці були отримані П. Чебишевим, А. Марковим, А. Ляпуновим, П. Буняковським, зокрема, у вигляді граничних теорем.

Сучасна теорія ймовірностей була створена в працях Р. Мізеса, Е. Бореля, С. Бернштейна, А. Колмогорова, А. Хінчина, Б. Гнеденка, П. Леві, Ю. Прохорова, та інших вчених.

Теорія ймовірностей – *це розділ математики, в якому вивчаються математичні моделі випадкових (стохастичних) явищ*.

З погляду ймовірнісників, усі природні та соціальні явища можна класифікувати на такі групи:

(д) *Детерміновані*. За вченням картезіанців, для таких явищ досить лише вказати початкові умови – і ми знатимемо все про їх майбутнє, розв'язуючи відповідні рівняння динаміки. На практиці детермінованих явищ немає.

(нн) *Недетерміновані неповторювані*. Наприклад: наявність життя на Марсі. Висловлювання щодо його імовірності не можна обґрунтувати на підставі інших спостережень – адже цей феномен унікальний.

(нпн) *Недетерміновані повторювані з непередбачуваною поведінкою частот*. Приклад: для прогнозу поведінки послідовних цифр числа π статистика не може бути застосована.

(нпс) *Недетерміновані повторювані зі* **стійкістю частот** – для них відносна частота події має певну границю при нескінченному зростанні кількості спостережень.

Явища з останньої групи називаються *випадковими*. Їх існування на практиці доведено всім досвідом застосувань теорії ймовірностей і може бути проілюстровано дослідами видатних математиків, які проводили серії підкидань монети та обчислювали частоти реверса:

Ж. Бюффон	4000 підкидань	частота реверса = 0.5080
П. Морган	4800 підкидань	частота реверса = 0.5005
К. Пірсон	24000 підкидань	частота реверса = 0.5005
В. Феллер	10000 підкидань	частота реверса = 0.4979

Для побудови теорії ймовірностей використовувалися такі підходи.

(1) **Суб'єктивний підхід,** згідно з яким імовірність є нормованою мірою впевненості експерта. Труднощі виникають у зв'язку із формалізацією підходу.

(2) **Класичний підхід** виходить з ідеї симетрії, рівноможливості. Нерозрізненні події повинні мати однакову ймовірність – часто цього цілком досить для побудови теорії. Однак цей підхід не спроможний описати експеримент із нескінченною кількістю елементарних подій, наприклад, з вибором випадкової точки на відрізку.

(3) **Частотне визначення** ймовірності як границі частот, що існує згідно з постулюванням *стійкості частот*. Це цілком спроможна конструкція, однак доведення відносно складних теорем (на кшталт закону великих чисел) стають занадто складними для засвоєння студентами.

(4) **Аксіоматичний підхід:** із частотного означення (ймовірність є границею частот) отримуємо певні властивості ймовірностей, але не доводимо їх, а просто постулюємо як аксіоми з урахуванням *стійкості частот*.

Останній, аксіоматичний підхід, належить А.М. Колмогорову (Die Grundbegriffe der Wahrscheinlichkeitstheorie, Berlin, 1933) і сьогодні є загальновизнаним при побудові теорії ймовірностей.

Розділ містить матеріал семестрового курсу теорії ймовірностей, що розрахований на 54 години лекцій та 36 годин практичних занять. Для повноти додано посилання на основні результати теорії міри, які викладаються паралельно. Крім того, в тих же часових рамках викладаються також дві теми з теорії випадкових процесів: процес Пуассона і вінерівський процес.

1.1. Стохастичний експеримент, події та операції над ними

Теорія ймовірностей, як будь-який інший розділ чистої математики, побудована аксіоматично, тобто має певні основні вихідні поняття, які не визначаються всередині теорії, та аксіоми, що пов'язують ці поняття. Основними поняттями теорії ймовірностей є поняття *стохастичного експерименту, елементарної події* та *простору елементарних подій*. Отже, наступні висловлювання та означення слід розуміти як пояснення або інтерпретації, що лежать поза математикою.

Стохастичний експеримент – це певне випробування, спостереження чи дослід, результат якого не можна передбачити однозначно (наприклад, підкидання монети чи грального кубика), або ж можна передбачити (наприклад, неправильна відповідь на іспиті).

Певний фіксований результат стохастичного експерименту, який не можна виразити через сукупність інших результатів (чи спричинити ними), називається **елементарною подією.** Зокрема, різні елементарні події не можуть відбутися одночасно. Множина всіх елементарних подій називається **простором елементарних подій**.

Отже, *в результаті проведення стохастичного експерименту завжди відбувається одна і тільки одна елементарна подія з простору елементарних подій.*

Оскільки у теорії ймовірностей фізична природа елементарних подій та простору елементарних подій не має суттєвого значення, то *простором елементарних подій* може бути довільна абстрактна множина, а *елементарна подія* є будь-яким елементом цієї множини.

У теорії ймовірностей простір елементарних подій позначається символом грецького алфавіту Ω, а елементарна подія – через ω .

1.1.1. Приклади стохастичних експериментів та просторів елементарних подій

1. *Підкидання однієї монети. Простір елементарних подій* має вигляд $\Omega = \{A,P\}$, оскільки результат підкидання, що відрізняється від аверса A, або реверса P, вважається неможливим.

2. *Підкидання двох монет*. Оскільки фіксація результату підкидання двох монет потребує задання впорядкованої пари з двох елементів

– результатів підкидання першої та другої монети, то простір елементарних подій містить пари можливих варіантів підкидань однієї монети: $\Omega = \{$АА,АР,РА,РР$\}$.

3. *Підкидання гральної кості.* $\Omega = \{1,2,3,4,5,6\}$, оскільки гральна кость має шість граней, а падіння на ребро вважається неможливим.

4. *Підкидання монети до першого успіху.* Можна припустити, що успіх обов'язково має відбутися за якоїсь скінченної кількості підкидань, а тому $\Omega = \{$У, НУ, ННУ,...,НН...НУ,...$\}$. З тим же результатом можна ототожнити простір елементарних подій із множиною натуральних чисел $\mathbb{N} = \{n,\ n \geq 1\}$. Для реалізації можливості нескінченної кількості підкидань треба постулювати, що простір Ω містить спеціальну точку $\infty = \{$НН...Н...$\}$.

5. *Вибір випадкової точки на одиничному відрізку.* У цьому випадку $\Omega = [0,1]$, оскільки кожна така точка однозначно задається своєю координатою з відрізку $[0,1]$.

1.1.2. Випадкові події

Зі стохастичним експериментом можна пов'язати певне висловлювання про його результат. Оскільки для кожної елементарної події можна встановити, справедливе дане висловлювання чи ні, то довільному висловлюванню відповідає певна множина *елементарних подій*, а саме: така підмножина простору елементарних подій, для елементів якої справджується дане висловлювання. **Випадковими подіями** *називаються підмножини простору елементарних подій, що є відповідними прообразами певних висловлювань про результат стохастичного експерименту.*

Приклади

1. *Підкидання двох монет*, $\Omega = \{$АА,АР,РА,РР$\}$. Тоді
подія $\{$Перший аверс$\}$=$\{$АА,АР$\}$, $\{$Хоча б один аверс$\}$=$\{$АА,АР,РА$\}$.

2. *Підкидання монети до першого успіху*, $\Omega = \{$У, НУ,..,Н...НУ,...$\}$.
подія $\{$Знадобиться не більше трьох підкидань$\}$ = $\{$У,НУ,ННУ$\}$.

3. *Підкидання гральної кості.* $\Omega = \{1,2,3,4,5,6\}$. Тоді
подія $\{$Випаде непарна кількість очок$\}$ =$\{1,\ 3,\ 5\}$.

4. *Вибір випадкової точки на одиничному відрізку*, $\Omega = [0,1]$. Клас усіх випадкових подій має містити всі інтервали $[0,x)$ та їх перетворення.

Зауважимо, що не всі підмножини простору елементарних подій повинні бути випадковими подіями. Підмножин може бути більше, ніж можна сформулювати висловлювань, враховуючи зліченну кількість останніх. Інше обґрунтування необхідності обмеження класу підмножин дається теоремою Банаха – Тарського: тривимірну одиничну кулю можна розбити на скінченну кількість частин, які в результаті переміщень заповнюють кулю радіуса 2 – внаслідок невимірності цих частин.

1.1.3. Властивості класу випадкових подій

Клас усіх випадкових подій у даному стохастичному експерименті будемо позначати через $\mathfrak{F}$.

Оскільки клас висловлювань замкнений відносно операцій заперечення, кон'юнкції, диз'юнкції та інших теоретико-множинних операцій і містить також суперечливі висловлювання, то природно прийняти відповідні властивості класу $\mathfrak{F}$ всіх випадкових подій – підмножин Ω.

1. Множина $\Omega \subset \Omega$ ("щось та відбудеться") є випадковою подією: $\Omega \equiv \{\omega : \omega \in \Omega\} \in \mathfrak{F}$, яка називається **універсальною.**

2. Множина $\varnothing$, що не містить жодної елементарної події ("нічого не відбудеться"), є випадковою подією: $\varnothing \equiv \{\omega : \omega \notin \Omega\} \in \mathfrak{F}$, і називається **неможливою** подією.

3. Належність елементарної події ω випадковій події A, позначення $\omega \in A$, відображається висловлюваннями: ω **сприяє** A, або ж подія A відбувається при даній елементарній події ω.

4. Справедливість включення $A \subset B$ визначає, що подія A **спричиняє подію** B (або ж міститься в ній). Альтернативна інтерпретація: якщо відбувається A, то відбувається і B.

5. Для події A її **доповнення** (або **заперечення**) $\overline{A} \equiv \{\omega : \omega \notin A\}$ є випадковою подією: $\overline{A} \in \mathfrak{F}$, яка полягає в тому, що не відбудеться A.

6. Для двох подій A, B їх **об'єднання** $A \cup B \equiv \{\omega : \omega \in A$ або $\omega \in B\}$ є випадковою подією, яка полягає в тому, що відбудеться A або B.

7. Для двох випадкових подій A, B їх **переріз** (або ж **перетин**) – множина $A \cap B \equiv \{\omega : \omega \in A$ та $\omega \in B\}$ є випадковою подією, яка полягає в тому, що події A і B відбудуться одночасно.

8. Якщо $A \cap B = \varnothing$, то події A, B називаються **несумісними** (або ж такими, що **не перетинаються**).

9. Для двох подій A, B їх **різниця** $A \setminus B \equiv \{\omega : \omega \in A$ та $\omega \notin B\}$ є випадковою подією, яка полягає в тому, що відбудеться A і одночасно не відбудеться B.

10. Для двох подій A, B їх **симетрична різниця** $A \Delta B$ є випадковою подією, яка полягає в тому, що з подій A, B відбудеться точно одна подія,

тобто $A \Delta B = (A \backslash B) \cup (B \backslash A)$.

11. Події послідовності $(A_n, n \geq 1)$ називаються **попарно несумісними**, якщо $A_i \cap A_j = \varnothing$ для всіх $i \neq j$.

12. Для послідовності випадкових подій $(A_n, n \geq 1)$ їх зліченне **об'єднання** $\cup A_n \equiv \{\omega : \exists n \geq 1 : \omega \in A_n\}$ є випадковою подією: $\cup A_n \in \mathfrak{F}$, яка полягає в тому, що відбудеться хоча б одна подія із цієї послідовності.

13. Зліченний **переріз** $\cap A_n \equiv \{\omega : \omega \in A_n, \ \forall n \geq 1\}$ є випадковою подією, яка полягає в тому, що відбудуться всі події із даної послідовності.

14. Нехай $(A_n, \, n \geq 1)$ – монотонно неспадна послідовність подій, тобто $A_n \subset A_{n+1}, \ \forall n \geq 1$. Їх **монотонною границею** A (що позначається як $A_n \uparrow A \equiv \lim A_n$) є випадкова подія, яка полягає в тому, що відбудеться принаймні одна подія даної послідовності (а отже, і кожна, починаючи з деякого номера), тобто $A = \cup A_n$. У такому випадку кажуть, що події A_n **монотонно збігаються** до A.

15. Нехай $(A_n, \, n \geq 1)$ – монотонно незростаюча послідовність подій, тобто $A_{n+1} \subset A_n, \, \forall n \geq 1$. Їх **монотонною границею** A є випадкова подія, яка полягає в тому, що відбудуться одночасно всі події послідовності, тобто $A = \cap A_n$ (це позначається як $A_n \downarrow A \equiv \lim A_n$). У такому випадку кажуть, що події A_n **монотонно збігаються** до A.

16. Для послідовності подій $(A_n, n \geq 1)$ їх **верхня границя** $\overline{\lim} \ A_n$ є випадковою подією, яка полягає в тому, що події даної послідовності відбудуться **нескінченно часто,** тобто відбудуться всі події з деякої нескінченної підпослідовності A_{n_k}. Цю подію можна зобразити у вигляді: $\overline{\lim} A_n = \cap_{m \geq 1} \cup_{n \geq m} A_n$. Дійсно, елементарна подія ω належить правій частині рівності тоді й тільки тоді, коли для довільного номера $m \geq 1$ знайдеться номер $n \geq m$ такий, що $\omega \in A_n$, тобто коли відбудеться нескінченна кількість серед подій A_n.

17. Для послідовності подій $(A_n, n \geq 1)$ їх **нижня границя** $\underline{\lim} \ A_n$ є випадковою подією, яка полягає в тому, що відбудуться всі події даної послідовності **починаючи з деякого номера**. Ця подія позначається через: $\underline{\lim} A_n = \cup_{m \geq 1} \cap_{n \geq m} A_n$. Дійсно, елементарна подія ω належить правій частині рівності тоді й тільки тоді, коли знайдеться номер $m \geq 1$ такий, що $\omega \in A_n$ при всіх $n \geq m$, тобто коли відбудуться всі події A_n, починаючи з деякого номера.

Між теоретико-ймовірнісними поняттями випадкових подій і операцій над ними, та теоретико-множинними поняттями множин і їх перетворень, є пряма відповідність, що ілюструється у такій таблиці.

Позначення	*У теорії ймовірностей*	*У теорії множин*
Ω	Простір елементарних подій, універсальна подія	Універсальна множина
ω	Елементарна подія	Елемент множини
$\varnothing$	Неможлива подія	Порожня множина
A	Випадкова подія	Множина
$\omega \in A$	ω сприяє події A	ω належить A
$\overline{A}$	Заперечення події A	Доповнення множини A
$A \subset B$	Подія A спричиняє B	A міститься в B
$A \cap B$	Відбудуться події A та B	Переріз множин A і B
$\cap_{n\ge 1} A_n$	Відбудуться всі події A_n одночасно	Переріз множин A_n
$A \cap B = \varnothing$	Події A і B несумісні	A і B не перетинаються
$A \cup B$	Відбудеться подія A або B	Об'єднання A і B
$\cup_{n\ge 1} A_n$	Відбудеться хоча б одна з подій A_n	Об'єднання множин A_n
$A\backslash B$	Відбудеться подія A, та не B	Різниця множин A та B
$A\Delta B$	Відбудеться точно одна з подій A, B	Симетрична різниця A, B
$A_n \uparrow A$	Події A_n не спадають та мають границю A	$A_n \subset A_{n+1},\ \cup_{n\ge 1} A_n = A$
$A_n \downarrow A$	Події A_n не зростають та мають границю A	$A_{n+1} \subset A_n,\ \cap_{n\ge 1} A_n = A$
$\overline{\lim} A_n$	Події A_n відбуваються нескінченно часто	$\cap_{m\ge 1} \cup_{n\ge m} A_n$
$\underline{\lim} A_n$	Події A_n відбуваються всі, починаючи з деякого номера	$\cup_{m\ge 1} \cap_{n\ge m} A_n$

Вправи

(1) Довести, що: (а) $A\backslash B = A\backslash(A \cap B)$, (б) $A\Delta B = (A \cup B)\backslash(A \cap B)$, (в) $(A\Delta B)\Delta C = A\Delta(B\Delta C)$, (г) $\overline{A \cup B} = \overline{A} \cap \overline{B}$, (д) $\overline{A\Delta B} = \overline{A}\Delta\overline{B}$, (е) $(A \cap B) \cap C = A \cap (B \cap C)$, (є) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$.

(2) Для довільної послідовності випадкових подій $(A_n, n \ge 1)$ має місце зображення $\cup_{n\ge 1} A_n = \cup_{n\ge 1} B_n$, з попарно несумісними подіями вигляду: $B_1 = A_1$, $B_n = A_n \backslash \cup_{k=1}^{n-1} A_k, n > 1$.

(3) Обчислити $\underline{\lim} A_n$, $\overline{\lim} A_n$ у випадку, коли: (а) $\Omega = \mathbb{R}$, а випадкові події $A_n = (-\infty, a_n)$, (б) $A_{2n+1} = A,\ A_{2n} = B$, (в) $A_n \uparrow A$, (г) $A_n \downarrow A$.

1.2. Аксіоми теорії ймовірностей та властивості ймовірності

Перелік наведених вище властивостей класу $\mathfrak{F}$ випадкових подій можна подовжити. Одночасно доцільно відшукати ті з них, що породжують всі інші. Тим самим приходимо до таких аксіом класу випадкових подій.

1.2.1. Аксіоми класу випадкових подій

Нехай $\mathfrak{F}$ – клас усіх випадкових подій у деякому стохастичному експерименті із простором елементарних подій Ω. Тоді, як відмічено у розділі про властивості класу випадкових подій, цей клас має такі властивості:

(F1 – нормованість) $\Omega \in \mathfrak{F}$.

(F2 – доповнення) З $A \in \mathfrak{F}$ *випливає, що* $\overline{A} \in \mathfrak{F}$.

(F3 – зліченні об'єднання) Якщо $A_n \in \mathfrak{F}$ *– зліченна послідовність, то* $\cup A_n \in \mathfrak{F}$.

Клас підмножин *універсальної* множини Ω, який задовольняє умови *(F1) – (F3)*, називається **сигма-алгеброю** (чи σ-**алгеброю**) підмножин Ω.

Клас $\mathfrak{F}$ називається **алгеброю,** якщо замість умови *(F3)* виконується слабша умова

(F3′– об'єднання) З $A, B \in \mathfrak{F}$ *випливає* $A \cup B \in \mathfrak{F}$.

Очевидно, ця властивість еквівалентна умові

(F3″– скінченні об'єднання) З $A_k \in \mathfrak{F}, k = \overline{1, n}$, *випливає* $\cup_{k=1}^{n} A_k \in \mathfrak{F}$.

Приклад. Для довільної події $A \subset \Omega$ клас $\{\varnothing, A, \overline{A}, \Omega\}$ є алгеброю і одночасно σ-алгеброю.

Першу групу аксіом теорії ймовірностей можна сформулювати так: **"Клас $\mathfrak{F}$ усіх випадкових подій є сигма-алгеброю".**

Для побудови сигма-алгебр, виходячи із заданих алгебр, використовують такий метод.

Означення. *Сигма-алгебра* $\mathfrak{F}$ **породжена** *класом випадкових подій* $\mathfrak{A}$, *що позначається як* $\mathfrak{F} \equiv \sigma[\mathfrak{A}]$, *якщо вона є найменшою з тих сигма-алгебр, що містять цей клас:*

$$\mathfrak{F} = \bigcap\nolimits_{S \supset \mathfrak{A},\ S \quad \sigma-\text{алгебра}} S.$$

Теорема (про монотонний клас). *Якщо клас множин* $\mathfrak{F}$ *містить алгебру* $\mathfrak{A}$ *та замкнений відносно монотонної збіжності* $A_n \uparrow A$, *то він містить породжену сигма-алгебру* $\sigma[\mathfrak{A}]$.

Доведення пропонується як вправа.

Вправи

(1) Довести, що з наведених аксіом випливають всі вказані вище властивості класу випадкових подій.

(2) Визначення породженої сигма-алгебри коректне, оскільки хоча б одна вказана сигма-алгебра S існує (клас $S = 2^\Omega$ усіх підмножин Ω), причому переріз будь-якої кількості сигма-алгебр є сигма-алгеброю.

(3) Клас $\mathfrak{A}$ є алгеброю тоді й тільки тоді, коли (а) $\Omega \in \mathfrak{A}$, (б) з $A, B \in \mathfrak{A}$ випливає $A \backslash B \in \mathfrak{A}$.

(4) Випадкові події $(H_n, n \geq 1)$ попарно несумісні. Довести тотожність
$\sigma[\{H_n, n \geq 1\}] = \{\cup_{i \subset I} H_i,\ I \subset \mathbb{N}\}$.

(5) Для σ-алгебри $\mathfrak{F}$ і $C \notin \mathfrak{F}$ $\sigma[\mathfrak{F} \cup \{C\}] = \{(A_1 \cap C) \cup (A_2 \cap \overline{C}),\ A_k \in \mathfrak{F}\}$.

(6) Клас підмножин $\mathbb{R}$ вигляду $F \cap G$, де F – відкрита, а G – замкнена множина, утворює алгебру, але не є сигма-алгеброю.

(7) Нехай $\Omega = \{1, 2, 3, 4, 5, 6\}$. Знайти всі елементи найменшої алгебри, яка містить події $A = \{1, 2\}, B = \{4, 5, 6\}$.

(8) Для зліченного простору $\Omega = \{\omega_n, n \geq 1\}$ описати найменші алгебру та сигма-алгебру, які містять усі одноточкові множини.

1.2.2. Аксіоми ймовірності

У теорії ймовірностей із кожною випадковою подією пов'язують числову міру її вірогідності – ймовірність. Імовірність розглядається як границя *відносних частот*, тобто відношень кількості сприятливих для висловлювання випробувань до кількості всіх випробувань. Можна відмітити такі основні властивості частот:

(*невід'ємність*) Частота(Довільного_висловлювання)≥ 0.

(*нормованість*) Частота(Універсального_висловлювання)$= 1$,
адже воно справджується в кожному випробуванні.

(*адитивність*) Частота(А або Б)=Частота(А)+Частота(Б),
якщо висловлювання А,Б не можуть справджуватися одночасно.

Справедливість адитивності засвідчує наступна діаграма:

ОООАОБББААООООАААБООО
OOOXOXXXXXOOOOXXXXOOO

Кількість символів А та Б дорівнює (1+2+3)+(3+1) = 6+4, кількість об'єднуючих символів Х дорівнює 1 + 5 + 4 = 10 = 6 + 4.

Отже, ймовірність як границя частот має бути *невід'ємною, нормованою та адитивною функцією випадкових подій* (як образів висловлювань щодо результату експерименту). Звідси приходимо до означення.

Означення. Імовірністю *називається числова функція* $\mathbf{P}$ *на класі* $\mathfrak{F}$ *всіх випадкових подій з такими властивостями:*

(P1 – невід'ємність*)* $\mathbf{P}(A) \geq 0, \ \forall A \in \mathfrak{F}$.

(P2 – нормованість*)* $\mathbf{P}(\Omega) = 1$.

(P3 – сигма-адитивність*) Для послідовності подій* $(A_n, n \geq 1)$*, що попарно несумісні (тобто* $A_i \cap A_j = \varnothing, \forall i \neq j$*), справедлива рівність*

$$\mathbf{P}\left(\bigcup\nolimits_{n\geq 1} A_n\right) = \sum\nolimits_{n\geq 1} \mathbf{P}(A_n).$$

Для обґрунтування останньої аксіоми (*P3*) зауважимо, що її твердження еквівалентне такій парі тверджень:

(P3′ – адитивність*) Для подій* A, B*, що несумісні,*

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

(P4 – неперервність в нулі*). Для довільної послідовності подій* (A_n) *таких, що* $A_{n+1} \subset A_n$*, та* $\cap_{n\geq 1} A_n = \varnothing$*, має місце збіжність*

$$\lim\nolimits_{n\to\infty} \mathbf{P}(A_n) = 0.$$

Перша властивість повністю відповідає властивості частот, а заперечення другої виглядало б не досить обґрунтованим.

Означення. *Функція* P *називається* адитивною *ймовірністю, якщо замість аксіоми (P3) виконується слабша умова (P3′).*

За індукцією неважко довести (**вправа**), що властивість адитивності *(P3′)* еквівалентна скінченній адитивності:

(P3″ – скінченна адитивність*) Для довільної скінченної послідовності подій* $(A_k,\ 1 \leq k \leq n)$*, що попарно несумісні,*

$$\mathbf{P}\left(\bigcup\nolimits_{1\leq k\leq n} A_k\right) = \sum\nolimits_{1\leq k\leq n} \mathbf{P}(A_k).$$

Другу групу аксіом теорії ймовірностей можна сформулювати так: **"Імовірність є невід'ємною нормованою сигма-адитивною функцією на класі всіх випадкових подій".**

Отже, аксіоматика теорії ймовірностей зводиться до постулювання існування ймовірнісного простору $(\Omega,\ \mathfrak{F},\ \mathbf{P})$, що має такі елементи:

Ω – *простір елементарних подій* (абстрактна множина),

$\mathfrak{F}$ – клас усіх випадкових подій – підмножин Ω, які утворюють *сигма-алгебру* з властивостями *(F1 – F3)*, а

$\mathbf{P}$ – *імовірність* на сигма-алгебрі подій $\mathfrak{F}$ із властивостями *(P1 – P3)*.

1.2.3. Властивості ймовірності

З *аксіом теорії ймовірностей* випливають такі властивості ймовірності.

Теорема (про основні властивості ймовірностей). *Справедливі такі властивості ймовірності.*

1. Імовірність неможливої події: $\mathbf{P}(\varnothing) = 0$. Дійсно, за означенням події $A_k \equiv \varnothing, k = \overline{1,n}$, *попарно несумісні*, а тому додатність $\mathbf{P}(\varnothing)$ суперечила б тотожності $\mathbf{P}(\varnothing) = \mathbf{P}(\cup_{k=1}^{n}\varnothing) = \sum_{k=1}^{n}\mathbf{P}(\varnothing)$, що є наслідком адитивності або ж сигма-адитивності (тут $n \le \infty$). Зауважимо також, що з $\mathbf{P}(A) = 0$ не випливає, що A – *неможлива* подія.

2. Імовірність доповнення події $A \in \mathfrak{F}$:

$$\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A),$$

оскільки $\Omega = A \cup \overline{A}$, де доданки *несумісні*, отже $1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup \overline{A}) = \mathbf{P}(A) + \mathbf{P}(\overline{A})$.

3. Імовірність вкладеної різниці: для $A, B \in \mathfrak{F}$, $A \subset B$

$$\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A),$$

оскільки $B = A \cup (B \setminus A)$, де доданки несумісні: $\mathbf{P}(B) = \mathbf{P}(A \cup (B \setminus A)) = \mathbf{P}(A) + \mathbf{P}(B \setminus A)$.

4. Монотонність імовірності: для $A, B \in \mathfrak{F}$, $A \subset B$

$$\mathbf{P}(A) \le \mathbf{P}(B),$$

виводиться з невід'ємності та попереднього пункту:

$$\mathbf{P}(B) - \mathbf{P}(A) = \mathbf{P}(B \setminus A) \ge 0.$$

5. Множина значень імовірності:

$$\{\mathbf{P}(A),\ A \in \mathfrak{F}\} \subset [0,1],$$

оскільки $\varnothing \subset A \subset \Omega$ і має місце *монотонність імовірності.*

6. Імовірність об'єднання двох подій $A, B \in \mathfrak{F}$:

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B),$$

оскільки $A \cup B = A \cup (B \setminus A \cap B)$, де доданки *несумісні*, а ймовірність другого визначається властивістю *ймовірності вкладеної різниці*:

$$\mathbf{P}(A \cup B) = \mathbf{P}(A \cup (B \setminus A \cap B)) = \mathbf{P}(A) + \mathbf{P}(B \setminus A \cap B) =$$
$$\mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B).$$

7. Формула включення–виключення для подій $(A_k, k = \overline{1,n})$:

$$\mathbf{P}(\bigcup_{1\le k\le n} A_k) = \sum_{1\le\ i\le\ n} \mathbf{P}(A_i) - \sum_{1\le i<j\le n} \mathbf{P}(A_i \cap A_j) +$$

$$\sum_{1\le i<j<k\le n} \mathbf{P}(A_i \cap A_j \cap A_k) - ... + (-1)^{n-1}\ \mathbf{P}(A_1 \cap ... \cap A_n).$$

При $n = 2$ це є властивість 6, а для $n > 2$ формула доводиться за індукцією з властивості *ймовірності об'єднання*. Нехай формула справедлива для заданого n. Позначимо $B = \cup_{k=1}^n A_k$, тоді

$$\mathbf{P}(\bigcup_{k=1}^{n+1} A_k) = \mathbf{P}(B \cup A_{n+1}) = \mathbf{P}(B) + \mathbf{P}(A_{n+1}) - \mathbf{P}(B \cap A_{n+1}) =$$

$$\sum_{1\le k\le n} \mathbf{P}(A_k) - \sum_{1\le k<j\le n} \mathbf{P}(A_k \cap A_j) + ... + \mathbf{P}(A_{n+1}) - \mathbf{P}(\cup_{k=1}^n (A_k \cap A_{n+1}))$$

$$= \sum_{1\le k\le n+1} \mathbf{P}(A_k) - \sum_{1\le k<j\le n} \mathbf{P}(A_k \cap A_j) + ... - \sum_{1\le k\le n} \mathbf{P}(A_k \cap A_{n+1}) + ...,$$

звідки за індукцією дістанемо дане твердження.

8. Напівадитивність імовірності:

$$\mathbf{P}\left(\bigcup_{1\le k\le n} A_k\right) \le \sum_{1\le k\le n} \mathbf{P}(A_k).$$

Для доведення позначимо $B_1 = A_1,\ B_k = A_k \setminus \cup_{j=1}^{k-1} A_j$. Тоді події B_k попарно несумісні, $B_k \subset A_k$ і $\cup_{\ 1\le k\le n} B_k = \cup_{\ 1\le k\le n} A_k$. Звідси за *адитивністю* та *монотонністю ймовірностей*

$$\mathbf{P}(\bigcup\nolimits_{1\le k\le n} A_k) = \mathbf{P}(\bigcup\nolimits_{1\le k\le n} B_k) = \sum_{k\le n} \mathbf{P}(B_k) \le \sum_{k\le n} \mathbf{P}(A_k).$$

9. Неперервність імовірності.

(а) Якщо $A_n \uparrow A$, то $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$.

(б) Якщо $A_n \downarrow A$, то $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$.

Для доведення (а) позначимо $B_1 = A_1,\ B_n = A_n \setminus A_{n-1},\ n > 1$. Тоді B_n *попарно несумісні*, $\cup_{k\le n} B_k = A_n$, $\cup B_k = A$,

$$\mathbf{P}(A) = \sum \mathbf{P}(B_n) = \lim_{n\to\infty} \sum_{k\le n} \mathbf{P}(B_k) = \lim_{n\to\infty} \mathbf{P}(A_n),$$

де остання рівність виводиться з *адитивності*, а збіжність є монотонною.

Для доведення (б) застосуємо (а) для доповнень: $\overline{A_n} \uparrow \overline{A}$, $1 - \mathbf{P}(A_n) = \mathbf{P}(\overline{A_n}) \uparrow \mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$.

10. Зліченна напівадитивність імовірності:

$$\mathbf{P}\left(\bigcup A_k\right) \le \sum \mathbf{P}(A_k)$$

випливає із *напівадитивності* та *неперервності ймовірності*:

$$\mathbf{P}\left(\bigcup A_k\right) = \lim_{n\to\infty} \mathbf{P}\left(\bigcup_{1\le k\le n} A_k\right) \le \lim_{n\to\infty} \sum_{1\le k\le n} \mathbf{P}(A_k).$$

11. Еквівалентність **неперервності в нулі та сигма-адитивності** для *адитивної* ймовірності. *Для того, щоб адитивна ймовірність* $\mathbf{P}$ *на алгебрі подій* $\mathfrak{A}$ *була сигма-адитивною, необхідно і достатньо, щоб вона була неперервною в нулі на* $\mathfrak{A}$,*тобто для будь-якої послідовності подій* $A_n \in \mathfrak{A}$ *таких, що* $A_n \downarrow \varnothing$, *повинна мати місце збіжність* $\mathbf{P}(A_n) \downarrow 0$.

Доведення. Якщо функція $\mathbf{P}$ сигма-адитивна, то вона неперервна в нулі за властивістю *неперервності ймовірності*.

Нехай функція $\mathbf{P}$ адитивна на алгебрі $\mathfrak{A}$ та неперервна в нулі, а $A_n \in \mathfrak{A}$ довільні події, що *попарно несумісні*. Позначимо $B_n = \cup_{k>n} A_k$. Тоді $B_n \downarrow \varnothing$ та $\mathbf{P}(B_n) \downarrow 0$. Оскільки внаслідок *скінченної адитивності* $\mathbf{P}(\cup A_k) = \mathbf{P}(\cup_{k\le n} A_k) + \mathbf{P}(B_n) = \sum_{k\le n} \mathbf{P}(A_k) \ + \mathbf{P}(B_n) \ \to \sum_{k\ge 1} \mathbf{P}(A_k), n \to \infty$, то ймовірність $\mathbf{P}$ сигма-адитивна на алгебрі $\mathfrak{A}$.

12. **Продовження неперервної ймовірності** на сигма-алгебру.

Для того, щоб адитивна ймовірність $\mathbf{P}$ *на алгебрі подій* $\mathfrak{A}$ *мала продовження до сигма-адитивної ймовірності на породженій сигма-алгебрі* $\mathfrak{F} = \sigma[\mathfrak{A}]$*, необхідно і достатньо, щоб вона була неперервною в нулі на алгебрі* $\mathfrak{A}$

Доведення. Якщо ймовірність $\mathbf{P}$ сигма-адитивна, то вона неперервна в нулі за властивістю неперервності.

Нехай ймовірність $\mathbf{P}$ адитивна на алгебрі $\mathfrak{A}$ та неперервна в нулі. Тоді $\mathbf{P}$ *сигма-адитивна* на алгебрі $\mathfrak{A}$, як показано у попередній властивості. За цієї умови існування та єдиність сигма-адитивного продовження на σ-алгебру $\mathfrak{F}$ стверджується у **теоремі Каратеодорі про продовження міри,** яка доводиться в курсі теорії міри та інтеграла.

Вправи

(1) Нехай $A_k \in \mathfrak{F}$, $\mathbf{P}(A_k) = 1, k \ge 1$. Довести, що $\mathbf{P}(\cap_{k\ge 1} A_k) = 1$.

(2) Довести (а) рівність $\mathbf{P}(A\Delta B) = \mathbf{P}(A) + \mathbf{P}(B) - 2\,\mathbf{P}(A \cap B)$, (б) нерівність трикутника $\mathbf{P}(A\Delta B) \le \mathbf{P}(A\Delta C) + \mathbf{P}(C\Delta B)$, (в) нерівність різниць $\mathbf{P}((\cup A_i)\Delta(\cup B_i)) \le \sum_i \mathbf{P}(A_i \Delta B_i)$.

(3) Для довільних подій A, B довести нерівність

$$\mathbf{P}(A \cup B)\mathbf{P}(A \cap B) \le \mathbf{P}(A)\mathbf{P}(B).$$

(4) Довести (а) нерівність Буля: $\mathbf{P}(A \cap B) \ge 1 - \mathbf{P}(\overline{A}) - \mathbf{P}(\overline{B})$, (б) нерівність $\mathbf{P}(\bigcup_{k=1}^n A_k) \ge \mathbf{P}(\bigcap_{k=1}^n A_k) + 1 - n$.

(5) Послідовність подій A_n збігається до границі A, якщо $A = \underline{\lim} A_n = \overline{\lim} A_n$. Довести, що тоді $\mathbf{P}(A) = \lim \mathbf{P}(A_n)$.

(6) Довести, що для ймовірності $\mathbf{P}$ на алгебрі $\mathfrak{A}$ функція

$$\mathbf{P}^*(B) = \inf(\textstyle\sum_{n\ge 1} \mathbf{P}(A_n), \ A_n \in \mathfrak{A}, \cup_{n\ge 1} A_n \supset B)$$

є неспадною і сигма-адитивною за $B \subset \Omega$, та є ймовірністю на $\sigma[\mathfrak{A}]$.

(7) Нехай $\mathfrak{F} = \sigma[\mathfrak{A}]$, а $\mathbf{P}$ – імовірність на $\mathfrak{F}$. Довести, що для довільної $A \in \mathfrak{F}$ знайдуться $A_n \in \mathfrak{A}$ такі, що $\mathbf{P}(A\Delta A_n) \to 0, n \to \infty$.

(8) Розглянемо алгебру $\mathfrak{A}_+$ підмножин $\mathbb{R}_+\backslash\{0\}$, що породжується системою напівінтервалів вигляду $(a, b]$, де $0 \le a < b$. Визначимо функцію $m(A) = 1\!\!1_{0\in Cl(A)}$ для $A \in \mathfrak{A}_+$, де $Cl(A)$ – замикання множини A. Довести, що функція m адитивна, але не є сигма-адитивною.

(9) Нехай Ω – нескінченна множина, а $\mathfrak{F}$ містить всі скінченні підмножини Ω та їх доповнення. Довести, що функція $\nu(A) = 1\!\!1_{|A|=\infty}$ є адитивною та не є сигма-адитивною на $\mathfrak{F}$.

(10) Нехай $\mathfrak{F}$ – сигма-алгебра, $\mathbf{P}$ – імовірність на $\mathfrak{F}$ і $C \subset \Omega, C \notin \mathfrak{F}$. Як подовжити $\mathbf{P}$ до ймовірності на $\sigma[\mathfrak{F} \cup \{C\}]$?

(11) Нехай $(A_k, k = \overline{1,n})$ – випадкові події. Визначимо $S_0 = 1$, та

$S_k = \sum_{1\le i_1 < \ldots < i_k \le n} \mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_k})$.

(а) Нехай B_m – подія, що полягає у тому, що одночасно відбудеться точно m подій з сукупності (A_k). Довести, що $\mathbf{P}(B_m) = \sum_{k=m}^{n}(-1)^{k-m}C_k^m S_k$.

(б) Нехай C_m – подія, що полягає у тому, що одночасно відбудеться принаймні m подій з сукупності (A_k). Довести, що $\mathbf{P}(C_m) = \sum_{k=m}^{n}(-1)^{k-m}C_{k-1}^{m-1} S_k$.

(в) Довести для всіх $m \le n/2$ нерівності Бонфероні:

$\sum_{k=1}^{2m}(-1)^{k-1}S_k \le \mathbf{P}(\cup_{k=1}^{n} A_k) \le \sum_{k=1}^{2m-1}(-1)^{k-1}S_k$.

(г) Довести тотожність Пуанкаре $\mathbf{P}(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n}(-1)^{k-1}S_k$.

(д) Довести нерівності Фреше: $S_{k+1}/C_n^{k+1} \le S_k/C_n^k$ при $k < n$.

(е) Довести нерівності Гумбела: $(C_n^{k+1} - S_{k+1})/C_{n-1}^k \le (C_n^k - S_k)/C_{n-1}^{k-1}$.

1.3. Приклади ймовірнісних просторів

Для застосувань імовірнісних методів на практиці необхідно будувати *ймовірнісні простори*. Зокрема, необхідним елементом розв'язання будь-якої ймовірнісної задачі є побудова відповідного ймовірнісного простору. Тому розглянемо конкретні приклади таких просторів.

1.3.1. Класичне означення ймовірностей

Нехай простір $\Omega = \{\omega_1, ..., \omega_n\}$ – скінченний, а клас подій $\mathfrak{F}$ містить всі підмножини Ω.

Припустимо, що всі елементарні події рівноймовірні (рівноможливі, симетричні, статистично нерозрізненні). Тоді їх імовірності мають бути однаковими і в сумі повинні становити 1. Отже, $\mathbf{P}(\{\omega_k\}) = 1/n$, $k = \overline{1,n}$. Оскільки кожна подія є об'єднанням одноточкових елементарних подій, то ймовірність будь-якої події має дорівнювати сумі доданків $1/n$ по всіх m елементарних подіях, що їй сприяють, тобто дорівнювати відношенню m/n. Звідси дістанемо таке означення.

Означення. *Будемо говорити, що для даного ймовірнісного простору* $(\Omega,\ \mathfrak{F},\ \mathbf{P})$ *зі скінченним простором елементарних подій* Ω *прийняте* **класичне означення ймовірностей,** *якщо для будь-якої події* $A \subset \Omega$ *її ймовірність дорівнює відношенню кількості елементарних подій, що сприяють* A*, до кількості усіх елементарних подій:*

$$\mathbf{P}(A) = |A|\ /\ |\Omega|\,,$$

де $|A|$ *– кількість елементів (потужність) множини* A*.*

Зауваження. Якщо у формулюванні задачі явно не вказано на вибір означення ймовірностей, однак $|\Omega|\ < \infty$ і присутні ключові слова **випадково, навмання** ... – то треба обрати саме класичне означення.

Приклади

1. *Підкидання симетричної гральної кості*. Імовірність парної кількості очок дорівнює $|\{2,4,6\}|\,/6 = 3/6 = 1/2$.

2. *Підкидання навмання двох симетричних монет*. Імовірність хоча б одного аверса $|\{\text{АР,РА,АА}\}|\,/4 = 3/4$.

3. *Випробування до першого успіху* – не може бути описане класичним означенням. Дійсно, припустимо рівноможливість зліченної кількості випробувань. Якщо спільна (однакова для різних елементарних подій) ймовірність додатна, то їх зліченна сума нескінченна, якщо ж вона нульова – то і сума нульова. У кожному випадку сума ймовірностей елементарних подій не дорівнює одиниці. Це суперечить *аксіомам теорії ймовірностей.*

4. *Послідовне підкидання* n симетричних монет як модель випадкового блукання. Простір елементарних подій має вигляд

$$\Omega = \{\omega = (\varepsilon_1\varepsilon_2...\varepsilon_n),\ \varepsilon_k \in \{+1,-1\}\} = \{+1,-1\}^n,$$

де $\varepsilon_k = +1$, якщо k -е випробування завершилося аверсом, і $\varepsilon_k = -1$ у протилежному випадку. Всього є 2^n елементарних подій, з них призводять до k аверсів та решті $n-k$ реверсів події, що відповідають сполученням із n елементів по k. Ця кількість дорівнює C_n^k. Оскільки монета симетрична, то всі елементарні події вважаються рівноможливими, імовірність кожної дорівнює 2^{-n}.

Розглянемо таку модель блукання. Припустимо, що частинка в момент k змінює свою координату (стрибає) на величину ε_k. Позначимо через

$S_n = \varepsilon_1 + ... + \ \varepsilon_n$ результуюче положення частинки. Тоді ймовірність наявності k додатних та $n-k$ від'ємних стрибків дорівнює

$$\mathbf{P}(S_n = k - (n-k)) = C_n^k \ / \ 2^n, \ k = \overline{0,n}.$$

Зокрема, при $n = 2m$ та $k = m$ імовірність повернення в 0 дорівнює

$$\mathbf{P}(S_{2m} = 0) = 2^{-2m} C_{2m}^m.$$

5. *Задача про збіг.* Навмання розкладено n листів до n конвертів з адресами. Простір елементарних подій збігається з множиною усіх перестановок порядку n, а саме – перестановок листів по відношенню до правильних адрес.

Розглянемо події $A = \{$хоча б один лист дійшов за призначенням$\}$, $A_k = \{k$-ий лист дійшов за призначенням$\}$. За *формулою включення - виключення*

$$\mathbf{P}(A) = \mathbf{P}(\cup_{k=1}^n A_k) =$$
$$\sum\nolimits_{1 \le i \le n} \mathbf{P}(A_i) - \sum\nolimits_{1 \le i < j \le n} \mathbf{P}(A_i \cap A_j) + .. + (-1)^{n-1} \ \mathbf{P}(A_1 \cap ... \cap A_n) =$$
$$C_n^1 \frac{(n-1)!}{n!} - C_n^2 \frac{(n-2)!}{n!} + ... + (-1)^{n-1} C_n^n \frac{0!}{n!} =$$
$$1 \ - \frac{1}{2!} \ + \frac{1}{3!} - ... + (-1)^{n-1} \frac{1}{n!} \to 1 - \frac{1}{e}, \ n \to \infty.$$

Вправи

(1) У групі n студентів. Яка ймовірність того, що хоча б у двох з них однакові дні народження ? Для яких n ця ймовірність не менша за 0.5 ?

(2) Кожна з n різних частинок потрапляє в одну з m комірок. (а) Знайти ймовірність того, що для кожного $k = \overline{1,m}$ у k-ту комірку потрапить n_k частинок. (б) Знайти ймовірність потрапляння у фіксовану комірку k частинок. За якої умови існує границя цієї ймовірності при $n, m \to \infty$? Знайти цю границю. (в) Яка ймовірність того, що буде зайнято рівно l комірок ?

(3) Кожна з n ідентичних (нерозрізнимих) частинок потрапляє в одну з m комірок. Рівноможливими є всі розміщення, що відрізняються кількістю частинок у комірках. Знайти ймовірності з пунктів (а) – (в) попередньої вправи.

(4) Кожна з n ідентичних частинок потрапляє в одну з m комірок. Рівноможливими є всі розміщення, що задовольняють забороні Паулі - у кожній комірці може бути не більше однієї частинки. Знайти ймовірності з пунктів (а) – (в) вправи 2.

1.3.2. Дискретний імовірнісний простір

Не завжди можна вважати, що елементарні події рівноможливі. Тоді застосовують таке узагальнення.

Нехай простір $\Omega = \{\omega_1, ..., \omega_n, ...\}$ – дискретний (скінченний або зліченний), клас випадкових подій $\mathfrak{F}$ містить всі підмножини Ω, і задана певна *ймовірність* $\mathbf{P}$ на $\mathfrak{F}$. Тоді визначені ймовірності окремих *елементарних подій* $p_k = \mathbf{P}(\{\omega_k\})$.

Вправа. Сигма-алгебра випадкових подій $\mathfrak{F} = 2^\Omega$ тоді й тільки тоді, коли вимірні всі одноточкові множини: $\{\omega_n\} \in \mathfrak{F}$ для всіх n.

Оскільки для довільної зліченної множини $A = \bigcup_{k:\ \omega_k \in A}\{\omega_k\}$, то за властивістю *сигма-адитивності*

$$\mathbf{P}(A) = \sum\nolimits_{k:\ \omega_k \in A} p_k\ , \quad \text{де} \quad p_k = \mathbf{P}(\{\omega_k\}).$$

Числова послідовність $(p_k, k \geq 1)$ задовольняє умови $p_k \geq 0, \sum p_k = 1$ і називається **дискретним розподілом** імовірностей.

Отже, для кожної ймовірності $\mathbf{P}$ на $\mathfrak{F}$ визначений дискретний розподіл на одноточкових множинах.

Навпаки, для кожного *дискретного розподілу* $(p_k,\ k \geq 1)$ наведена формула для ймовірності $\mathbf{P}(A)$ задає загальне означення ймовірностей у **дискретному ймовірнісному просторі.**

Теорема (про сигма-адитивність дискретної ймовірнісної міри). *Для кожного дискретного розподілу ймовірностей* (p_k) *функція*

$$\mathbf{P}(A) = \sum\nolimits_{k:\ \omega_k \in A} p_k,\ A \in \mathfrak{F},$$

є адитивною ймовірністю та неперервною в нулі, а тому є ймовірністю.

Доведення. Невід'ємність та нормованість очевидні. Адитивність є наслідком адитивності суми.

Можна вважати, що $\Omega = \mathbb{N}$. Якщо $A_n \downarrow \varnothing$, то $a_n \equiv \inf A_n \uparrow \infty$, та

$$\mathbf{P}(A_n) = \sum\nolimits_{k:\ \omega_k \in A_n} p_k \leq \sum\nolimits_{k \geq a_n} p_k \to 0,\ n \to \infty.$$

Еквівалентність *неперервності в нулі та сигма-адитивності* встановлено у теоремі *про основні властивості ймовірностей* $\boxdot$

Приклади

1. *Підкидання монети до першого успіху*. Простір елементарних подій має вигляд $\Omega = \{\omega_n = (\text{Н...НУ}),\ n \geq 1\}$. Співставимо елементарну подію ω_n з елементарною подією в серії з n підкидань монети, що містить $n-1$ реверс (неуспіх) та 1 аверс (успіх). Виходячи з цього співставлення,

природно постулювати, що $p_n = \mathbf{P}(\{\omega_n\}) = 2^{-n}, n \geq 1$. Ця послідовність дійсно є *дискретним розподілом*. Тому ймовірність парної кількості підкидань до першого успіху дорівнює $\sum_{k=1}^{\infty} 2^{-2k} = 1/3$.

2. Нехай при підкиданні монети, окрім випадіння аверсу (А) та реверсу (Р), можливе падіння на ребро (О). У цьому випадку простір елементарних подій має вигляд $\Omega = \{\omega_n = (r_1, ...r_{n-1}, \text{А}),\ r_k \in \{\text{Р,О}\},\ n \geq 1\}$. З *рівноможливості* елементарних подій випливає, що $\mathbf{P}(\{\omega_n\}) = 3^{-n}$. Імовірність отримати хоча б один успіх дорівнює $\sum_{n=1}^{\infty} 2^{n-1}3^{-n} = 1$, а парної кількості підкидань до першого успіху – $\sum_{k=1}^{\infty} 2^{2k-1}3^{-2k} = 2/5$.

3. Нехай p, q, r – задані числа з відрізку $[0, 1]$. За яких умов знайдеться *ймовірнісний простір* та події A, B на ньому такі, що $\mathbf{P}(A) = p$, $\mathbf{P}(B) = q$, $\mathbf{P}(A \cap B) = r$?

Розглянемо такі події:

$$H_1 = A \backslash B,\ H_2 = B \backslash A,\ H_3 = A \cap B,\ H_4 = \overline{A \cup B}.$$

Очевидно, що в довільному ймовірнісному просторі події H_k *попарно несумісні*, $\cup_{k=1}^4 H_k = \Omega$, $\sigma[\{A, B\}] = \sigma[\{H_1, ..., H_4\}]$. Отже, події H_k можна розглядати (і будувати) як елементарні, а для цього їх імовірності мають утворювати *дискретний розподіл* імовірностей. Оскільки

$$\mathbf{P}(H_1) = p - r,\ \mathbf{P}(H_2) = q - r,\ \mathbf{P}(H_3) = r,\ \mathbf{P}(H_4) = 1 - (p + q - r),$$

то шукана необхідна і достатня умова має вигляд:

$$0 \leq r \leq p,\ r \leq q,\ p + q - r \leq 1.$$

Вправи

(1) Довести, що множина $\{\mathbf{P}(A),\ A \in \mathfrak{F}\} \subset [0, 1]$ є замкненою.

(2) Нехай $\Omega = \{\omega_n, n \geq 1\}$ – дискретний простір з розподілом ймовірностей $p_n = \mathbf{P}(\{\omega_n\})$. Довести, що рівність $\{\mathbf{P}(A), A \subset \Omega\} = [0, 1]$ виконується тоді й тільки тоді, коли $p_n \leq \sum_{k>n} p_k$ для всіх $n \geq 1$.

(3) Нехай $(\Omega,\ \mathfrak{F},\ \mathbf{P})$ – довільний імовірнісний простір. Випадкова подія A називається атомом, якщо $\mathbf{P}(A) > 0$ та для будь-якої події з $B \subset A$ випливає або $\mathbf{P}(B) = 0$, або $\mathbf{P}(A \backslash B) = 0$. Атоми A, B вважаються однаковими, якщо $\mathbf{P}(A \Delta B) = 0$. Довести такі твердження:. (а) Якщо A, B є атомами, то або $\mathbf{P}(A \cap B) = 0$, або $\mathbf{P}(A \Delta B) = 0$. (б) Множина всіх атомів на даному просторі не більш ніж злічення. (в) Існує розбиття $\Omega = \cup_{n \geq 0} A_n$ на попарно несумісні події такі, що $A_n, n \geq 1$, є атомами, а A_0 не містить жодного атома. (г) Для довільного $p \in [0, \mathbf{P}(A_0)]$ знайдеться подія A така, що $\mathbf{P}(A) = p$.

(4) Нехай $\Omega = \{\omega_n, n \geq 1\}$ – дискретний простір з імовірністю $\mathbf{P}$, а послідовність множин $\Omega_n \subset \Omega$ така, що $\Omega_n \uparrow \Omega, n \to \infty$. Довести, що для кожної події $A \subset \Omega$ $\mathbf{P}(A) = \lim_{n \to \infty} \mathbf{P}_n(A)$, де $\mathbf{P}_n(A) \equiv \mathbf{P}(A \cap \Omega_n)/\mathbf{P}(\Omega_n)$.

1.3.3. Геометричне означення ймовірностей

На прикладі задачі про випадковий вибір точки на відрізку $[a, b)$ бачимо, що існують стохастичні експерименти з незліченною кількістю елементарних подій. Хоча окремі елементарні події тут виглядають рівноможливими, класичне означення застосувати неможливо – всі ймовірності елементарних подій мають бути нульовими, тому що сума незліченної кількості додатних чисел нескінченна. Тому природно перейти від імовірностей окремих точок до ймовірностей інтервалів.

Нехай $\Omega = [a, b)$, сигма-алгебра подій $\mathfrak{F}$ містить всі інтервали $[x, y)$, і на $\mathfrak{F}$ задана деяка ймовірність $\mathbf{P}$. Припущення про рівноможливість окремих точок можна проінтерпретувати так, що ймовірність потрапляння випадкової точки до певного інтервалу залежить лише від його довжини, а не від розташування цього інтервалу. З цього припущення, адитивності, нормованості та *неперервності ймовірності* випливає, що

$$\mathbf{P}([x, y)) \ = (y - x)/(b - a) = |[x, y)| \ / \ |[a, b)| \,.$$

Дійсно, імовірності $\mathbf{P}([a + (b - a)\frac{k-1}{n}, \ a + (b - a)\frac{k}{n}))$ при $k = \overline{1, n}$ мають не залежати від k та в сумі становити 1, отже кожна з них дорівнює $1/n$. Звідси виводимо, що $\mathbf{P}([a, a + (b - a)\frac{k}{n})) = \frac{k}{n}$. Обираючи підпослідовність $k/n \uparrow (y - a)/(b - a)$, із *неперервності ймовірності* отримуємо рівність

$$\mathbf{P}([a, y)) = |[a, y)| \ / \ |[a, b)| \,.$$

У загальному випадку за припущення рівноможливості приходимо до такого означення.

Означення. *Нехай простір $\Omega \subset \mathbb{R}^d$ є обмеженою множиною евклідового простору $\mathbb{R}^d$, а клас $\mathfrak{F}$ містить усі підмножини Ω із визначеним d-вимірним об'ємом. Будемо говорити, що у просторі $(\Omega, \ \mathfrak{F}, \ \mathbf{P})$ прийнято* **геометричне означення ймовірностей**, *якщо для всіх $A \in \mathfrak{F}$*

$$\mathbf{P}(A) = |A| \ / \ |\Omega| \,,$$

де $|A|$ – d-вимірний об'єм множини A, тобто довжина при $d = 1$, площа при $d = 2$, об'єм при $d = 3$ і так далі.

Зауваження. Якщо у формулюванні задачі явно не вказано на вибір означення ймовірностей, але $\Omega \subset \mathbb{R}^d$ і присутні ключові слова **випадково, навмання** – то треба обрати саме геометричне означення.

Приклади

1. *Задача про зустріч*. Дві особи домовилися про зустріч між 12-ю та 1-ю годиною пополудні. Відомо, що моменти їх приходу до місця зустрічі

– випадкові. Прийшовши до місця, кожен чекає на іншого протягом чверті години, а потім залишає місце зустрічі. Яка ймовірність того, що зустріч відбудеться?

Простір елементарних подій має вигляд

$$\Omega = [0,1] \times [0,1] = \{\omega = (\omega_1, \omega_2) : \omega_i \in [0,1]\}.$$

Ключове слово "випадковими" слід трактувати на користь прийняття геометричного означення ймовірностей. Подія, що відповідає зустрічі, має вигляд

$$A = \{\omega = (\omega_1, \omega_2) \in \Omega : |\omega_1 - \omega_2| \le 1/4\}.$$

Тому $\mathbf{P}(A) = |A| \,/\, |\Omega| = (1 - (3/4)^2)/1 = 7/16$.

2. *Парадокс Бертрана*. Випадково обирається хорда одиничного кола. Подія A полягає в тому, що довжина хорди не перевищує 1. Розглянемо такі підходи до обчислення ймовірності цієї події.

(а) Хорда обирається за *випадковим положенням її вершин*. Унаслідок інваріантності довжини відносно поворотів одну вершину можна зафіксувати. Тому $\Omega = [0, 2\pi]$, множина елементарних подій, які сприяють події A, містить дві дуги довжини $\pi/3$, а ймовірність A дорівнює

$$\mathbf{P}(A) = (\pi/3 + \pi/3)/2\pi = 1/3.$$

(б) Хорда визначається *випадковою точкою перетину із заданим радіусом*. Простір $\Omega = [0,1]$. Сприяють події A точки радіуса, що ближчі до кола ніж точка перетину із перпендикулярною хордою одиничної довжини. Тому

$$\mathbf{P}(A) = 1 - \sqrt{3}/2.$$

(в) Хорда визначається *випадковою точкою перетину з перпендикулярним до цієї хорди радіусом* всередині кола. Простір елементарних подій Ω містить всі точки всередині кола, а елементарні події з A утворюють кільце внутрішнього радіуса $\sqrt{3}/2$. Тому

$$\mathbf{P}(A) = (\pi - \pi(\sqrt{3}/2)^2)/\pi = 1/4.$$

Різні відповіді на наведені задачі вказують на те, що одне й те саме "текстове" формулювання задачі може призводити до різних імовірнісних просторів і, як наслідок, до різних теоретико-ймовірнісних задач. Тому парадокс Бертрана є цілком уявним.

3. *Задача Бюффона*. Голка *навмання* кидається на площину із системою паралельних прямих на однаковій відстані $2a$ одна від одної. Подія A полягає в тому, що має місце перетин голки із однією з прямих.

Простір елементарних подій задається парами (x, φ), де $x \in [0, a]$ – відстань від центра голки до найближчої прямої, а $\varphi \in [0, \pi]$ – кут між голкою та напрямом прямих, тобто $\Omega = [0, a] \times [0, \pi]$. Оскільки Ω є підмножиною евклідового простору, а вибір виконується навмання, то слід прийняти *геометричне означення ймовірностей*.

Перетин голки із найближчою прямою можливий тоді й тільки тоді, коли відстань x не перевищує довжини $l \sin \varphi$ проекції голки на ортогональний до прямих напрям. За припущення $2\,l \leq 2\,a$ відповідна множина елементарних подій містить всі пари (x, φ) такі, що $x \leq l \sin \varphi$. Обчислюючи за геометричним означенням площу вказаної множини, знаходимо

$$\mathbf{P}(A) = \int_0^\pi l \sin \varphi \; d\varphi \; / \; \pi a = 2\,l / \pi a.$$

Цей результат дає можливість статистичного оцінювання числа π через частоту перетинів $\nu_n(A)$ з формули $\pi \approx 2\,ln \; / \; a \; \nu_n(A)$.

Вправи

(1) У сфері радіусу R навмання обрано N точок. Знайти ймовірність того, що відстань до центру найближчої точки не перевищуватиме r. За яких умов границя цієї ймовірності при $N, R \to \infty$ є додатною ?

(2) Яка ймовірність того, що з трьох навмання узятих відрізків довжини не більше a можна побудувати трикутник ?

(3, Лаплас) Знайти ймовірність перетину голки довжини $l < 1$, що підкинута навмання, зі сторонами довжини 1 клітчастого паперу на столі.

(4) Якої товщини має бути монета, щоб ймовірність її падіння на ребро після випадкового підкидання дорівнювала $1/3$?

1.4. Умовні ймовірності

Умовні ймовірності використовуються у випадку, коли в ході стохастичного експерименту стає відомою певна інформація про його проміжні результати.

Приклад. Нехай у просторі $\Omega = \{\omega_1, ..., \omega_n\}$ прийнято *класичне означення ймовірностей*, $A, B \subset \Omega$, $n_A = |A|$, $n_B = |B|$, $n_{AB} = |A \cap B|$. За означенням, $\mathbf{P}(A) = n_A / n$. Припустимо, стає відомим, що подія B вже відбулася. Тоді кількість усіх можливих у подальшому *елементарних подій* зменшиться з n до n_B, а тих, що сприяють A, з n_A до n_{AB}. Отже, в новому (умовному) стохастичному експерименті ймовірність A дорівнює

$$\mathbf{P}_B(A) = n_{AB} \; / \; n_B = (n_{AB}/n) \; / \; (n_B/n) = \mathbf{P}(A \cap B) \; / \; \mathbf{P}(B).$$

1.4.1. Означення та властивості

Означення. *Нехай $A, B \in \mathfrak{F}$ і $\mathbf{P}(B) > 0$.* Умовною ймовірністю *події A за умови події B називається відношення ймовірності їх перерізу до ймовірності умови:*

$$\mathbf{P}(A \mid B) \equiv \mathbf{P}(A \cap B) \ / \ \mathbf{P}(B).$$

Приклади

1. *Зміна шансів на хоча б одну шістку при двох підкиданнях кості, якщо відома сума очок.* Імовірність хоча б однієї шістки при 2 підкиданнях кості дорівнює $1 - (5/6)^2 = 11/36$. Якщо ж стає відомим, що сума очок дорівнює 8, то відповідна умовна ймовірність дорівнює

$|\{(26), (62)\}| \ / \ |\{(26), (35), (44), (53), (62)\}| = 2/5 > 11/36.$

2. *Вибір двох тузів із колоди без повернення.* Імовірність обрати другим туза за умови, що першою картою без поверення обрано туза, дорівнює $3/51$, на відміну від безумовної ймовірності $4/52$.

Зауваження. Якщо текстове формулювання задачі містить зворот "Відомо, що ..." щодо результату стохастичного експерименту, а задача полягає у знаходженні ймовірності певної події, то мається на увазі знаходження саме *умовної ймовірності* вказаної події, за умови, що відбулася наведена на початку подія.

Теорема (про властивості умовної ймовірності). *Нехай $B \in \mathfrak{F}$ і $\mathbf{P}(B) > 0$. Тоді*

(а) $\mathbf{P}(\,\cdot \mid B) \geq 0$,

(б) $\mathbf{P}(\Omega \mid B) = \mathbf{P}(B \mid B) = 1$,

(в) $\mathbf{P}(\cup A_n \mid B) = \sum \mathbf{P}(A_n \mid B)$ *для попарно несумісних $A_n \in \mathfrak{F}$.*

Доведення

(а) Невід'ємність є наслідком означення та невід'ємності ймовірності.

(б) Нормованість випливає з рівності $\mathbf{P}(B \cap B) = \mathbf{P}(B)$.

(в) За означенням умовної ймовірності внаслідок *сигма-адитивності* ймовірності з попарної несумісності подій $A_n \cap B$ отримуємо:

$$\mathbf{P}(\cup A_n \mid B) = \mathbf{P}((\cup A_n) \cap B)/\mathbf{P}(B) = \mathbf{P}(\cup(A_n \cap B))/\mathbf{P}(B) =$$

$$\sum \mathbf{P}(A_n \cap B)/\mathbf{P}(B) = \sum \mathbf{P}(A_n \mid B) \ \boxdot$$

Наслідок. *Умовна ймовірність $\mathbf{P}(A \mid B)$ як функція події A має всі основні властивості ймовірностей, що випливають з аксіом та викладені в теоремі про основні властивості ймовірностей.* Дійсно, твердження (а),(б),(в) теореми повністю відповідають *аксіомам теорії ймовірностей (P1),(P2),(P3)*.

1.4.2. Імовірність перерізу випадкових подій

Умовна ймовірність визначається через вихідну, початкову ймовірність, що одночасно описує як проміжний, так і заключний етапи стохастичного експерименту. При формулюванні складних стохастичних експериментів часто діють зворотним чином: спочатку задають імовірності проміжних результатів та умовні ймовірності заключних результатів експерименту, а вже через них обчислюють сумісні ймовірності всіх подій.

Наприклад: вибір 2 тузів без повернення можна описати за допомогою ймовірностей вибору тузів чи інших карт на першому етапі та умовних імовірностей вибору тузів на другому кроці за умовою, що першим обрано туза чи іншу карту. Тому наступна теорема не є тавтологією.

Теорема (про ймовірність перерізу подій). *Нехай* $A, B \in \mathfrak{F}$, *причому* $\mathbf{P}(B) > 0$. *Тоді*

$$\mathbf{P}(A \cap B) = \mathbf{P}(B) \cdot \mathbf{P}(A \mid B).$$

Доведення очевидне $\boxdot$

Зауваження. У випадку, коли $\mathbf{P}(B) = 0$, наведена формула також справедлива, оскільки за *монотонністю ймовірності* $\mathbf{P}(A \cap B) \le \mathbf{P}(B) = 0$, а права частина завжди нульова. Тому доцільно визначити $\mathbf{P}(A \mid B) = 0$ у випадку, коли $\mathbf{P}(B) = 0$.

Теорема (про ймовірність перерізу n подій). *Нехай* $A_k \in \mathfrak{F}$ *такі, що* $\mathbf{P}(A_k) > 0$, $k = \overline{0, n}$. *Тоді*

$$\mathbf{P}\left(\bigcap\nolimits_{0 \le k \le n} A_k\right) = \mathbf{P}(A_0) \prod\nolimits_{1 \le k \le n} \mathbf{P}(A_k \mid A_0 \cap \ldots \cap A_{k-1}).$$

Доведення. Формула виводиться з теореми *про ймовірність перерізу подій* за індукцією.

Вправи

(1) При підкиданні n гральних костей випала принаймні одна шістка. Яка ймовірність того, що їх не менше k ?

(2) Кидають три гральних кості. Яка ймовірність того, що випаде принаймні одна шістка, якщо відомо, що всі три очки, що випали, є різними ? Як ця умовна ймовірність відрізняється від безумовної ?

1.4.3. Формула повної ймовірності

Розглянемо деякий двохетапний стохастичний експеримент (наприклад, послідовний вибір 2 карт без повернення). Фіксуючи результат першого етапу, ми не отримаємо елементарну подію – тому що наразі невідомий результат другого етапу. Одночасно події, що пов'язані з результатами першого етапу, мають всі характерні властивості елементарних подій: вони *попарно несумісні* та в *об'єднанні* складають весь простір Ω.

Означення. *Послідовність подій* $(H_k,\ k \geq 1)$ *називається* **повною групою подій** (гіпотез), *якщо* $H_i \cap H_j = \varnothing$ *при* $i \neq j$, *та* $\cup H_k = \Omega$. *Еквівалентне означення полягає в тому, що* $(H_k,\ k \geq 1)$ *утворюють* **розбиття** *простору* Ω, *тобто в результаті проведення стохастичного експерименту відбувається одна і тільки одна подія з повної групи:* $\forall \omega \in \Omega\ \exists! n \geq 1 : \omega \in H_n$.

Теорема (формула повної ймовірності). *Нехай* $(H_k,\ k \geq 1)$ – *повна група подій. Для кожної події* $A \in \mathfrak{F}$ *виконується рівність*

$$\mathbf{P}(A) = \sum\nolimits_{k\geq 1} \mathbf{P}(A \mid H_k)\, \mathbf{P}(H_k).$$

Доведення. Оскільки $A = \cup_{\,k\geq 1}(A \cap H_k)$, де події в правій частині *попарно несумісні*, то за властивістю *сигма-адитивності* та за теоремою *про ймовірність перерізу подій*

$$\mathbf{P}(A) = \sum\nolimits_{k\geq 1} \mathbf{P}(A \cap H_k) = \sum\nolimits_{k\geq 1} \mathbf{P}(A \mid H_k)\, \mathbf{P}(H_k)\ \boxdot$$

Зауваження. Дана формула справедлива і у випадку, коли $\mathbf{P}(H_k) = 0$ для деяких k, оскільки за означенням добуток нуля на будь-яке число є нульовим.

Приклади

1. *Випадковий вибір кулі з випадково обраної урни*. Маємо n урн, у k -й урні k білих куль та $n - k$ чорних. Навмання обирається урна, а з неї – випадкова куля. Яка ймовірність того, що куля біла ? Для розв'язання позначимо через A відповідну подію та визначимо *повну групу подій* $H_k = \{$обрано k -у урну$\}$. Тоді за *класичним означенням імовірностей* отримуємо $\mathbf{P}(H_k) = 1/n$ та $\mathbf{P}(A \mid H_k) = k/n$. Отже, шукана ймовірність дорівнює $\mathbf{P}(A) = \sum_{k=1}^{n}(k/n)/n = (n+1)/2n$.

2. *Який білет більш щасливий: обраний першим чи останнім студентом ?* Екзамен складають n студентів, які послідовно обирають без повернення по одному білету з первісної загальної кількості m. Ця кількість містить $k \leq m$ щасливих білетів. Позначимо через $A_n,\ n \geq 1$, подію, що полягає у виборі щасливого білета n-м студентом. Тоді кожна з пар подій $(A_n, \overline{A_n})$ є повною групою подій. За *класичним означенням імовірностей* $\mathbf{P}(A_1) = k/m$. Тому за *формулою повної ймовірності*

$\mathbf{P}(A_2) = \mathbf{P}(A_2 \mid A_1)\mathbf{P}(A_1) + \mathbf{P}(A_2 \mid \overline{A_1})\mathbf{P}(\overline{A_1}) = \frac{k-1}{m-1}\frac{k}{m} + \frac{k}{m-1}\frac{m-k}{m} = \frac{k}{m}$.

За індукцією доводимо, що $\mathbf{P}(A_j) = k/m$ при $j \leq m$.

Вправи

(1) У кожній з m урн знаходяться k білих та $n-k$ чорних куль. З першої урни навмання обрали кулю та переклали у другу, потім випадково обрану кулю з другої урни переклали у третю і так далі. Знайти ймовірність витягнути білу кулю з останньої урни.

(2) Імовірність того, що у випадково обраній сім'ї n дітей, дорівнює αp^n при $n \geq 1$, та $1 - \alpha p/(1-p)$ при $n = 0$, а ймовірність народження хлопчика – q. (а) Знайти ймовірність того, що сім'я має k дітей – хлопчиків. (б) Відомо, що у сім'ї хоча б один хлопчик. Яка ймовірність того, що хлопчиків не менше двох ?

(3) Урна містить n пронумерованих куль. З неї навмання з поверненням k разів обирають кулю та записують її номер. Яка ймовірність того, що серед записаних є точно m різних номерів ?

1.4.4. Формула Байєса

Наступна формула дає можливість послідовно уточнювати ймовірності *повної групи подій* за результатами спостережень на підставі проміжних результатів стохастичного експерименту.

Теорема (формула Байєса). *Нехай* $(H_k,\ k \geq 1)$ – *повна група подій, і* $B \in \mathfrak{F}$ *із* $\mathbf{P}(B) > 0$. *Тоді*

$$\mathbf{P}(H_k \mid B) = \frac{\mathbf{P}(B \mid H_k)\,\mathbf{P}(H_k)}{\sum_{j\geq 1} \mathbf{P}(B \mid H_j)\,\mathbf{P}(H_j)}.$$

Доведення. За означенням *умовної ймовірності* та за *формулою повної ймовірності*

$$\mathbf{P}(H_k \mid B) = \mathbf{P}(H_k \cap B)/\mathbf{P}(B) = \mathbf{P}(B \mid H_k)\mathbf{P}(H_k)\ /\ \sum \mathbf{P}(B \mid H_j)\,\mathbf{P}(H_j)\ \boxdot$$

Приклади

1. *Оцінка невідомого складу урни за кольором навмання обраної кулі.* Урна містить n куль, із них $k \in [0, n]$ білі (гіпотеза H_k), інші – чорні. Всі апріорні припущення щодо вмісту урни рівноможливі. З урни навмання обрано кулю, яка виявилася білою. Як треба змінити ймовірності гіпотез щодо складу урни?

Нехай A – подія, що полягає у виборі білої кулі. За умовою рівноможливості гіпотез $\mathbf{P}(H_k) = 1/(n+1)$,

$$\mathbf{P}(H_k \mid A) = \tfrac{k}{n}\tfrac{1}{n+1}\left(\sum_{k=0}^{n} \tfrac{k}{n}\tfrac{1}{n+1}\right)^{-1} = \tfrac{2k}{n(n+1)}.$$

2. *Відсоток виробництва бракованої продукції одного з двох заводів*. На ринку представлено продукцію двох телевізійних заводів. Перший

продукує 10000 телевізорів на рік, другий – 80000. Перший завод має два проценти бракованої продукції, другий – 0.5 процента. Навмання обраний телевізор виявився бракованим. Який з виробників є для нього найбільш імовірним ?

Позначимо через $H_k, k = 1, 2$, гіпотезу щодо походження телевізора з k-го заводу, A – подію, що полягає в тому, що навмання обраний телевізор є бракованим. Тоді

$\mathbf{P}(H_1) = 10000/(10000 + 80000) = 1/9,$

$\mathbf{P}(H_1 \mid A) = \frac{0.02 \cdot 1/9}{0.02 \cdot 1/9 + 0.005 \cdot 8/9} = 1/3, \ \mathbf{P}(H_2 \mid A) = 2/3.$

Вправи

(1, Льюіс Керрол) В урні знаходиться одна куля невідомого кольору – біла чи чорна. В урну поклали білу кулю. Після цього навмання обрана з урни куля виявилася білою. Яка ймовірність того, що первісно в урні була біла куля ?

(2) Урна містить $m > 3$ білих куль та n чорних. Випадково втрачено одну кулю. Для перевірки складу урни з неї навмання витягли без повернення дві кулі. Вони виявилися білими. Яка ймовірність того, що втрачена куля була білою ?

1.5. Незалежні випадкові події

У теорії ймовірностей використовується специфічне поняття незалежності, яке можна назвати статистичною незалежністю (на відміну від причинної). Дві події вважають незалежними, якщо наявна інформація про одну з них не змінює шансів для іншої. Як вказував А.М. Колмогоров, насамперед поняття незалежності відрізняє предмет теорії ймовірностей від теорії міри.

1.5.1. Незалежні події та їх перетворення

Означення. *Нехай* $A, B \in \mathfrak{F}$ *і* $\mathbf{P}(B) > 0$. *Подія* A *не залежить від події* B, *якщо* $\mathbf{P}(A \mid B) = \mathbf{P}(A)$.

З означення *умовної ймовірності* отримуємо еквівалентне, але більш загальне та симетричне означення *попарної незалежності*.

Означення. *Дві події* A, B *називаються* незалежними, *якщо*

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

Зауваження. При формулюванні складних стохастичних експериментів часто вказують імовірності результатів окремих їх етапів, а також постулюють незалежність подій з різних етапів. Тому наведене означення незалежності використовується як певна теорема **про ймовірність перерізу незалежних подій**, що дозволяє обчислити ймовірність перерізу.

Приклади

1. *Повторне підкидання кості*. Нехай подія A_k полягає у випаданні шістки у k-му підкиданні. Тоді *класичне означення ймовірностей* зводиться до постулювання тотожності $\mathbf{P}(A_1 \cap A_2) = 1/36$. Еквівалентним підходом є прийняття *рівноможливості* $\mathbf{P}(A_k) = 1/6$ та постулювання *незалежності* подій A_k.

2. *Події, що пов'язані із координатами випадкової точки в квадраті*. Нехай $\Omega = [0,1]^2$, прямокутник $A = A_1 \times A_2$. За умови випадковості згідно з *геометричним означенням імовірності* $\mathbf{P}(A) = |A| \,/\, |\Omega|$. Якщо ж постулювати випадковість вибору координат точки та *незалежність* координат, маємо еквівалентний вираз $\mathbf{P}(A) = (|A_1| \,/1) \times (|A_2| \,/1) = |A| \,/1$.

Теорема (про перетворення незалежних подій).

(а) Довільна подія A не залежить від Ω та $\varnothing$.

(б) Якщо події A і B незалежні, то $\overline{A}$ і B також незалежні.

(в) Якщо подія A не залежить від B_1 та від B_2, причому $B_1 \subset B_2$, то A і $B_2 \setminus B_1$ також незалежні.

(г) Якщо подія A не залежить від подій $B_n, n \geq 1$, що є попарно несумісними: $B_i \cap B_j = \varnothing, \forall i \neq j$, то події A і $\cup_{n\geq 1} B_n$ – незалежні.

Доведення

(а) $\mathbf{P}(A \cap \Omega) = \mathbf{P}(A) = \mathbf{P}(A)\ \mathbf{P}(\Omega)$.

(б) $\mathbf{P}(\overline{A} \cap B) = \mathbf{P}(B \setminus A \cap B) = \mathbf{P}(B) - \mathbf{P}(A \cap B) = \mathbf{P}(B) - \mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(\overline{A})\mathbf{P}(B)$.

(в) $\mathbf{P}(A \cap (B_2 \setminus B_1)) = \mathbf{P}(A \cap B_2 \setminus A \cap B_1) = \mathbf{P}(A \cap B_2) - \mathbf{P}(A \cap B_1) = \mathbf{P}(A)(\mathbf{P}(B_2) - \mathbf{P}(B_1)) = \mathbf{P}(A)\ \mathbf{P}(B_2 \setminus B_1)$.

(г) $\mathbf{P}(A \cap (\cup_{n\geq 1} B_n)) = \mathbf{P}(\cup_{n\geq 1}(A \cap B_n)) = \sum_{n\geq 1} \mathbf{P}(A \cap B_n) = \sum_{n\geq 1} \mathbf{P}(A)\mathbf{P}(B_n) = \mathbf{P}(A) \sum_{n\geq 1} \mathbf{P}(B_n) = \mathbf{P}(A)\mathbf{P}(\cup_{n\geq 1} B_n)$ $\boxdot$

1.5.2. Незалежність подій з послідовності, приклад Бернштейна

Поняття незалежності можна поширити на послідовності подій.

Означення. *Події з послідовності* $(A_k,\ k = \overline{1,n})$ *називаються* **попарно незалежними**, *якщо для всіх $i \neq j$ події A_i та A_j незалежні.*

Більш загальне поняття *незалежності* стосується ймовірностей одночасної появи довільної кількості подій.

Означення. *Події з послідовності* $(A_k,\ k=\overline{1,n})$ *називаються* **незалежними в сукупності,** *якщо для довільних натуральних* $m \in (1,n]$, *і довільних натуральних* $1 \le k_1 < k_2 < ... < k_m \le n$

$$\mathbf{P}(A_{k_1} \cap A_{k_2} \cap ... \cap A_{k_m}) = \prod\nolimits_{1\le\ i\ \le m} \mathbf{P}(A_{k_i}).$$

Очевидно, що з *незалежності в сукупності* випливає *попарна незалежність* подій послідовності $(A_k,\ k=\overline{1,n})$ – умови другого означення містяться у першому при $m=2$. Приклад С.Н.Бернштейна доводить хибність оберненого твердження – *з попарної незалежності не випливає незалежність у сукупності.*

Розглянемо правильний тетраедр із 4 гранями. Перші три пофарбовано в синій, жовтий та червоний кольори відповідно. Остання грань розділена на три рівні частини, які теж пофарбовані в синій, жовтий та червоний кольори відповідно.

Тетраедр навмання підкинули і він впав на одну з граней. Нехай С, Ж та Ч – випадкові події, які полягають у тому, що на нижній грані тетраедра присутній синій, жовтий та червоний колір відповідно. Тоді

$\mathbf{P}(\text{С}) = \mathbf{P}(\text{Ж}) = \mathbf{P}(\text{Ч}) = 2/4 = 1/2$,

$\mathbf{P}(\text{С}\cap\text{Ж}) = \mathbf{P}(\text{Ж}\cap\text{Ч}) = \mathbf{P}(\text{С}\cap\text{Ч}) = 1/4 = \mathbf{P}(\text{С}\cap\text{Ж}\cap\text{Ч}) \ne 1/8$.

Отже, події С,Ж,Ч незалежні попарно, але залежні в сукупності.

Вправи

(1) Подія A не залежить від себе самої тоді й тільки тоді, коли $\mathbf{P}(A) \in \{0,1\}$.

(2) Скільки існує пар незалежних подій у ймовірнісному просторі з простою кількістю елементарних подій при класичному означенні ймовірностей ?

(3) Обчислити ймовірність об'єднання незалежних у сукупності подій через імовірності окремих подій.

(4) Відомі ймовірності трьох незалежних подій $\mathbf{P}(A_i) = p_i, i = \overline{1,3}$. Після проведення випробування виявилось, що якісь дві події з A_i відбулися, а третя – ні. Знайти ймовірність того, що за цих умов відбулася подія A_1.

(5) Гравці A і B змагаються на шаховому турнірі з імовірностями перемоги у кожній партії a та b відповідно та ймовірністю нічиєї $c = 1-a-b$. Результати партій незалежні. Турнір виграє той, хто пережене супротивника хоча б на два очки. (а) Знайти ймовірність перемоги A. (б) Яка ймовірність того, що турнір не завершиться? (в) Розв'язати попередні завдання для кількості партій $r > 2$.

(6) Випадкові події $(A_k, k=\overline{1,n})$ незалежні у сукупності. (а) Довести, що події з породженої алгебри $\mathfrak{A}[(A_k, k=\overline{1,n-1})]$ не залежать від A_n. (б) Позначимо $A^\delta = A$ при $\delta = 1$, та $A^\delta = \overline{A}$ при $\delta = 0$. Довести, що для довільних $\delta_k \in \{0,1\}$ випадкові події $(A_k^{\delta_k}, k=\overline{1,n})$ незалежні у сукупності.

1.6. Дискретні випадкові величини

Поняття випадкової події дозволяє вивчати якісні (дихотомічні) наслідки стохастичних експериментів. На практиці важливо вивчати також кількісні результати (наприклад: кількість аверсів при двох підкиданнях монети). Для визначення такої кількості необхідно кожній *елементарній події* поставити у відповідність певне число – кількісний результат експерименту, тобто визначити функцію на просторі елементарних подій. Тому розглянемо таке означення.

Означення. **Дискретною випадковою величиною** *називається відображення $\xi : \Omega \to \mathbb{R}$, яке має скінченну або зліченну множину значень: $\xi(\Omega) \equiv \{\xi(\omega),\ \omega \in \Omega\} = \{x_1, ...x_n, ..\}$, причому для кожного можливого значення x_n його прообраз: $\{\omega : \xi(\omega) = x_n\}$ є випадковою подією, тобто*

$$\{\omega : \xi(\omega) = x_n\} \in \mathfrak{F}, \forall x_n.$$

Остання умова (**вимірність**) є суттєвою, тому що при її порушенні ймовірності деяких висловлювань щодо значень ξ не були б визначеними.

Зауваження. Множина значень $\{x_n\}$ може бути числовою, чи множиною векторів, або навіть підмножиною довільного лінійного простору.

У подальшому будемо вважати, що всі значення $\{x_n\}$ є *різними*.

1.6.1. Розподіл дискретної величини

Означення. **Розподілом** *дискретної випадкової величини $\xi(\omega)$ зі значеннями $\{x_n\}$ називається послідовність імовірностей*

$$(p_n = p_\xi(x_n) \equiv \mathbf{P}(\xi = x_n),\ n \geq 1).$$

Ця послідовність є невід'ємною і в сумі становить 1, тобто є *дискретним розподілом* імовірностей.

Зауваження. Оскільки аргументом будь-якої випадкової величини $\xi(\omega)$ завжди є елементарна подія ω, то при запису випадкової величини цей аргумент часто не вказують і просто пишуть ξ. Крім того, при запису ймовірностей подій, пов'язаних із випадковою величиною, для скорочення часто опускають вказівку на елементарну подію:

$$\{\xi = x\} \equiv \{\omega : \xi(\omega) = x\}.$$

Зокрема, за домовленістю $\mathbf{P}(\xi = x) \equiv \mathbf{P}(\{\omega : \xi(\omega) = x\})$.

Розподіл *дискретних випадкових величин* зображують у таблиці:

$$\left[\begin{array}{cccccc} \xi & x_1 & x_2 & ... & x_n & ... \\ \mathbf{P} & p_1 & p_2 & ... & p_n & ... \end{array}\right].$$

Теорема (про розподіл функції від дискретної величини). *Нехай ξ – дискретна випадкова величина з розподілом $(p_n = p_\xi(x_n))$ та значеннями $\{x_n\}$. Тоді для довільної функції g справедлива рівність*

$$\mathbf{P}(g(\xi) \in B) = \sum\nolimits_{n:\ g(x_n) \in B} p_\xi(x_n).$$

Зауважимо, що значення $g(x_n)$ можуть бути однаковими.

Доведення є наслідком зображення

$$\{g(\xi) \in B\} = \bigcup\nolimits_{n:\ g(x_n) \in B} \{\xi = x_n\},$$

де події справа попарно несумісні, оскільки всі значення x_n – різні $\boxdot$

1.6.2. Індикаторна та проста випадкові величини

Означення. *Нехай A – випадкова подія. Її* індикаторною величиною $1\!\!1_A$ *називається дво-значна випадкова величина*

$$1\!\!1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A. \end{cases}$$

Якщо інтерпретувати значення 1 як "успіх", а 0 – як "неуспіх", то спостереження величини $1\!\!1_A$ називають випробуванням Бернуллі, а її розподіл називається розподілом Бернуллі :

$$\mathbf{P}(1\!\!1_A = x) = p^x (1-p)^{1-x},\ x \in \{0, 1\}.$$

Означення. *Дискретна випадкова величина називається* простою *величиною, якщо вона має скінченну множину значень.*

Очевидно, що сума, добуток, функції від простих величин – прості.

Якщо ξ – проста випадкова величина з різними значеннями $\{x_1, ..., x_n\}$, то $C_k = \{\xi = x_k\} \in \mathfrak{F}$ є *повною групою подій* і має місце зображення

$$\xi = \sum\nolimits_{k=1}^{n} x_k 1\!\!1_{C_k}.$$

Вправа. Довести, що: $1\!\!1_{A \cap B} = 1\!\!1_A\ 1\!\!1_B$, $1\!\!1_{A \cup B} = 1\!\!1_A + 1\!\!1_B - 1\!\!1_{A \cap B}$, $1\!\!1_{A \Delta B} = |1\!\!1_A - 1\!\!1_B|$, $1\!\!1_{\cup A_n} = \sum 1\!\!1_{A_n}$ для попарно несумісних A_n, $1\!\!1_{\overline{A}} = 1 - 1\!\!1_A$, $1\!\!1_A = \lim 1\!\!1_{A_n}$, якщо $A_n \uparrow A$ або $A_n \downarrow A$, $\overline{\lim}_{n\to\infty} A_n = \{\sum 1\!\!1_{A_n} = \infty\}$.

1.6.3. Схема стохастичних випробувань Бернуллі

Значна кількість *дискретних розподілів* пов'язана з таким *стохастичним експериментом.*

Означення. *Схемою* випробувань Бернуллі *називається стохастичний експеримент, що зводиться до послідовності:*

(а) незалежних у сукупності випробувань,

(б) кожне з яких є дихотомічним – закінчується одним із двох результатів: успіх або неуспіх, причому

(в) імовірність успіху не залежить від номера випробування.

Для побудови *ймовірнісного простору*, який відповідає схемі Бернуллі, позначимо через n кількість випробувань, p – імовірність успіху в одному випробуванні, $q = 1-p$ – імовірність неуспіху. Якщо ототожнити значення 1 з успіхом та 0 із неуспіхом, то елементарна подія матиме вигляд

$$\omega = (\omega_1, ..., \omega_n),\ \omega_k \in \{0, 1\}.$$

Отже, шуканий простір елементарних подій є дискретним, множина елементарних подій дорівнює $\Omega = \{\omega\} = \{0, 1\}^n$, *сигма-алгебра* подій містить усі підмножини: $\mathfrak{F} = 2^\Omega$.

Для знаходження ймовірностей зафіксуємо $\varepsilon = (\varepsilon_1, ..., \varepsilon_n) \in \{0, 1\}^n$ та розглянемо одноточкову подію

$$\{\varepsilon\} = \{\omega : \omega = \varepsilon\} = \bigcap\nolimits_{k=1}^{n} \{\omega : \omega_k = \varepsilon_k\}.$$

За умовою події під знаком перерізу *незалежні в сукупності*, оскільки стосуються різних випробувань. Крім того, імовірність результату ε_k у k-му випробуванні дорівнює

$$\mathbf{P}(\{\omega : \omega_k = \varepsilon_k\}) = p\mathbb{I}_{\{\varepsilon_k=1\}} + q\mathbb{I}_{\{\varepsilon_k=0\}} = p^{\varepsilon_k} q^{1-\varepsilon_k}.$$

Звідси за зазначеною вище *незалежністю у сукупності* отримуємо

$$\mathbf{P}(\{\omega : \omega = \varepsilon\}) = \mathbf{P}\left(\bigcap\nolimits_{k=1}^{n} \{\omega : \omega_k = \varepsilon_k\}\right) =$$
$$\prod\nolimits_{k=1}^{n} \mathbf{P}(\{\omega : \omega_k = \varepsilon_k\}) = \prod\nolimits_{k=1}^{n} p^{\varepsilon_k} q^{1-\varepsilon_k} = p^{\nu_n(\varepsilon)} q^{n-\nu_n(\varepsilon)},$$

де функція $\nu_n(\omega)$ дорівнює загальній кількості успіхів (частоті успіхів), що містяться в елементарній події $\omega = (\omega_1, ..., \omega_n)$:

$$\nu_n(\omega) = \sum\nolimits_{k=1}^{n} \omega_k = |\{k : \omega_k = 1\}|\,.$$

Як в усякому *дискретному ймовірнісному просторі*, імовірність будь-якої події дорівнює сумі ймовірностей сприятливих елементарних подій.

1.6.4. Біноміальний розподіл

Означення. *Дискретна випадкова величина ξ має біноміальний розподіл із параметрами n та p, позначення $\xi \simeq B(n,p)$, якщо вона збігається з загальною кількістю успіхів у схемі випробувань Бернуллі з n випробуваннями та ймовірністю успіху p в одному випробуванні.*

Для обчислення розподілу величини ξ зауважимо, що за означенням $\xi(\omega) = \nu_n(\omega)$, де ν_n – визначена вище *кількість успіхів* у *випробуваннях Бернуллі*. Тому

$$\mathbf{P}(\xi = k) = \sum\nolimits_{\omega:\ \xi(\omega)=k} \mathbf{P}(\{\omega\}) = \sum\nolimits_{\omega:\ \nu_n(\omega)=k} p^{\nu_n(\omega)} q^{n-\nu_n(\omega)} =$$

$$|\{\omega : \xi(\omega) = k\}|\ p^k\ q^{n-k}.$$

Оскільки кількість елементарних подій у події $\{\xi = k\}$ дорівнює кількості сполучень (k-елементних підмножин n-елементної множини) C_n^k, то

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

Ця послідовність називається **біноміальним розподілом** імовірностей.

Теорема (про властивості біноміальних імовірностей). *Біноміальні ймовірності $p_n(k)$ не спадають при $k \le np + p$, та не зростають при $k > np + p$, тому* ***найбільш імовірне значення біноміального розподілу*** *дорівнює $k = [np + p]$.*

Доведення. Обчислимо послідовні відношення

$$p_n(k)\ /\ p_n(k-1) = 1 + (np + p - k)\ /\ k(1-p).$$

Права частина менша за 1 тоді й тільки тоді, коли $k > np + p$, звідки випливає необхідний висновок $\boxdot$

Вправи

(1, задача Банаха про сірникові коробки) Один математик завжди носить із собою по коробці сірників у кожній кишені. Кожного разу для виймання сірника він навмання обирає кишеню. Знайти ймовірність того, що у момент, коли він вперше витягне порожню коробку, інша міститиме точно k сірників.

(2, задача Бертрана про балотування) Кандидат A набрав при голосуванні a голосів, а кандидат B – b голосів, причому $a > b$. Голосування $a + b$ виборців проводиться послідовно, один за одним. Яка ймовірність того, що депутат A лідирував весь час ?

(3) Яка ймовірність того, що в нескінченій серії випробувань Бернуллі a успіхів поспіль вперше відбудуться раніше, ніж b неуспіхів ?

(4) У якому випадку найбільш імовірне значення біноміального розподілу досягається у двох точках ?

1.6.5. Поліноміальний розподіл

Означення. Поліноміальною схемою випробувань *називається послідовність з* n *незалежних випробувань, результат кожного з яких належить множині* $\overline{1,k}$ *та має дискретний розподіл* $p = (p_i, i = \overline{1,k})$.

Результатом таких випробувань є вектор $\widehat{\nu}(n) = (\widehat{\nu}_{n1}, ..., \widehat{\nu}_{nk})$, що містить частоти $\widehat{\nu}_{ni}$ для i-го результату у n випробуваннях. З незалежності випробувань випливає формула для поліноміального розподілу

$$\mathbf{P}(\widehat{\nu}_{n1} = i_1, ...\widehat{\nu}_{nk} = i_k) = \frac{n!}{i_1!...i_k!} p_1^{i_1}...p_k^{i_k}.$$

1.6.6. Геометричний розподіл

Означення. *Дискретна випадкова величина* ξ *має геометричний розподіл із параметром* p, *позначення* $\xi \simeq G(p)$, *якщо вона збігається з кількістю випробувань до першого успіху в нескінченній послідовності випробувань Бернуллі з імовірністю успіху* p *в одному випробуванні.*

За властивістю *неперервності ймовірності*, імовірності подій у схемі з необмеженою кількістю випробувань є граничними при $n \to \infty$ для схеми з n випробуваннями. Тому для кожного $1 \le k < n$

$$\mathbf{P}_n(\xi = k) = \sum\nolimits_{\omega_{k+1},..\omega_n \in \{0,1\}} \mathbf{P}_n(\{(\overline{0...01\omega_{k+1}...\omega_n})\}) =$$

$$\mathbf{P}_n(\bigcup\nolimits_{\omega_{k+1},..\omega_n \in \{0,1\}} \{(\overline{0...01\omega_{k+1}...\omega_n})\}) = \mathbf{P}_k(\{(\overline{0..01})\}) = (1-p)^{k-1}p.$$

Отже, геометричний розподіл імовірностей має вигляд

$$\mathbf{P}(\xi = k) = (1-p)^{k-1}p, \ k = 1, 2, ...$$

Іноді розглядають також величину, що дорівнює кількості неуспіхів до першого успіху, з розподілом $\mathbf{P}(\xi = k) = (1-p)^k p, \ k = 0, 1, ...$

Випадкова величина ξ має розподіл Паскаля з параметром $a > 0$, якщо вона має *геометричний розподіл* з параметром $p = 1/(1+a)$.

1.6.7. Негативний біноміальний розподіл

Означення. *Дискретна випадкова величина* ξ *має* негативний біноміальнй розподіл *із параметрами* p, r, *позначення* $\xi \simeq G(p, r)$, *якщо вона збігається з кількістю випробувань до* r*-го успіху в нескінченній* *послідовності випробувань Бернуллі з імовірністю успіху p в одному випробуванні*. Розподіл має вигляд

$$\mathbf{P}(\xi = k) = C_{k-1}^{r-1}\ (1-p)^{k-r} p^r,\ k = r, r+1, ...$$

Зауважимо, що кожна сприятлива елементарна подія довжини k містить $k-r$ неуспіхів та r успіхів, причому останнім має бути успіх. Тому ймовірність такої події дорівнює $(1-p)^{k-r}p^r$, а загальна кількість їх збігається з кількістю розміщень $r-1$ успіху на перших $k-1$ місці.

1.6.8. Гіпергеометричний розподіл

Розглянемо такий стохастичний експеримент. В озері плаває всього N риб, із них n щук, а інші – карасі. Рибалка виловив m риб. Дискретна випадкова величина ξ, що дорівнює кількості щук у вилові, має *гіпергеометричний розподіл* із параметрами N, n, m, позначення $\xi \simeq H(N, n, m)$.

Для обчислення розподілу величини ξ приймемо класичне означення ймовірностей та зауважимо, що кількість всіх елементарних подій дорівнює кількості способів обрати m риб із загального числа N, тобто C_N^m. Кількість елементарних подій, що сприяють події $\{\xi = k\}$, очевидно, дорівнює за основним правилом комбінаторики $C_n^k\ C_{N-n}^{m-k}$ – добуткові кількості способів обрати k щук із загальної кількості n та $m-k$ карасів із кількості $N-n$. Отже, гіпергеометричний розподіл має вигляд

$$\mathbf{P}(\xi = k) = \frac{C_n^k C_{N-n}^{m-k}}{C_N^m},\ k = \overline{0, \min(n, m)}.$$

Зауваження. Якщо проінтерпретувати N як загальну кількість виробів, що виготовлені на підприємстві, n – як кількість бракованих серед них, m – як обсяг вибіркової партії, що була перевірена відділом контролю якості, то випадкова кількість виявлених бракованих виробів матиме гіпергеометричний розподіл. Цей факт використовують, зокрема, при статистичному контролі якості продукції.

1.6.9. Розподіл Пуассона

Означення. *Дискретна випадкова величина ξ має* розподіл Пуассона *з параметром $\lambda > 0$, позначення $\xi \simeq \Pi(\lambda)$, якщо*

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

Нормованість пуассонівських імовірностей випливає з формули розкладу в ряд Тейлора функції $\exp(\lambda)$. Розподіл Пуассона широко використовується в теорії ймовірностей та статистиці для моделювання випадкових потоків подій, таких як кількість телефонних викликів за певний час, кількість захворювань за період, кількість випадково обраних точок у певному об'ємі тощо. Одночасно він є граничним для *біноміального розподілу*.

Вправи

(1) Довести, що негативний біноміальний розподіл задовольняє таку рекурентну тотожність: $\mathbf{P}(\xi = k) = \mathbf{P}(\xi = k-1)(k-1)/(k-r)$.

(2) Найбільш імовірне значення k величини ξ з гіпергеометричним розподілом дорівнює $(n+1)(m+1)/(N+2) - 1$. *Вказівка*: відношення послідовних імовірностей має вигляд $(k+1)(N-n-m+k+1)/(n-k)(m-k)$.

(3) Величини ξ_{Nnm} мають гіпергеометричний розподіл $\xi_{Nnm} \simeq H(N, n, m)$. Довести, що $\mathbf{P}(\xi_{Nnm} = k) \to C_m^k p^k (1-p)^{m-k}$ при $N \to \infty, n/N \to p,\ k \geq 0$.

(4) Довести, що розподіл Пуассона задовольняє таку рекурентну тотожність: $\mathbf{P}(\xi = k) = \mathbf{P}(\xi = k-1)\lambda/k$.

(5) Знайти найбільш імовірне значення для розподілу Пуассона.

1.6.10. Граничні теореми Пуассона і Муавра – Лапласа

Схемою серій випробувань Бернуллі *називається подвійна послідовність випробувань, у якій в n-ій серії (підпослідовності) міститься n випробувань Бернуллі з імовірністю успіху p_n в одному випробуванні.*

Теорема (гранична теорема Пуассона). *Розглянемо схему серій випробувань Бернуллі, в якій n-та серія містить n випробувань з імовірністю успіху p_n. Позначимо через ν_n – кількість успіхів у n-ій серії. Припустимо, що одночасно $n \to \infty$ і $p_n \to 0$ так, що $np_n \to \lambda > 0$. Тоді*

$$\lim_{n\to\infty} \mathbf{P}(\nu_n = k) = \frac{\lambda^k}{k!}\, e^{-\lambda},\ \forall k = 0, 1, ...$$

Доведення. Обчислимо

$$\mathbf{P}(\nu_n = k) = C_n^k\, p_n^k\, (1-p_n)^{n-k} = a_{nk}\, \frac{(np_n)^k}{k!}\, \left(1 - \frac{np_n}{n}\right)^n,$$

де $a_{nk} = (1-p_n)^{-k} \prod_{1 \leq i < k} (1 - \frac{i}{n}) \to 1$, $n \to \infty$, за умовою. Крім того, при кожному фіксованому k має місце збіжність

$$(np_n)^k \to \lambda^k, \quad \left(1 - \frac{np_n}{n}\right)^n \to e^{-\lambda}, n \to \infty \boxdot$$

Приклад. За 1 хв сервер виконав 2000 з'єднань у мережі, з імовірністю 0.001 кожне з них було перерване через перевантаження, незалежно від інших. Тоді ймовірність того, що перервалися принаймні 2 з'єднання, наближено дорівнює $\mathbf{P}(\nu_{2000} \geq 2) \approx 1 - e^{-2} - \frac{2^1}{1!}e^{-2} \approx 0.594$.

Теорема Пуассона діє у випадку, коли ймовірність успіху p_n прямує до нуля. Якщо вона відділена від нуля, слід застосувати таку теорему.

Теорема (локальна гранична теорема Муавра – Лапласа). *Розглянемо схему серій випробувань Бернуллі, в якій n-та серія містить n випробувань з імовірністю успіху p, $q = 1 - p$. Позначимо через ν_n – кількість успіхів у n-ій серії. Припустимо, що числа n і k одночасно прямують до нескінченності так, що нормоване та центроване значення k є обмеженим:*

$$x_{nk} \equiv (k - np)\,/\sqrt{npq} = O(1),\ n, k \to \infty.$$

*Тоді має місце **асимптотична еквівалентність** (тобто відношення правої та лівої частин прямує до одиниці):*

$$\mathbf{P}(\nu_n = k) \sim \frac{1}{\sqrt{npq}}\varphi(x_{nk}),\ n, k \to \infty,$$

*де функція $\varphi(x) = \exp\left(-x^2/2\right)/\sqrt{2\pi},\ x \in \mathbb{R}$, є так званою стандартною **нормальною щільністю розподілу**.*

Доведення теореми спирається на формулу Стірлінга

$$n! \sim \sqrt{2\pi n}(n/e)^n.$$

Зауважимо, що в даній граничній схемі має місце одночасна збіжність $n \to \infty$, $k \to \infty$, $n - k \to \infty$, причому $\varepsilon \equiv (k - np)/n = O(1/\sqrt{n})$. Тому з виразу для *біноміального розподілу* ймовірностей $p_n(k)$ підстановкою $k = n(p + \varepsilon)$ дістанемо

$$\ln\left(\sqrt{2\pi\ npq}\ p_n(k)\right) = \tfrac{1}{2}\ln npq + (n + \tfrac{1}{2})\ln n - n-$$

$$(k + \tfrac{1}{2})\ln k + k - (n - k + \tfrac{1}{2})\ln(n - k) + n - k + k\ln p + (n - k)\ln q + o(1) =$$

$$\tfrac{1}{2}\ln npq + \tfrac{1}{2}\ln\tfrac{n}{k(n-k)} + n\ln n - k\ln k - (n - k)\ln(n - k) +$$

$$k\ln p + (n - k)\ln q + o(1) =$$

$$\tfrac{1}{2}\ln npq + \tfrac{1}{2}\ln\tfrac{n}{n(p+\varepsilon)n(q-\varepsilon)} + n\ln n\ -$$

$$-n(p+\varepsilon)\ln n(p+\varepsilon) - n(q-\varepsilon)\ln n(q-\varepsilon) + n(p+\varepsilon)\ln p + n(q-\varepsilon)\ln q + o(1) =$$

$$-n(p + \varepsilon)\ln(1 + \varepsilon/p) - n(q - \varepsilon)\ln(1 - \varepsilon/q) + o(1) =$$

$$-n(p+\varepsilon)(\varepsilon/p-(\varepsilon/p)^2/2)-n(q-\varepsilon)(-\varepsilon/q-(\varepsilon/q)^2/2)+o(1)=$$
$$-n\varepsilon-n\varepsilon^2/p+n\varepsilon^2/2p+n\varepsilon-n\varepsilon^2/q+n\varepsilon^2/2q+o(1)=$$
$$-n\varepsilon^2/2pq+o(1)=-x_{nk}^2/2+o(1),\ n\to\infty\ \boxdot$$

Теорема (інтегральна гранична теорема Муавра – Лапласа). *За припущень локальної теореми при довільних $a<b$ для випадкових величин $\xi_n\equiv(\nu_n-np)/\sqrt{npq}$ має місце при $n\to\infty$ збіжність*

$$\mathbf{P}\left(a<\xi_n<b\right)\to\Phi(b)-\Phi(a),\ n\to\infty,$$

*до **стандартної нормальної функції розподілу** $\Phi(x)\equiv\int_{-\infty}^{x}\varphi(y)dy$.*

Доведення виводиться з *локальної граничної теореми Муавра- Лапласа*, оскільки за теоремою *про розподіл функції від дискретної величини* ймовірність у лівій частині дорівнює сумі

$$\sum\nolimits_{k:\ (k-np)/\sqrt{npq}\ \in\ (a,b)}p_n(k)=\sum\nolimits_{k=np+a\sqrt{npq}}^{np+b\sqrt{npq}}(\sqrt{npq}\ p_n(k))\ /\ \sqrt{npq},$$

що є інтегральною сумою для $\int_a^b\varphi(y)dy$ $\boxdot$

Зауважимо, що інтегральна теорема є наслідком більш загальної *класичної центральної граничної теореми*, яка буде розглядатися нижче.

Приклад. При 720 підкиданнях гральної кості розглядатимемо як успіх випадання шістки. Для такої схеми *випробувань Бернуллі*

$$np=120,\quad \sqrt{npq}=\sqrt{720\cdot 1/6\cdot 5/6}=10,$$

і справедлива наближена рівність

$$\mathbf{P}(90<\nu_{720}<150)=\mathbf{P}(-3<\frac{\nu_{720}-120}{10}<3)\approx\Phi(3)-\Phi(-3)\approx 0.997.$$

Зауваження. Випадкова величина $\xi_n\equiv(\nu_n-np)h_n$ пропорційна кроку $h_n=1/\sqrt{npq}$ та є дискретною, тому догранична ймовірність у *інтегральній граничній теоремі Муавра – Лапласа* змінюється стрибками у точках вигляду $x_{nk}=(k-np)h_n, k\in\mathbb{Z}_+$, на відміну від граничної ймовірності, що є неперервною за a,b.

Для уникнення похибок у таких випадках використовують таку **поправку на неперервність:** *якщо дискретна випадкова величина ξ має крок h (тобто $\xi\in h\mathbb{Z}$), при її апроксимації притримуються правил:*

$$\mathbf{P}(\xi=x)\equiv\mathbf{P}(x-h/2<\xi<x+h/2),$$
$$\mathbf{P}(\xi\le x)\equiv\mathbf{P}(\xi<x+h/2),\quad \mathbf{P}(\xi<x)\equiv\mathbf{P}(\xi\le x-h/2).$$

Наприклад, для $\xi_n = (\nu_n - np)h_n$:

$$\mathbf{P}(\nu_n = k) = \mathbf{P}(\xi_n = x_{nk}) \equiv$$

$$\mathbf{P}(x_{nk} - h_n/2 < \xi_n < x_{nk} + h_n/2) \approx \Phi(x_{nk} + h_n/2) - \Phi(x_{nk} - h_n/2).$$

Аналогічно, при апроксимації дискретного розподілу $\mathbf{P}(\xi_n = x_{nk})$ наближення $h_n\varphi(x_{nk})$ уточнюють **поправкою Шеппарда** вигляду

$$\mathbf{P}(\nu_n = k) \approx \Phi(x_{nk} + h_n/2) - \Phi(x_{nk} - h_n/2) \approx h_n\varphi(x_{nk}) + \varphi''(x_{nk})h_n^3/24,$$

де використано розклад у ряд Тейлора функції Φ порядку 3.

Вправи

(1) При $np = \lambda$ сума модулів різниць між імовірностями $\mathbf{P}(\nu_n = k)$ та їх граничними значеннями у граничній теоремі Пуассона не перевищує $2\,\lambda^2/n$.

(2) Величини $(\xi_n, n \geq 1)$ незалежні, $\mathbf{P}(\xi_k = 1) = p_k = 1 - \mathbf{P}(\xi_k = 0)$, $S_n = \xi_1 + ... + \xi_n$, $\mu_n = \sum_{k=1}^n p_k$, $\sigma_n^2 = \sum_{k=1}^n p_k(1-p_k)$, $x_{nk} = (k - \mu_n)/\sigma_n$. Довести, що за умов $\sigma_n \to \infty$, $x_{nk} = O(1), k, n \to \infty$, справедливе асимптотичне зображення $\mathbf{P}(S_n = k) \sim (2\pi\sigma_n^2)^{-1/2}\exp(-x_{nk}^2/2), k, n \to \infty$.

(3) Довести аналог (а) теореми Пуассона при $n/N \to p, n \to \infty$, (б) Муавра–Лапласа для числа успіхів при виборі без повернення (для гіпергеометричного розподілу числа успіхів).

1.7. Загальне означення випадкової величини та вектора

Для формулювання загального означення слід структурувати клас множин у просторі значень випадкової величини.

1.7.1. Борелєва сигма-алгебра, борелєві множини

Розглянемо на числовій осі $\mathbb{R}$ клас **напівінтервалів**

$$\mathfrak{I}(\mathbb{R}) = \{[a,b),\ -\infty \leq a < b \leq \infty\}.$$

Скінченні об'єднання попарно несумісних напівінтервалів є алгеброю підмножин $\mathbb{R}$ (**вправа**):

$$\mathfrak{A}(\mathbb{R}) = \{A = \cup_{k=1}^n [a_k, b_k),\ -\infty \leq a_1 < b_1 < a_2 < ... < b_n \leq \infty\}.$$

Нагадаємо, що переріз будь-якого класу σ-алгебр завжди є *σ-алгеброю*, яка називається найменшою σ-алгеброю з даного класу.

Означення. Борелєвою сигма-алгеброю $\mathfrak{B}(\mathbb{R})$ *називається найменша σ-алгебра, яка містить алгебру* $\mathfrak{A}(\mathbb{R})$, (*або клас* $\mathfrak{I}(\mathbb{R})$, *що еквівалентно). Інакше кажучи, σ-алгебра* $\mathfrak{B}(\mathbb{R})$ *породжена алгеброю* $\mathfrak{A}(\mathbb{R})$: $\mathfrak{B}(\mathbb{R}) = \sigma[\mathfrak{A}(\mathbb{R})]$. *Будь-який елемент* $B \in \mathfrak{B}(\mathbb{R})$ *називається* борелєвою множиною, *або ж* борелівською множиною.

Вправи

(1) Борелєвими є всі інтервали та відрізки, одноточкові множини, всі відкриті множини (як зліченні об'єднання відкритих інтервалів) та замкнені множини (як доповнення відкритих). Сигма-алгебра, що породжена відкритими підмножинами $\mathbb{R}$, збігається з борелєвою сигма-алгеброю.

(2) Берівська сигма-алгебра, що породжується класом неперервних функцій: $\sigma[\{x : f(x) < a\},\ a \in \mathbb{R}, f \in C(\mathbb{R})]$ збігається з борелєвою.

(3) Нехай $\mathbf{P}$ – імовірність на $\mathfrak{B}(\mathbb{R})$. Довести, що для довільного $\varepsilon > 0$ та для довільної множини $A \in \mathfrak{B}(\mathbb{R})$ знайдуться компактна множина C та відкрита множина D такі, що $C \subset A \subset D$ та $\mathbf{P}(D \backslash C) \leq \varepsilon$.

(4) Розглянемо простір функцій $\Omega = \{x : \mathbb{R} \to \{0, 1\}\}$ з топологією поточкової збіжності. Довести, що цей простір є компактним. Визначимо функції $x_n \in \Omega$ рівностями $x_n(t) = [2^n t] - 2[2^{n-1}t]$, де $[x]$ – ціла частина x. Нехай x – довільна гранична точка послідовності x_n, що належить Ω внаслідок компактності. Довести, що $x(b + t) = x(t)$ та $x(b - t) = 1 - x(t)$ для всіх бінарно-раціональних чисел b та всіх t, що не є такими. Вивести звідси, що функція x не є борелєвою.

(5) Приклад неборелєвої множини, пов'язаний з мірою Лебега L. Нехай O – коло одиничної довжини з множиною $\mathbb{Q}$ раціональних точок на ньому. Оберемо за аксіомою вибору по одному елементу з кожного класу еквівалентності при циклічному додаванні з фактор-множиною $\mathbb{Q}$. Отримана множина A є шуканою, оскільки $\cup_{q \in \mathbb{Q}}\{a + q,\ a \in A\} = O$, причому множини в об'єднанні попарно несумісні і повинні були б мати однакову довжину (міру Лебега), що неможливо.

Аналогічно, система n-вимірних кутів

$$\mathfrak{I}(\mathbb{R}^n) = \{A = (-\infty, a) = \prod\nolimits_{k=1}^{n}(-\infty, a_k),\ a_k \in \mathbb{R} \cup \infty\}$$

у n-вимірному просторі *породжує* алгебру $\mathfrak{A}(\mathbb{R}^n)$ скінченних об'єднань напіввідкритих паралелепіпедів, що не перетинаються.

Означення. *Борелєвою σ-алгеброю* $\mathfrak{B}(\mathbb{R}^n)$ *у* $\mathbb{R}^n$ *називається найменша сигма-алгебра, яка містить алгебру* $\mathfrak{A}(\mathbb{R}^n)$.

Зауважимо, що прямий добуток $B_1 \times ... \times B_n$ множин $B_k \in \mathfrak{B}(\mathbb{R})$ є борелєвою множиною. Дійсно, клас множин

$$\{B_k \in \mathfrak{B}(\mathbb{R}) : \mathbb{R}^{k-1} \times B_k \times \mathbb{R}^{n-k} \in \mathfrak{B}(\mathbb{R}^n)\}$$

містить всі *напівінтервали* і є σ-алгеброю, тому збігається з $\mathfrak{B}(\mathbb{R})$. Отже,

$$B_1 \times ... \times \ B_n = \bigcap_{k=1}^{n} \left(\mathbb{R}^{k-1} \times B_k \times \mathbb{R}^{n-k}\right) \in \mathfrak{B}(\mathbb{R}^n).$$

Означення. *Функція* $g : \mathbb{R}^k \to \mathbb{R}^n$ *називається* борелєвою*, якщо*

$$\{x : g(x) \in B_n\} \in \mathfrak{B}(\mathbb{R}^k), \ \forall B_n \in \mathfrak{B}(\mathbb{R}^n).$$

Зокрема, всі неперервні функції є борелєвими. Дійсно, прообраз відкритого інтервалу для неперервної функції є відкритою множиною – отже, борелєвою. Оскільки клас множин, прообразами яких є борелєві множини, є σ-алгеброю та містить всі відкриті напівінтервали $(-\infty, a)$, то цей клас збігається з борелєвою σ-алгеброю.

1.7.2. Загальне означення випадкової величини

Існують стохастичні експерименти, в результаті яких можна спостерігати незліченну кількість числових або навіть векторних значень. Тому розглянемо таке узагальнення поняття *дискретної випадкової величини*.

Означення. Випадковою величиною *називається така числова функція* $\xi : \Omega \to \mathbb{R}$, *що для довільної борелєвої множини* $B \subset \mathbb{R}$ *її прообраз:* $\{\omega : \xi(\omega) \in B\}$ *є випадковою подією, тобто:*

$$\{\omega : \xi(\omega) \in B\} \in \mathfrak{F}, \ \forall B \in .\mathfrak{B}(\mathbb{R}).$$

Функції ξ, які задовольняють наведену умову, називаються вимірними відносно сигма-алгебри $\mathfrak{F}$. Отже, для випадкових величин (вимірних функцій) можна говорити про ймовірність події $\{\xi \in B\}$.

Тут і надалі буде використовуватись скорочений запис – аргумент у функції $\xi(\omega)$ будемо опускати, а фігурні дужки, що містять випадкові величини, означатимуть множину відповідних елементарних подій:

$$\{\xi \in B\} \equiv \{\omega : \xi(\omega) \in B\}.$$

Узагальненою випадковою величиною (або ж невласною) будемо називати функцію зі значеннями в множині $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, яка задовольняє ту саму умову *вимірності*.

Приклади недискретних випадкових величин.

1. *Випадкова точка на відрізку*. В експерименті з вибором точки на одиничному відрізку маємо $\Omega = [0, 1]$. Функція $\xi(\omega) = \omega$ набуває незліченну кількість значень.

2. *Час до зустрічі*. В задачі про зустріч (при *геометричному означенні ймовірностей*) час до зустрічі двох осіб має континуум можливих значень.

Цей час як функція елементарної події $\omega = (\omega_1, \omega_2)$, де $\omega_i \in [0, 1]$ – момент приходу на місце зустрічі i-ї особи, дорівнює: $\xi(\omega) = \max(\omega_1, \omega_2)$.

Приклад функції, що не є випадковою величиною: індикаторна величина $\xi = \mathbb{1}_C$ множини, яка не є випадковою подією: $C \notin \mathfrak{F}$.

1.7.3. Критерій вимірності, вимірність суперпозиції

Теорема (про критерій вимірності скалярної функції). *Нехай $\xi : \Omega \to \mathbb{R}$ – довільна функція. Вона є випадковою величиною тоді й тільки тоді, коли прообраз кожного напівінтервалу $(-\infty, x)$ є випадковою подією: $\{\xi < x\} \in \mathfrak{F}$, $\forall x \in \mathbb{R}$.*

Доведення. З властивостей операції обчислення прообразу випливає, що клас борелєвих множин

$$\mathfrak{B}_\xi^* = \{B \in \mathfrak{B}(\mathbb{R}) : \{\xi \in B\} \in \mathfrak{F}\}$$

є *сигма-алгеброю*, оскільки прообрази доповнення та об'єднання множин збігаються з відповідними доповненням чи об'єднанням прообразів. З іншого боку, за означенням *випадкової величини* клас $\mathfrak{B}_\xi^*$ містить всі *напівінтервали*. Оскільки клас напівінтервалів *породжує* борелєву σ-алгебру, то $\mathfrak{B}_\xi^* = \mathfrak{B}(\mathbb{R})$ і твердження теореми є наслідком вказаного означення $\boxdot$

У багатьох застосуваннях доводиться мати справу з функціями від випадкових величин. Тому важливим є твердження про суперпозицію.

Теорема (про вимірність суперпозиції). *Нехай ξ – випадкова величина, а $g : \mathbb{R} \to \mathbb{R}$ – борелєва функція. Тоді суперпозиція $g(\xi)$ є випадковою величиною.*

Доведення. Обчислимо $\{g(\xi) < x\} = \{\xi \in g^{(-1)}((-\infty, x))\} \in \mathfrak{F}$, оскільки $g^{(-1)}((-\infty, x))$ – *борелєва множина* за означенням g $\boxdot$

1.7.4. Випадковий вектор

Означення. **Випадковим вектором** *у просторі $\mathbb{R}^n$ називається така функція $\xi : \Omega \to \mathbb{R}^n$, що для довільної борелєвої множини B її прообраз $\{\xi \in B\}$ є випадковою подією: $\{\omega : \xi(\omega) \in B\} \in \mathfrak{F}$, $\forall B \in \mathfrak{B}(\mathbb{R}^n)$.*

Вправа. Як і для величин, довести, що функція $\xi = (\xi_1, ...\xi_n) : \Omega \to \mathbb{R}^n$ є випадковим вектором тоді й тільки тоді, коли для довільного кута $\{\xi < x\} \equiv \{\xi_1 < x_1, ..., \xi_n < x_n\} \in \mathfrak{F}$ при всіх $x = (x_1, ..., x_n) \in \mathbb{R}^n$.

Теорема (про вимірність функції від випадкового вектора). *Нехай ξ – випадковий вектор у $\mathbb{R}^n$, а $g : \mathbb{R}^n \to \mathbb{R}^m$ – борелєва функція. Тоді суперпозиція $g(\xi)$ є випадковим вектором у $\mathbb{R}^m$.*

Доведення цілком аналогічне доведенню теореми *про вимірність суперпозиції* для скалярних величин.

Зокрема, кожна координата випадкового вектора є *випадковою величиною*, оскільки координатна функція $\pi_k : \mathbb{R}^n \to \mathbb{R}$ неперервна і є борелєвою. Виконується й обернене твердження: вектор, утворений з n випадкових величин, є *випадковим вектором* у $\mathbb{R}^n$. Дійсно, прообраз n-вимірного *кута* є перерізом прообразів кутів для координатних відображень:

$$\{\xi < x\} = \bigcap_{k=1}^{n}\{\xi_k < x_k\} \in \mathfrak{F}.$$

Теорема (про перетворення випадкових величин). *Якщо ξ, η – випадкові величини, то $|\xi|$, $\xi + \eta$, $\xi - \eta$, $\xi \cdot \eta$, ξ/η, $\max(\xi, \eta)$ також є випадковими величинами.*

Доведення випливає з неперервності (отже, борелєвості) функцій $|x|$, $x + y$, $x - y$, $x \cdot y$, x/y, $\max(x, y)$ від *випадкового вектора* (ξ, η) $\boxdot$

1.7.5. Вимірність границі випадкових величин

Теорема (про вимірність границі випадкових величин). *Нехай $(\xi_n, n \geq 1)$ – випадкові величини такі, що границя $\xi(\omega) = \lim \xi_n(\omega)$ існує при всіх $\omega \in \Omega$. Тоді ξ – випадкова величина.*

Доведення випливає з тотожності

$$\{\lim \xi_n < x\} = \bigcup_{k\geq 1}\bigcup_{m\geq 1}\bigcap_{n\geq m}\{\xi_n < x - 1/k\}.$$

Її справедливість обґрунтовується тим, що при існуванні границі $\lim \xi_n$ вона збігається з верхньою границею $\overline{\lim}\xi_n$, а остання строго менша за число x тоді й тільки тоді, коли для деякого $k < \infty$ всі члени послідовності менші за $x - 1/k$, починаючи з деякого номера $\boxdot$

Наслідок. Якщо $(\xi_n,\ n \geq 1)$ – випадкові величини, то $\overline{\lim}\xi_n$, $\underline{\lim}\xi_n$ також є випадковими величинами (можливо, *невласними*).

Доведення ґрунтується на двох попередніх теоремах і випливає з того, що вказані граничні значення є *монотонними границями* випадкових величин. Зокрема, $\overline{\lim}_{n\to\infty}\xi_n = \lim_{N\to\infty} \lim_{m\to\infty} \max_{N\leq\ n\ \leq N+m}\ \xi_n$ $\boxdot$

Вправи

(1) Якщо ξ, η – випадкові величини, то $\{\xi = \eta\}$ – випадкова подія.

(2) Якщо $(\xi_n, n \geq 1)$ випадкові величини, то $\inf \xi_n$, $\sup \xi_n$ є також випадковими величинами, можливо, невласними.

1.7.6. Апроксимація простими величинами

В подальшому символом $x_n \uparrow x, n \to \infty$, будемо позначати **монотонно неспадну збіжність** числової послідовності x_n до x, тобто такі властивості: (1) $x_n \leq x_{n+1}$ при всіх n та (2) $\lim_{n\to\infty} x_n = x$. Аналогічний зміст матиме запис $x_n \downarrow x$. Запис $[a]$ означає цілу частину числа a.

Теорема (про апроксимацію випадкових величин простими).

(а) Нехай ξ – невід'ємна випадкова величина. Тоді існує послідовність простих величин ξ_n така, що $0 \leq \xi_n(\omega) \uparrow \xi(\omega)$, $n \to \infty$, $\forall \omega \in \Omega$.

(б) Нехай ξ – довільна випадкова величина. Тоді існує послідовність простих величин ξ_n таких, що $\lim_{n\to\infty} \xi_n = \xi$ та $|\xi_n| \ \leq |\xi|$ при всіх ω.

Доведення. Розглянемо таку послідовність дійсних функцій на $\mathbb{R}_+$:

$$\varphi_n(x) \equiv \sum\nolimits_{k\,=\,0}^{n2^n-1} k2^{-n}\, 1\!\!\mathrm{I}_{x\,\in\,[k2^{-n},\,(k+1)2^{-n})} + n\, 1\!\!\mathrm{I}_{n\,\leq\,x} = \min(2^{-n}[2^n x],\ n).$$

З означення індикаторної функції виводимо, що

(1) $0 \leq \varphi_n(x) \leq x$,

(2) $0 \leq x - \varphi_n(x) \leq 2^{-n}$ при $x < n$, та

(3) $\varphi_n(x) \leq \varphi_{n+1}(x)$ при всіх $x \geq 0$, $n \geq 1$.

Остання властивість монотонності є наслідком того, що при переході від розбиття з кроком 2^{-n} до кроку 2^{-n-1} ліва частина графіка функції φ_n на відрізку $[k2^{-n}, (k+1)2^{-n})$ не змінюється, а права зростає.

З (1),(2) і (3) остаточно виводимо, що для кожного $x \geq 0$

$$0 \leq \varphi_n(x) \uparrow x,\ n \to \infty.$$

(а) Покладемо $\xi_n = \varphi_n(\xi)$. Тоді ξ_n є *простою* випадковою величиною, оскільки φ_n – *борелєва* функція та має скінченну множину значень. З останнього співвідношення для φ_n випливає шукане твердження (а) теореми: $0 \leq \xi_n = \varphi_n(\xi) \uparrow \xi, n \to \infty$, для всіх $\omega \in \Omega$.

(б) Визначимо **додатну та від'ємну частини** величини ξ :

$$\xi^+ = \max(\xi, 0), \quad \xi^- = \max(-\xi, 0),$$

невід'ємні випадкові величини за теоремою *про вимірність суперпозиції.*

Нехай $\xi_n^\pm$ – прості величини, які внаслідок (а) апроксимують випадкові величини ξ^+ та ξ^- відповідно. Тоді $\xi_n = \xi_n^+ - \xi_n^-$ є шуканою простою величиною, оскільки $\xi = \xi^+ - \xi^- = \lim \xi_n^+ - \lim \xi_n^-, = \lim(\xi_n^+ - \xi_n^-,) = \lim \xi_n$, причому $|\xi_n| = \xi_n^+ + \xi_n^- \leq \xi^+ + \xi^- = |\xi|$ ⊡

1.8. Функція розподілу випадкової величини та її властивості

Сукупність імовірностей подій вигляду $\{\xi = x\}$ не завжди повністю описує величину ξ . Наприклад, для випадкової точки на відрізку всі ці ймовірності нульові. Тому в загальному випадку на відміну від дискретного для визначення розподілу випадкової величини використовують прообрази інтервалів.

Означення. Функцією розподілу *випадкової величини ξ називається функція $F_\xi(x)$ дійсного аргументу x, яка задається рівністю*

$$F_\xi(x) = \mathbf{P}(\xi < x), \ x \in \mathbb{R}.$$

1.8.1. Властивості функції розподілу

Теорема (про основні властивості функції розподілу). *Нехай випадкова величина ξ має функцію розподілу $F \equiv F_\xi$. Тоді ця функція*

(1) неспадна: $F(x) \le F(y)$, $\forall\ x \le y$,

(2) нормована: $F(-\infty) = 0,\ F(+\infty) = 1$,

(3) неперервна зліва: $F(x - 0) = F(x)$, $\forall\ x \in \mathbb{R}$.

Зауваження. Тут та надалі символом $F(x - 0)$ позначається монотонна границя $F(x - 0) = \lim_{y\uparrow x,\ y<x} F(y)$. Аналогічно визначаються границі $F(-\infty) = \lim_{y\downarrow -\infty} F(y),\ F(+\infty) = \lim_{y\uparrow +\infty} F(y)$. Існування та єдиність границь у (2) та (3) випливає з монотонності (1) функції розподілу.

Альтернативне означення функції розподілу має вигляд $\mathbf{P}(\xi \le x)$, що відрізняється властивістю неперервності справа.

Доведення

(1) Нерівність є наслідком *монотонності ймовірності*, тому що при $x < y$ має місце включення $\{\xi < x\} \subset \{\xi < y\}$.

(2) Границя монотонної функції не залежить від вибору апроксимуючої послідовності аргументів. Тому y можна вважати цілозначними. Оскільки послідовність подій $\{\xi < n\}$ монотонно не зростає при $n \downarrow -\infty$, а $\cap_{\,n<0}\{\xi < n\} = \varnothing$, то $\{\xi < n\} \downarrow \varnothing$. Отже, за *неперервністю ймовірності*

$$F(-\infty) = \lim_{n\downarrow -\infty} F(n) = \lim_{n\downarrow -\infty} \mathbf{P}(\xi < n) = \mathbf{P}(\varnothing) = 0.$$

Аналогічно, з $\{\xi < n\} \uparrow \Omega$ при $n \uparrow \infty$ дістанемо

$$F(+\infty) = \lim_{n\uparrow \infty} F(n) = \lim_{n\uparrow \infty} \mathbf{P}(\xi < n) = \mathbf{P}(\Omega) = 1.$$

(3) Клас відкритих напівінтервалів $(-\infty, x), x \in \mathbb{R}$, є неперервним зліва для монотонної збіжності: $(-\infty, y) \uparrow (-\infty, x)$ при $y \uparrow x$, оскільки $(-\infty, x) = \cup_{n\geq 1}(-\infty, x-1/n)$. Тому має місце *монотонна збіжність*:

$$\{\xi \in (-\infty, y)\} = \{\xi < y\} \uparrow \{\xi < x\} = \{\xi \in (-\infty, x)\} \text{ при } y \uparrow x.$$

Отже, за *неперервністю ймовірностей*

$$F(x-0) = \lim_{n\to\infty} F(x-1/n) = \lim_{n\to\infty} \mathbf{P}(\xi < x-1/n) = \mathbf{P}(\xi < x) = F(x)\boxdot$$

Теорема (про властивості функції розподілу). *Нехай величина ξ має функцію розподілу F_ξ. Тоді*

(а) $\mathbf{P}(a \leq \xi < b) = F_\xi(b) - F_\xi(a)$,

(б) $\mathbf{P}(\xi = x) = F_\xi(x+0) - F_\xi(x)$,

(в) $\mathbf{P}(\xi \geq x) = 1 - F_\xi(x)$,

(г) $\mathbf{P}(\xi \leq x) = F_\xi(x+0)$,

(д) $\mathbf{P}(\xi > x) = 1 - F_\xi(x+0)$.

Доведення

(а) З формули для *ймовірності вкладеної різниці* подій отримуємо

$$\mathbf{P}(a \leq \xi < b) = \mathbf{P}(\{\xi < b\} \backslash \{\xi < a\}) = \mathbf{P}(\xi < b) - \mathbf{P}(\xi < a).$$

(б) Зі співвідношення $\{\xi = x\} = \cap_{n>1}\{x \leq \xi < x + 1/n\}$, *неперервності ймовірності* та твердження (а) дістанемо при $n \to \infty$:

$$\mathbf{P}(\xi = x) = \lim \mathbf{P}(x \leq \xi < x + 1/n) = \lim(F_\xi(x+1/n) - F_\xi(x)).$$

(в) Є наслідком означення функції розподілу та формули про *ймовірність доповнення.*

(г,д) Виводяться з означення *функції розподілу* та (б) $\boxdot$

Означення. *Будь-яка функція $F : \mathbb{R} \to \mathbb{R}$, яка задовольняє умови (1), (2) і (3), називається функцією розподілу.*

Означення. *Нехай F – довільна функція розподілу. Назвемо* **адитивною мірою Лебега – Стілтьєса,** *породженою F (або, коротше, адитивною мірою F), таку функцію F на алгебрі $\mathfrak{A}(\mathbb{R})$ підмножин $\mathbb{R}$ вигляду*

$$A = \bigcup\nolimits_{k=1}^{n} [a_k, b_k), \quad \text{де } -\infty \leq a_1 < b_1 < a_2 < ... < b_n \leq \infty,$$

значення якої задається рівністю

$$F(A) = \sum\nolimits_{k=1}^{n} \left(F(b_k) - F(a_k)\right).$$

Теорема (про адитивну міру Лебега – Стілтьєса). *Адитивна міра F є невід'ємною, нормованою та скінченно-адитивною функцією на $\mathfrak{A}(\mathbb{R})$. Зокрема, F напівадитивна та монотонна.*

Доведення. Невід'ємність є очевидним наслідком *монотонності* (невід'ємності приростів) функції розподілу F, нормованість виводиться з *нормованості* F. Для доведення адитивності зауважимо, що з несумісності множин $A, B \in \mathfrak{A}(\mathbb{R})$ випливає *попарна несумісність* інтервалів, які складають ці множини. Тому сума приростів F на інтервалах з об'єднання $A \cup B$ перегрупуванням зводиться до суми відповідних сум для A та B, що і доводить адитивність. Згідно з теоремою *про основні властивості ймовірностей напівадитивність* та монотонність є наслідками невід'ємності та адитивності $\boxdot$

1.8.2. Неперервність у нулі міри Лебега – Стілтьєса

Теорема (про неперервність у нулі міри Лебега – Стілтьєса). *Адитивна міра Лебега – Стілтьєса* F *є неперервною в нулі на* $\mathfrak{A}(\mathbb{R})$, *і як наслідок, сигма-адитивною на* $\mathfrak{A}(\mathbb{R})$.

Доведення. Нехай $A_n \in \mathfrak{A}(\mathbb{R})$, $A_n \downarrow \varnothing$. Оберемо довільне $\varepsilon > 0$.

1. Оберемо $c > 0$ так, щоб $F(\overline{[-c, c)}) = F(-c) + 1 - F(c) < \varepsilon$. Це можливо за умови *нормованості* функції розподілу.

2. Позначимо $B_n = A_n \cap [-c, c) \in \mathfrak{A}(\mathbb{R})$. Тоді за монотонністю адитивної міри F

$$F(A_n \backslash B_n) = F(A_n \cap \overline{[-c, c)}) \le F(\overline{[-c, c)}) < \varepsilon.$$

3. Оскільки $B_n \in \mathfrak{A}(\mathbb{R})$, то $B_n = \cup_{k=1}^{m_n} [a_{nk}, b_{nk})$, де інтервали попарно несумісні. Оберемо $c_{nk} \in [a_{nk}, b_{nk})$ так, щоб

$$\sum\nolimits_{k=1}^{m_n} (F(b_{nk}) - F(c_{nk})) \le \varepsilon 2^{-n}.$$

Це можливо, оскільки функція розподілу F *неперервна зліва* в кожній точці b_{nk}, отже остання сума прямує до нуля, коли $c_{nk} \uparrow b_{nk}$ для всіх k.

Визначимо множини

$$C_n \equiv \bigcup\nolimits_{k=1}^{m_n} [a_{nk}, c_{nk}) \subset D_n \equiv \bigcup\nolimits_{k=1}^{m_n} [a_{nk}, (c_{nk} + b_{nk})/2] \subset B_n.$$

За побудовою $C_n \in \mathfrak{A}(\mathbb{R})$, та

$$F(B_n \backslash C_n) = \sum\nolimits_{k=1}^{m_n} (F(b_{nk}) - F(c_{nk})) \le \varepsilon 2^{-n}.$$

4. Множина $D_n \subset B_n \subset A_n$ замкнена та обмежена – отже, компактна. Послідовність $D_n \subset B_n \subset [-c, c]$ належить компактові $[-c, c]$, причому є нецентрованою: $\cap D_n \subset \cap A_n = \varnothing$. За теоремою про нецентровану послідовність компактних підмножин компакту (що еквівалентна існуванню

скінченного підпокриття відкритого покриття $[-c,c]\backslash D_n$ компакту $[-c,c]$) існує скінченне m таке, що $\cap_{k=1}^m D_k = \varnothing$. Тому

$$\cap_{k=1}^n C_k \subset \cap_{k=1}^n D_k \subset \cap_{k=1}^m D_k = \varnothing, \ \forall n \geq m.$$

5. Обчислимо при $n \geq m$ з урахуванням того, що $B_n \subset B_k$ при $n \geq k$, а міра F *напівадитивна*:

$$F(B_n) = F(B_n \setminus \ \cap_{k=1}^n C_k) = F(\cup_{k=1}^n (B_n \backslash C_k)) \leq$$

$$F(\cup_{k=1}^n (B_k \backslash C_k)) \leq \sum_{k=1}^n F(B_k \backslash C_k) \leq \sum_{k\geq 1} \varepsilon 2^{-k} = \varepsilon.$$

6. Залишилося зазначити, що

$$F(A_n) \leq F(A_n \backslash B_n) + F(B_n) \leq 2\,\varepsilon, \quad \forall n \geq m.$$

Отже, $F(A_n) \to 0$ при $n \to \infty$, тобто міра F *неперервна в нулі*.

Еквівалентність *неперервності в нулі та сигма-адитивності* для адитивних мір доведена у теоремі *про основні властивості ймовірностей* ⊡

Теорема (про існування міри Лебега – Стілтьєса). *Нехай F – довільна функція розподілу. На борелєвій сигма-алгебрі $\mathfrak{B}(\mathbb{R})$ існує єдина невід'ємна, нормована та сигма-адитивна міра F така, що*

$$F((-\infty, x)) = F(x), \ \forall x \in \mathbb{R}.$$

Доведення. За теоремою *про адитивну міру Лебега – Стілтьєса* побудуємо *адитивну* ймовірність F на алгебрі, як це зроблено вище. На підставі її σ-адитивності на алгебрі $\mathfrak{A}(\mathbb{R})$ застосуємо *теорему Каратеодорі про продовження міри* з алгебри на *породжену* σ-алгебру $\mathfrak{B}(\mathbb{R})$ ⊡

Означення. *Імовірнісна міра F на $\mathfrak{B}(\mathbb{R})$, що побудована за останньою теоремою, називається* мірою Лебега – Стілтьєса, *яка породжена функцією розподілу F.*

1.8.3. Обчислення ймовірностей через функцію розподілу

Теорема (про обчислення ймовірностей, пов'язаних із випадковою величиною). *Нехай випадкова величина ξ має функцію розподілу F_ξ. Тоді для довільної борелєвої множини $B \in \mathfrak{B}(\mathbb{R})$ виконується рівність*

$$\mathbf{P}(\xi \in B) = F_\xi(B),$$

де F_ξ – міра Лебега – Стілтьєса, що породжена функцією розподілу F_ξ.

Доведення. За означенням сформульована рівність має місце для напівінтервала $B = (-\infty, x)$. З іншого боку, вирази в обох частинах як функції множини B є *сигма-адитивними* мірами. Для лівої частини це випливає з властивостей прообразу для відображення ξ та σ-адитивності *ймовірності* $\mathbf{P}$, а для правої частини виконується за означенням. Тому за *теоремою Каратеодорі про продовження міри* вказані міри збігаються на всій *породженій* σ-алгебрі $\mathfrak{B}(\mathbb{R})$ $\boxdot$

Враховуючи останню теорему, **розподілом** випадкової величини ξ будемо називати відповідну міру Лебега – Стілтьєса F_ξ на $\mathfrak{B}(\mathbb{R})$.

1.8.4. Функції від випадкової величини

Теорема (про обчислення розподілу функції від випадкової величини). *Якщо випадкова величина ξ має функцію розподілу $F_\xi(x)$, а функція $g : \mathbb{R} \to \mathbb{R}$ – борелєва, то для всіх $B \in \mathfrak{B}(\mathbb{R})$*

$$\mathbf{P}(g(\xi) \in B) = F_\xi(g^{(-1)}(B)),$$

де $g^{(-1)}(\cdot)$ – прообраз відображення g.

Доведення є очевидним наслідком теореми *про обчислення ймовірностей, пов'язаних із випадковою величиною*, оскільки за означенням прообразу відображення $\{g(\xi) \in B\} = \{\xi \in g^{(-1)}(B)\}$ $\boxdot$

Наслідок (про збереження однакової розподіленості). *Якщо випадкові величини $\xi_k, k = 1, 2,$ **однаково розподілені,** тобто мають однакові функції розподілу, а g – борелєва функція, то величини $g(\xi_k)$ також однаково розподілені.*

Доведення. Якщо F – спільний *розподіл* випадкових величин ξ_k (тобто породжена *міра Лебега – Стілтьєса*), то

$$\mathbf{P}(g(\xi_1) \in B) = F(g^{(-1)}(B)) = \mathbf{P}(g(\xi_2) \in B),\ \forall B \in \mathfrak{B}(\mathbb{R})\ \boxdot$$

Наслідок (про розподіл лінійного перетворення випадкової величини). *Для лінійної функції $g(x) = a+bx,\ b > 0$, **функція розподілу лінійного перетворення** $\zeta = a + b\xi$ має вигляд* $\mathbf{P}(\zeta < x) = \mathbf{P}(\xi < (x-a)/b) = F_\xi((x-a)/b)$. **Доведення** очевидне $\boxdot$

1.9. Абсолютно неперервні величини

Дискретна функція розподілу (*функція розподілу дискретної випадкової величини*) є сталою в кожному околі своєї точки неперервності та збігається з сумою своїх стрибків на інтервалі $(-\infty, x)$. Функція розподілу *простої* величини чисто розривна та кусково-стала. Дійсно, якщо величина ξ набуває значень $\{x_n, n \geq 1\}$ із імовірностями $\{p_n, n \geq 1\}$, то:

$$F_\xi(x) = \mathbf{P}\left(\bigcup\nolimits_{n:\ x_n<x}\{\xi = x_n\}\right) = \sum\nolimits_{n:\ x_n<x} p_n.$$

Вправи

(1) Функція розподілу має не більш ніж зліченну множину точок розриву.

(2) Довести, що неперервна функція розподілу рівномірно неперервна.

(3) Для випадкових величин ξ_1, ξ_2 відомі ймовірності $\mathbf{P}(\max(\xi_1, \xi_2) = a)$, $\mathbf{P}(\min(\xi_1, \xi_2) = a)$, $\mathbf{P}(\xi_1 = a)$. Обчислити ймовірність $\mathbf{P}(\xi_2 = a)$.

(4) Точка $x \in \mathbb{R}$ називається точкою зростання функції розподілу F, якщо $F(x+\varepsilon) > F(x-\varepsilon)$ для всіх $\varepsilon > 0$. Множина точок зростання називається носієм цієї функції. (а) Побудуйте приклад дискретної функції розподілу, носій якої збігається з $\mathbb{R}$. (б) Довести, що носій довільної функції розподілу є замкненою множиною. (в) Якщо функція розподілу F неперервна, то її носій є довершеною множиною, тобто у будь-якому околі довільної його точки міститься хоча б одна інша його точка. (г) Для довільної замкненої множини $C \subset \mathbb{R}$ побудувати функцію розподілу, що має носій C.

1.9.1. Абсолютно неперервні функції розподілу

Наступний клас містить виключно неперервні функції розподілу.

Означення. *Випадкова величина ξ та її функція розподілу F_ξ називаються* абсолютно неперервними, *якщо існує функція $f_\xi(x)$ така, що*

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(y)dy,\ \ \forall x \in \mathbb{R}.$$

Інтеграл у цьому означенні слід розуміти як:
(а) інтеграл Рімана – Стілтьєса для кусково-неперервної функції $f_\xi(x)$,
(б) інтеграл Лебега – Стілтьєса для вимірної функції $f_\xi(x)$.

Означення. *Функція $f_\xi(x)$ називається* щільністю *розподілу випадкової величини ξ та функції розподілу F_ξ.*

Зауваження. Поняття інтеграла Лебега як часткового випадку *математичного сподівання* разом із його основними властивостями викладене нижче в розділі про загальне означення математичного сподівання.

З *нормованості* та *монотонності* функції розподілу випливають такі властивості щільності:

(1 –невід'ємність) $f_\xi(x) \geq 0$ та

(2 –нормованість) $\int_{-\infty}^{\infty} f_\xi(y)dy = 1$.

Будь-яка невід'ємна інтегровна нормована функція є щільністю певної функції розподілу, оскільки з властивостей інтегралу (Рімана або Лебега) випливають характеристичні властивості (1 – 3) *функції розподілу*.

Якщо функція розподілу диференційовна, то з формули Ньютона – Лейбніца (для інтеграла Рімана) або ж із теореми Радона – Нікодима (для інтеграла Лебега) випливає, що щільність (майже всюди) збігається з похідною функції розподілу

$$f_\xi(x) = F'_\xi(x) = dF_\xi(x)/dx.$$

В кожній точці неперервності x за теоремою про середнє має місце зображення $\mathbf{P}(x \le \xi < x + h) = \int_x^{x+h} f_\xi(y)dy = hf_\xi(x) + o(h), h \to 0$, тому $f_\xi(x)$ можна інтерпретувати як "щільність імовірності" в околі точки x.

1.9.2. Обчислення ймовірностей через щільність розподілу

Теорема (про обчислення ймовірностей, пов'язаних із неперервною величиною). *Якщо випадкова величина ξ абсолютно неперервна з щільністю $f_\xi(x)$, то*

$$\mathbf{P}(\xi \in B) = \int_B f_\xi(x)dx, \ \forall B \in \mathfrak{B}(\mathbb{R}).$$

Доведення. Шукана рівність виконується за означенням для напівінтервалів $B = (-\infty, x)$. З іншого боку, обидві частини рівності як функції B є *сигма-адитивними* мірами. Тому за *теоремою Каратеодорі про продовження міри* вони збігаються на породженій σ-алгебрі $\mathfrak{B}(\mathbb{R})$ $\boxdot$

Наслідок (про щільність лінійного перетворення випадкової величини). *Нехай випадкова величина ξ має щільність розподілу f_ξ. Тоді лінійне перетворення $\zeta = a + b\xi$ має щільність $f_\zeta(x) = b^{-1} f_\xi((x-a)/b)$.*

Доведення виводиться з означення щільності та з формули заміни змінної в інтегралі $y = (u-a)/b$:

$$\mathbf{P}(\zeta < x) = \mathbf{P}(\xi < (x-a)/b) = \int_{-\infty}^{(x-a)/b} f_\xi(y)dy = b^{-1} \int_{-\infty}^{x} f_\xi((u-a)/b)du \ \boxdot$$

1.9.3. Класифікація функцій розподілу

Одночасно з класами *абсолютно неперервних* та *дискретних функцій розподілу* існує екзотичний клас **сингулярних функцій розподілу,** які є неперервними, але множина їх точок зростання має нульову довжину.

Очевидно, що вказані три класи *розподілів* не перетинаються. Виявляється, що їх опукла оболонка вичерпує весь клас функцій розподілу.

Теорема (теорема Лебега про зображення функції розподілу). *Для довільної функції розподілу F існують: дискретна F_d, абсолютно неперервна F_a і сингулярна F_s функції розподілу та три числа $\alpha, \beta, \gamma \geq 0$, такі, що $\alpha + \beta + \gamma = 1$ та*

$$F(x) = \alpha F_d(x) + \beta F_a(x) + \gamma F_s(x),\ \forall x \in \mathbb{R}.$$

Доведення не наводиться.

Вправи

(1) Якщо $\mathbb{Q} = \{x_n.n \geq 1\}$ множина раціональних точок, то функція $F(x) = \sum_{n\geq 1:\ x_n<x} 2^{-n}$ є дискретною функцією розподілу, що не є кусково-сталою.

(2) Функція Кантора визначається формулою $F(x) = \sum_{k:\ x_k=2} 2^{-k}$ на щільній множині D точок $x \in [0.1]$, які мають раціональний розклад у тернарній системі зчислення такого вигляду $x = 0, x_1...x_n...2222..$, де $x_k \in \{0, 1, 2\}$, та продовжується за неперервністю на $[0, 1]$. Довести неперервність F на D, однозначність продовження та сингулярність цієї функції.

(3) Нехай f – обмежена щільність розподілу, що дорівнює нулю поза інтервалом $[0, 1]$, $g = f/\sup f$, а (ξ, η) – випадкова точка всередині квадрату $[0, 1]^2$. Довести, що випадкова величина ξ за умови, що $\eta \leq f(\xi)$, має щільність f.

1.9.4. Рівномірний розподіл

Означення. *Випадкова величина ξ має* рівномірний розподіл *на відрізку $[a, b]$, позначення $\xi \simeq U(a, b)$, якщо її щільність є сталою всередині цього відрізку, та дорівнює нулю поза ним.*

Отже, ймовірність потрапляння величини у множину всередині відрізку як інтеграл від щільності пропорційна довжині цієї множини і не залежить від її положення – тобто виконується умова рівноймовірності значень. З умови *нормованості* виводимо, що *щільність* розподілу рівномірної на $[a, b]$ випадкової величини дорівнює

$$f_\xi(x) = (b - a)^{-1} 1\!\!\mathrm{I}_{x \in [a,b]},$$

а *функція розподілу* має вигляд

$$F_\xi(x) = \begin{cases} 0, & x < a, \\ (x - a)/(b - a), & x \in [a, b], \\ 1, & x > b. \end{cases}$$

1.9.5. Нормальний (Гауссів) розподіл

(а) **Означення.** *Випадкова величина ξ має* стандартний нормальний розподіл, *позначення $\xi \simeq N(0,1)$, якщо її щільність дорівнює*

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2), \quad \forall x \in \mathbb{R},$$

тобто функція розподілу має вигляд

$$F_\xi(x) = \Phi(x) \equiv \int_{-\infty}^{x} \varphi(y)dy.$$

Зауважимо, що саме такі щільність та функція розподілу є граничними в *інтегральній граничній теоремі Муавра – Лапласа.* Часто цей розподіл називають також гауссовим розподілом. Нормованість: $\Phi(\infty) = 1$ є наслідком виразу для інтеграла Пуассона, що відомий з курсу математичного аналізу.

(б) **Означення.** *Випадкова величина ξ має* нормальний розподіл *із параметрами μ і $\sigma > 0$ (що позначається як $\xi \simeq N(\mu, \sigma^2)$), якщо її можна зобразити у вигляді лінійного перетворення $\xi = \mu + \sigma\,\zeta$, де ζ* – стандартна нормальна *величина, тобто має стандартний нормальний розподіл.* Нормальна щільність розподілу *величини ξ має вигляд*

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \forall x \in \mathbb{R}.$$

Це рівняння можна взяти за інше означення величини ξ.

Щільність $f_\xi(x)$ обчислюється за наслідком *про розподіл лінійного перетворення випадкової величини.*

1.9.6. Логнормальний розподіл

Означення. *Випадкова величина ξ має* логнормальний розподіл, *позначення $\xi \simeq LN(\mu, \sigma^2)$, якщо її логарифм має нормальний розподіл з відповідними параметрами:* $\ln \xi \simeq N(\mu, \sigma^2)$.

Випадкову величину ξ можна зобразити у вигляді $\xi = \exp(\zeta)$, де $\zeta \simeq N(\mu, \sigma^2)$. Оскільки нормальний розподіл є граничним для сум великого числа незалежних величин, то логнормальний розподіл є граничним у мультиплікативних схемах, що містять добутки незалежних факторів.

1.9.7. Функція інтенсивності розподілу

Нехай випадкова величина ξ має *щільність* розподілу $f(x)$ і *функцію розподілу* $F(x)$. Значення $f(x)$ (за припущенням неперервності f) можна інтерпретувати як границю

$$f(x) = \lim_{h\to 0} h^{-1}\mathbf{P}(x \le \xi < x+h) = \lim_{h\to 0} h^{-1}\int_x^{x+h} f(y)dy.$$

На практиці одночасно з безумовною ймовірністю $\mathbf{P}(x \le \xi < x+h)$ важливе значення відіграє *умовна ймовірність* $\mathbf{P}(x \le \xi < x+h \mid \xi \ge x)$. Саме вона вказує на частку відмов за одиницю часу серед тих приладів, які не відмовили на початок поточного періоду, в той час як безумовна ймовірність задає частку відмов серед усіх приладів, що спостерігалися з самого початку випробувань. Розглянемо такий приклад. Нехай ξ – число років до відмови електричної лампи і $\mathbf{P}(\xi = 1) = 0.9$, $\mathbf{P}(\xi = 2) = 0.1$. У який рік експлуатації лампи перегоряють інтенсивніше ? Для відповіді на питання припустимо, що на початку було 1000 ламп. Тоді у середньому

Рік	Початково	Відмов	Імовірність	Залишок	Інтенсивність
1	1000	900	900/1000=0.9	100	900/1000=0.9
2	100	100	100/1000=0.1	0	100/ 100=1.0

і роком найбільш інтенсивних відмов є другий рік, а не перший.

Відповідним умовним аналогом щільності є функція, що дорівнює границі для умовних імовірностей

$$\lambda(x) = \lim_{h\downarrow 0}\ h^{-1}\mathbf{P}(x \le \xi < x+h \mid \xi \ge x).$$

За означенням *умовної ймовірності*

$$\mathbf{P}(x \le \xi < x+h \mid \xi \ge x) = \mathbf{P}(x \le \xi < x+h,\ \xi \ge x)/\mathbf{P}(\xi \ge x) =$$

$$\mathbf{P}(x \le \xi < x+h)\ /\ \mathbf{P}(\xi \ge x) = h\ f(x/(1-F(x)) + o(h),\ h \to 0.$$

Звідси приходимо до такого означення.

Означення. **Функцією інтенсивності** *невід'ємної випадкової величини, що має щільність $f(x)$ і функцію розподілу $F(x)$, називається функція*

$$\lambda(x) = f(x)/(1-F(x)),\ x \ge 0.$$

За означенням для всіх точок неперервності x щільності f

$$\mathbf{P}(x \le \xi < x+h \mid \xi \ge x) = h\lambda(x) + o(h),\ h \to 0.$$

Теорема (про обчислення розподілу через функцію інтенсивності). *Нехай випадкова величина ξ невід'ємна, абсолютно неперервна і* *має функцію інтенсивності* $\lambda(x)$*. Тоді її функція розподілу та щільність мають при* $x \geq 0$ *вигляд*

$$F(x) = 1 - \exp\left(-\int_0^x \lambda(y)dy\right), \quad f(x) = \lambda(x)\exp\left(-\int_0^x \lambda(y)dy\right).$$

Доведення. Обчислимо інтеграл

$$\int_0^x \lambda(y)dy = \int_0^x (1 - F(y))^{-1} f(y)dy = \int_0^x (1 - F(y))^{-1} dF(y) =$$
$$\int_{F(0)}^{F(x)} (1-u)^{-1} du = -\ln(1 - F(x)),$$

із заміною змінної $u = F(y)$ та урахуванням того, що щільність $f(x)$ є (майже скрізь) похідною для функції розподілу $F(x)$. Розв'язком цього рівняння є наведена рівність для $F(x)$. Із *монотонності* та *нормованості* функції розподілу виводимо характеристичні властивості інтенсивності

$$\lambda(x) \geq 0, \ \int_0^\infty \lambda(x)dx = \infty.$$

Далі, заміною змінної $u = \int_0^x \lambda(y)dy$ обчислимо інтеграл

$$\int_0^t \lambda(x)\exp\left(-\int_0^x \lambda(y)dy\right)dx = 1 - \exp\left(-\int_0^t \lambda(y)dy\right) = F(t),$$

звідки за означенням щільності дістанемо зображення для щільності $\boxdot$

1.9.8. Показниковий (експоненційний) розподіл

Означення. *Випадкова величина* ξ *має* показниковий розподіл *(або* експоненційний розподіл*) із параметром* $\lambda > 0$*, позначення* $\xi \simeq Exp(\lambda)$*, якщо вона абсолютно неперервна, невід'ємна і має сталу функцію інтенсивності, що дорівнює* λ*.*

Отже, щільність і функція розподілу величини ξ дорівнюють

$$f(x) = \lambda\ \exp(-\lambda x), \ \ F(x) = 1 - \exp(-\lambda x), \ \forall x \geq 0.$$

Теорема (про відсутність післядії для показникового розподілу). *Показникова випадкова величина* ξ *має властивість* ***відсутності післядії****, тобто для довільних* $t, s \geq 0$ *має місце тотожність*

$$\mathbf{P}(\xi \geq t + s \mid \xi \geq t) = \mathbf{P}(\xi \geq s).$$

Цю властивість можна проінтерпретувати так: за умови "виживання" показникова величина повністю забуває своє минуле.

Доведення теореми є очевидним наслідком мультиплікативності експоненти $\exp(-\lambda x)$ та означення *умовної ймовірності:*

$$\mathbf{P}(\xi \geq t + s \mid \xi \geq t) = \mathbf{P}(\xi \geq t + s, \xi \geq t)\ /\ \mathbf{P}(\xi \geq t) =$$

$$\mathbf{P}(\xi \geq t+s) \;/\; \mathbf{P}(\xi \geq t) = \exp(-\lambda(t+s)) \;/ \exp(-\lambda t) = \mathbf{P}(\xi \geq s) \; \boxdot$$

Вправа. Довести, що в класі невід'ємних та необмежених величин властивість відсутності післядії мають лише показникові величини. *Вказівка*: відсутність післядії еквівалентна мультиплікативності $Q(t+s) = Q(t)\;Q(s)$ для ймовірності виживання $Q(t) = \mathbf{P}(\xi \geq t)$, а єдиним монотонним розв'язком рівняння $Q(t+s) = Q(t)\;Q(s)$ з $Q(0) = 1$ є експонента.

1.9.9. Розподіл Вейбула

Важливими є розподіли зі змінною інтенсивністю.

Означення. *Випадкова величина ξ має* розподіл Вейбула *з параметрами $\lambda > 0$ і $a > 0$, якщо її функція інтенсивності є степеневою:*

$$\lambda(x) = \lambda a(\lambda x)^{a-1}.$$

Отже, при $x > 0$ щільність та функція розподілу Вейбула мають вигляд

$$f(x) = \lambda a(\lambda x)^{a-1} \exp(-(\lambda x)^a), \;\; F(x) = 1 - \exp(-(\lambda x)^a).$$

Функція розподілу Вейбула використовується в теорії стохастичних моделей як більш реальний замінник показникової функції розподілу, оскільки властивість відсутності післядії у більшості застосувань не є прийнятною. Дійсно, більш реалістичний характер поведінки функції інтенсивності полягає в тому, що на початковому етапі вона спадає (ефект початкових конструктивних відмов), лише потім наступає період відносної сталості, що змінюється етапом неухильного зростання (ефект старіння).

1.9.10. Розподіл Гомпертца

Означення. *Випадкова величина ξ має* розподіл Гомпертца *з параметрами $\lambda, \mu > 0$ і $a > 0$, якщо її функція інтенсивності є показниковою:*

$$\lambda(x) = \lambda + \mu \exp(\alpha x), \;\; x \geq 0.$$

Розподіл Гомпертца використовується у демографії як математична модель тривалості життя людини.

1.9.11. Бета-розподіл

Означення. *Випадкова величина ξ має* бета-розподіл *на $(0,1)$ з параметрами $\alpha, \beta > 0$, якщо її щільність розподілу має вигляд:*

$$f(x) = x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha,\beta), \;\; x \in (0,1),$$

де $B(\alpha, \beta)$ – повна бета-функція.

При $\alpha = \beta = 1$ даний розподіл збігається з *рівномірним розподілом*. У інших ипадках він відбиває можливу помірну варіативність щільності.

1.9.12. Розподіл Парето

Зазначені вище функції розподілу прямують до одиниці при $x \to \infty$ з експоненційною швидкістю. Для розподілу Парето ця швидкість є степеневою, що є суттєвим у деяких моделях.

Означення. *Випадкова величина ξ має* розподіл Парето, *якщо для деяких $\lambda > 0, \alpha > 0$ її функція розподілу при $x \geq 1/\lambda$ дорівнює*

$$\mathbf{P}(\xi < x) = 1 - (\lambda x)^{-\alpha}.$$

1.9.13. Розподіл Коші

Означення. *Випадкова величина ξ має* розподіл Коші, *якщо її можна зобразити у вигляді $\xi = tg\ \varphi$, де випадковий кут φ є рівномірно розподіленим на відрізку* $(-\pi/2, \pi/2)$.

Оскільки $\varphi \in (-\pi/2, \pi/2)$, то

$$\mathbf{P}(\xi < x) = \mathbf{P}(\varphi < arctg\ x) = \frac{1}{2} + \frac{1}{\pi} arctg\ x,\ x \in \mathbb{R},$$

і щільність ξ має вигляд $f_\xi(x) = \frac{1}{\pi}\frac{1}{1+x^2},\ x \in \mathbb{R}$.

Вправи

(1) Випадкова величина ξ має неперервну функцію розподілу $F(x)$. Довести, що випадкова величина $F(\xi)$ рівномірно розподілена на відрізку $[0, 1]$.

(2) Випадкова величина ξ рівномірно розподілена на $[0, 1]$. Довести, що випадкова величина $-\ln \xi$ має показниковий розподіл.

(3) Випадкова величина ξ рівномірно розподілена на $[0, 1]$. Для довільної функції розподілу F позначимо ліву обернену $F^{(-1)}(y) \equiv \sup(x : F(x) < y)$. Довести, що випадкова величина $F^{(-1)}(\xi)$ має функцію розподілу F.

(4) Випадкова величина ξ має показниковий розподіл з параметром λ. (а) Довести, що при $\alpha > 0$ величина $\xi^{1/\alpha}$ має розподіл Вейбула з параметрами $\lambda^{1/\alpha}, \alpha$. (б) Знайти розподіл випадкової величини $[\xi]$.

(5) Випадкова величина ξ має розподіл Коші. Довести, що величини: (а) $1/\xi$, (б) $2\xi/(1-\xi^2)$, (в) $(3\xi - \xi^3)/(1 - 3\xi^2)$ мають розподіл Коші.

1.10. Математичне сподівання дискретної випадкової величини

Дискретний розподіл або *функція розподілу* випадкової величини є досить складними функціональними характеристиками. Тому доцільно розглянути більш прості числові характеристики випадкових величин. Перша з них – математичне сподівання або ж середнє значення, характеристика положення.

Розглянемо такий **приклад**. Нехай *проста* випадкова величина ζ набуває значень $x_1, ..., x_m$ з імовірностями $p_1, ..., p_m$. Припустимо, що спостерігаються n *незалежних* реалізацій $\zeta_1, ..., \zeta_n$ величини ζ . Тоді сукупне середнє цих значень можна подати у вигляді

$$\mu_n = \frac{1}{n}\sum\nolimits_{k=1}^{n} \zeta_k = \frac{1}{n}\sum\nolimits_{k=1}^{n}\sum\nolimits_{i=1}^{m} x_i \mathbb{I}_{\{\zeta_k = x_i\}} =$$

$$\frac{1}{n}\sum\nolimits_{i=1}^{m} x_i \sum\nolimits_{k=1}^{n} \mathbb{I}_{\{\zeta_k = x_i\}} = \sum\nolimits_{i=1}^{m} x_i \frac{\nu_n(x_i)}{n},$$

де $\nu_n(x_i)$ – кількість тих спостережень із числа n, що дорівнюють x_i. За припущенням *стійкості частот*, з якого ми постулювали поняття ймовірності, $\nu_n(x_i)/n \to \mathbf{P}(\zeta = x_i) = p_i$ при $n \to \infty$. Тому $\mu_n \to \sum x_i\, p_i,\ n \to \infty$. Отже, сума в правій частині є граничною для вибіркових середніх і має відігравати суттєве значення.

Означення. *Нехай ξ – дискретна випадкова величина з множиною різних значень $\{x_n,\ n \ge 1\}$ та розподілом $(p_n \equiv \mathbf{P}(\xi = x_n), n \ge 1)$.* Математичним сподіванням дискретної величини ξ *називається сума ряду*

$$\mathbf{E}\xi \equiv \sum\nolimits_{n \ge 1} x_n\, p_n,$$

за умови, що ряд збігається абсолютно (для простих величин це так).

Зауваження. Враховуючи наведену вищу асимптотичну властивість, а також для скорочення, математичне сподівання випадкової величини інколи називають її середнім значенням. Хоча з наведених далі прикладів неважко зробити висновок, що середнє, як значення, може ніколи не набуватися.

Відмітимо також, що дане означення поширюється на випадок, коли *дискретна випадкова величина* ξ набуває векторні значення: $\{x_n\} \subset \mathbb{R}^d$.

Зауваження. Позначення $\mathbf{E}\xi$ походить від Expectation. Деякі автори позначають математичне сподівання величини ξ через $\mathrm{M}\xi$. Є.Б. Динкін та В.М. Шуренков запропонували використовувати символ $\mathbf{P}\xi$.

1.10.1. Математичне сподівання функції від дискретної величини

Теорема (про обчислення математичного сподівання функції від дискретної величини). *Нехай ξ – дискретна випадкова величина з множиною різних значень $\{x_n,\ n \geq 1\}$, а g – числова функція. Тоді*

$$\mathbf{E}g(\xi) = \sum\nolimits_{n\geq 1} g(x_n)\,\mathbf{P}(\xi = x_n).$$

Зауважимо, що на відміну від означення математичного сподівання серед значень $g(x_n)$ можуть бути однакові.

Доведення. Випадкова величина $\zeta = g(\xi)$ дискретна. Позначимо через $\{z_n\} = \{g(x_j), j \geq 1\}$ множину всіх різних її значень. Тоді за означенням

$$\mathbf{E}g(\xi) = \sum\nolimits_n z_n\,\mathbf{P}(g(\xi) = z_n) =$$
$$\sum\nolimits_n z_n\,\mathbf{P}(\cup_{\ j:\ g(x_j)\ =\ z_n}\{\xi = x_j\}) = \sum\nolimits_n z_n\ \sum\nolimits_{j:\ g(x_j)\ =\ z_n}\mathbf{P}(\xi = x_j) =$$
$$\sum\nolimits_n\sum\nolimits_{j:\ g(x_j)\ =\ z_n} g(x_j)\mathbf{P}(\xi = x_j) = \sum\nolimits_j g(x_j)\mathbf{P}(\xi = x_j)\ \boxdot$$

1.10.2. Властивості математичного сподівання дискретної величини

Теорема (про властивості математичного сподівання дискретної величини). *Нехай ξ, η – дискретні випадкові величини. Математичне сподівання має такі властивості:*

*(а – **позитивність**) Якщо $\xi \geq 0$, то $\mathbf{E}\xi \geq 0$.*

*(б – **однорідність**) $\mathbf{E}(c\xi) = c\mathbf{E}\xi$.*

*(в – **адитивність**) $\mathbf{E}(\xi + \eta) = \mathbf{E}\xi + \mathbf{E}\eta$.*

*(г – **монотонність**) Якщо $\xi \geq \eta$, то $\mathbf{E}\xi \geq \mathbf{E}\eta$.*

*(д – **неперервність знизу**) Нехай ξ_n, η – прості величини такі, що $0 \leq \xi_n \uparrow \xi \equiv \lim \xi_n \geq \eta$ при всіх ω. Тоді $\lim_{n\to\infty} \mathbf{E}\xi_n \geq \mathbf{E}\eta$.*

Доведення

Властивість (а) очевидна, оскільки всі значення ξ невід'ємні.

Твердження (б) випливає з означення, оскільки величина $c\xi$ набуває значень $\{cx_1, ..., cx_n, ...\}$ із тими самими ймовірностями.

Для доведення (в) припустимо, що ξ набуває різні значення $\{x_i\}$, а η набуває різних значень $\{y_j\}$. За теоремою *про обчислення математичного сподівання функції від дискретної величини* (ξ, η), яка дорівнює сумі $g((\xi, \eta)) = \xi + \eta$:

$$\mathbf{E}(\xi+\eta) = \sum_{i,j}(x_i+y_j)\,\mathbf{P}(\xi=x_i,\eta=y_j) =$$
$$\sum_{i,j} x_i\,\mathbf{P}(\xi=x_i,\eta=y_j) + \sum_{i,j} y_j\,\mathbf{P}(\xi=x_i,\eta=y_j) =$$
$$\sum_i x_i \sum_j \mathbf{P}(\xi=x_i,\eta=y_j) + \sum_j y_j \sum_i \mathbf{P}(\xi=x_i,\eta=y_j) =$$
$$\sum_i x_i\,\mathbf{P}(\xi=x_i) + \sum_j y_j\,\mathbf{P}(\eta=y_j) = \mathbf{E}\xi + \mathbf{E}\eta.$$

Рівність кратних та повторних сум випливає з абсолютної збіжності результуючих рядів. Передостання рівність є наслідком *сигма-адитивності* ймовірності та таких тотожностей: $\{\xi = x_i\} = \cup_j\{\xi = x_i, \eta = y_j\}$, і $\{\eta = y_j\} = \cup_i\{\xi = x_i, \eta = y_j\}$, де доданки справа *попарно несумісні*.

Монотонність (г) є очевидним наслідком (а) та (в):

$$0 \le \mathbf{E}(\xi-\eta) = \mathbf{E}\xi - \mathbf{E}\eta.$$

Для доведення (д) зауважимо, що існування границі випливає з монотонності послідовності $\mathbf{E}\xi_n$ внаслідок (г).

Розглянемо для довільного $\varepsilon > 0$ випадкові події $A_n = \{\xi_n \ge \eta - \varepsilon\}$. За умовою $A_n \uparrow \Omega$, звідки за *неперервністю ймовірності* $\mathbf{P}(A_n) \uparrow 1, n \to \infty$. Оскільки величина η *проста*, то її найбільше значення $m = \max_{\omega\in\Omega}\eta(\omega)$ є скінченною сталою.

Для всіх ω за означенням A_n справедлива нерівність

$$\xi_n \ge (\eta - \varepsilon)\mathbb{1}_{A_n} \ge \eta\mathbb{1}_{A_n} - \varepsilon =$$
$$\eta - \eta\mathbb{1}_{\overline{A}_n} - \varepsilon \ge \eta - \mathbb{1}_{\overline{A}_n}\max\eta - \varepsilon = \eta - m\,\mathbb{1}_{\overline{A}_n} - \varepsilon,$$

отже, за *монотонністю* математичного сподівання маємо при $n \to \infty$

$$\lim \mathbf{E}\xi_n \ge \lim \mathbf{E}(\eta - m\,\mathbb{1}_{\overline{A}_n} - \varepsilon) = \mathbf{E}\eta - m\,\lim \mathbf{P}(\overline{A}_n) - \varepsilon = \mathbf{E}\eta - \varepsilon,$$

оскільки $\mathbf{P}(\overline{A}_n) \to 0$. Переходячи в останній нерівності до границі $\varepsilon \to 0$, доводимо потрібну нерівність $\boxdot$

1.10.3. Приклади обчислення математичного сподівання дискретних величин

1. *Індикаторна величина*
$\mathbf{E}\mathbb{1}_A = 1 \cdot \mathbf{P}(A) + 0 \cdot \mathbf{P}(\overline{A}) = \mathbf{P}(A)$.
2. Рівномірний розподіл
Нехай $\mathbf{P}(\xi = k) = 1/n, k = \overline{1,n}$. *Тоді* $\mathbf{E}\xi = \sum_{k=1}^n k/n = (n+1)/2$.
3. *Біноміальний розподіл*
Нехай $\xi \simeq B(n,p)$. *Тоді* $\mathbf{E}\xi = np$. Наведемо три способи обчислення.
(а) $\mathbf{E}\xi = \sum_{k=0}^n kC_n^k\, p^k q^{n-k} = np\sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!}\, p^{k-1}q^{n-k} = np.$

(б) Розглянемо **генератрису**: $\varphi(z) \equiv \mathbf{E}z^\xi = \sum_{k=0}^n z^k C_n^k \; p^k q^{n-k} = (pz+q)^n$. Оскільки $\mathbf{E}\xi = \varphi'(z)\mid_{z=1}$, то $\mathbf{E}\xi = np(pz+q)^{n-1}\mid_{z=1} = np$.

(в) Нехай $\chi_k = 1\!\!\mathrm{I}_{Y_k}$ – *індикаторна величина* успіху в k-му *випробуванні Бернуллі*. Тоді $\mathbf{E}\chi_k = p$ і $\xi = \sum_{k=1}^n \chi_k$, отже, за *адитивністю* математичного сподівання $\mathbf{E}\xi = \sum_{k=1}^n \mathbf{E}\chi_k = np$.

4. *Геометричний розподіл. Для $\xi \simeq G(p)$ за методом генератрис* $\mathbf{E}\xi = \sum_{k\ge 1} k\, q^{k-1} p = p\left(\sum_{k\ge1} q^k\right)_q' = p/(1-q)^2 = 1/p$.

5. *Розподіл Пуассона. Для $\xi \simeq \Pi(\lambda)$ за означенням* $\mathbf{E}\xi = \sum_{k\ge 0} k\, \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k\ge1} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda$.

Вправи

(1) Знайти математичне сподівання числа листів, що дійдуть до адресата, у задачі про збіг із розділу про дискретний імовірнісний простір.

(2) k куль послідовно кидають навмання у n урн. Знайти математичне сподівання числа непорожніх урн.

(3) Довести, що математичне сподівання випадкової величини з гіпергеометричним розподілом та параметрами N, n, m дорівнює nm/N.

(4) Довести, що для цілозначної невід'ємної величини
$\mathbf{E}\xi = \sum_{n\ge1} \mathbf{P}(\xi \ge n) = \sum_{n\ge 0} \mathbf{P}(\xi > n)$.

(5) Нехай $\mathbf{P}(B) > 0$, а ξ – дискретна випадкова величина. Визначимо умовне математичне сподівання $\mathbf{E}(\xi \mid B) \equiv \mathbf{E}(\xi 1\!\!\mathrm{I}_B)/\mathbf{P}(B)$. Довести, що: (а) для нього виконуються всі властивості математичного сподівання дискретної величини, (б) $\mathbf{E}(\xi \mid B) = \sum_{n\ge1} x_n \mathbf{P}(\xi = x_n \mid B)$, де $\{x_n, n \ge 1\}$ – значення ξ.

(6) Довести, що для простої невід'ємної величини ξ
$\lim_{n\to\infty} \mathbf{E}\xi^{n+1}/\mathbf{E}\xi^n = \lim_{n\to\infty} \sqrt[n]{\mathbf{E}\xi^n} = \max_\omega \xi(\omega)$.

1.11. Загальне означення математичного сподівання

1.11.1. Невід'ємні випадкові величини

Нагадаємо, що запис $\xi_n \uparrow \xi$ означає монотонну збіжність випадкових величин: $\xi_n \le \xi_{n+1}$, $\lim_{n\to\infty} \xi_n = \xi$, $\forall \omega \in \Omega$.

Означення. *Нехай $\xi \ge 0$.* **Математичним сподіванням** *невід'ємної випадкової величини ξ назвемо монотонну границю*

$$\mathbf{E}\xi = \lim_{n\to\infty} \mathbf{E}\xi_n \le \infty,$$

де ξ_n – прості величини такі, що $0 \le \xi_n \uparrow \xi, n \to \infty$.

Зауваження. Існування вказаної послідовності ξ_n випливає з теореми *про апроксимацію випадкових величин простими*, а існування скінченої або нескінченної границі математичних сподівань $\mathbf{E}\xi_n \uparrow \mathbf{E}\xi$, $n \to \infty$, є наслідком монотонності цієї послідовності. Остання виводиться з твердження монотонності теореми *про властивості математичного сподівання дискретної величини*.

1.11.2. Коректність визначення математичного сподівання

Теорема (про коректність визначення математичного сподівання). *Границя* $\lim \mathbf{E}\xi_n$ *в означенні математичного сподівання існує, скінченна або нескінченна, і не залежить від вибору простих величин* ξ_n *таких, що* $0 \le \xi_n \uparrow \xi$. *Математичне сподівання* $\mathbf{E}\xi$ *визначене коректно.*

Доведення. Існування границі обґрунтовано в зауваженні. Для доведення її єдиності припустимо, що одночасно з ξ_n прості величини ξ_n' такі, що $0 \le \xi_n' \uparrow \xi$. Тоді $0 \le \xi_n \uparrow \xi \ge \xi_k'$ для кожного k. Отже, за властивістю (д) *неперервності знизу математичного сподівання дискретної величини* має місце нерівність $\lim_{n\to\infty} \mathbf{E}\xi_n \ge \mathbf{E}\xi_k'$. Переходячи тут до границі $k \to \infty$, отримуємо $\lim_{n\to\infty} \mathbf{E}\xi_n \ge \lim_{k\to\infty} \mathbf{E}\xi_k'$. Переставивши ξ_n' і ξ_n місцями, виводимо рівність $\lim \mathbf{E}\xi_n = \lim \mathbf{E}\xi_k'$, незважаючи на скінченність чи нескінченність границь $\boxdot$

Теорема (про інваріантне зображення математичного сподівання). *Справедлива рівність*

$$\mathbf{E}\xi = \sup\left(\mathbf{E}\eta : \ \eta \ \textit{проста}, \ 0 \le \eta \le \xi\right).$$

Доведення. Якщо η проста, $0 \le \eta \le \xi$, а $0 \le \xi_n \uparrow \xi$, ξ_n – прості, то величини $\max(\eta, \xi_n) \uparrow \xi$ і також прості. Звідси $\mathbf{E}\eta \le \mathbf{E}\max(\eta, \xi_n) \uparrow \mathbf{E}\xi$ за *монотонністю математичного сподівання дискретних величин* та за теоремою *про коректність визначення математичного сподівання*, отже $\mathbf{E}\eta \le \mathbf{E}\xi$. Тому верхня межа в правій частині рівності теореми не перевищує $\mathbf{E}\xi$.

З іншого боку, ця верхня межа досягається на деякій послідовності за означенням верхньої межі і відповідна границя не менша за $\mathbf{E}\xi$ $\boxdot$

1.11.3. Інтегровні невід'ємні випадкові величини

Означення. *Невід'ємна випадкова величина* ξ *називається* інтегровною, *якщо* $\mathbf{E}\xi < \infty$.

Теорема (про властивості інтегровних невід'ємних величин).
(а) Якщо $0 \le \xi \le \zeta$ *і* ζ *інтегровна, то* ξ *інтегровна і* $\mathbf{E}\xi \le \mathbf{E}\zeta$.
(б) Якщо $0 \le \xi, \zeta$ *інтегровні, то сума* $\xi + \zeta$ *також інтегровна і*

$$\mathbf{E}(\xi + \zeta) = \mathbf{E}\xi + \mathbf{E}\zeta.$$

Доведення

(а) Випливає з теореми *про інваріантне зображення математичного сподівання*, оскільки відповідні множини простих величин є вкладеними:

$$\{\eta : \eta \text{ проста}, 0 \le \eta \le \xi\} \subset \{\eta : \eta \text{ проста}, 0 \le \eta \le \zeta\},$$

а операції обчислення *математичного сподівання дискретної величини* та обчислення верхньої межі числової множини монотонні.

(б) Якщо $0 \le \xi_n \uparrow \xi$ та $0 \le \zeta_n \uparrow \zeta, n \to \infty$, де ξ_n, ζ_n – прості, то $0 \le \xi_n + \zeta_n \uparrow \xi + \zeta$, причому сума $\xi_n + \zeta_n$ є *простою* випадковою величиною. Отже, з теорем *про коректність визначення математичного сподівання* та *про властивості математичного сподівання дискретної величини*

$$\mathbf{E}(\xi + \zeta) = \lim \mathbf{E}(\xi_n + \zeta_n) = \lim \mathbf{E}\xi_n + \lim \mathbf{E}\zeta_n = \mathbf{E}\xi + \mathbf{E}\zeta < \infty$$

за властивістю границі суми збіжних числових послідовностей ⊡

1.11.4. Математичне сподівання знакозмінних випадкових величин

Нагадаємо, що *додатною та від'ємною частинами* знакозмінної величини ξ називаються невід'ємні випадкові величини

$$\xi^+ = \max(\xi, 0), \ \xi^- = \max(-\xi, 0),$$

З означення отримуємо (**вправа**): $\xi = \xi^+ - \xi^-$, $\xi^+ \cdot \xi^- = 0$, $\xi^+ + \xi^- = |\xi|$.

Означення. *Знакозмінна випадкова величина* ξ *називається* інтегровною, *якщо невід'ємна величина* $|\xi| = \xi^+ + \xi^-$ *є інтегровною. За цієї умови* математичним сподіванням ξ *називається число*

$$\mathbf{E}\xi = \mathbf{E}\xi^+ - \mathbf{E}\xi^-.$$

Зауваження. З теореми *про властивості інтегровних невід'ємних величин* виводимо, що інтегровність ξ еквівалентна одночасній інтегровності випадкових величин $\xi^{\pm}$.

1.11.5. Властивості математичного сподівання

Означення. *Висловлювання щодо результату стохастичного експерименту виконується* майже напевне *(скорочення:* м.н.*), якщо ймовірність множини сприятливих елементарних подій дорівнює одиниці.*

Теорема (про властивості математичного сподівання).

*(1 – **нормованість**)* $\mathbf{E}c = c$.

*(2 – **центрованість**) Якщо* $\xi = 0$ *м.н., то* $\mathbf{E}\xi = 0$.

*(3 – **невід'ємність**) Якщо* $\xi \geq 0$ *м.н., то* $\mathbf{E}\xi \geq 0$.

*(4 – **додатність**) Якщо* $\xi \geq 0$, *і* $\mathbf{E}\xi = 0$, *то* $\xi = 0$ *м.н.*

*(5 – **адитивність**) Якщо* ξ, ζ *– інтегровні, то* $\mathbf{E}(\xi + \zeta) = \mathbf{E}\xi + \mathbf{E}\zeta$.

*(6 – **однорідність**)* $\mathbf{E}(c\xi) = c\mathbf{E}(\xi)$.

*(7 – **монотонність**) Якщо* $\xi \leq \zeta$ *м.н., то* $\mathbf{E}\xi \leq \mathbf{E}\zeta$.

*(8 – **інваріантність**) Якщо* $\xi = \zeta$ *м.н., то* $\mathbf{E}\xi = \mathbf{E}\zeta$.

*(9 – **опуклість**)* $|\mathbf{E}\xi| \leq \mathbf{E}\,|\xi|$.

Зауважимо, що властивість лінійності математичного сподівання полягає в *однорідності* та *адитивності.*

Доведення

Рівність (1) очевидна, адже стала є *простою* випадковою величиною.

При доведенні (2) можна вважати, що $\xi \geq 0$, оскільки в загальному випадку з $\xi = 0$ *м.н.*та зі включень $\{\xi = 0\} \subset \{\xi^{\pm} = 0\}$ випливає $\xi^{\pm} = 0$ *м.н.*, а тоді за означенням $\mathbf{E}\xi = \mathbf{E}\xi^{+} - \mathbf{E}\xi^{-} = 0$. Далі, з $0 \leq \zeta \leq \xi$, де ζ – *проста*, а $\xi = 0$ *м.н.*, виводимо, що $\zeta = 0$ *м.н.* Тоді $\zeta = \sum_{k=1}^{n} c_k \mathbb{1}_{A_k}$, де $c_1 = 0, \mathbf{P}(A_1) = 1, \mathbf{P}(A_k) = 0, \forall k > 1$, і $\mathbf{E}\zeta = 0 \cdot 1 + 0 = 0$. Тому за теоремою *про інваріантне зображення математичного сподівання* $\mathbf{E}\xi = 0$.

В умовах (3) маємо $\xi^{-} = 0$ *м.н.* і внаслідок (2) $\mathbf{E}\xi^{-} = 0$. Оскільки $\xi^{+} \geq 0$, з *невід'ємності* математичного сподівання дискретної величини та теореми *про інваріантне зображення математичного сподівання* дістанемо $\mathbf{E}\xi^{+} \geq 0$. Тому $\mathbf{E}\xi = \mathbf{E}\xi^{+} - \mathbf{E}\xi^{-} \geq 0$.

Для доведення (4) припустимо, що $\xi \neq 0$ *м.н.* Тоді $\mathbf{P}(\xi \geq \varepsilon) > 0$ для деякого $\varepsilon > 0$, тому що

$$\{\xi \geq 0\} \cap \{\xi \neq 0\} = \{\xi > 0\} = \cup_{k>0}\{\xi > 1/k\}.$$

Оскільки $\xi \geq \varepsilon \ \mathbb{1}_{\{\xi \,\geq\, \varepsilon\}}$ для всіх ω, то за теоремою *про інваріантне зображення математичного сподівання*

$$\mathbf{E}\xi \geq \mathbf{E}\varepsilon \ \mathbb{1}_{\{\xi \,\geq\, \varepsilon\}} = \varepsilon \ \mathbf{P}(\xi \geq \varepsilon) > 0,$$

що суперечить умові.

Для доведення (5) зауважимо, що у випадку невід'ємних $\zeta, \xi \geq 0$ *адитивність* доведена вище. Для знакозмінних величин запишемо тотожність

$$\xi^+ + \zeta^+ - (\xi + \zeta)^+ = \ \xi^- + \zeta^- - (\xi + \zeta)^- \equiv \eta.$$

Оскільки $(x+y)^+ \leq x^+ + y^+$, то $0 \leq \eta \leq \xi^+ + \zeta^+$, отже η – невід'ємна і *інтегровна* за теоремою *про властивості інтегровних невід'ємних величин.* З *адитивності* математичного сподівання для невід'ємних величин випливає рівність $\mathbf{E}(\xi^\pm + \zeta^\pm - \eta) = \mathbf{E}(\xi^\pm + \zeta^\pm) - \mathbf{E}\eta$.

З урахуванням означення η звідси остаточно виводимо

$$\mathbf{E}(\xi + \zeta) = \mathbf{E}(\xi + \zeta)^+ - \mathbf{E}(\xi + \zeta)^- =$$
$$\mathbf{E}(\xi^+ + \zeta^+ - \eta) - \mathbf{E}(\xi^- + \zeta^- - \eta) =$$
$$\mathbf{E}(\xi^+ + \zeta^+) - \mathbf{E}\eta - \mathbf{E}(\xi^- + \zeta^-) + \mathbf{E}\eta =$$
$$\mathbf{E}\xi^+ + \mathbf{E}\zeta^+ - \mathbf{E}\xi^- - \mathbf{E}\zeta^- = \mathbf{E}\xi + \mathbf{E}\zeta.$$

Для $c \geq 0, \xi \geq 0$ доведення (6) випливає зі збіжності $0 \leq c\xi_n \uparrow c\xi$ та *однорідності математичного сподівання дискретних величин*:

$$\mathbf{E}(c\xi) = \lim \mathbf{E}(c\xi_n) = \lim c\mathbf{E}(\xi_n) = c\mathbf{E}\xi.$$

У загальному випадку скористаємося *адитивністю* (5) та тотожністю $c\xi = c^+\xi^+ - c^-\xi^+ - c^+\xi^- + c^-\xi^-$.

Доведення твердження (7). За умовою $\zeta - \xi \geq 0$ *м.н.*, тому внаслідок (3) та (5, 6) $\mathbf{E}\zeta - \mathbf{E}\xi = \mathbf{E}(\zeta - \xi) \geq 0$.

Доведення (8) випливає з твердження (7), якщо його застосувати до нерівностей $\xi \leq \zeta, \ -\xi \leq -\zeta$.

Властивість (9) отримуємо з (7) та нерівностей $-|\xi| \leq \xi \leq |\xi|$ ⊡

1.11.6. Перехід до границі під знаком математичного сподівання

У даному розділі збіжність випадкових величин будемо розуміти як поточкову, тобто збіжність при всіх $\omega \in \Omega$.

Теорема (теорема Лебега про монотонну збіжність). *Нехай величини* (ξ_n) *такі, що:* $0 \leq \xi_n \uparrow \xi, n \to \infty$, *для всіх* ω. *Тоді* $\mathbf{E}\xi_n \uparrow \mathbf{E}\xi \leq \infty$.

Доведення. За теоремою *про апроксимацію випадкових величин простими* оберемо для кожного n *прості* $0 \leq \zeta_{nk} \uparrow \xi_n$ при $k \to \infty$. Позначимо

$$\zeta_k = \max\nolimits_{n \leq k} \zeta_{nk}.$$

Тоді ζ_k – прості невід'ємні випадкові величини, та не спадають:

$$\zeta_k = \max_{n\le k} \zeta_{nk} \le \max_{n\le k} \zeta_{n,k+1} \le \max_{n\le k+1} \zeta_{n,k+1} = \zeta_{k+1}.$$

Крім того, з нерівності $\zeta_{nk} \le \xi_n$ та монотонності ξ_n виводимо, що

$$\zeta_k = \max_{n\le k} \zeta_{nk} \le \max_{n\le k} \ \xi_n = \xi_k.$$

Тому

$$\zeta_{nk} \le \zeta_k \le \xi_k \le \xi, \ \forall n \le k, \ \forall \omega.$$

Позначимо монотонну границю $\zeta = \lim \zeta_k$. Тоді з останніх нерівностей при $k \to \infty$ отримуємо $\xi_n \le \zeta \le \xi$, звідки при $n \to \infty$ маємо $\zeta = \xi$. За теоремою *про коректність визначення математичного сподівання* $\lim \mathbf{E}\zeta_k = \mathbf{E}\zeta = \mathbf{E}\xi$. Отже, з тих же нерівностей за *монотонністю* математичного сподівання

$$\mathbf{E}\xi = \lim \mathbf{E}\zeta_k \le \lim \mathbf{E}\xi_k \le \mathbf{E}\xi,$$

тобто $\lim \mathbf{E}\xi_k = \mathbf{E}\xi$, незалежно від того, скінченне $\mathbf{E}\xi$ чи ні $\boxdot$

Теорема (теорема Лебега про мажоровану збіжність). *Нехай величини* (ξ_n) *мажоруються інтегровною величиною:* $|\xi_n| \le \eta$, *та* $\mathbf{E}\eta < \infty$.

*(а –**Нерівності Фату**). Тоді*

$$\mathbf{E} \ \underline{\lim}_{n\to\infty}\xi_n \le \underline{\lim}_{n\to\infty}\mathbf{E}\xi_n \le \overline{\lim}_{n\to\infty}\mathbf{E}\xi_n \le \mathbf{E} \ \overline{\lim}_{n\to\infty}\xi_n.$$

*(б –**Мажорована збіжність**) Якщо додатково* $\xi_n \to \xi, \ n \to \infty$, *то*

$$\mathbf{E}\xi_n \to \mathbf{E}\xi, \quad \mathbf{E}\,|\xi_n - \xi| \to 0, \ n \to \infty.$$

Зауваження. Для справедливості лівої *нерівності Фату* достатньо виконання умови невід'ємності: $\xi_n \ge 0$.

Доведення

(а) Доведемо ліву нерівність. Позначимо $\zeta_n = \inf_{\ k\ge n} \xi_k$.

За означенням нижньої границі $\zeta_n \uparrow \underline{\lim}\xi_n$ і $\zeta_n \ge \inf_{\ k\ge n}(-\eta) = -\eta$. Отже, $0 \le \zeta_n + \eta \uparrow \underline{\lim} \ \xi_n + \eta$. За *адитивністю* математичного сподівання та за *теоремою Лебега про монотонну збіжність*

$$\mathbf{E}\underline{\lim}\xi_n + \mathbf{E}\eta = \mathbf{E}(\underline{\lim}\xi_n + \eta) = \mathbf{E}\lim(\zeta_n + \eta) = \lim \mathbf{E}(\zeta_n + \eta) =$$

$$\lim \mathbf{E}\zeta_n + \mathbf{E}\eta = \underline{\lim}\mathbf{E}\zeta_n + \mathbf{E}\eta \le \underline{\lim}\mathbf{E}\xi_n + \mathbf{E}\eta,$$

оскільки $\mathbf{E}\zeta_n \le \mathbf{E}\xi_n$ за *монотонністю* математичного сподівання.

Звідси випливає ліва нерівність Фату. Дане доведення справедливе також для довільних невід'ємних величин ξ_n, якщо в ньому обрати $\eta \equiv 0$.

Права нерівність Фату виводиться з лівої після підстановки $-\xi_n$ замість ξ_n та $\underline{\lim}(-\xi_n) = -\overline{\lim}\xi_n$, $\overline{\lim}(-\xi_n) = -\underline{\lim}\xi_n$.

Перше твердження (б) випливає з (а), якщо врахувати, що за означенням границі числової послідовності $\underline{\lim}\xi_n = \overline{\lim}\xi_n = \xi$, звідки

$$\mathbf{E}\xi = \mathbf{E}\,\underline{\lim}\xi_n \le \underline{\lim}\,\mathbf{E}\xi_n \le \overline{\lim}\,\mathbf{E}\xi_n \le \mathbf{E}\,\overline{\lim}\xi_n = \mathbf{E}\xi,$$

тобто всі нерівності тут є рівностями.

Для доведення другого твердження пункту (б) теореми застосуємо перше до *мажоровано збіжної* послідовності

$$\xi_n' = |\xi_n - \xi| \to 0,\ 0 \le \xi_n' \le 2\,\eta \ \boxdot$$

Вправи

(1) Випадкова величина ξ інтегровна тоді й тільки тоді, коли при деякому $h > 0$ абсолютно збігається ряд $I(\xi, h) \equiv \sum_{n\in\mathbb{Z}} nh\ \mathbf{P}(nh \le \xi < nh + h)$. Ця збіжність не залежить від $h > 0$, причому $\mathbf{E}\xi = \lim_{h\to 0} I(\xi, h)$.

(2) Довести справедливість формули для математичного сподівання дискретної (не простої) випадкової величини.

(3) Нехай $\xi_n \to \xi$ і $\eta_n \to \eta, n \to \infty$, м.н., та $|\xi_n| \le \eta_n$, причому $\mathbf{E}\eta_n \to \mathbf{E}\eta$. Довести, що $\mathbf{E}\xi_n \to \mathbf{E}\xi, n \to \infty$.

(4) Нехай ξ – інтегровна випадкова величина. Довести абсолютну неперервність інтеграла Лебега: для кожного $\varepsilon > 0$ знайдеться $\delta > 0$ таке, що з $A \in \mathfrak{F}$, $\mathbf{P}(A) \le \delta$ випливає, що $|\mathbf{E}\xi 1\!\!\mathrm{I}_A| \le \varepsilon$.

(5) Випадкова величина ξ називається узагальнено інтегровною за умови, коли $\min(\mathbf{E}\xi^+, \mathbf{E}\xi^-) < \infty$. В цьому разі $\mathbf{E}\xi \equiv \mathbf{E}\xi^+ - \mathbf{E}\xi^-$ визначено коректно. Вивести для таких величин твердження теореми про властивості математичного сподівання.

(6) Довести, що $\mathbf{P}(\underline{\lim}A_n) \le \underline{\lim}\ \mathbf{P}(A_n) \le \overline{\lim}\ \mathbf{P}(A_n) \le \mathbf{P}(\overline{\lim}A_n)$.

(7) Довести σ-адитивність математичного сподівання: для інтегровної величини ξ та попарно несумісних $A_n \in \mathfrak{F} : \mathbf{E}\xi 1\!\!\mathrm{I}_{\cup_{n\ge 1}A_n} = \sum_{n\ge 1} \mathbf{E}\xi 1\!\!\mathrm{I}_{A_n}$.

(8) Довести, що в теоремі про монотонну збіжність умову $0 \le \xi_1$ можна замінити на $\eta \le \xi_1$ з інтегровною η.

(9) Навести приклад послідовності $(\xi_n, n \ge 1)$ випадкових величин таких, що $\xi_n \to 0, n \to \infty$, але $\mathbf{E}\xi_n \ge 1$ внаслідок порушення умови мажорованості.

1.11.7. Абстрактний інтеграл Лебега

Якщо ξ – інтегровна або невід'ємна випадкова величина на *ймовірнісному просторі* $(\Omega, \mathfrak{F}, \mathbf{P})$, то для її *математичного сподівання* використовують також таке позначення:

$$\int_\Omega \xi(\omega)\mathbf{P}(d\omega) \equiv \mathbf{E}\xi.$$

Вимірним простором будемо називати пару, що складається з абстрактної множини Ω з виділеною *сигма-алгеброю* $\mathfrak{F}$ її підмножин, або ж трійку, з додатковим третім елементом – мірою μ на $\mathfrak{F}$.

Означення. **Інтегралом Лебега** за мірою μ *від вимірної функції* $\xi(\omega)$ *на вимірному просторі* $(\Omega, \mathfrak{F}, \mu)$ *називається число*

$$\int_{\Omega} \xi(\omega)\mu(d\omega) \equiv \mu(\Omega) \int_{\Omega} \xi(\omega)\mathbf{P}(d\omega),$$

у припущенні невід'ємності або інтегровності ξ, *де ймовірність* $\mathbf{P}$ *на* $\mathfrak{F}$ *визначається як* $\mathbf{P}(A) = \mu(A)/\mu(\Omega)$. *Інтегровна величина* ξ *називається* **інтегровною за мірою** μ.

Інтегралом від ξ *на множині* $A \in \mathfrak{F}$:

$$\int_{A} \xi(\omega)\mathbf{P}(d\omega)$$

називається інтеграл на Ω *від функції* $\xi(\omega)1\!\!1_A(\omega)$.

Зауваження. Інтеграл Лебега має всі властивості, що викладені в теоремах: *про коректність визначення математичного сподівання*, *про інваріантне зображення математичного сподівання*, *про властивості інтегровних невід'ємних величин*, *про властивості математичного сподівання*, в *теоремі Лебега про монотонну збіжність*, *теоремі Лебега про мажоровану збіжність*. Те саме стосується інтегралів Лебега – Стілтьєса та за мірою Лебега нижче.

1.11.8. Інтеграли Лебега – Стілтьєса та за мірою Лебега

Означення. *Нехай* F *– нормована міра Лебега – Стілтьєса на* $\mathfrak{B}(\mathbb{R})$, *а* $g(x)$ *– борелєва функція.* **Інтеграл Лебега – Стілтьєса** *від функції* g *за мірою* F *визначається як математичне сподівання випадкової величини* $g : \Omega \to \mathbb{R}$ *на ймовірнісному просторі* $(\Omega, \mathfrak{F}, \mathbf{P}) \equiv (\mathbb{R}, \mathfrak{B}(\mathbb{R}), F)$, *тобто як монотонна границя інтегралів від апроксимуючих простих борелєвих функцій із лінійним продовженням на знакозмінні борелєві функції.*

Обидва позначення для інтеграла Лебега – Стілтьєса

$$\int_{-\infty}^{\infty} g(x)dF(x) = \int_{\mathbb{R}} g(x)F(dx)$$

мають однаковий зміст.

Інтегралом Лебега – Стілтьєса від борелєвої функції g за довільною скінченною мірою μ на $\mathfrak{B}(\mathbb{R})$ називається добуток числа $\mu(\mathbb{R})$ на інтеграл за нормованою мірою $\mu(\cdot)/\mu(\mathbb{R})$.

Інша конструкція такого інтегралу використовує поняття **невласної функції розподілу** – *неспадної* обмеженої *неперервної зліва* функції, що відрізняється від *функції розподілу* лише властивістю *нормованості*, тобто на додатний множник. Доводиться, що кожна міра на $\mathfrak{B}(\mathbb{R})$ породжується приростами деякої невласної функції розподілу. Інтеграл Лебега за цією мірою називається інтегралом Лебега – Стілтьєса.

Нагадаємо, що **інтегралом Рімана – Стілтьєса** на $[a, b]$ від кусково-неперервної функції g називається спільна границя верхніх та нижніх інтегральних сум вигляду

$$\lim_{d(\pi)\to 0} \sum_{k=0}^{n} g_k(F(a_{k+1}) - F(a_k)) \equiv (LS) \int_{[a,b]} g(x)F(dx),$$

де $\pi = \{a = a_0 < a_1 < .. < a_{n+1} = b\}$ – розбиття відрізку $[a, b]$ з діаметром $d(\pi)$, а g_k – відповідно верхні або нижні межі функції g на відрізках $[a_k, a_{k+1}]$ цього розбиття.

Теорема (про збіжність інтегралів Рімана – Стілтьєса та Лебега – Стілтьєса). *Нехай F – неспадна обмежена неперервна зліва функція, функція g – кусково-неперервна на інтервалі $[a, b]$, і її точки розриву відрізняються від точок розриву F. Тоді інтеграл Рімана – Стілтьєса та інтеграл Лебега – Стілтьєса збігаються:*

$$(RS) \int_a^b g(x)dF(x) = (LS) \int_{[a,b]} g(x)F(dx).$$

Доведення не наводиться.

Означення. **Мірою Лебега** *на класі обмежених борелєвих множин $B \subset [a, b]$ називається функція $L(B) \equiv (b - a)F_{ab}(B)$, що подовжується на всі $B \in \mathfrak{B}(\mathbb{R})$ як $L(B) = \lim_{m,n\to\infty} L(B \cap [-m, n])$. Тут функція рівномірного розподілу дорівнює:*

$$F_{ab}(x) = \begin{cases} 0, \ x < a, \\ (x - a)/(b - a), \ x \in [a, b], \\ 1, \ x > b. \end{cases}$$

Означення. *Нехай $g(x)$ – борелєва функція.* **Інтеграл за мірою Лебега** *на відрізку $[a, b]$ від g визначається як нормований інтеграл Лебега – Стілтьєса за мірою $(b - a)F_{ab}(\cdot)$:*

$$\int_a^b g(x)dx \equiv (b - a) \int_{-\infty}^{\infty} g(x)dF_{ab}(x).$$

Зауваження. Можна довести, що для кусково-неперервних функцій g інтеграл за мірою Лебега збігається з інтегралом Рімана.

Вправа. Довести, що міра Лебега інваріантна відносно зсувів.

1.11.9. Теореми Фубіні та Радона – Нікодима.

Теорема Фубіні застосовується для обчислення кратних інтегралів.

Означення. *Вимірний простір* $(\Omega, \mathfrak{F}, \mu)$ *називається* прямим добутком *вимірних просторів* $(\Omega_k, \mathfrak{F}_k, \mu_k)$, $k = 1, 2$, *якщо:*

(а) $\Omega = \Omega_1 \times \Omega_2 \equiv \{\omega = (\omega_1, \omega_2),\ \omega_k \in \Omega_k\}$.

(б) $\mathfrak{F} = \mathfrak{F}_1 \otimes \mathfrak{F}_2 \equiv \sigma[\{A_1 \times A_2, \quad A_1 \in \mathfrak{F}_1, A_2 \in \mathfrak{F}_2\}]$.

(в) $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2),\ \forall A_k \in \mathfrak{F}_k$.

Теорема (теорема Фубіні про кратний та повторні інтеграли). *Нехай простір* $(\Omega, \mathfrak{F}, \mu) = (\Omega_1, \mathfrak{F}_1, \mu_1) \otimes (\Omega_2, \mathfrak{F}_2, \mu_2)$ *є прямим добутком, а* $\xi(\omega) = \xi(\omega_1, \omega_2)$ *випадкова величина на* $(\Omega, \mathfrak{F}, \mu)$

Якщо величина ξ *інтегровна за мірою* μ, *то функція* $\xi(\cdot, \omega_2)$ *для* μ_2*-майже всіх* ω_2 *інтегровна за мірою* μ_1, *величина* $\int_{\Omega_1} \xi(\omega_1, \omega_2)\mu_1(d\omega_1)$ *є інтегровною за мірою* μ_2 *і справедлива рівність*

$$\int_{\Omega} \xi(\omega)\mu(d\omega) = \int_{\Omega_2} \left[\int_{\Omega_1} \xi(\omega_1, \omega_2)\mu_1(d\omega_1)\right] \mu_2(d\omega_2).$$

Для невід'ємної величини ξ *має місце обернене: зі збіжності повторного інтегралу випливає інтегровність* ξ *та збіжність обох інтегралів.*

Доведення не наводиться.

Означення. *Числова функція* $\mu : \mathfrak{F} \to \mathbb{R}$ *на вимірному просторі* $(\Omega, \mathfrak{F})$ *називається* сигма-скінченною *мірою, якщо існує послідовність* $C_n \in \mathfrak{F}$ *таких, що кожне зі звужень* $\mu(\cdot \cap C_n)$ *є мірою на* $\mathfrak{F}$. *Інтеграл Лебега за такою мірою* μ *визначається як границя інтегралів за звуженнями.*

Прикладами *сигма-скінченних* мір є точкова міра та *міра Лебега.*

Означення. *Нехай* ν *– міра, а* μ *– сигма-скінченна міра на просторі* $(\Omega, \mathfrak{F})$. *Міра* ν абсолютно неперервна відносно міри μ, *позначення* $\nu \ll \mu$, *якщо для кожної множини* $A \in \mathfrak{F}$ *із* $\mu(A) = 0$ *випливає* $\nu(A) = 0$.

Теорема (теорема Радона – Нікодима). *Нехай* $(\Omega, \mathfrak{F}, \mu)$ *– вимірний простір зі сигма-скінченною мірою* μ, *а міра* ν *задана на* $\mathfrak{F}$ *і абсолютно неперервна відносно міри* $\mu : \nu \ll \mu$. *Тоді знайдеться* $\mathfrak{F}$*-вимірна інтегровна за мірою* μ *функція* f *на* Ω *така, що*

$$\nu(A) = \int_A f(\omega)\mu(d\omega),\ \forall A \in \mathfrak{F}.$$

Ця функція f *називається* **щільністю міри** ν *відносно* μ *та позначається через*

$$f(\omega) = \frac{d\nu}{d\mu}(\omega).$$

Доведення не наводиться.

1.11.10. Обчислення математичного сподівання

Теорема (про обчислення математичного сподівання функції від випадкової величини). *Нехай випадкова величина ξ має функцію розподілу $F_\xi(x)$, що породжує відповідну міру Лебега – Стілтьєса $F_\xi(\cdot)$ на $\mathfrak{B}(\mathbb{R})$, а $g(x)$ – довільна борелєва функція. Випадкова величина $g(\xi)$ інтегровна тоді й тільки тоді, коли функція $g(x)$ інтегровна за мірою $F_\xi(\cdot)$, і за цим припущенням*

$$\mathbf{E}g(\xi) = \int_{-\infty}^{\infty} g(x)\, dF_\xi(x).$$

Зокрема, дане твердження справедливе при $g(x) \equiv x$.

Зауваження. Це твердження можна розглядати як теорему про заміну змінної ω на x за допомогою відображення ξ в *інтегралі Лебега*

$$\int_\Omega g(\xi(\omega))\mathbf{P}(d\omega) = \int_\mathbb{R} g(x)\mathbf{P}(\xi^{(-1)}(dx)) = \int_\mathbb{R} g(x)F_\xi(dx).$$

Доведення. Нехай $g(x) = \sum c_k 1\!\!1_{x\in B_k}$ *проста* функція. Тоді тотожність (завжди визначених) інтегралів випливає з *лінійності математичного сподівання* та *інтеграла Лебега – Стілтьєса*, а також із теореми *про обчислення ймовірностей,пов'язаних із випадковою величиною* ξ :

$$\mathbf{E}g(\xi) = \sum g(c_k)\mathbf{P}(\xi \in B_k) = \sum g(c_k)F_\xi(B_k) = \int_{-\infty}^{\infty} g(x)\, dF_\xi(x).$$

Нехай $g(x)$ – невід'ємна *борелєва* функція. Розглянемо *прості* борелєві функції $g_n(x) = \varphi_n(g(x))$, де дійсні функції φ_n визначені в теоремі *про апроксимацію випадкових величин простими.* З цієї теореми випливає поточкова *монотонна збіжність* $0 \le g_n(x) \uparrow g(x), n \to \infty$. Шукану рівність для $g(x)$ дістанемо граничним переходом $n \to \infty$ з рівності, що вже доведена для простих функцій $g_n(x)$, за *теоремами Лебега про монотонну збіжність* для математичного сподівання та для інтеграла Лебега – Стілтьєса. Зауважимо, що рівність є справедливою як для інтегровних, так і для неінтегровних невід'ємних g. Тому функції $g(\xi)$ та $g(x)$ є *інтегровними* одночасно.

Для знакозмінних g використаємо *лінійність* математичного сподівання та *інтеграла Лебега – Стілтьєса*:

$$\mathbf{E}g(\xi) = \mathbf{E}g^+(\xi) - \mathbf{E}g^-(\xi) = \int_{-\infty}^{\infty} g^+(x)\, dF_\xi(x) - \int_{-\infty}^{\infty} g^-(x)\, dF_\xi(x) \boxdot$$

Теорему *про обчислення математичного сподівання функції від дискретної величини* наведено вище.

Теорема (про обчислення математичного сподівання функції від неперервної величини). *Нехай випадкова величина ξ має щільність $f_\xi(x)$, а $g(x)$ – довільна борелєва функція. Випадкова величина $g(\xi)$ інтегровна тоді й тільки тоді, коли функція $g(x)f_\xi(x)$ інтегровна за мірою Лебега на $\mathbb{R}$, і має місце тотожність*

$$\mathbf{E}g(\xi) = \int_{-\infty}^{\infty} g(x)\ f_\xi(x)\ dx.$$

Зауваження. Якщо функції $g(x),\ f_\xi(x)$ кусково неперервні, то *інтеграл Лебега* в правій частині можна обчислювати як інтеграл Рімана.

Доведення. Розглянемо клас $\mathfrak{L}$ борелєвих функцій g, для яких має місце рівність $\int_{-\infty}^{\infty} g(x)dF_\xi(x) = \int_{-\infty}^{\infty} g(x)f_\xi(x)\ dx$.За теоремою *про обчислення ймовірностей, пов'язаних із неперервною величиною*, клас $\mathfrak{L}$ містить всі індикаторні функції $g(x) = 1\!\!\mathrm{I}_B(x)$ для борелєвих B. За лінійністю він містить всі прості функції $g(x)$. За *теоремами Лебега про монотонну збіжність* та про мажоровану збіжність клас $\mathfrak{L}$ замкнений відносно монотонної та *мажорованої збіжності.* Тому $\mathfrak{L}$ містить всі інтегровні функції.

Отже, остаточно дана теорема випливає з теореми *про обчислення математичного сподівання функції від випадкової величини* $\boxdot$

Розглянемо приклади обчислення математичного сподівання.

1. *Рівномірний розподіл:* для $\xi \simeq U(a,b)$:
   $\mathbf{E}\xi = \int_a^b x(b-a)^{-1}dx = (a+b)/2.$
2. *Показниковий розподіл:* для $\xi \simeq Exp(\lambda)$:
   $\mathbf{E}\xi = \int_0^\infty x\ \lambda \exp(-\lambda x)dx = 1/\lambda.$
3. *Нормальний розподіл:* для $\xi \simeq N(\mu, \sigma^2)$:
   $\mathbf{E}\xi = \mathbf{E}(\mu + \sigma\ \zeta) = \mu + \sigma\ \mathbf{E}\zeta = \mu,$

де ζ – *стандартна нормальна* величина, $\mathbf{E}\zeta = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx = 0.$

Наслідок: параметр μ нормального розподілу збігається з математичним сподіванням відповідної нормальної величини.

4. *Розподіл Коші.* Відповідна випадкова величина не є *інтегровною*, оскільки інтеграл $\int_{-\infty}^{\infty} x\frac{1}{1+x^2}dx$ не є абсолютно збіжним.

Вправи

(1) Довести, що суму абсолютно збіжного ряду можна подати у вигляді інтеграла Лебега за лічильною мірою $\lambda(B) = \sum_{n\in\mathbb{Z}} 1\!\!\mathrm{I}_{n\in B}$.

(2) Довести, що клас борелєвих функцій, які є інтегровними за нормованою мірою Лебега – Стілтьєса, містить всі обмежені та деякі необмежені функції.

(3) Невід'ємна випадкова величина ξ інтегровна тоді й тільки тоді, коли (а) $\int_0^\infty (1 - F_\xi(x))dx < \infty$, або (б) $\exists a > 0 : \int_a^\infty (-\ln F_\xi(x))dx < \infty$.

(4) Випадкова величина $|\xi|^\alpha$ з $\alpha > 0$, інтегровна тоді й тільки тоді, коли $\mathbf{E}\,|\xi|^\alpha = \int_{-\infty}^{\infty} |x|^{\alpha-1}\,(F_\xi(-x) + 1 - F_\xi(x))dx < \infty$.

(5) Довести, що для інтегровних випадкових величин ξ, η :
$\mathbf{E}\xi = \int_0^\infty (F_\xi(-x) + 1 - F_\xi(x))dx$,
$\mathbf{E}\xi - \mathbf{E}\eta = \int_{-\infty}^{\infty} (\mathbf{P}(\eta < x \le \xi) - \mathbf{P}(\xi < x \le \eta))\mathrm{sign}(x)dx$.

(6) Нехай $\mathbf{P}(|\xi| \ge x) = o(x^\alpha), x \to \infty$, для деякого $\alpha > 0$. Довести, що: (а) $\mathbf{E}\,|\xi|^\beta < \infty$ для всіх $\beta \in (0, \alpha)$, (б) твердження (а) не має місця при $\beta = \alpha$.

(7) Випадкова величина ξ така, що $\mathbf{P}(|\xi| \ge \alpha n) = o(\mathbf{P}(|\xi| \ge n), n \to \infty$, для кожного $\alpha > 1$. Довести, що $\mathbf{E}\,|\xi|^\beta < \infty$ при всіх $\beta > 0$.

(8) Нехай G – неспадна диференційовна функція на $[a, b]$ з похідною g, а f – інтегровна за мірою Лебега функція на $[G(a), G(b)]$. Довести формулу заміни змінної $x = G(y)$: $\int_{G(a)}^{G(b)} f(x)dx = \int_a^b f(G(y))g(y)dy$.

1.12. Дисперсія та її властивості

Математичне сподівання випадкової величини можна назвати характеристикою її "положення", оскільки воно є границею середніх арифметичних послідовних спостережень. Порівнюючи 2 лотерейні білети з однаковими середніми виграшами (перший виграє 10$ з імовірністю 0.1, другий 1000$ з імовірністю 0.001), приходимо до висновку, що одного середнього для порівняння випадкових величин замало – необхідна також характеристика ступеня "розсіяння" ймовірнісної маси.

Означення. Дисперсією *інтегровної випадкової величини ξ називається середній квадрат відхилення від середнього:*

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

Корінь квадратний $\sigma_\xi = \sqrt{\mathbf{D}\xi}$ називається квадратичним відхиленням (*або ж* стандартним відхиленням) *випадкової величини ξ.*

Від Variation йде також інше позначення дисперсії через $\mathbb{V}\xi$.

Зауваження. Дисперсія $\mathbf{D}\xi$ скінченна тоді й тільки тоді, коли ξ є квадратично інтегровною, тобто $\mathbf{E}\xi^2 < \infty$. Дійсно, внаслідок нерівності $2\,|\xi| \le 1 + \xi^2$ з квадратичної інтегровності випливає інтегровність ξ.

1.12.1. Властивості дисперсії

Теорема (про властивості дисперсії). *За умови квадратичної інтегровності випадкової величини ξ :*

(a) $\mathbf{D}\xi \ge 0$,

(б) $\mathbf{D}\xi = 0$ *тоді й тільки тоді, коли* $\xi = \mathbf{E}\xi$ *майже напевне,*
(в) $\mathbf{D}\xi = \mathbf{E}\xi^2 - (\mathbf{E}\xi)^2$,
(г) $\mathbf{D}(a\xi + b) = a^2\mathbf{D}\xi$,
(д) $\mathbf{D}\xi = \min_{c\in\mathbb{R}} \mathbf{E}(\xi - c)^2$, $\mathbf{E}\xi = arg \min_{c\in\mathbb{R}} \mathbf{E}(\xi - c)^2$.

Доведення

(а) Випливає з *позитивності* математичного сподівання, оскільки величина $(\xi - \mathbf{E}\xi)^2 \geq 0$.

(б) Виводиться з невід'ємності квадратичної функції та з теореми *про властивості математичного сподівання*, а саме, з властивостей *невід'ємності* і *додатності*.

Для доведення (в) обчислимо за *лінійністю* математичного сподівання $\mathbf{D}\xi = \mathbf{E}(\xi^2 - 2\xi\mathbf{E}\xi + (\mathbf{E}\xi)^2) = \mathbf{E}\xi^2 - (\mathbf{E}\xi)^2$.

(г) За означенням дисперсії та *лінійністю* математичного сподівання

$$\mathbf{D}(a\xi + b) = \mathbf{E}(a\xi + b - \mathbf{E}(a\xi + b))^2 = \mathbf{E}(a\xi - a\mathbf{E}\xi)^2 = a^2\mathbf{D}\xi.$$

Для доведення (д) обчислимо

$$\mathbf{E}(\xi - c)^2 = \mathbf{E}(\xi - \mathbf{E}\xi + \mathbf{E}\xi - c)^2 =$$
$$\mathbf{E}(\xi - \mathbf{E}\xi)^2 + (\mathbf{E}\xi - c)^2 + 2\mathbf{E}(\xi - \mathbf{E}\xi)(\mathbf{E}\xi - c) =$$
$$\mathbf{E}(\xi - \mathbf{E}\xi)^2 + (\mathbf{E}\xi - c)^2 \geq \mathbf{E}(\xi - \mathbf{E}\xi)^2 = \mathbf{D}\xi,$$

причому рівність має місце тоді й тільки тоді, коли $c = \mathbf{E}\xi$ ⊡

Теорема (про обчислення дисперсії). *Нехай величина* ξ *має функцію розподілу* $F_\xi(x)$. *Тоді*

$$\mathbf{D}\xi = \int_{-\infty}^{\infty} (x - \mathbf{E}\xi)^2 \, dF_\xi(x) = \int_{-\infty}^{\infty} x^2 \, dF_\xi(x) - (\mathbf{E}\xi)^2.$$

Якщо існує щільність розподілу $f_\xi(x)$, *то*

$$\mathbf{D}\xi = \int_{-\infty}^{\infty} (x - \mathbf{E}\xi)^2 \, f_\xi(x)dx = \int_{-\infty}^{\infty} x^2 \, f_\xi(x)dx - (\mathbf{E}\xi)^2.$$

Доведення є наслідком теореми *про обчислення математичного сподівання функції від випадкової величини*, а саме для функції $g(x) = (x - \mathbf{E}\xi)^2$ ⊡

1.12.2. Приклади обчислення дисперсії

1. *Індикаторна величина:*

$$\mathbf{D}\mathbb{I}_A = 1 \cdot \mathbf{P}(A) + 0 \cdot \mathbf{P}(\overline{A}) - (\mathbf{P}(A))^2 = \mathbf{P}(A)(1 - \mathbf{P}(A)).$$

2. *Біноміальний розподіл:* $\xi \simeq B(n, p)$, $\mathbf{D}\xi = npq$.

Розглянемо *генератрису*

$$\varphi(z) = \mathbf{E}z^{\xi} = \sum\nolimits_{k=0}^{n} z^k C_n^k\, p^k q^{n-k} = (pz+q)^n.$$

За формулою *про обчислення математичного сподівання функції від дискретної величини*

$$\mathbf{E}\xi(\xi-1) = \sum\nolimits_{k=0}^{n} k(k-1) C_n^k\, p^k q^{n-k} = \varphi''(z)\,|_{z=1},$$

$$\mathbf{D}\xi = \mathbf{E}\xi(\xi-1) + \mathbf{E}\xi - (\mathbf{E}\xi)^2 =$$
$$n(n-1)p^2(pz+q)^{n-2}\,|_{z=1} + np - (np)^2 = npq.$$

3. *Розподіл Пуассона:* $\xi \simeq \Pi(\lambda)$, $\mathbf{D}\xi = \lambda$.

$$\mathbf{D}\xi = \mathbf{E}\xi(\xi-1) + \mathbf{E}\xi - (\mathbf{E}\xi)^2 =$$
$$\sum\nolimits_{k\ge 0} k(k-1)\,\frac{\lambda^k}{k!}e^{-\lambda} + \lambda - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

4. *Рівномірний розподіл:* $\xi \simeq U(a,b)$, $\mathbf{D}\xi = (b-a)^2/12$.

$$\mathbf{D}\xi = \int_a^b x^2 \frac{1}{b-a} dx - (a+b)^2/4 = (b-a)^2/12.$$

5. *Показниковий розподіл:* $\xi \simeq Exp(\lambda)$, $\mathbf{D}\xi = 1/\lambda^2$.

$$\mathbf{D}\xi = \int_0^{\infty} x^2\, \lambda \exp(-\lambda x) dx - 1/\lambda^2 = 1/\lambda^2.$$

6. *Нормальний розподіл:* $\xi \simeq N(\mu,\sigma^2)$, $\mathbf{D}\xi = \sigma^2$.
За формулою (г) про дисперсію лінійного перетворення

$$\mathbf{D}\xi = \mathbf{D}(\mu + \sigma\,\zeta) = \sigma^2 \mathbf{D}\zeta = \sigma^2,$$

де ζ – *стандартна нормальна* величина, для якої заміною $x^2 = y$ виводимо

$$\mathbf{D}\zeta = \mathbf{E}\zeta^2 = \int_{-\infty}^{\infty} x^2 (2\pi)^{-1/2} \exp(-x^2/2) dx = \Gamma(3/2)\pi^{-1/2} = 1.$$

Наслідок: параметр σ^2 *нормального розподілу* збігається з дисперсією відповідної нормальної величини.

1.13. Імовірнісні нерівності

1.13.1. Нерівності Чебишева

Теорема (нерівність Маркова). *Нехай g – додатна неспадна функція, а величина $\xi \ge 0$. Тоді для довільної сталої $c > 0$*

$$\mathbf{P}(\xi \ge c) \le \frac{\mathbf{E}g(\xi)}{g(c)}.$$

Доведення. З невід'ємності і монотонності функції g виводимо нерівність $g(\xi) \geq g(c)\mathbb{1}_{\{\xi \,\geq\, c\}}$ для всіх $\omega \in \Omega$. Звідси за *монотонністю* математичного сподівання дістанемо шукану нерівність:

$$\mathbf{E}g(\xi) \geq \mathbf{E}g(c)\mathbb{1}_{\{\xi \,\geq\, c\}} = g(c)\mathbf{P}(\xi \ \geq c) \ \boxdot$$

Теорема (нерівність Чебишева). *Для кожного $\varepsilon > 0$ виконується нерівність*

$$\mathbf{P}(|\xi - \mathbf{E}\xi| \geq \varepsilon) \leq \mathbf{D}\xi \ / \ \varepsilon^2.$$

Доведення. Із *загальної нерівності Чебишева* для невід'ємної випадкової величини $(\xi - \mathbf{E}\xi)^2$ та функції $g(x) \equiv x^+$ отримуємо

$$\mathbf{P}(|\xi - \mathbf{E}\xi| \geq \varepsilon) = \mathbf{P}((\xi - \mathbf{E}\xi)^2 \geq \varepsilon^2) \leq \mathbf{E}(\xi - \mathbf{E}\xi)^2/\varepsilon^2 = \mathbf{D}\xi \ / \ \varepsilon^2 \ \boxdot$$

Правило трьох сигма. Нерівність Чебишева дає загрублену оцінку відхилення випадкової величини від свого математичного сподівання, оскільки виконується для довільного розподілу та не враховує специфіку нормального розподілу. Якщо ж величина ξ має *нормальний розподіл*, а $\sigma_\xi = \sqrt{\mathbf{D}\xi}$ – її *квадратичне відхилення*, то при $\varepsilon = 3\sigma_\xi$ ліва частина нерівності Чебишева наближено дорівнює 0.003. Останньою ймовірністю часто можна знехтувати. Тому, наприклад, у техніці вважають, що випадкові похибки при обробці деталей завжди не перевищують $3\sigma_\xi$. Це і є правило трьох сигма.

1.13.2. Нерівності Йенсена та Ляпунова

Теорема (нерівність Йенсена). *Нехай функція $g(x)$ опукла донизу (як x^2), а випадкові величини ξ та $g(\xi)$ інтегровні. Тоді*

$$g(\mathbf{E}\xi) \leq \mathbf{E}g(\xi).$$

Доведення. Для довільної опуклої донизу функції g в кожній точці x існує *опорна пряма*, графік якої лежить цілком під графіком функції:

$$g(x) \geq g(x_0) + k(x_0)(x - x_0), \quad \forall x \in \mathbb{R}.$$

Підставимо в цю нерівність $x = \xi$, $x_0 = \mathbf{E}\xi$ та скористаємося *монотонністю* математичного сподівання:

$$\mathbf{E}g(\xi) \geq \mathbf{E}(g(\mathbf{E}\xi) + k(\mathbf{E}\xi)(\xi - \mathbf{E}\xi)) = g(\mathbf{E}\xi) + k(\mathbf{E}\xi)\mathbf{E}(\xi - \mathbf{E}\xi) = g(\mathbf{E}\xi) \ \boxdot$$

Теорема (нерівність Ляпунова). *Якщо $1 \leq s \leq t$, то*

$$(\mathbf{E}\,|\xi|^s)^{1/s} \leq (\mathbf{E}\,|\xi|^t)^{1/t}.$$

Доведення. Оскільки функція $g(x) = x^{t/s}$, $x \geq 0$, опукла донизу, то потрібна нерівність випливає з *нерівності Йенсена*

$$(\mathbf{E}\,|\xi|^s)^{t/s} = g(\mathbf{E}\,|\xi|^s) \leq \mathbf{E}g(|\xi|^s) = \mathbf{E}\,|\xi|^t \ \boxdot$$

1.13.3. Нерівності Гельдера та Коші

Теорема (нерівність Гельдера). *Якщо числа* $p, q > 0$ *спряжені, тобто:* $1/p + 1/q = 1$, *то*

$$\mathbf{E}\,|\xi\eta| \leq (\mathbf{E}\,|\xi|^p)^{1/p}\,(\mathbf{E}\,|\eta|^q)^{1/q}.$$

Доведення. Підставимо в елементарну нерівність $xy \leq x^p/p + y^q/q$ вирази $x = |\xi|\,/(\mathbf{E}\,|\xi|^p)^{1/p}$ та $y = |\eta|\,/(\mathbf{E}\,|\eta|^q)^{1/q}$ та використаємо *монотонність* математичного сподівання:

$$\mathbf{E}\,|\xi\eta|\,/(\mathbf{E}\,|\xi|^p)^{1/p}(\mathbf{E}\,|\eta|^q)^{1/q} \leq$$

$$\mathbf{E}\,|\xi|^p\,/p\mathbf{E}\,|\xi|^p + \mathbf{E}\,|\eta|^q\,/q\mathbf{E}\,|\eta|^q = 1/p + 1/q = 1 \ \boxdot$$

Теорема (нерівність Коші). *Для квадратично інтегровних* ξ, η

$$(\mathbf{E}\xi\eta)^2 \leq \mathbf{E}\xi^2\,\mathbf{E}\eta^2.$$

Доведення. Випливає з *нерівності Гельдера* при $p = q = 2$ $\boxdot$

Зауважимо, що за припущенням невиродженості ξ, тобто $\mathbf{E}\xi^2 > 0$, рівність у нерівності Коші має місце тоді й тільки тоді, коли $\eta = c\xi$ *майже напевне* для деякої сталої c. *Достатність* тут перевіряється підстановкою: $(\mathbf{E}\xi c\xi)^2 = c^2(\mathbf{E}\xi^2)^2 = \mathbf{E}\xi^2\,\mathbf{E}(c\xi)^2$.

Для доведення *необхідності* скористаємося властивістю *додатності* з теореми *про властивості математичного сподівання*, та такою тотожністю при виборі $c = \sqrt{\mathbf{E}\eta^2/\mathbf{E}\xi^2}$:

$$\mathbf{E}(\eta - c\xi)^2 = 2\left(\mathbf{E}\eta^2 - c\mathbf{E}\xi\eta\right) = 2\,c\left(\sqrt{\mathbf{E}\xi^2\,\mathbf{E}\eta^2} - \mathbf{E}\xi\eta\right) \ \boxdot$$

1.13.4. Нерівність Мінковського

Теорема (нерівність Мінковського). *Якщо число* $p \geq 1$, *то*

$$(\mathbf{E}\,|\xi + \eta|^p)^{1/p} \leq (\mathbf{E}\,|\xi|^p)^{1/p} + (\mathbf{E}\,|\eta|^p)^{1/p}.$$

Зауваження. Величину $\|\xi\| \equiv (\mathbf{E}\,|\xi|^p)^{1/p}$ можна розглядати як норму у лінійному просторі $L_p(\mathbf{P}) = \{\xi : \mathbf{E}\,|\xi|^p < \infty\}$ інтегровних у степені p

випадкових величин. Нерівність Мінковського гарантує необхідну умову напівадитивності для цієї норми.

Доведення. Нерівність при $p = 1$ вже доведена. Для $p > 1$ застосуємо нерівність Гельдера в правій частині нерівності

$$\mathbf{E}\,|\xi+\eta|^p \le \mathbf{E}\,|\xi|\,|\xi+\eta|^{p-1} + \ \mathbf{E}\,|\eta|\,|\xi+\eta|^{p-1} \le$$
$$(\mathbf{E}\,|\xi|^p)^{1/p}(\mathbf{E}\,|\xi+\eta|^{pq-q})^{1/q} + (\mathbf{E}\,|\eta|^p)^{1/p}(\mathbf{E}\,|\xi+\eta|^{pq-q})^{1/q} =$$
$$\left((\mathbf{E}\,|\xi|^p)^{1/p} + (\mathbf{E}\,|\eta|^p)^{1/p}\right)(\mathbf{E}\,|\xi+\eta|^p)^{1/q} \ \boxdot$$

Вправи

(1) Довести для відповідних α нерівність $\mathbf{P}(\xi \ge x) \le \exp(-ax)\mathbf{E}\exp(a\xi)$.

(2) Нехай Φ – стандартна нормальна функція розподілу, а φ – її щільність. Довести при $x > 0$ справедливість включень (а) $\Phi(-x)/\varphi(x) \in [x^{-1} - x^{-3}, x^{-1}]$, (б) $\Phi(-x)/\varphi(x) \in [2/(\sqrt{x^2+4}+x),\ 2/(\sqrt{x^2+2}+x)]$.

(3) Нехай $(p_k, k = \overline{1,n})$ – дискретний розподіл імовірностей, а $x_k > 0$. Довести нерівність $\prod_{k=1}^n x_k^{p_k} \le \sum_{k=1}^n x_k p_k$.

(4) Для якої величини ξ нерівність Чебишева може бути рівністю ?

(5) Для заданих x, m, s^2 знайти верхню межу ймовірностей $\mathbf{P}(\xi \ge x)$ за умови фіксованих $\mathbf{E}\xi = m$, $\mathbf{D}\xi = s^2$.

(6) Випадкова величина $\xi \ge 0$ і $\mathbf{E}\xi^2 < \infty$. Довести при $\varepsilon \in (0,1)$ нерівність $\mathbf{P}(\xi > \varepsilon\mathbf{E}\xi) \ge (1-\varepsilon)^2(\mathbf{E}\xi)^2/(\mathbf{E}\xi^2)$.

(7) За умови $|\xi| \le c$ довести, що $\mathbf{P}(|\xi - \mathbf{E}\xi| \ge \varepsilon) \ge (\mathbf{E}\xi^2 - \varepsilon^2)/c^2$.

(8) Довести, що $\mathbf{P}(\xi - \mathbf{E}\xi > x\sigma_\xi) \le 1/(1+x^2)$ при $x > 0$.

(9) Для інтегровної величини ξ довести, що: (а) при $t > 0$
$\mathbf{P}(|\xi - \mathbf{E}\xi| > t\mathbf{E}\,|\xi|) \le 1/t$, (б) $\mathbf{P}(|\xi| > t) = o(1/t), t \to \infty$.

(10) Нехай $p, q, r > 0$ такі, що $r^{-1} = p^{-1} + q^{-1}$, а $\xi, \eta \ge 0$. Довести нерівність $(\mathbf{E}(\xi\eta)^r)^{1/r} \le (\mathbf{E}\xi^p)^{1/p}(\mathbf{E}\eta^q)^{1/q}$.

(11) Нехай випадкова величина ν_n має біноміальний розподіл з параметрами n, p, а функція $H(\theta, p) = -(1-\theta)\ln\frac{1-\theta}{1-p} - \theta\ln\frac{\theta}{p}$. Довести для всіх $\theta \in (p, 1)$ нерівність $\mathbf{P}(\nu_n \ge n\theta) \le \exp(nH(\theta,p))$. Вивести звідси при $p = 1/2$ нерівність $\mathbf{P}(|\nu_n - n/2| \ge n\varepsilon) \le \exp(-n\varepsilon^2/2)$.

1.14. Сумісна функція розподілу випадкового вектора та її властивості

Означення. Сумісною функцією розподілу *(або ж* сукупною функцією розподілу*)* *випадкового вектора* $\xi = (\xi_1, ..., \xi_n)$ *називається дійсна* *функція* $F_\xi(x) : \mathbb{R}^n \to \mathbb{R}$, *яка в точці* $x = (x_1, ..., x_n)$ *дорівнює*

$$F_\xi(x) \equiv F_\xi(x_1, ..., x_n) = \mathbf{P}(\xi < x) \equiv \mathbf{P}(\xi_1 < x_1, ..., \xi_n < x_n).$$

У даному розділі порівняння векторів та інші операції будемо проводити покоординатно. Зокрема, запис $x < y$ еквівалентний $x_k < y_k$ при всіх $k = \overline{1, n}$. Запис $y \uparrow x$ визначає, що $y_k \uparrow x_k$ при кожному k.

1.14.1. Загальні властивості сумісної функції розподілу

Теорема (про властивості сумісної функції розподілу). *Нехай випадковий вектор* ξ *має сумісну функцію розподілу* F. *Тоді ця функція:*
(1) неспадна: $F(x) \le F(y)$ *при всіх* $x \le y$,
(2) нормована: $F(x) \to 0$ *при* $x \downarrow -\infty$ *так, що* $\min_k x_k \downarrow -\infty$, *та*
$F(x) \to 1$ *при* $x \uparrow +\infty$ *так, що* $\min_k x_k \uparrow +\infty$,
(3) неперервна зліва: $F(x - 0) \equiv \lim_{y\uparrow x,\ y<x} F(y) = F(x)$ *для всіх* x.
Зауваження. Збіжність у (2),(3) випливає з монотонності (1).
Доведення
(1) випливає з *монотонності ймовірності*, оскільки $\{\xi < x\} \subset \{\xi < y\}$,
(2) виводиться з *неперервності ймовірності*, оскільки

$$\{\xi < x(m)\} \downarrow \varnothing \text{ при } x(m) \downarrow -\infty \text{ так,що } \min_k x_k(m) \downarrow -\infty, m \to \infty,$$

$$\{\xi < x(m)\} \uparrow \Omega \text{ при } x(m) \uparrow \infty \text{ так,що } \min_k x_k(m) \uparrow +\infty,\ m \to \infty.$$

(3) доводиться аналогічно: $\{\xi < y\} \uparrow \{\xi < x\}$ при $y \uparrow x$ $\boxdot$

1.14.2. Невід'ємність приростів

На відміну від *функції розподілу* скалярної випадкової величини, *сумісна функція розподілу* має ще одну характеристичну властивість. Для її формулювання позначимо через Π_x *кут*

$$\Pi_x = (-\infty, x) = \Pi_{k=1}^n (-\infty, x_k), \quad x = (x_1, ..., x_n).$$

Для початку припустимо, що розмірність $n = 2$. Зауважимо, що паралелепіпед $[a, b) = [a_1, b_1) \times [a_2, b_2)$ зображується у вигляді вкладеної різниці вкладених різниць кутів

$$[a, b) = (\Pi_b \setminus \Pi_{b'}) \setminus (\Pi_{a'} \setminus \Pi_a),$$

де точки $b' = (a_1, b_2)$, $a' = (a_2, b_1)$.

Якщо випадковий вектор ξ має сумісну функцію розподілу F, то з властивості *ймовірності вкладеної різниці* подій і наведеного зображення виводимо, що

$$\mathbf{P}(\xi \in [a,b)) = \mathbf{P}(\xi \in \Pi_b \setminus \Pi_{b'}) - \mathbf{P}(\xi \in \Pi_{a'} \setminus \Pi_a) =$$

$$\mathbf{P}(\xi \in \Pi_b) - \mathbf{P}(\xi \in \Pi_{b'}) - \mathbf{P}(\xi \in \Pi_{a'}) + \mathbf{P}(\xi \in \Pi_a) = \Delta_{[a,b)}F,$$

де за означенням $\Delta_{[a,b)}F \equiv F(b) - F(b') - F(a') + F(a)$.

Вираз $\Delta_{[a,b)}F$ називається при $n = 2$ приростом сумісної функції розподілу на прямокутнику $[a,b)$.

У загальному випадку $n > 2$ правило чергування знаків при значеннях функції у вершинах паралелепіпеда $[a,b)$ визначається аналогічно.

Означення. *Нехай функція* $F = F(x_1, ..., x_n) : \mathbb{R}^n \to \mathbb{R}$. *Її* k*-м* **частковим приростом** *на* $[a_k, b_k)$ *називається різниця*

$$\Delta^k_{[a_k,b_k)}F = F(x_1, ..., b_k, ..., x_n) - F(x_1, ..., a_k, ..., x_n),$$

а **приростом на паралелепіпеді** $[a,b) \equiv \prod_{k=1}^n [a_k, b_k)$ – *результат послідовних часткових приростів*

$$\Delta_{[a,b)}F = \Delta^1_{[a_1,b_1)} ... \Delta^{n-1}_{[a_{n-1},b_{n-1})} \Delta^n_{[a_n,b_n)}F.$$

Теорема (про ймовірність значення всередині паралелепіпеда). *Якщо випадковий вектор* ξ *має сумісну функцію розподілу* F, *то*

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

Доведення проводиться за індукцією, з індуктивним припущенням

$$\mathbf{P}(\{\xi_1 \in [a_1,b_1) ..., \xi_{k-1} \in [a_{k-1}, b_{k-1})\} \cap A) =$$

$$\Delta^1_{[a_1,b_1)} ... \Delta^{k-1}_{[a_{k-1},b_{k-1})} \mathbf{P}(\{\xi_1 < x_1, ..., \xi_{k-1} < x_{k-1}\} \cap A),$$

де A – довільна подія. Справедливість індуктивного переходу виводиться послідовними підстановками у останню рівність $A = \{\xi_k \in [a_k, b_k)\} \cap A$ та врахування у правій частині тотожності для довільних подій A, B :

$$\mathbf{P}(B \cap \{\xi_k \in [a_k, b_k)\ \} \cap A) = \Delta^k_{[a_k,b_k)} \mathbf{P}(B \cap \{\xi_k < x_k\} \cap A) \ \boxdot$$

Теорема (про прирости сумісної функції розподілу). *Нехай* ξ – *випадковий вектор, а* $F = F_\xi$ *його сумісна функція розподілу. Тоді:*

(4) для довільного паралелепіпеда $[a,b) = \prod_{k=1}^n [a_k, b_k)$

$$\Delta_{[a,b)}F = \mathbf{P}(\xi \in [a,b)) \geq 0.$$

Доведення випливає з теореми про ймовірність значення всередині паралелепіпеда та з *невід'ємності* ймовірності $\boxdot$

Означення. *Будь-яка функція* $F = F(x_1, ..., x_n) : \mathbb{R}^n \to \mathbb{R}$, *яка задовольняє умови (1-3),(4), називається* сумісною функцією розподілу.

Вправи

(1) Нехай $G_1, ..., G_n$ – функції розподілу. Тоді $F(x_1, ..., x_n) = \prod_{k=1}^n G_k(x_k)$ є сумісною функцією розподілу.

(2) Функції (а) $F(x_1, x_2) = 1/2 + \pi^{-1} \arctan \max(x_1, x_2)$, (б) $F(x_1, x_2) = \min(x_1 + x_2, 1)1\!\!\mathrm{I}_{x_1 \ge 0,\ x_2 \ge 0}$ задовольняють умови (1)-(3), однак не є сумісними функціями розподілу.

(3) Нехай F, G – функції розподілу. Довести, що функції

(а) $F(x)G(y)(1 + \alpha(1 - F(x))(1 - G(y)))$ при $\alpha \in (-1, 1)$,

(б) $\alpha^{-1} \ln(1 + (\exp(\alpha F(x)) - 1)(\exp(\alpha G(y)) - 1)/(\exp(\alpha) - 1))$ при $\alpha > 0$

є сумісними функціями розподілу, причому координати відповідного випадкового вектора мають функції розподілу F та G. Узагальнити це твердження для n функцій розподілу.

(4) Для функцій розподілу F, G довести, що: (а) $V(x, y) \equiv \min(F(x), G(y))$ є сумісною функцією розподілу, (б) функції розподілу координат відповідного випадкового вектора збігаються з F та G. (в) Якщо сумісна функція розподілу U задовольняє умову (б), то $U(x, y) \le V(x, y)$ для всіх x, y.

1.14.3. Міра Лебега – Стілтьєса в евклідовому просторі

Означення. *Нехай F – довільна сумісна функція розподілу. Назвемо адитивною мірою Лебега – Стілтьєса в $\mathbb{R}^n$, що породжена F (або, коротше, адитивною мірою F), таку функцію F на алгебрі об'єднань попарно несумісних паралелепіпедів*

$$\mathfrak{A}(\mathbb{R}^n) = \left\{ A = \bigcup\nolimits_{k=1}^m [a(k), b(k)), \quad [a(i), b(i)) \cap [a(j), b(j)) = \varnothing, i \ne j \right\},$$

значення якої задаються сумою приростів на паралелепіпедах:

$$F(A) = \sum\nolimits_{k=1}^m \Delta_{[a(k),\, b(k))} F.$$

Теорема (про адитивну міру в евклідовому просторі). *Адитивна міра F є невід'ємною, нормованою та скінченно-адитивною функцією на $\mathfrak{A}(\mathbb{R}^n)$. Зокрема, F напівадитивна.*

Доведення аналогічне доведенню теореми *про адитивну міру Лебега – Стілтьєса* для одновимірного випадку $\boxdot$

Теорема (про неперервність у нулі міри в евклідовому просторі). *Адитивна міра F є неперервною в нулі на $\mathfrak{A}(\mathbb{R}^n)$, і як наслідок, σ-адитивна на алгебрі $\mathfrak{A}(\mathbb{R}^n)$.*

Доведення не відрізняється від доведення теореми *про неперервність у нулі міри Лебега – Стілтьєса* в одновимірному випадку $\boxdot$

Теорема (про міру Лебега – Стілтьєса в евклідовому просторі). *Для сумісної функції розподілу F у $\mathbb{R}^n$ на борелєвій σ-алгебрі $\mathfrak{B}(\mathbb{R}^n)$ існує єдина невід'ємна, нормована і σ-адитивна міра F така, що*

$$F((-\infty, x)) = F(x), \ \forall x \in \mathbb{R}^n.$$

Доведення. Достатньо визначити адитивну міру Лебега – Стілтьєса на алгебрі, як це зроблено вище, та на підставі її σ-адитивності застосувати *теорему Каратеодорі про продовження міри* $\boxdot$

Означення. *Побудована за теоремою про міру Лебега – Стілтьєса в евклідовому просторі міра F називається* мірою Лебега – Стілтьєса, що породжена сумісною функцією розподілу *F. Цю міру будемо також називати* розподілом *випадкового вектора ξ.*

Теорема (про обчислення ймовірностей, пов'язаних із випадковим вектором). *Нехай випадковий вектор ξ має сумісну функцію розподілу F_ξ. Тоді для кожної множини $B \in \mathfrak{B}(\mathbb{R}^n)$ виконується рівність*

$$\mathbf{P}(\xi \in B) = F_\xi(B),$$

де $F_\xi(B)$ – міра Лебега – Стілтьєса, що породжена сумісною функцією розподілу $F_\xi(x)$.

Доведення. За означенням шукана рівність має місце для кутів $B = (-\infty, x)$. З іншого боку, вирази в обох частинах як функції множин B є σ-адитивними мірами. Тому за *теоремою Каратеодорі про продовження міри* вони збігаються на всій *породженій сигма-алгебрі* $\mathfrak{B}(\mathbb{R}^n)$ $\boxdot$

Приклад. Нехай випадковий вектор $\xi = (\xi_1, ..., \xi_n)$ має сумісну функцію розподілу F_ξ. Оскільки $\{\xi_k < x_k\} = \{\xi \in \mathbb{R}^{k-1} \times (-\infty, x_k) \times \mathbb{R}^{n-k}\}$, то маргінальна функція розподілу величини ξ_k дорівнює

$$F_{\xi_k}(x_k) = \mathbf{P}(\xi_k < x_k) = F_\xi(\infty, ...\infty, x_k, \infty, ..., \infty).$$

1.14.4. Сумісна щільність

Означення. *Випадковий вектор $\xi = (\xi_1, ..., \xi_n)$ та його сумісна функція розподілу F_ξ називаються* абсолютно неперервними, *якщо існує* *невід'ємна вимірна функція $f_\xi(x)$ така, що*

$$F_\xi(x) = \int_{(-\infty,x)} f_\xi(y)dy, \ \forall x \in \mathbb{R}^n.$$

Функція $f_\xi(x)$ називається сумісною щільністю *випадкового вектора ξ та сумісної функції розподілу F_ξ.*

З *нормованості* та монотонності функції розподілу випливають такі властивості щільності розподілу:

(1 – невід'ємність) $f_\xi(x) \geq 0$ і

(2 – нормованість) $\int_{\mathbb{R}^n} f_\xi(y)dy = 1.$

Будь-яка невід'ємна вимірна нормована функція є *сумісною щільністю* певної сумісної функції розподілу, оскільки з властивостей інтеграла (Рімана або Лебега) випливатимуть характеристичні властивості (1)-(4) сумісної функції розподілу, а отже, може бути побудована відповідна *міра Лебега – Стілтьєса, що породжена сумісною функцією розподілу.*

Якщо сумісна функція розподілу диференційовна, то з формули Ньютона – Лейбніца (для інтеграла Рімана) або ж із *теореми Радона – Нікодима* (для інтеграла Лебега) випливає, що сумісна щільність (майже всюди) збігається з похідною функції розподілу:

$$f_\xi(x_1, ..., x_n) = \frac{\partial}{\partial x_1} ... \frac{\partial}{\partial x_n} F_\xi(x_1, ..., x_n).$$

Теорема (про обчислення ймовірностей, пов'язаних із неперервним вектором). *Якщо випадковий вектор ξ має сумісну щільність f_ξ, то для всіх борелєвих $B \in \mathfrak{B}(\mathbb{R}^n)$:*

$$\mathbf{P}(\xi \in B) = \int_B f_\xi(x)dx.$$

Доведення. Шукана рівність виконується для кутів $B = (-\infty, x)$ за означенням. З іншого боку, обидві частини рівності як функції множини B є *сигма-адитивними* мірами. Тому за *теоремою Каратеодорі про продовження міри* вони збігаються на породженій σ-алгебрі $\mathfrak{B}(\mathbb{R}^n)$ $\boxdot$

Наслідок (про єдиність сумісної щільності). *Сумісна щільність визначається сумісною функцією розподілу однозначно з точністю до рівності майже всюди за мірою Лебега.*

Доведення. Якщо функції $f_i(x)$ одночасно є щільностями, з означення та попередньої теореми виводимо, що

$$\int_B (f_1(x) - f_2(x))dx = 0, \ \forall B \in \mathfrak{B}(\mathbb{R}^n).$$

Оберемо $B^{\pm} = \{x : (f_1 - f_2)^{\pm}(x) > 0\}$, та підставимо в останню рівність. За властивістю інтеграла Лебега отримаємо $L(B^{\pm}) = 0$ $\boxdot$

Вправи

(1) Нехай $\mathbf{P}$ – довільна ймовірність на $\mathfrak{B}(\mathbb{R}^n)$. Довести, що для кожних $\varepsilon > 0$ та $B \in \mathfrak{B}(\mathbb{R}^n)$ знайдуться компактна множина $C \subset B$ та обмежена відкрита множина $A \supset B$ такі, що $\mathbf{P}(A \backslash C) \le \varepsilon$.

(2) Нехай сумісна функція розподілу F має сумісну щільність f. Довести, що приріст на паралелепіпеді $[a, b)$ дорівнює інтегралу.

$\Delta_{[a,b)}F = \int_{a_1}^{b_1} ... \int_{a_n}^{b_n} f(x_1, ..., x_n)dx_1...dx_n$.

(3) Якщо випадковий вектор $\xi = (\xi_1, ..., \xi_n)$ має сумісну щільність f, то його координати мають маргінальні щільності

$f_{\xi_k}(x) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, ...x_{k-1}, x, x_{k+1}, .., x_n)dx_1...dx_{k-1}dx_{k+1}...dx_n$.

1.14.5. Функції від випадкового вектора

Теорема (про обчислення ймовірностей і математичного сподівання функції від випадкового вектора). *Якщо випадковий вектор ξ має сумісну функцію розподілу $F_\xi(x)$, а $g : \mathbb{R}^n \to \mathbb{R}^m$ борелєва функція, то*

$$\mathbf{P}(g(\xi) \in B) = \int_{g^{(-1)}(B)} dF_\xi(x), \ \forall B \in \mathfrak{B}(\mathbb{R}^m),$$

$$\mathbf{E}g(\xi) = \int_{\mathbb{R}^m} g(x) \, dF_\xi(x).$$

За умови існування сумісної щільності ці інтеграли Лебега – Стільтьєса можна замінити на відповідні інтеграли за мірою Лебега з сумісною щільністю.

Доведення повністю аналогічне доведенню теореми *про обчислення ймовірностей, пов'язаних із випадковою величиною* та *про обчислення математичного сподівання функції від випадкової величини* $\boxdot$

Приклад. Точки ξ, η випадково обрано на одиничному відрізку. Яка ймовірність події $A = \{\xi \cdot \eta < 1/2\}$?

З умови випадковості робимо висновок, що вектор (ξ, η) *рівномірно розподілений* на квадраті $[0, 1]^2$, тобто має *сумісну щільність* $f(x, y) = 1\!\!1_{(x,y)\in[0,1]^2}$. Позначимо $B = \{(x, y) \in [0, 1]^2 : x \cdot y < 1/2\}$. Тоді $\mathbf{P}(A) = \mathbf{P}((\xi, \eta) \in B) = \int_B f(x, y)dxdy = 1/2 + \int_{1/2}^{1} dx/2x = (1 + \ln 2)/2$.

1.15. Числові характеристики випадкових векторів

Означення. Математичним сподіванням випадкового вектора *з інтегровними координатами* $\xi = (\xi_1, ..., \xi_n)$ *називається вектор, складений з математичних сподівань координат* ξ :

$$\mathbf{E}\xi = (\mathbf{E}\xi_1, ..., \mathbf{E}\xi_n).$$

Зауваження. Для математичного сподівання випадкового вектора виконуються всі основні властивості математичного сподівання – *позитивність*, *монотонність*, *лінійність* тощо.

1.15.1. Коваріація та кореляція випадкових величин

Означення. *Нехай* ξ, η *– квадратично інтегровні величини. Їх* коваріацією *називається число*

$$\mathrm{Cov}(\xi, \eta) \equiv \mathbf{E}(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta) = \mathbf{E}\xi\eta - \mathbf{E}\xi\mathbf{E}\eta,$$

що є скінченною за нерівністю Коші:

$$|\mathrm{Cov}(\xi, \eta)| \le \sqrt{\mathbf{E}(\xi - \mathbf{E}\xi)^2\mathbf{E}(\eta - \mathbf{E}\eta)^2} = \sqrt{\mathbf{D}\xi\ \mathbf{D}\eta} = \sigma_\xi\sigma_\eta.$$

Коефіцієнтом кореляції *випадкових величин* ξ, η *називається безрозмірна величина*

$$\rho(\xi, \eta) = \frac{\mathrm{Cov}(\xi, \eta)}{\sigma_\xi\sigma_\eta},$$

яка внаслідок нерівності Коші набуває значень з інтервалу $[-1, 1]$.

Величини ξ, η *називаються* некорельованими*, якщо* $\mathrm{Cov}(\xi, \eta) = 0$.

Якщо простір *квадратично інтегровних* центрованих величин

$$L_2^0(\mathbf{P}) \equiv \{\xi :\ \mathbf{E}\xi^2 < \infty,\ \mathbf{E}\xi = 0\}$$

розглядати як гільбертів простір зі скалярним добутком $(\xi, \eta) = \mathbf{E}\xi\eta$, то $\|\xi\|^2 = \mathbf{D}\xi$ і $\rho(\xi, \eta) = \cos(\xi\wedge\eta)$. *Некорельованість* величин інтерпретується як їх ортогональність у даному просторі.

Вправи

(1) Довести теорему про обчислення ймовірностей та математичного сподівання від випадкового вектора, через його сумісну щільність.

(2) Довести, що коваріація $\mathrm{Cov}(\xi, \eta)$ є білінійною функцією від ξ, η.

(3) Довести, що клас всіх випадкових величин, що є некорельованими з заданою системою випадкових величин, є лінійним простором.

(4) Довести тотожність $\mathbf{D}(\xi+\eta)=\mathbf{D}(\xi)+2\,\mathrm{Cov}(\xi,\eta)+\mathbf{D}(\eta)$.

(5) Випадковий вектор $\xi=(\xi_1,...,\xi_n)$ має сумісну щільність f_ξ, а функція $g:U\to V$ неперервно диференційовна та взаємно однозначно відображає відкриту множину $U\subset\mathbb{R}^n$ на відкриту множину $V\subset\mathbb{R}^n$, причому для всіх $x\in U$ якобіан $J_g(x)\equiv\det(\partial g/\partial x)(x)>0$. Довести, що випадковий вектор $\eta\equiv g(\xi)$ має сумісну щільність $f_\eta(y)=f_\xi(g^{-1}(y))(J_g(g^{-1}(y)))^{-1}\mathbb{I}_{y\in V}$. Розглянути приклади: (а) лінійного перетворення, (б) переходу до полярних координат.

(6) Нехай $\mathbf{E}\xi_k=0$, $\mathbf{D}\xi_k=1$, $k=1,2$, та $\rho(\xi_1,\xi_2)=\rho$. Довести нерівність $\mathbf{E}\max(\xi_1^2,\xi_2^2)\le 1+\sqrt{1+\rho^2}$.

(7) Вивести з нерівності Коші, що рівність $\rho(\xi,\eta)=\pm1$ має місце тоді й тільки тоді, коли $\eta=a\pm b\xi$ м.н. для деяких сталих $a,b>0$.

(8) Випадкові величини $\xi_k, k=\overline{1,3}$, обмежені: $|\xi_k|\le 1$. Довести нерівність Белла: $|\mathbf{E}\xi_1\xi_2-\mathbf{E}\xi_1\xi_3|\le 1-\mathbf{E}\xi_2\xi_3$.

(9) Випадкові величини $\xi_k, k=1,2$, центровані і нормовані: $\mathbf{E}\xi_k=0$, і $\mathbf{D}\xi_k=1$, та мають кореляцію $\rho=\mathbf{E}\xi_1\xi_2$. Довести такий аналог нерівності Чебишева: $\mathbf{P}(\{|\xi_1|\ge\varepsilon\}\cup\{|\xi_2|\ge\varepsilon\})\le\left(1+\sqrt{1-\rho^2}\right)/\varepsilon^2$.

1.15.2. Коваріаційна матриця випадкового вектора

Означення. *Нехай $\xi=(\xi_1,...,\xi_n)$ – випадковий вектор з квадратично інтегровними координатами. Його* коваріаційною матрицею *називається матриця з коваріаціями його координат:*

$$\mathrm{Cov}(\xi)\equiv(\mathrm{Cov}(\xi_i,\xi_j),\ i,j=\overline{1,n}).$$

Теорема (про властивості коваріаційної матриці). *Коваріаційна матриця V =Cov(ξ) випадкового вектора ξ симетрична та **невід'ємно визначена**, тобто:*

$$V=V',\quad c'Vc\ge 0,\quad \forall c\in\mathbb{R}^n.$$

Доведення. Симетричність очевидна. *Невід'ємна визначеність* випливає з теореми *про властивості дисперсії* та наступної теореми ⊡

Теорема (про дисперсію лінійної форми від випадкового вектора). *Нехай $\xi=(\xi_1,...,\xi_n)$ – випадковий вектор із коваріаційною матрицею Cov(ξ), стала $c\in\mathbb{R}^n$, а $\eta=c'\xi\equiv\sum_{k=1}^n c_k\xi_k$ – лінійна форма від ξ. Тоді виконується рівність*

$$\mathbf{D}\eta=c'\mathrm{Cov}(\xi)c=\sum\nolimits_{i,\ j}c_i\mathrm{Cov}(\xi_i,\xi_j)c_j.$$

Доведення. Піднесемо до квадрату та скористаємося *лінійністю* математичного сподівання:

$$\mathbf{D}\eta = \mathbf{E}(\eta - \mathbf{E}\eta)^2 = \mathbf{E}(\sum c_k(\xi_k - \mathbf{E}\xi_k))^2 =$$
$$\mathbf{E}\sum_{i,j} c_i(\xi_i - \mathbf{E}\xi_i)c_j(\xi_j - \mathbf{E}\xi_j) = \sum_{i,j} c_i \mathrm{Cov}(\xi_i, \xi_j) c_j \ \boxdot$$

Теорема (про коваріаційну матрицю лінійного перетворення). *Нехай $\xi = (\xi_1, ..., \xi_n)$ – випадковий вектор із коваріаційною матрицею $Cov(\xi)$, матриця $C \in \pounds(\mathbb{R}^n, \mathbb{R}^m)$, а випадковий вектор*

$$\eta = C\xi = \left(\sum\nolimits_{k=1}^{n} C_{ik}\xi_k, \ i = \overline{1,m}\right)$$

утворений відповідним лінійним перетворенням ξ. Тоді

$$\mathrm{Cov}(\eta) = C \ \mathrm{Cov}(\xi) \ C'.$$

Доведення. Обчислимо за означенням коваріації та за *лінійністю* математичного сподівання

$$\mathrm{Cov}(\eta)_{i,j} = \mathbf{E}(\eta_i - \mathbf{E}\eta_i)(\eta_j - \mathbf{E}\eta_j) = \mathbf{E}\sum\nolimits_k C_{ik}(\xi_k - \mathbf{E}\xi_k)\sum\nolimits_l C_{jl}(\xi_l - \mathbf{E}\xi_l) =$$
$$\sum\nolimits_{k,l} C_{ik}\mathrm{Cov}(\xi_k, \xi_l)(C')_{lj} = (C \ \mathrm{Cov}(\xi) \ C')_{ij} \ \boxdot$$

Вправи

(1) Коваріаційна матриця $\mathrm{Cov}(\xi)$ вектора ξ має власні числа λ_k. Довести, що $\max_{x\in\mathbb{R}^n} \mathbf{D}(x'\xi)/x'x = \max_{1\le k\le n} \lambda_k$.

(2) Взаємною коваріаційною матрицею квадратично інтегровних випадкових векторів ξ, η називається матриця: $\mathrm{Cov}(\xi, \eta) \equiv \mathbf{E}(\xi - \mathbf{E}\xi)\cdot(\eta - \mathbf{E}\eta)'$, де добуток утворений попарними добутками відповідних координат. Довести, що коваріаційна матриця складеного вектора (ξ, η) дорівнює $\begin{pmatrix} \mathrm{Cov}(\xi) & \mathrm{Cov}(\xi, \eta) \\ \mathrm{Cov}(\xi, \eta)' & \mathrm{Cov}(\eta) \end{pmatrix}$.

1.16. Незалежні випадкові величини

Кожна випадкова величина породжує певні випадкові події, що є прообразами *борелєвих множин*.. Незалежність випадкових величин означає, що всі такі породжені події *незалежні*.

Означення. *Випадкові величини $(\xi_1, ..., \xi_n)$* незалежні в сукупності, *якщо всі породжені ними випадкові події незалежні в сукупності, тобто для довільних $B_k \in \mathfrak{B}(\mathbb{R})$*

$$\mathbf{P}(\xi_1 \in B_1, ..., \xi_n \in B_n) = \prod\nolimits_{k=1}^{n} \mathbf{P}(\xi_k \in B_k).$$

Зауваження (про незалежність підмножин). *Якщо випадкові величини з деякої множини незалежні у сукупності, то за означенням випадкові величини з будь-якої її підмножини також незалежні.*

Означення. *Випадкові величини* $(\xi_1, ..., \xi_n)$ попарно незалежні, *якщо для всіх* $k < j$ *величини* ξ_k *та* ξ_j *незалежні.*

Ясно, що з незалежності у сукупності випливає незалежність попарна. Наведений вище приклад Бернштейна свідчить, що обернене твердження не виконується.

1.16.1. Критерій незалежності випадкових величин

Теорема (про критерій незалежності випадкових величин). *Випадкові величини* $(\xi_1, ..., \xi_n)$ *незалежні в сукупності тоді й тільки тоді, коли їх сумісна функція розподілу розпадається для всіх* $(x_1, ..., x_n) \in \mathbb{R}^n$ *у добуток*

$$F_{(\xi_1,...,\xi_n)}(x_1, ..., x_n) \equiv \mathbf{P}(\xi_1 < x_1, ..., \xi_n < x_n) =$$
$$= \prod_{k=1}^{n} \mathbf{P}(\xi_k < x_k) \equiv \prod_{k=1}^{n} F_{\xi_k}(x_k).$$

Доведення. *Необхідність* умови теореми очевидна, оскільки напівінтервали $B_k = (-\infty, x_k)$ є борелєвими множинами.

Достатність. Доведення проведемо за індукцією з таким індуктивним твердженням: для всіх $B_i \in \mathfrak{B}(\mathbb{R}), x_i \in \mathbb{R}$

$$\mathbf{P}(\{\xi_1 \in B_1, ..., \xi_{k-1} \in B_{k-1}\} \cap \{\xi_k \in B_k\} \cap \{\xi_{k+1} < x_{k+1}, ..., \xi_n < x_n\}) =$$
$$\left(\prod_{i=1}^{k-1} \mathbf{P}(\xi_i \in B_i)\right) \mathbf{P}(\xi_k \in B_k) \prod_{i=k+1}^{n} \mathbf{P}(\xi_i < x_i)$$

При $k = 0$ рівність виконується за умовою. Треба її довести при $k = n$.

Нехай рівність виконується для $k-1$. Тоді вона має місце при значенні k для множин $B_k = (-\infty, x_k)$. Зауважимо, що обидві частини такої рівності є сигма-адитивними мірами за $B_k \in \mathfrak{B}(\mathbb{R})$. Тому рівність має місце для напівінтервалів $B_k = [a_k, b_k)$, для множин з алгебри $B_k \in \mathfrak{A}(\mathbb{R})$, та за *теоремою Каратеодорі про продовження міри* для всіх $B_k \in \mathfrak{B}(\mathbb{R})$ $\boxdot$

Теорема (про критерій незалежності абсолютно неперервних величин). *Нехай випадковий вектор* $\xi = (\xi_1, ..., \xi_n)$ *має сумісну щільність* $f_\xi(x_1, .x_n)$. *Випадкові величини* $(\xi_1, ..., \xi_n)$ *незалежні в сукупності тоді й тільки тоді, коли*

$$f_\xi(x_1, ..., x_n) = \prod_{k=1}^{n} f_{\xi_k}(x_k), \ \forall x_k \in \mathbb{R}.$$

Доведення. За умови *абсолютної неперервності* умова теореми *про критерій незалежності випадкових величин* еквівалентна рівності

$$\int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} f_\xi(y_1, ..., y_n) dy_1 ... dy_n = \prod_{k=1}^{n} \int_{-\infty}^{x_k} f_{\xi_k}(y_k) dy_k =$$

$$\int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} \prod_{k=1}^{n} f_{\xi_k}(y_k) dy_1 ... dy_n, \quad \forall x_k \in \mathbb{R},$$

де остання рівність є наслідком *теореми Фубіні про кратний та повторні інтеграли*. Звідси за означенням сумісної щільності та з наслідку *про єдиність сумісної щільності* виводимо сформульоване твердження $\boxdot$

1.16.2. Перетворення незалежних величин

Теорема (про перетворення незалежних величин). *Нехай випадкові величини* $(\xi_1, ..., \xi_n)$ *незалежні в сукупності, а* $g_1,...,g_n$ *– борелєві функції. Тоді випадкові величини*

$$(g_1(\xi_1), ..., g_n(\xi_n))$$

також незалежні в сукупності.

Доведення. Обчислимо при $B_k \in \mathfrak{B}(\mathbb{R})$

$$\mathbf{P}(g_1(\xi_1) \in B_1, ..., g_n(\xi_n) \in B_n) = \mathbf{P}(\xi_1 \in g_1^{(-1)}(B_1), ..., \xi_n \in g_n^{(-1)}(B_n)) =$$

$$\prod_{k=1}^{n} \mathbf{P}(\xi_k \in g_k^{(-1)}(B_k)) = \prod_{k=1}^{n} \mathbf{P}(g_k(\xi_k) \in B_k) \ \boxdot$$

Теорема (про векторні перетворення незалежних величин). *Нехай випадкові величини* $(\xi_1, ..., \xi_n)$ *незалежні в сукупності,* $k < n$, *а функції* $f : \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^{n-k} \to \mathbb{R}$ *– борелєві. Тоді незалежними є величини:*

$$f(\xi_1, ..., \xi_k), \quad g(\xi_{k+1}, ..., \xi_n).$$

Доведення. За критерієм незалежності досить довести, що

$$\mathbf{P}((\xi_1, ...\xi_k) \in B, (\xi_{k+1}, ...\xi_n) \in C) = \mathbf{P}((\xi_1, ...\xi_k) \in B)\mathbf{P}((\xi_{k+1}, ...\xi_n) \in C)$$

для всіх $B \in \mathfrak{B}(\mathbb{R}^k)$, $C \in \mathfrak{B}(\mathbb{R}^{n-k})$. Дійсно, тут для довільних $x, y \in \mathbb{R}$ можна обрати $B = f^{(-1)}((-\infty, x))$ і $C = g^{(-1)}((-\infty, y))$, та скористатись теоремою *про критерій незалежності випадкових величин*.

Вказана рівність для прямокутників $B = \prod_{i=1}^{k} B_i$, $C = \prod_{j=k+1}^{n} C_j$ випливає з означення незалежності.

Обидві частини наведеної рівності є мірами на $\mathfrak{B}(\mathbb{R}^k)$ та $\mathfrak{B}(\mathbb{R}^{n-k})$ за аргументами B і C відповідно. Для множин B, C із алгебр $\mathfrak{A}(\mathbb{R}^k)$, $\mathfrak{A}(\mathbb{R}^{n-k})$, що є об'єднаннями попарно несумісних прямокутників, виводимо цю рівність з *адитивності* міри. Нарешті, для *борелєвих* B, C наведена рівність справедлива, оскільки відповідні класи множин B, C замкнені відносно *монотонної збіжності* за властивістю *неперервності ймовірностей* (мір) $\boxdot$

Вправи

(1) Дискретні випадкові величини $(\xi_1, ..., \xi_n)$ незалежні в сукупності тоді й тільки тоді, коли $\mathbf{P}(\xi_1 = x_1, ..., \xi_n = x_n) = \prod_{k=1}^{n} \mathbf{P}(\xi_k = x_k)$ для всіх x_k.

(2) Випадкові величини $(\xi_1, ..., \xi_n)$ незалежні в сукупності тоді й тільки тоді, коли $\mathbf{E}g_1(\xi_1)...g_n(\xi_n) = \prod_{k=1}^n \mathbf{E}g_k(\xi_k)$ для всіх $g_k \in C_b(\mathbb{R})$.

(3) Нехай F_k – функції розподілу зі щільностями f_k, і $|a| < 1$. Довести, що (а) функція $f(x,y) = f_1(x)f_2(y)(1 + a(2\,F_1(x) - 1)(2\,F_2(y) - 1))$ є сумісною щільністю розподілу деякого випадкового вектора (ξ_1, ξ_2), (б) величини ξ_k мають щільності f_k та є залежними при $a \neq 0$.

(4) Знайти функцію розподілу максимуму та мінімуму незалежних однаково розподілених випадкових величин через їх спільну функцію розподілу.

(5) Випадкові величини $\xi_1, \xi_2 \simeq Exp(1)$ незалежні та мають однаковий показниковий розподіл із параметром 1. Довести що величини (а) $\min(\xi_1, \xi_2)$ та $|\xi_1 - \xi_2|$ незалежні однаково розподілені, (б) $\xi_1/(\xi_1 + \xi_2)$ та $\xi_1 + \xi_2$ незалежні. (в) Знайти відповідні сумісні розподіли. (г) Узагальнити ці твердження на випадок n незалежних однаково розподілених показникових величин.

(6) Випадкові величини ξ_k незалежні та мають показникові розподіли з параметрами $\lambda_k, k = 1, 2$. Знайти функцію розподілу величин (а) $\xi_1/(\xi_1 + \xi_2)$, (б) $\xi_1 + \xi_2$, (в) ξ_1/ξ_2, (г) $(\xi_1 + \xi_2)/\xi_1$.

(7) Випадкові величини ξ_1, ξ_2 незалежні та мають стандартний нормальний розподіл. Довести, що відношення ξ_1/ξ_2 має розподіл Коші.

(8, М.Й.Ядренко) Випадкові величини ξ, η незалежні та мають такі щільності: $f_\xi(x) = 1\!\!1_{|x|<1}/\pi\sqrt{1 - x^2}$, $f_\eta(x) = x \exp(-x^2/2) 1\!\!1_{x>0}$. Довести, що добуток $\xi\eta$ має нормальний розподіл.

(9) Узагальнити теорему про векторні перетворення незалежних величин для скінченого числа функцій.

1.17. Математичне сподівання добутку незалежних величин

1.17.1. Математичне сподівання добутку

Теорема (про математичне сподівання добутку незалежних величин). *Якщо випадкові величини ξ, η незалежні та інтегровні, то добуток $\xi\eta$ є інтегровним і має місце рівність*

$$\mathbf{E}\xi\eta = \mathbf{E}\xi\ \mathbf{E}\eta.$$

Доведення

(а) Нехай $\xi = 1\!\!1_A$, $\eta = 1\!\!1_B$. Тоді події A, B незалежні і

$$\mathbf{E}\xi\eta = \mathbf{E}\ 1\!\!1_A 1\!\!1_B = \mathbf{E}1\!\!1_{A\cap B} = \mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B) = \mathbf{E}\xi\ \mathbf{E}\eta.$$

(б) Нехай величини ξ, η незалежні і *прості*, $\xi = \sum x_i 1\!\!1_{A_i}$, $\eta = \sum y_j 1\!\!1_{B_j}$. Тоді випадкова величина $\xi\eta$ дорівнює скінченній сумі

$$\xi\eta = \sum x_i 1\!\!1_{A_i} \sum y_j 1\!\!1_{B_j} = \sum\nolimits_{i,j} x_i y_j 1\!\!1_{A_i} 1\!\!1_{B_j} = \sum\nolimits_{i,j} x_i y_j \ 1\!\!1_{A_i \cap B_j}.$$

Звідси за *лінійністю* математичного сподівання

$$\mathbf{E}(\xi\eta) = \sum\nolimits_{i,j} x_i y_j \ \mathbf{E} \ 1\!\!1_{A_i \cap B_j} = \sum\nolimits_{i,j} x_i y_j \mathbf{P}(A_i \cap B_j) =$$

$$\sum\nolimits_{i,j} x_i y_j \mathbf{P}(A_i) \mathbf{P}(B_j) = \sum\nolimits_i x_i \mathbf{P}(A_i) \sum\nolimits_j y_j \mathbf{P}(B_j) = \mathbf{E}\xi \ \mathbf{E}\eta.$$

(в) Якщо $\xi, \eta \geq 0$, за теоремою *про апроксимацію випадкових величин простими* побудуємо прості апроксимуючі величини $0 \leq \xi_n \uparrow \xi$, $0 \leq \eta_n \uparrow \eta$, де за визначенням: $\xi_n = \varphi_n(\xi)$, $\eta_n = \varphi_n(\eta)$ з борелєвою функцією φ_n. Тоді $0 \leq \xi_n \eta_n \uparrow \xi\eta$, причому величини ξ_n, η_n незалежні за теоремою *про перетворення незалежних величин.*

З твердження (б) та *теореми Лебега про монотонну збіжність* із рівняння $E\xi_n\eta_n = E\xi_n E\eta_n$ граничним переходом дістанемо рівність теореми. Зауважимо, що умова інтегровності в цій частині не використовується.

(г) Для знакозмінних інтегровних незалежних ξ, η за теоремою *про перетворення незалежних величин* величини $\xi^{\pm}, \eta^{\pm}$ незалежні. Тому за *лінійністю* та внаслідок інтегровності добутків $\xi^{\pm}\eta^{\pm}$ за пунктом (в):

$$\mathbf{E}\xi\eta = \mathbf{E}(\xi^+\eta^+ - \xi^-\eta^+ - \xi^+\eta^- + \xi^-\eta^-) =$$

$$\mathbf{E}\xi^+\mathbf{E}\eta^+ - \mathbf{E}\xi^-\mathbf{E}\eta^+ - \mathbf{E}\xi^+\mathbf{E}\eta^- + \mathbf{E}\xi^-\mathbf{E}\eta^- = \mathbf{E}\xi\mathbf{E}\eta \ \boxdot$$

Наслідок (про некорельованість незалежних величин). *Із попарної незалежності випадкових величин випливає їх некорельованість.*

Доведення є наслідком теореми *про математичне сподівання добутку незалежних величин,* оскільки за означенням *коваріації* справедливі рівності $\mathrm{Cov}(\xi, \eta) = \mathbf{E}\xi\eta - \mathbf{E}\xi \ \mathbf{E}\eta = \mathbf{E}\xi \ \mathbf{E}\eta - \mathbf{E}\xi \ \mathbf{E}\eta = 0$ $\boxdot$

1.17.2. Дисперсія суми незалежних величин

Теорема (про дисперсію суми незалежних величин). *Нехай випадкові величини* $(\xi_1, ..., \xi_n)$ *попарно незалежні. Тоді*

$$\mathbf{D}\left(\sum\nolimits_{k=1}^n \xi_k\right) = \sum\nolimits_{k=1}^n \mathbf{D}(\xi_k).$$

Доведення. Обчислимо за означенням дисперсії та за *лінійністю* математичного сподівання

$$\mathbf{D}\left(\sum\nolimits_{k=1}^n \xi_k\right) = \mathbf{E}\left(\sum\nolimits_{k=1}^n (\xi_k - \mathbf{E}\xi_k)\right)^2 = \sum\nolimits_{i,j} \mathbf{E}(\xi_i - \mathbf{E}\xi_i)(\xi_j - \mathbf{E}\xi_j) =$$

$$\sum\nolimits_{i,j}(\mathbf{E}(\xi_i\xi_j) - \mathbf{E}\xi_i\mathbf{E}\xi_j) = \sum\nolimits_{i\neq j} 0 + \sum\nolimits_{i=j}(\mathbf{E}(\xi_i\xi_j) - \mathbf{E}\xi_i\mathbf{E}\xi_j) =$$
$$\sum\nolimits_{i}(\mathbf{E}\xi_i^2 - (\mathbf{E}\xi_i)^2) = \sum\nolimits_{k=1}^{n} \mathbf{D}\xi_k,$$

де при $i \neq j$ використано теорему *про математичне сподівання добутку незалежних величин*, та теорему *про властивості дисперсії* $\boxdot$

Приклад. Дисперсія *біноміального розподілу*. Нехай $\chi_k = 1\!\!\mathrm{I}_{Y_k}$ – *індикаторна величина* успіху в k-му *випробуванні Бернуллі*. За умовою ці величини незалежні. Тоді загальна кількість успіхів дорівнює сумі $\xi = \sum_{k=1}^{n} \chi_k$, отже, з теореми *про дисперсію суми незалежних величин* та формули для дисперсії індикаторної величини випливає рівність $\mathbf{D}\xi = \sum_{k=1}^{n} \mathbf{D}\chi_k = np(1-p)$.

1.17.3. Математичне сподівання функції від незалежних величин

На відміну від добутку, для загальної функції від незалежних величин обчислення математичного сподівання зводиться до знаходження повторного інтегралу.

Теорема (про математичне сподівання функції від незалежних величин). *Нехай випадкові величини ξ, η незалежні, мають функції розподілу F_ξ, F_η, а $g : \mathbb{R}^2 \to \mathbb{R}$ невід'ємна борелева функція. Тоді*

$$\mathbf{E}g(\xi,\eta) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g(x,y) dF_\xi(x) \right) dF_\eta(y).$$

Доведення

(1) Нехай $g(x,y) = 1\!\!\mathrm{I}_{(x,y)\in B\times C}$ для $B, C \in \mathfrak{B}(\mathbb{R})$. Тоді обидві частини рівності дорівнюють $F_\xi(B)F_\eta(C)$, оскільки $g(x,y) = 1\!\!\mathrm{I}_{x\in B}1\!\!\mathrm{I}_{y\in C}$, а за теоремою *про перетворення незалежних величин* множники $1\!\!\mathrm{I}_{\xi\in B}1\!\!\mathrm{I}_{\eta\in C}$ – незалежні.

(2) Унаслідок (1) та лінійності за g тотожності теореми її задовольняють функції $g(x,y) = 1\!\!\mathrm{I}_{(x,y)\in D}$ для довільних множин D з алгебри $\mathfrak{A}(\mathbb{R}^2)$. Оскільки обидві частини тотожності неперервні за D, то вона виконується для всіх простих невід'ємних борелевих функцій $g : \mathbb{R}^2 \to \mathbb{R}$.

(3) Для довільної невід'ємної борелевої функції g оберемо прості g_n так, щоб $0 \leq g_n \uparrow g$ поточково. Тоді для кожного $y \in \mathbb{R}$ мають місце монотонні поточкові збіжності

$$\int_{-\infty}^{\infty} g_n(x,y) dF_\xi(x) \uparrow \int_{-\infty}^{\infty} g(x,y) dF_\xi(x),\ n \to \infty,$$
$$g_n(\xi,\eta) \uparrow g(\xi,\eta),\ n \to \infty.$$

Тому твердження теореми випливає з *теореми Лебега про монотонну збіжність* для інтеграла Лебега та для інтеграла Лебега – Стілтьєса, якщо доведене твердження (2) застосувати до функцій g_n ⊡

Вправи

(1) Для незалежних ξ, η довести, що:
$\mathbf{P}(\xi < \eta) = \int_{-\infty}^{\infty} F_\xi(x) dF_\eta(x)$.

(2) Довести для функцій розподілу F_1, F_2 тотожність
$F_1(a)F_2(a) = \int_{(-\infty,a)} F_1(x)dF_2(x) + \int_{(-\infty,a)} F_2(x+0)dF_1(x)$.

(3) Довести, що $\mathbf{E}\xi\eta = \mathbf{E}\xi\mathbf{E}\eta$ для незалежних невід'ємних величин.

(4) Випадкові величини ξ, η незалежні, а борелєва функція g невід'ємна. Довести тотожність $\mathbf{E}g(\xi,\eta) = \mathbf{E}(\mathbf{E}(g(x,\eta) \mid_{x=\xi})$.

(5) Для незалежних випадкових величин ξ, η з відомим математичними сподіваннями та дисперсіями обчислити дисперсію добутку $\mathbf{D}(\xi\eta)$.

(6) Випадкова величина ξ рівномірно розподілена на $[-1, 1]$. Тоді величини ξ та ξ^2 некорельовані, але залежні.

(7) Обмежені випадкові величини ξ, η є незалежними тоді й тільки тоді, коли $\mathbf{E}\xi^m\eta^n = \mathbf{E}\xi^m\mathbf{E}\eta^n$ для всіх $m, n \geq 1$.

(8) Величини $\xi_1, ..., \xi_n$ незалежні, однаково розподілені, та $\mathbf{E}\xi_1 = \mathbf{E}\xi_1^3 = 0$. Довести, що випадкові величини $\mu = \frac{1}{n}\sum_{k=1}^n \xi_k$ та $\sigma^2 = \frac{1}{n}\sum_{k=1}^n (\xi_k - \mu)^2$ – некорельовані.

(9) Випадкові величини ξ_1 та ξ_2 є незалежними тоді й тільки тоді, коли $\text{Cov}(g_1(\xi_1), g_2(\xi_2)) = 0$ для всіх борелєвих функцій g_k таких, що величини $g_k(\xi_k)$ квадратично інтегровні.

(10) Випадкова ламана на площині виходить з початку координат, утворена n відрізками одиничної довжини, причому кути між сусідніми відрізками незалежні та дорівнюють $\pm\alpha$ з імовірністю $1/2$. Обчислити математичне сподівання квадрата відстані міх початковою та останньою вершиною ламаної.

(11) Розглянемо узагальнення випробувань Бернуллі, у якому ймовірність успіху в k-му випробуванні дорівнює p_k. Нехай ξ_n – кількість успіхів у n випробуваннях. (а) Знайти $\mathbf{E}\xi_n, \mathbf{D}\xi_n$. (б) Найбільше значення $\mathbf{D}\xi_n$ за умови $n^{-1}\sum_{k=1}^n p_k = p$ досягається при $p_k \equiv p$.

(12) Випадкові вектори $(\xi_1, ...\xi_n)$ з $\mathbb{R}^d$ називаються незалежними в сукупності, якщо відповідні прообрази довільних борелєвих множин є незалежними у сукупності випадковими подіями. Довести: (а) теорему про критерій незалежності випадкових векторів, (б) критерій незалежності абсолютно неперервних векторів, (в) теореми про перетворення незалежних векторів та (г) про математичне сподівання функції від незалежних випадкових векторів.

1.18. Розподіл суми незалежних величин. Гама-розподіл

1.18.1. Розподіл суми незалежних величин

Означення. Згорткою функцій розподілу F, G *називається функція*

$$F * G(a) = \int_{-\infty}^{\infty} F(a-y)\, dG(y).$$

Теорема (про функцію розподілу суми незалежних величин). *Нехай випадкові величини ξ, η незалежні та мають функції розподілу F_ξ, F_η. Тоді функція розподілу $F_{\xi+\eta}$ суми $\xi+\eta$ дорівнює згортці:*

$$F_{\xi\ +\ \eta} = F_\xi\ *\ F_\eta.$$

Доведення. Для довільної множини $B \in \mathfrak{B}(\mathbb{R}^2)$ позначимо через $B_y \equiv \{x : (x,y) \in B\}$ переріз множини B на рівні y.

Нехай $B^a = \{(x,y) : x+y < a\}$. Тоді $\mathbf{P}(\xi+\eta < a) = \mathbf{P}((\xi,\eta) \in B^a)$ і переріз $B_y^a = \{x : x < a-y\}$. Застосуємо теорему *про математичне сподівання функції від незалежних величин* для функції $g(x,y) = 1\!\!\mathrm{I}_{(x,y)\in B^a}$:

$$\mathbf{P}(\xi+\eta < a) = \mathbf{P}((\xi,\eta) \in B^a) = \mathbf{E}g(\xi,\eta) =$$

$$\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} 1\!\!\mathrm{I}_{(x,y)\in B^a}\ dF_\xi(x)\right) dF_\eta(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1\!\!\mathrm{I}_{x\in B_y^a}\ dF_\xi(x) dF_\eta(y) =$$

$$\int_{-\infty}^{\infty} F_\xi(B_y^a)\, dF_\eta(y) = \int_{-\infty}^{\infty} F_\xi(a-y)\, dF_\eta(y) = F_\xi * F_\eta(a)\ \boxdot$$

Наслідок (про властивості згортки).

(а) *Згортка функцій розподілу є функцією розподілу.*

(б) *Згортка є комутативною та асоціативною операцією.*

Доведення очевидне, оскільки ліва частина останньої рівності є функцію розподілу та симетрична відносно ξ, η $\boxdot$

Означення. Згорткою щільностей розподілу f, g *називається інтегральне перетворення*

$$f * g(a) = \int_{-\infty}^{\infty} f(a-y)\ g(y) dy.$$

Теорема (про щільність розподілу суми незалежних величин). *Нехай величини ξ, η незалежні та мають щільності розподілу f_ξ, f_η. Тоді щільність розподілу $f_{\xi+\eta}$ суми $\xi+\eta$ існує і дорівнює згортці $f_\xi * f_\eta$.*

Доведення. Заміною порядку інтегрування за *теоремою Фубіні про кратний та повторні інтеграли* досить перевірити, що при всіх $a \in \mathbb{R}$:

$$F_\xi * F_\eta(a) = \int_{-\infty}^{\infty} F_\xi(a-y)\, dF_\eta(y) = \int_{-\infty}^{a} F_\xi(a-y)\, f_\eta(y) dy =$$

$$\int_{-\infty}^{\infty}\left(\int_{-\infty}^{a-y} f_\xi(x)f_\eta(y)dx\right)dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\xi(x)f_\eta(y)1\!\!1_{x+y\,<\,a}\,dx\,dy =$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\xi(u-y)f_\eta(y)1\!\!1_{u\,<\,a}\,du\,dy = \int_{-\infty}^{a} f_\xi * f_\eta(u)\,du,$$

де використано також теорему *про обчислення математичного сподівання функції від неперервної величини* через її щільність та заміну змінних $u = x + y,\ y = y$ у кратному інтегралі ⊡

Вправи

(1) Для незалежних дискретних випадкових величин ξ, η

$\mathbf{P}(\xi + \eta = a) = \sum_y \mathbf{P}(\xi = a - y)\mathbf{P}(\eta = y),$

де сума поширюється на множину значень величини η.

(2) Прості випадкові величини ξ_1, ξ_2 незалежні, однаково розподілені, а їх розподіл не є рівномірним. Довести, що

$\max(\mathbf{P}(\xi_1 + \xi_2 = a),\ a \in \mathbb{R}) < \max(\mathbf{P}(\xi_1 = a),\ a \in \mathbb{R}).$

(3, П.Леві) Довести, що для кожного натурального $n > 1$, що не є простим, знайдеться пара незалежних величин ξ_1, ξ_2 таких, що всі ймовірності з розподілу $\mathbf{P}(\xi_1 + \xi_2 = k), 0 \le k < n$, однакові і додатні.

(4) Якщо ξ, η – незалежні величини з розподілами Пуассона з параметрами λ, μ відповідно,.то (а) сума $\xi + \eta$ має розподіл Пуассона з параметром $\lambda + \mu$, (б) $\mathbf{P}(\xi_1 = k \mid \xi_1 + \xi_2 = n) = C_n^k p^k (1-p)^{n-k}$, де $p = \lambda/(\lambda + \mu)$.

(5) Випадкові величини ξ_1, ξ_2 незалежні та мають геометричний розподіл: $\mathbf{P}(\xi_i = k) = q^k(1-q), k \ge 0$. (а) Довести, що $\mathbf{P}(\xi_1 = k \mid \xi_1 + \xi_2 = n) = 1/(n+1), k = \overline{0, n}$. (б) Знайти сумісний розподіл величин ξ_1 та $\max(\xi_1, \xi_2)$.

(6) Для довільних щільностей f, g згортка $f * g(x)$ неперервна.

(7) Щільність $f(x)$ інтегровна у квадраті на $\mathbb{R}$. Довести, що згортка $f * f(x)$ обмежена.

(8) Якщо величини $(\xi_k, k = \overline{1, n})$ незалежні у сукупності та мають відповідні функції розподілу (F_k), то функція розподілу суми $\xi_1 + ... + \xi_n$ дорівнює згортці $F_1 * F_2 * ... * F_n$, причому остання не залежить від порядку обчислення інтегралів.

(9) Випадкові величини ξ_k незалежні та мають рівномірний на $[0, 1]$ розподіл, а $U_n(x) = \mathbf{P}(\xi_1 + ... + \xi_n < x)$. Довести, що: (а) $U'_{n+1}(x) = U_n(x) - U_n(x-1)$, (б) $U_n(x) = (n!)^{-1} \sum_{k=0}^{n} (-1)^k C_n^k (x-k)^n 1\!\!1_{k<x}$, (в) щільність суми $\xi_1 + ... + \xi_n$ дорівнює $((n-1)!)^{-1} \sum_{k=0}^{n} (-1)^k C_n^k (x-k)^{n-1} 1\!\!1_{k<x}$.

(10) Випадкові величини $\xi_1, ..., \xi_n$ незалежні та рівномірно розподілені на відрізку $[0, 1]$. Знайти щільність величини $\prod_{k=1}^{n} \xi_k$.

(11) Нехай ξ, η – незалежні величини такі, що величини $\xi + \eta$ та ξ мають однакові функції розподілу. Довести, що $\eta = 0$ м.н.

(12) Випадкові величини ξ_1, ξ_2 незалежні однаково розподілені, і $\min(\xi_1, \xi_2)$ має показниковий розподіл. Довести, що ξ_k мають показниковий розподіл.

(13) Випадкові величини $\xi_1, ..., \xi_n, \eta$ незалежні у сукупності, $\mathbf{P}(\xi_k = 0) = \mathbf{P}(\xi_k = 1) = 1/2$, а η рівномірно розподілена на $[0,1]$. Довести, що величина $\xi = \sum_{k=1}^{n} 2^{-k}\xi_k + 2^{-n}\eta$ також рівномірно розподілена на $[0,1]$.

(14, Роббінс) Випадкові величини $\xi_1, ..., \xi_n$ незалежні у сукупності, а $\zeta_k = \xi_1 + .. + \xi_k$ – їх частинні суми. Довести нерівності

$$\mathbf{P}(\zeta_1 \le x_1, ..., \zeta_n \le x_n) \ge \prod_{k=1}^{n} \mathbf{P}(\zeta_k \le x_k),\ \forall x_1, ..., x_n.$$

(15) Випадкові величини $\xi_1, ..., \xi_n$ незалежні у сукупності та мають показникові розподіли з попарно різними параметрами $\lambda_k, k = \overline{1,n}$. Довести, що щільність суми $\xi_1 + ... + \xi_n$ має вигляд $f_n(x) = \sum_{k=1}^{n} c_{nk} \exp(-\lambda_k x), x \ge 0$. Знайти рекурентні за n рівняння для коефіцієнтів c_{nk} та обчислити ці коефіцієнти.

(16) Випадкові величини ξ_1, ξ_2 незалежні та мають показниковий розподіл $Exp(\lambda)$. Тоді (а) різниця $\zeta = \xi_1 - \xi_2$ має розподіл Лапласа зі щільністю $f_\zeta(x) = \frac{1}{2}\lambda \exp(-\lambda |x|)$, (б) випадкові величини $\xi_1 + \xi_2$ та ξ_1/ξ_2 також незалежні.

1.18.2. Розподіл Ерланга

Теорема (про розподіл Ерланга). *Нехай величини $(\xi_n,\ n \ge 1)$ незалежні в сукупності, однаково розподілені і мають показниковий розподіл із параметром λ. Тоді сума $\zeta_n = \xi_1 + ... + \xi_n$ має* ***розподіл Ерланга*** *порядку n з параметром λ, тобто її щільність дорівнює:*

$$f_n(x) = \lambda(\lambda x)^{n-1} \exp(-\lambda x)/(n-1)!,\ x \ge 0,$$

Доведення. Тут та нижче при $x < 0$ всі щільності $f_n(x) = 0$.

За умовою $f_1(x) = \lambda \exp(-\lambda x), x \ge 0$.

Оскільки сума $\zeta_n = \zeta_{n-1} + \xi_n = (\xi_1 + ... + \xi_{n-1}) + \xi_n$, містить незалежні доданки за теоремою *про векторні перетворення незалежних величин*, із теореми *про щільність розподілу суми незалежних величин* дістанемо рівняння

$$f_n(x) = (f_{n-1} * f_1)(x) = \int_{-\infty}^{x} f_{n-1}(x-y)\ f_1(y)dy =$$

$$\int_0^x f_{n-1}(x-y)\ \lambda \exp(-y)dy = \int_0^x f_{n-1}(y)\ \lambda \exp(-\lambda x + \lambda y)dy.$$

Переходячи до функцій $g_n(x) = f_n(x)\exp(\lambda x)$, виводимо при $n \ge 2$ рекурентні рівняння

$$g_n(x) = \lambda \int_0^x g_{n-1}(y)dy,\quad g_1(x) = \lambda,\ \forall x \ge 0,$$

звідки за індукцією знаходимо

$$g_n(x) = \lambda(\lambda x)^{n-1}/(n-1)!,\quad f_n(x) = \lambda(\lambda x)^{n-1} \exp(-\lambda x)/(n-1)!\ \boxdot$$

1.18.3. Розподіли гама та хі-квадрат

Заміною в означенні *розподілу Ерланга* факторіалу на повну гама-функцію

$$\Gamma(\alpha) \equiv \int_0^\infty x^{\alpha-1}\exp(-x)dx,$$

отримаємо більш загальну сім'ю розподілів.

Означення. *Невід'ємна випадкова величина ζ має* гама-розподіл *із параметрами $\lambda, \alpha > 0$, тобто $\zeta \simeq \Gamma(\lambda,\alpha)$, якщо вона має щільність*

$$f(x) = \lambda(\lambda x)^{\alpha-1}\exp(-\lambda x)\ /\ \Gamma(\alpha),\ x \geq 0.$$

При $\alpha = n \in \mathbb{N}$ гама-розподіл збігається з *розподілом Ерланга*.

Теорема (про інваріантність гама-розподілів відносно згортки). *Нехай випадкові величини $\zeta_1 \simeq \Gamma(\lambda,\alpha)$ і $\zeta_2 \simeq \Gamma(\lambda,\beta)$ незалежні. Тоді їх сума також має гама-розподіл:*

$$\zeta_1 + \zeta_2 \simeq \Gamma(\lambda, \alpha+\beta).$$

Доведення. З теореми *про щільність розподілу суми незалежних величин* виводимо, що

$$f_{\zeta_1+\zeta_2}(x) = f_{\zeta_1} * f_{\zeta_2}(x) =$$

$$\int_0^x \lambda(\lambda y)^{\alpha-1}\exp(-\lambda y)\ \lambda(\lambda x - \lambda y)^{\beta-1}\exp(-\lambda x + \lambda y)dy/\Gamma(\alpha)\Gamma(\beta) =$$

$$\lambda^{\alpha+\beta}\ x^{\alpha+\beta-1}\ \exp(-\lambda x)\ \int_0^1 (1-t)^{\alpha-1}t^{\beta-1}dt\ /\ \Gamma(\alpha)\Gamma(\beta) =$$

$$= \lambda(\lambda x)^{\alpha+\beta-1}\exp(-\lambda x)\ /\ \Gamma(\alpha+\beta),\ x \geq 0,$$

де зроблено заміну $t = y/x$ та використані означення та властивість повної бета-функції $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ $\boxdot$

Частковим випадком гама-розподілу є розподіл хі-квадрат, який широко застосовується в статистиці.

Означення. *Випадкова величина χ_n^2 має* ***хі-квадрат розподіл*** *з n ступенями свободи, якщо її можна подати у вигляді*

$$\chi_n^2 = \zeta_1^2 + ... + \zeta_n^2,$$

де $(\zeta_k,\ k \geq 1)$ – незалежні в сукупності стандартні нормальні величини.

Теорема (про хі-квадрат розподіл). *Нехай випадкова величина χ_n^2 має хі-квадрат розподіл та n ступенів свободи. Тоді $\chi_n^2 \simeq \Gamma(1/2,\ n/2)$.*

Доведення. З теореми *про обчислення розподілу функції від випадкової величини* виводимо рівність $\zeta_1^2 \simeq \Gamma(1/2, 1/2)$:

$$\mathbf{P}(\zeta_1^2 < x) = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}}\exp(-y^2/2)dy = 2\int_0^{\sqrt{x}} \frac{1}{\sqrt{2\pi}}\exp(-y^2/2)dy =$$

$$\int_0^x \frac{1}{\sqrt{2\pi}} u^{-1/2}\ \exp(-u/2)du = \frac{1}{\Gamma(1/2)} \int_0^x (1/2)(u/2)^{-1/2}\ \exp(-u/2)du.$$

Отже, при $n = 1$ теорема доведена. Оскільки величини ζ_k^2 незалежні в сукупності, то доданки в сумі $\chi_{n+1}^2 = \chi_n^2 + \zeta_{n+1}^2$ незалежні за теоремою *про векторні перетворення незалежних величин.* Якщо $\chi_n^2 \simeq \Gamma(1/2, n/2)$, то за теоремою *про інваріантність гама-розподілів відносно згортки* отримуємо рівність $\chi_{n+1}^2 \simeq \Gamma(1/2, n/2) + \Gamma(1/2, 1/2) \simeq \Gamma(1/2, n/2 + 1/2)$, що доводить теорему за індукцією ⊡

Вправи

(1) Довести, що функція розподілу Ерланга порядку n з параметром λ має вигляд $F_n(x) = 1 - (1 + \lambda x + ... + (\lambda x)^{n-1}/(n-1)!) \exp(-\lambda x),\ x \geq 0.$

(2) Вивести з теореми про інваріантність гама-розподілів відносно згортки, що математичне сподівання $\mu(\lambda, \alpha) = \mathbf{E}\Gamma(\lambda, \alpha)$ та дисперсія $\sigma^2(\lambda, \alpha) = \mathbf{D}\Gamma(\lambda, \alpha)$ є адитивними додатними функціями параметра α і тому є лінійними функціями. Обчислити звідси $\mathbf{E}\Gamma(\lambda, \alpha) = \alpha/\lambda$, $\mathbf{D}\Gamma(\lambda, \alpha) = \alpha/\lambda^2$. Знайти $\mathbf{E}\Gamma^r(\lambda, \alpha)$.

(3) Довести формулу Ліувілля перетворення кратних інтегралів

$$\int_0^\infty ... \int_0^\infty f(x_1 + ... + x_n) x_1^{\alpha_1 - 1} ... x_n^{\alpha_n - 1} dx_1 ... dx_n = \frac{\Gamma(\alpha_1)...\Gamma(\alpha_n)}{\Gamma(\alpha_1 + ... + \alpha_n)} \int_0^\infty f(y) y^{\alpha_1 + ... + \alpha_n - 1} dy.$$

(4) Випадкові величини $(\zeta_n, n \geq 1)$ незалежні однаково розподілені зі спільним показниковим розподілом $Exp(\lambda)$. (а) Знайти розподіл випадкової величини $\nu = \inf(n \geq 1 : \zeta_1 + ... + \zeta_n \geq 1)$. (б) Довести, що для кожного $k > 0$ величини ξ_ν та $\xi_{\nu+k}$ незалежні. (в) Знайти розподіл $\xi_{\nu+k}$.

(5) Випадкові величини $\zeta_k \simeq \Gamma(\lambda, \alpha_k), k = 1, 2$, незалежні. Довести, що величини $\zeta_1 + \zeta_2$ і $\zeta_1 /\ (\zeta_1 + \zeta_2)$ незалежні, та знайти їх розподіли.

1.19. Нормальні випадкові вектори

У даному розділі, як і в інших, всі вектори розглядаються як вектористовпчики. Символ $'$ означає транспонування векторів та матриць.

Означення. *n-вимірний випадковий вектор $\zeta = (\zeta_1, ..., \zeta_n)'$ називається* стандартним нормальним вектором, *позначення $\zeta \simeq N_n(0, I)$, якщо випадкові величини $\zeta_1, ..., \zeta_n$ є незалежними в сукупності стандартними нормальними величинами: $\zeta_k \simeq N(0, 1)$.*

Зауваження. З теореми *про критерій незалежності абсолютно неперервних величин* та з означення *стандартної нормальної* величини випливає, що вектор ζ є стандартним нормальним тоді й тільки тоді, коли його *сумісна щільність* має вигляд

$$f_\zeta(x) = \prod\nolimits_{k=1}^{n} (2\pi)^{-1/2} \exp(-x_k^2/2) =$$
$$= (2\pi)^{-n/2} \exp\left(-\sum\nolimits_{k=1}^{n} x_k^2/2\right) = (2\pi)^{-n/2} \exp\left(-x'x/2\right).$$

Означення. *n-вимірний випадковий вектор $\xi = (\xi_1, ..., \xi_n)'$ називається* **нормальним вектором,** *якщо його можна зобразити у вигляді лінійного невиродженого перетворення стандартного нормального вектора ζ, тобто якщо для деяких вектора $m \in \mathbb{R}^n$ та невиродженої матриці $A \in \mathfrak{L}_n(\mathbb{R}^n)$ виконується рівність $\xi = m + A\,\zeta$.*

Зауваження. Можна розглядати **узагальнені нормальні вектори**, для яких матриця A може бути виродженою. Вони не є *абсолютно неперервними*, однак відповідне звуження на підпростір повного рангу вже є *нормальним вектором* меншої розмірності. Більшість із наведених нижче властивостей, за винятком формули для щільності, мають місце і для узагальнених нормальних векторів.

1.19.1. Сумісна щільність нормального вектора

Теорема (про сумісну щільність нормального вектора). *Випадковий вектор ξ є n-вимірним нормальним вектором тоді й тільки тоді, коли його сумісна щільність має вигляд*

$$f_\xi(x) = (2\pi)^{-n/2} \left|\det V\right|^{-1/2} \exp(-(x-m)'V^{-1}(x-m)/2),\ x \in \mathbb{R}^n,$$

де $m \in \mathbb{R}^n$, а матриця $V \in \mathfrak{L}_n(\mathbb{R}^n)$ є симетричною додатно визначеною. Якщо $\xi = m + A\,\zeta$, де $\zeta \simeq N_n(0, I)$, то $m = m$, $V = AA'$.

Означення. *Якщо сумісна щільність випадкового вектора ξ має наведений вище вигляд, будемо писати $\xi \simeq N_n(m, V)$.*

Зауваження. Якщо $m = 0$ і $V = I$ – одинична матриця, то позначення $N_n(0, I)$ відповідає означенню *стандартного нормального вектора*.

Доведення. *Необхідність.* Нехай $\xi = m + A\ \zeta$, де ζ – стандартний нормальний вектор і $\det A \neq 0$.

Для обчислення щільності ξ розглянемо n-вимірний кут

$$\Pi_x = (-\infty, x) = \prod\nolimits_{k=1}^{n} (-\infty, x_k).$$

За теоремою *про обчислення ймовірностей, пов'язаних із неперервним вектором*

$$F_\xi(x) = \mathbf{P}(\xi < x) = \mathbf{P}(m + A\,\zeta \in \Pi_x) = \mathbf{P}(\zeta \in A^{-1}(\Pi_x - m)) =$$

$$\int_{A^{-1}(\Pi_x - m)} f_\zeta(y)dy = \int_{y:\, m+Ay\in\Pi_x} f_\zeta(y)dy = \int_{\Pi_x} f_\zeta(A^{-1}(u-m)) \left|\det A^{-1}\right| du,$$

де використана заміна змінної $u = m + Ay$ з якобіаном $|\det A^{-1}|$. Звідси за означенням *сумісної щільності*

$$f_\xi(u) = f_\zeta(A^{-1}(u-m)) \left|\det A^{-1}\right| =$$

$$|\det A|^{-1}\ (2\pi)^{-n/2}\ \exp(-(A^{-1}(u-m))' A^{-1}(u-m)/2) =$$

$$(2\pi)^{-n/2}\ |\det V|^{-1/2}\ \exp(-(u-m)' V^{-1}(u-m)/2),$$

де враховано визначення $V = AA'$ та рівність $|\det A| = |\det V|^{1/2}$. Симетричність V очевидна, додатна визначеність випливає з невиродженості матриці A та рівності $c'Vc = c'AA'c = |c'A|^2 > 0\ \forall c \neq 0$.

Достатність. Нехай вектор $m \in \mathbb{R}^n$, матриця V симетрична додатно визначена, а випадковий вектор ξ має наведену вище сумісну щільність. Зобразимо матрицю V у вигляді $V = A\ A'$ із деякою невиродженою матрицею A (це можливо після зведення V до діагональної форми). Розглянемо випадковий вектор $\zeta = -A^{-1}m + A^{-1}\xi \equiv m_1 + A_1\xi$. Тоді $\xi = m + A\ \zeta$, та аналогічно до наведених вище міркувань

$$f_\zeta(x) = f_\xi(A_1^{-1}(x - m_1)) \left|\det A_1^{-1}\right| = f_\xi(Ax + m)\, |\det A| =$$

$$(2\pi)^{-n/2} \exp(-x'x/2),$$

тобто ζ – *стандартний нормальний вектор* $\boxdot$

1.19.2. Параметри розподілу нормального вектора

Теорема (про інтерпретацію параметрів нормального розподілу). *Нехай $\xi \simeq N_n(m, V)$ – нормальний вектор. Тоді*

$$\mathbf{E}\xi = m,\ \mathrm{Cov}(\xi) = V.$$

Доведення. Очевидно, що за *лінійністю* $\mathbf{E}\xi = \mathbf{E}(m + A\ \zeta) = m + A\ \mathbf{E}\zeta = m + A0 = m$. Далі, за теоремою *про коваріаційну матрицю лінійного перетворення* випадкового вектора $\mathrm{Cov}(\xi) = \mathrm{Cov}(\xi - m) = \mathrm{Cov}(A\zeta) = A\mathrm{Cov}(\zeta)A' = AIA' = V$ $\boxdot$

1.19.3. Нормальний вектор на площині

Теорема (про щільність нормального розподілу на площині). *Для довільного **нормального вектора на площині** $\xi = (\xi_1, \xi_2) \simeq N_2(\cdot)$ існують п'ять сталих $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2 > 0$, $\sigma_2^2 > 0$, $\rho \in (-1, 1)$ такі, що сумісна* *щільність дорівнює*

$$f_\xi(x) = (2\pi\sigma_1\sigma_2)^{-1}(1-\rho^2)^{-1/2}\times$$
$$\exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x_1-\mu_1}{\sigma_1}\ \frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right),$$

де $\mu_k = \mathbf{E}\xi_k$, $\sigma_k^2 = \mathbf{D}\xi_k$, $k = 1, 2$, *а* ρ *– коефіцієнт кореляції* ξ_1 *і* ξ_2 *:*

$$\rho(\xi_1, \xi_2) = \mathrm{Cov}(\xi_1, \xi_2)/\sigma_1\sigma_2 = \mathbf{E}(\xi_1 - \mu_1)(\xi_2 - \mu_2)/\sigma_1\sigma_2 = \rho.$$

Зауваження. Вектор (ξ_1, ξ_2) позначають через $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

Доведення. Будь-яка невироджена симетрична додатно визначена матриця 2×2 має вигляд $V = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, де $\sigma_k > 0$, $|\rho| < 1$. З теореми *про сумісну щільність нормального вектора* та формули обертання матриці $V^{-1} = (1-\rho^2)^{-1}\begin{pmatrix} \sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2} \end{pmatrix}$ знаходимо наведений вираз для щільності. Подальше випливає з теореми *про інтерпретацію параметрів нормального розподілу* та означення *коефіцієнта кореляції* $\boxdot$

Теорема (про незалежність координат нормального вектора). *Нехай* $\xi = (\xi_1, \xi_2)' \simeq N_2(m, V)$. *Величини* ξ_1 *і* ξ_2 *незалежні тоді й тільки тоді, коли вони некорельовані:* $Cov(\xi_1, \xi_2) = 0$.

Доведення. Некорельованість незалежних величин доведена вище як наслідок теореми *про математичне сподівання добутку незалежних величин.* Навпаки, з некорельованості величин виводимо, що коефіцієнт кореляції $\rho = 0$. З теореми *про щільність нормального розподілу на площині* виводимо, що ця *щільність* розпадається в добуток маргінальних одновимірних щільностей. Тому сформульоване твердження випливає з теореми *про критерій незалежності абсолютно неперервних величин* $\boxdot$

Вправи

(1) Нехай $\xi = (\xi_1, \xi_2) \simeq N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ *–нормальний вектор на площині.* Довести, що сумісну щільність ξ можна зобразити у вигляді добутку $f_\xi(x_1, x_2) = f_{\xi_{12}}(x_1 \mid x_2) f_{\xi_2}(x_2)$, де $\xi_2 \simeq N(\mu_2, \sigma_2^2)$, а $f_{\xi_{12}}(x_1 \mid x_2)$ є щільністю відносно x_1 нормальної величини $\xi_{12} \simeq N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2))$. Остання щільність називається умовною щільністю ξ_1 за умови $\xi_2 = x_2$.

(2) Нехай (α_1, α_2) – незалежні рівномірно розподілені на $(0, 1)$ величини. Довести, що величини $\xi_1 = \sqrt{-2\ln\alpha_1}\sin(2\pi\alpha_2)$, $\xi_2 = \sqrt{-2\ln\alpha_1}\cos(2\pi\alpha_2)$ – незалежні стандартні нормальні.

(3) Якщо $(\zeta_k, k = \overline{1, 3})$ – незалежні стандартні нормальні випадкові величини, то $(\zeta_1 + \zeta_2\zeta_3)/\sqrt{1 + \zeta_3^2} \simeq N(0, 1)$.

(4) Якщо ζ_k – незалежні стандартні нормальні величини, а кут θ не залежить від них та рівномірно розподілений на $[0, 2\pi]$, то величина $\zeta_1 \cos\theta + \zeta_2 \sin\theta$ також стандартна нормальна.

(5) Для стандартного нормального вектора $(\zeta_1, \zeta_2) \simeq N_2$ (а) знайти щільність розподілу величини $\max(|\zeta_1|, |\zeta_2|)$, (б) знайти сумісну щільність полярних координат (R, φ) цього вектора, (в) довести, що випадкова величина $\zeta_1\zeta_2/\sqrt{\zeta_1^2 + \zeta_2^2}$ має нормальний розподіл та знайти його параметри.

(6) Нехай $(\xi_1, \xi_2) \simeq N_2(0, 0, 1, 1, \rho)$. Довести, що: (а) $\mathbf{P}(\xi_1 \cdot \xi_2 < 0) = (\arccos\rho)/\pi$, (б) $\mathbf{E}\max(\xi_1, \xi_2) = \sqrt{(1-\rho)/\pi}$, (в) щільність відношення ξ_1/ξ_2 дорівнює $\sqrt{1-\rho^2}/\ \pi(1 - 2\rho x + x^2)$.

(7) Нехай випадковий вектор (ξ_1, ξ_2) рівномірно розподілений на одиничному колі $O = \{(x_1, x_2) : x_1^2 + x_2^2 \le 1\}$, а $\xi = \xi_1^2 + \xi_2^2$. Довести, що величини $\zeta_k = \xi_k\sqrt{-2\xi^{-1}\ln\xi}$ є незалежними стандартними нормальними.

(8) Для стандартного нормального вектора (ζ_1, ζ_2) (а) полярні координати незалежні, квадрат полярного радіусу має хі-квадрат розподіл, полярний кут рівномірно розподілений, (б) величини $\zeta_1^2 + \zeta_2^2$ і ζ_1/ζ_2 незалежні.

(9) Для стандартного нормального вектора $(\zeta_{11}, \zeta_{12}, \zeta_{21}, \zeta_{22}) \simeq N_4$ знайти розподіл визначника $\det(\zeta_{ij}, i, j = 1, 2)$.

(10) Навести приклад некорельованих випадкових величин ξ_1, ξ_2, що мають нормальні розподіли, однак є залежними.

(11) Довести, що $\mathbf{E}\exp(t\,\|\xi\|^2) < \infty$ для $\xi \simeq N_d(m, V)$ та $t \ge 0$.

(12) Нехай $\xi \simeq N_d(m, AA')$. Довести, що для довільної $B \in \mathfrak{B}(\mathbb{R}^n)$
$\overline{\lim}_{n\to\infty} 2n^{-2}\ln\mathbf{P}(\xi/n \in B) \le -\inf_{x\in B}\|A^{-1}(x-m)\|^2$.

1.19.4. Лінійні перетворення нормальних векторів

Нагадаємо: матриця U називається ортонормованою, якщо $U^{-1} = U'$.

Теорема (про лінійні перетворення нормальних векторів).

(а) Якщо $\xi \simeq N_n(m, V)$, де стала $a \in \mathbb{R}^n$ і перетворення $T \in \mathfrak{L}_n(\mathbb{R}^n)$ невироджене, то випадковий вектор $\eta \equiv a + T\xi$ також є нормальним вектором, та $\eta \simeq N_n(a + Tm,\ TVT')$.

(б) Якщо $\zeta \simeq N_n(0, I)$ – стандартний нормальний вектор, а U – ортонормована матриця, то вектор $U\zeta \simeq N_n(0, I)$ теж стандартний нормальний.

(в) Якщо нормальні вектори $\xi_k \simeq N_{n_k}(m_k, V_k)$, $k = 1, 2$, незалежні, то складений ними вектор $(\xi_1, \xi_2) \simeq N_{n_1+n_2}$ також є нормальним.

(г) Нехай $T \in \mathfrak{L}(\mathbb{R}^n, \mathbb{R}^k)$ для деякого $k < n$, причому rang$(T) = k$. Якщо $\xi \simeq N_n(m, V)$, то $T\xi \simeq N_k(Tm,\ TVT')$.

Доведення

(а) Нехай $\xi = m + A\ \zeta$, де вектор ζ – стандартний нормальний. Тоді отримуємо $\eta = a + T\xi = a + Tm + TA\ \zeta$. Отже, за означенням нормального вектора $\eta \simeq N_n(a + Tm,\ TA(TA)') = N_n(a + Tm,\ TVT')$.

(б) Випливає з (а) та з означення стандартного нормального вектора, оскільки $\mathbf{E}U\zeta = U\mathbf{E}\zeta = 0$, $\mathrm{Cov}(U\zeta) = UIU' = I$.

(в) Нехай $\xi_k = m_k + A_k\zeta_k$, де ζ_k – стандартні нормальні вектори. Тоді вектори $\zeta_k = A_k^{-1}(\xi_k - m_k)$ незалежні за теоремою *про перетворення незалежних величин*. За означенням вектор (ζ_1, ζ_2) є стандартним нормальним, оскільки його координатами є незалежні *стандартні нормальні* величини. Отже, (ξ_1, ξ_2) є нормальним вектором як невироджене лінійне перетворення (ζ_1, ζ_2).

(г) Нехай $\xi = m + A\ \zeta$, де вектор ζ – стандартний нормальний. Тоді $\eta = T\xi = Tm + TA\zeta = Tm + B\zeta \in \mathbb{R}^k$, де $B = TA$, $\mathrm{rang}(B) = k$.

Оскільки розмірність ядра $\mathrm{Ker}(B) \equiv \{x : Bx = 0\}$ дорівнює $n - \mathrm{rang}(B) = n - k$, то в цьому підпросторі існує ортонормований базис $\{e_{k+1}, ..., e_n\}$ розмірності $n - k$. Нехай $\{e_1, ..., e_k\}$ його доповнення до ортонормованого базису в $\mathbb{R}^n$, що існує за теоремою Грама – Шмідта. Тоді матриця $U = (e_1, ..., e_n)$ є *ортонормованою* і за теоремою *про лінійні перетворення нормальних векторів* вектор $\theta \equiv U^{-1}\zeta = (\theta_1, ..., \theta_n)$ є стандартним нормальним. За означенням U та побудовою базису $BU = (Be_1, ..., Be_n) = (Be_1, ..., Be_k, 0, ..., 0)$. Тому $\eta = T\xi = Tm + B\zeta = Tm + BU\ \ U^{-1}\zeta = Tm + (Be_1, ..., Be_k, 0, ..., 0)\theta = (Be_1, ..., Be_k)\ \ \theta(k) \simeq N_k(Tm,\ \cdot)$, оскільки вектор $\theta(k) \equiv (\theta_1, ..., \theta_k)$, що утворений k *незалежними в сукупності стандартними нормальними* величинами, є стандартним нормальним k-вимірним, а $\mathrm{rang}(Be_1, ..., Be_k) = k$. *Коваріаційна матриця* вектора $T\xi$ обчислюється за теоремою *про коваріаційну матрицю лінійного перетворення* випадкового вектора: $\mathrm{Cov}(T\xi) = T\mathrm{Cov}(\xi)T' = TVT'$ ⊡

Вправи

(1) Знайти параметри розподілу вектора у пункті (в) теореми.

(2) Нехай $f_k(x_1, x_2), k = 1, 2,$ – щільності двовимірних нормальних розподілів з нульовими середніми, одиничними дисперсіями та різними кореляціями ρ_k. Довести, що випадковий вектор зі щільністю $(f_1(x_1, x_2) + f_2(x_1, x_2))/2$ не є нормальним вектором, однак його координати нормально розподілені.

(3) Нехай φ – стандартна нормальна щільність, а непарна функція ψ така, що $|\psi(x)| \le (2\pi e)^{-1/2}\mathbb{1}_{|x|\le 1}, \forall x$. Довести, що функція $\varphi(x)\varphi(y) + \psi(x)\psi(y)$ є щільністю розподілу двовимірного випадкового вектора, що не є нормальним, однак обидві його координати мають стандартний нормальний розподіл.

(4) Навести приклад нормальних випадкових величин ξ, η, сума яких не має нормального розподілу.

(5, нерівність Слепяна) Нехай $(\xi_1, ..., \xi_n)$ та $(\eta_1, ..., \eta_n)$ – нормальні вектори такі, що $\mathbf{E}\xi_k = \mathbf{E}\eta_k$, $\mathbf{D}\xi_k = \mathbf{D}\eta_k$, $\mathrm{Cov}(\xi_k, \xi_j) \le \mathrm{Cov}(\eta_k, \eta_j), \forall k, j = \overline{1, n}$. Тоді $\mathbf{P}(\max_{1\le k\le n} \xi_k < x) \le \mathbf{P}(\max_{1\le k\le n} \eta_k < x)$, $\forall\ x \in \mathbb{R}$.

(6) Нехай (ξ_1, ξ_2) нормальний вектор з $\mathbf{E}\xi_k = 0$ та $\mathrm{Cov}(\xi_1, \xi_2) \le 0$. Довести, що $\mathbf{P}(\xi_1 \ge x_1,\ \xi_2 \ge x_2) \le \mathbf{P}(\xi_1 \ge x_1)\ \mathbf{P}(\xi_2 \ge x_2)$, $\forall x_k \ge 0$.

(7) Нехай $(\xi_1, \xi_2) \simeq N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Довести, що $|\rho(f_1(\xi_1), f_2(\xi_2))| \le |\rho|$ за квадратичної інтегровності величин під знаком коефіцієнта кореляції.

(8) Нехай $\zeta_k, k = \overline{1, n}$, – незалежні стандартні нормальні величини. Вивести при $\sigma_k \in \mathbb{R}$ з теореми про лінійні перетворення нормальних векторів рівність розподілів величин $\sum_{k=1}^n (\zeta_k + \sigma_k)^2 \simeq \left(\zeta_1 + (\sum_{k=1}^n \sigma_k^2)^{1/2}\right)^2 + \sum_{k=2}^n \zeta_k^2$.

(9) Навести приклад, що вказує на суттєвість умови незалежності нормальних координат ξ_k для нормальності розподілу вектора ξ у теоремі про лінійні перетворення нормальних векторів, (в).

1.19.5. Незалежність і некорельованість нормальних величин

Наслідок (про незалежність лінійних форм нормального вектора). *Нехай $\xi \simeq N_n(m, V)$, та $c, d \in \mathbb{R}^n$, $rang(c, d) = 2$. Для того, щоб лінійні форми $c'\xi$ і $d'\xi$ були незалежні, необхідно й достатньо, щоб вони були некорельовані: $Cov(c'\xi,\ d'\xi) = c'Vd = 0$.*

Доведення. За теоремою *про лінійні перетворення нормальних векторів,* (г), вектор $(c'\xi,\ d'\xi)$ є нормальним вектором на площині. Тому твердження теореми випливає з теореми *про незалежність координат нормального вектора* на площині $\boxdot$

Зауваження. Твердження наслідку поширюється і на випадок довільної кількості лінійних форм від нормального вектора.

Наслідок (про нормальність суми незалежних нормальних векторів). *Нехай $\xi_k \simeq N_n(m_k, V_k)$ – незалежні нормальні випадкові вектори. Тоді $\xi_1 + \xi_2 \simeq N_n(m_1 + m_2, V_1 + V_2)$ має нормальний розподіл.*

Доведення. За теоремою *про лінійні перетворення нормальних векторів*, (в), складений вектор (ξ_1, ξ_2) є нормальним. Тому його лінійне перетворення $\xi_1 + \xi_2$ є нормальним вектором за теоремою *про лінійні перетворення нормальних векторів,* (г), оскільки лінійне відображення $T(x, y) = x + y : \mathbb{R}^{2n} \to \mathbb{R}^n$ має ранг n. Координати ξ_1 незалежні і *некорельовані* з координатами ξ_2 за теоремою *про математичне сподівання* *добутку незалежних величин*, тому

$$\mathbf{E}(\xi_1 + \xi_2) = m_1 + m_2,$$

$$\mathrm{Cov}(\xi_1 + \xi_2) = (\mathrm{Cov}(\xi_{1i} + \xi_{2i}, \xi_{1j} + \xi_{2j}),\ i, j = \overline{1, n}) =$$

$$(\mathrm{Cov}(\xi_{1i}, \xi_{1j}) + \mathrm{Cov}(\xi_{2i}, \xi_{2j}),\ i, j = \overline{1, n}) = \mathrm{Cov}(\xi_1) + \mathrm{Cov}(\xi_2) = V_1 + V_2,$$

отже, за теоремою *про інтерпретацію параметрів нормального розподілу* $\xi_1 + \xi_2 \simeq N_n(m_1 + m_2, V_1 + V_2)$ $\boxdot$

Вправи

(1) Якщо ζ_k – незалежні нормальні стандартні величини, то їх лінійні перетворення $\zeta_1 + \zeta_2$ та $\zeta_1 - \zeta_2$ – також незалежні нормальні величини.

(2) Нехай ζ – n-вимірний стандартний нормальний вектор, а лінійні підпростори $E_1, E_2 \subset \mathbb{R}^n$ такі що $E_1 \cap E_2 = \{0\}$. Довести, що проекції $\Pi_{E_1}\zeta, \Pi_{E_2}\zeta$ є незалежними нормальними векторами.

(3, теорема Максвелла) Нехай випадкові величини ξ_1, ξ_2 незалежні, а їх сумісна щільність не змінюється при поворотах вектора (ξ_1, ξ_2). Довести, що цей випадковий вектор є нормальним.

(4) Знайти лінійне перетворення $T : \mathbb{R}^n \to \mathbb{R}^{n-1}$ таке, що випадкові вектори $T(N_n(\mu, \sigma^2 I))$ та $N_{n-1}(0, \sigma^2 I)$ – однаково розподілені.

(5) Нехай $\xi \simeq N_n(\mu, V)$. Довести, що випадкова величина $(\xi - \mu)' V^{-1} (\xi - \mu)$ має хі-квадрат розподіл з n ступенями свободи.

(6) **Узагальнений нормальний вектор** визначається як лінійне перетворення $\xi = m + A\zeta$ стандартного нормального вектора з довільною матрицею $A \in \mathfrak{L}(\mathbb{R}^n, \mathbb{R}^n)$. (а) Довести, що вектор ξ є узагальненим нормальним тоді й тільки тоді, коли для кожного лінійного неперервного функціоналу $l : \mathbb{R}^n \to \mathbb{R}$ величина $l(\xi)$ має нормальний розподіл або є сталою м.н. (б) Довести, що розподіл ξ однозначно визначається вектором m та матрицею $V = AA'$. Довести для класу узагальнених нормальних векторів: (в) теорему про інтерпретацію параметрів, (г) теорему про лінійні перетворення нормальних векторів, (д) теорему про нормальність суми незалежних нормальних векторів. *Вказівка*: нехай T' – нульове продовження на $\mathbb{R}^n$ матриці ортонормованого базису простору власних векторів матриці A. Тоді $AT' = 0$, ідемпотентна матриця $J = TT'$ містить $n - \mathrm{rang}(A) = n - \mathrm{rang}(V)$ перших одиниць на діагоналі, причому для кожного $\varepsilon \neq 0$ матриця $A + \varepsilon T$ невироджена. Отже, вектор $\xi_\varepsilon = m + (A + \varepsilon T)\zeta = \xi + \varepsilon T\zeta$ є n-вимірним нормальним, $\xi_\varepsilon \to \xi, \varepsilon \to 0$, і оскільки $\mathbf{E}\xi_\varepsilon = m$, $\mathrm{Cov}(\xi_\varepsilon) = (A + \varepsilon T)(A + \varepsilon T)' = V + \varepsilon^2 J$, то розподіл $\xi_\varepsilon \simeq N_n(m, V + \varepsilon^2 J)$ повністю визначається параметрами ε та m, V. Нарешті, границя $\lim_{\varepsilon \to 0} \mathbf{P}(\xi_\varepsilon \in \Pi_x) = \mathbf{P}(\xi \in \Pi_x)$ залежить лише від m, V та x.

1.20. Збіжність за ймовірністю та її властивості

Означення. *Послідовність випадкових величин* $(\xi_n,\ n \geq 1)$ збігається за ймовірністю *до величини* ξ *(позначення* $\xi_n \to^P \xi$*), якщо*

$$\mathbf{P}(|\xi_n - \xi| \geq \varepsilon) \to 0,\ n \to \infty,\ \forall \varepsilon > 0.$$

Зауваження. Якщо одночасно $\xi_n \to^P \eta$, то $\mathbf{P}(|\eta - \xi| \geq \varepsilon) = 0$ для всіх $\varepsilon > 0$, звідки $\eta - \xi = 0$ *м.н.*, тобто границя за ймовірністю визначена однозначно з точністю до рівності *майже напевне*.

Означення. *Послідовність випадкових величин* $(\xi_n,\ n \geq 1)$ обмежена за ймовірністю*, якщо* $\lim_{c\to\infty} \sup_{n\geq 1} \mathbf{P}(|\xi_n| \geq c) = 0.$

Вправа. Доведіть, що клас усіх випадкових величин із метрикою Леві

$$\rho(\xi, \eta) = \inf(\varepsilon > 0 : \mathbf{P}(|\xi - \eta| \geq \varepsilon) \leq \varepsilon)$$

є повним метричним простором, збіжність в якому є збіжністю за ймовірністю. Вказівка. Якщо $\mathbf{P}(|\xi - \eta| \geq \varepsilon) \leq \delta$, то $\rho(\xi, \eta) \leq \max(\varepsilon, \delta)$. Звідси знаходимо $\rho(\xi, \eta) = \inf(\max(\varepsilon, \delta) > 0 : \mathbf{P}(|\xi - \eta| \geq \delta) \leq \varepsilon)$. Тому $\rho(\xi_n, \xi) \to 0$ тоді й тільки тоді, коли $\lim \mathbf{P}(|\xi_n - \xi| \geq \delta) = 0,\ \forall \delta > 0$. Нерівність трикутника випливає з нерівності $\mathbf{P}(|\xi - \eta| \geq \varepsilon + \delta) \leq \mathbf{P}(|\xi - \zeta| \geq \varepsilon) + \mathbf{P}(|\zeta - \eta| \geq \delta)$. Доведення повноти спирається на можливість виділення з фундаментальної за ймовірністю послідовності підпослідовності, що фундаментальна майже напевне.

1.20.1. Властивості збіжності за ймовірністю

Теорема (про властивості збіжності за ймовірністю). *Нехай має місце збіжність* $\xi_n \to^P \xi, n \to \infty$.

(а) Якщо $f \in C(\mathbb{R})$, *то* $f(\xi_n) \to^P f(\xi), n \to \infty$.

(б) Якщо $|\xi_n| \leq c$, *то* $\mathbf{E}\xi_n \to \mathbf{E}\xi$ *та* $\mathbf{E}\,|\xi_n - \xi| \to 0, n \to \infty$.

(в) Якщо $f \in C_b(\mathbb{R})$, *то* $\mathbf{E}f(\xi_n) \to \mathbf{E}f(\xi), n \to \infty$.

(г) Якщо $V(x)$ *додатна неперервна функція,* $V(x)/x \to \infty,\ x \to \infty$, *та* $\sup_{n\geq 1} \mathbf{E}V(|\xi_n|) < \infty$, *то* $\mathbf{E}\xi_n \to \mathbf{E}\xi$ *і* $\mathbf{E}\,|\xi_n - \xi| \to \infty$ *при* $n \to \infty$.

(д) Якщо також $\eta_n \to^P \eta$, *то* $a\xi_n + b\eta_n \to^P a\xi + b\eta$ *при* $a, b \in \mathbb{R}$.

Доведення

(а) Для довільного $\gamma > 0$ оберемо $c > 0$ так, щоб $\mathbf{P}(|\xi| > c) \leq \gamma$. Розглянемо модуль неперервності функції f у δ-околі відрізка $[-c, c]$:

$$w_{c,\delta}(f) = \sup_{|x|\leq c,\ |x-y|\leq \delta}\ |f(x) - f(y)|\,.$$

З неперервності f та компактності відрізка випливає, що $w_{c,\delta}\,(f) \to 0$ при $\delta \to 0$, оскільки неперервна функція рівномірно неперервна на компакті і при $\delta \le 1$ за означенням

$$w_{c,\delta}\,(f) \le \sup{}_{|x|,|y| \le c+1,\ |x-y| \le \delta}\ |f(x) - f(y)| \to 0,\ \delta \to 0.$$

Тому для довільного $\varepsilon > 0$ знайдеться $\delta > 0$ таке, що $w_{c,\delta}(f) < \varepsilon$. З останньої умови та означення $w_{c,\delta}(f)$ виводимо включення

$$\{|\xi_n - \xi| < \delta\} \cap \{|\xi| \le c\} \subset \{|f(\xi_n) - f(\xi)| \le w_{c,\delta}(f)\} \subset \{|f(\xi_n) - f(\xi)| \le \varepsilon\}.$$

Переходом до доповнень, звідси за вказаного $\delta > 0$ отримуємо включення

$$\{|f(\xi_n) - f(\xi)| \ge \varepsilon\} \subset \{|\xi| > c\} \cup \{|\xi_n - \xi|\ \ge \delta\},$$

Звідси за *напівадитивністю* ймовірності виводимо нерівності

$$\mathbf{P}(|f(\xi_n) - f(\xi)| \ge \varepsilon) \le \mathbf{P}(|\xi| > c) + \mathbf{P}(|\xi_n - \xi| \ge \delta) \le$$

$$\gamma + \mathbf{P}(|\xi_n - \xi| \ge \delta) \le 2\,\gamma$$

для досить великих n, де γ довільна додатна стала. Тому ліва частина прямує до нуля при $n \to \infty$.

(б) Нехай функція $g \in C(\mathbb{R})$ така, що $g(x) = 0$ при $|x| \le c$ та $g(x) > 0$ при $|x| > c$. Тоді $0 = g(\xi_n) \to^P g(\xi)$, отже, $g(\xi) = 0$ та $|\xi| \le c$ *майже напевне*. Тому для кожного $\varepsilon > 0$ *з імовірністю 1* справедлива нерівність

$$|\xi_n - \xi| \le \varepsilon + 2\,c\ \mathbb{1}_{\{|\xi_n - \xi|\ \ge\ \varepsilon\}},$$

і з *монотонності* математичного сподівання отримуємо

$$\overline{\lim}_{n\to\infty}\mathbf{E}\,|\xi_n - \xi| \le \varepsilon + 2\,c\ \overline{\lim}_{n\to\infty}\mathbf{P}(|\xi_n - \xi| \ge \varepsilon) = \varepsilon.$$

Зважаючи на довільність $\varepsilon > 0$, звідси виводимо твердження (б).

(в) За твердженням (а) $f(\xi_n) \to^P f(\xi)$, де величини $f(\xi_n)$ обмежені. Тому твердження (в) випливає з (б).

(г) Нехай $\tau_c(x)$ – непарна неперервна функція, яка при $x \ge 0$ дорівнює $\tau_c(x) = \min(x,\ c\ \max(c + 1 - x, 0))$. Зауважимо, що

(1) $\tau_c(x) = x,\ \forall\,|x|\ \le c,$
(2) $\tau_c(x) = 0,\ \forall\,|x|\ > c + 1,$
(3) $|\tau_c(x)| \le c, \quad \tau_c \in C_b.$

Тому за властивістю (в) мають місце збіжності при $n \to \infty$:

$$\tau_c(\xi_n) \to^P \tau_c(\xi),\ \mathbf{E}\tau_c(\xi_n) \to \mathbf{E}\tau_c(\xi).$$

Позначимо $\varepsilon(c) = \sup_{x\,\ge\,c} x/V(x)$. Тоді $\varepsilon(c) \to 0$ при $c \to \infty$ за умовою. Крім того, за властивістю (1) та означенням $\varepsilon(c)$ виконуються нерівності

$$|x - \tau_c(x)|\ \le |x|\ \ \mathbb{1}_{|x|\,>\,c} \le \varepsilon(c)V(|x|),\ \forall x \in \mathbb{R}.$$

Далі, позначимо $V_b(x) = \min(V(x), b)$. З неперервності V та обмеженості V_b випливає збіжність $\mathbf{E}V_b(|\xi_n|) \to \mathbf{E}V_b(|\xi|)$. Тому

$$\mathbf{E}V_b(|\xi|) = \lim_{n\to\infty} \mathbf{E}V_b(|\xi_n|) \le \sup_{n\ge1} \mathbf{E}V(|\xi_n|) < \infty.$$

Оскільки $V_b \uparrow V, b \to \infty$, то за *теоремою Лебега про монотонну збіжність* $\mathbf{E}V(|\xi|)$ також не перевищує останньої величини.

Тому з означення τ_c та попередніх оцінок дістанемо нерівність

$$\overline{\lim}_{n\to\infty} \mathbf{E}\,|\xi_n - \xi| \le \overline{\lim}_{n\to\infty} \mathbf{E}\,|\xi_n - \tau_c(\xi_n)| +$$
$$\overline{\lim}_{n\to\infty} \;|\mathbf{E}\tau_c(\xi_n) - \mathbf{E}\tau_c(\xi)| \; + \mathbf{E}\; |\tau_c(\xi) - \xi| \; \le$$
$$\varepsilon(c) \sup_{n\ge1} \mathbf{E}V(|\xi_n|) + 0 + \varepsilon(c)\mathbf{E}V(|\xi|).$$

Вибором c праву частину можна зробити як завгодно малою.

(д) Твердження є наслідком очевидного включення

$$\{|a\xi_n + b\eta_n - a\xi - b\eta| \ge \varepsilon\} \subset$$
$$\{|a|\,|\xi_n - \xi| \ge \varepsilon/2\} \cup \{|b|\,|\eta_n - \eta| \ge \varepsilon/2\} \boxdot$$

Вправи

(1) Збіжності $\xi_n \to^P \xi$ та $\mathbf{E}\exp(-|\xi_n - \xi|) \to 1, n \to \infty$, еквівалентні.

(2) Послідовність $(\xi_n, n \ge 1)$ обмежена за ймовірністю, а $\eta_n \to^P 0$ при $n \to \infty$. Довести, що $\xi_n \eta_n \to^P 0$.

(3, лема Слуцького) Нехай $\xi_n \to^P \xi$, $\eta_n \to^P \eta$, а $f \in C(\mathbb{R}^2)$. Довести, що $f(\xi_n, \eta_n) \to^P f(\xi, \eta),\ n \to \infty$.

(4) Якщо $\xi_n \le \zeta_n \le \eta_n$ м.н. та $\xi_n \to^P \zeta$, $\eta_n \to^P \zeta$, то $\zeta_n \to^P \zeta,\ n \to \infty$.

(5, лема Пратта) Нехай $\xi_{1n} \le \xi_{2n} \le \xi_{3n}$ м.н., $\xi_{in} \to^P \xi_i,\ n \to \infty, i = \overline{1,3}$, та $\mathbf{E}\xi_{1n} \to \mathbf{E}\xi_1$, $\mathbf{E}\xi_{3n} \to \mathbf{E}\xi_3$. Довести, що $\mathbf{E}\xi_{2n} \to \mathbf{E}\xi_2, n \to \infty$.

(6) Довести узагальнення теореми Лебега про мажоровану збіжність на випадок збіжності за ймовірністю.

1.20.2. Інші види збіжності випадкових величин

Означення. *Послідовність випадкових величин* $(\xi_n, n \ge 1)$ збігається з імовірністю 1 *до величини* ξ *або ж* збігається майже напевне, *позначення* $\xi_n \to^{P1} \xi$, *якщо*

$$\mathbf{P}(\{\omega : \exists \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\}) = 1.$$

Означення. *Послідовність випадкових величин* $(\xi_n\ , n \ge 1)$ *збігається до величини* ξ у середньому *порядку* q, *позначення* $\xi_n \to^{Lq} \xi$, *якщо збігаються відповідні середні:* $\mathbf{E}\,|\xi_n - \xi|^q \to 0$ *при* $n \to \infty$. *Збіжність* *у середньому порядку* 2 *називається також збіжністю* у середньому квадратичному.

Теорема (про співвідношення між різними видами збіжності). *Завжди справедливі такі імплікації між видами збіжності*

$$\boxed{P1} \Longrightarrow \boxed{P} \Longleftarrow \boxed{L_1} \Longleftarrow \boxed{L_q, q > 1}.$$

Вправа. Наведіть приклади, для яких інші імплікації є хибними. *Вказівка*. Нехай $(\Omega, \mathfrak{F}, \mathbf{P}) = ([0,1),\ \mathfrak{B}[0,1),\ L)$. Тоді

$(P \not\Rightarrow P1)$ $\xi_n = 1\!\!\mathrm{I}_{\omega \in [\sqrt{n}, \sqrt{n+1})\ mod\ 1}$, $(P \not\Rightarrow L_1)$ $\xi_n = n^2 1\!\!\mathrm{I}_{\omega \in [1/(n+1),\ 1/n)}$,

$(L_1 \not\Rightarrow L_2)$ $\xi_n = n 1\!\!\mathrm{I}_{\omega \in [0, 1/n^2)}$, $(P1 \not\Rightarrow L_1)$ $\xi_n = n 1\!\!\mathrm{I}_{\omega \in [0, 1/n)}$.

Доведення. За означенням границі рівність $\xi(\omega) = \lim_{n \to \infty} \xi_n(\omega)$ має місце з імовірністю 1 тоді й тільки тоді, коли

$$\begin{aligned} \exists O &\in \mathfrak{F} : \mathbf{P}(O) = 1,\ \forall \omega \in O, \forall \varepsilon > 0, \\ \exists N &= N(\varepsilon, \omega) < \infty :\ |\xi_n(\omega) - \xi(\omega)| < \varepsilon,\ \forall n \ge N. \end{aligned}$$

Тому імплікація $P1 \Longrightarrow P$ випливає з рівняння $\mathbf{P}(\overline{O}) = 0$, скінченності $N(\varepsilon, \omega)$ та включення

$$\{|\xi_n - \xi| \ge \varepsilon\} \subset \overline{O} \cup (O \cap \{n \le N(\varepsilon, \omega)\}) \downarrow \overline{O} \cup \varnothing,\ n \to \infty.$$

Друга імплікація $P \Longleftarrow L_1$ є наслідком *загальної нерівності Чебишева* з функцією $g(x) = x$ та випадковою величиною $|\xi_n - \xi|$:

$$\mathbf{P}(|\xi_n - \xi| \ge \varepsilon) \le \mathbf{E}\,|\xi_n - \xi|\ \ / \ \varepsilon \to 0,\ n \to \infty.$$

Нарешті, імплікація $L_1 \Longleftarrow L_q$ є наслідком *нерівності Ляпунова*

$$\mathbf{E}\,|\xi_n - \xi| \le (\mathbf{E}\,|\xi_n - \xi|^q)^{1/q}\ \boxdot$$

Зауваження. *Теорема Лебега про мажоровану збіжність*, *теорема Лебега про монотонну збіжність*, та теорема *про властивості збіжності за ймовірністю* дають достатні умови для справедливості інших імплікацій:

$$\boxed{P1} \Longrightarrow \boxed{L_1}, \qquad \boxed{P} \Longrightarrow \boxed{L_1}.$$

1.20.3. Критерій збіжності майже напевне

За означенням послідовність випадкових величин $(\xi_n,\ n \ge 1)$ *збігається майже напевне* до величини ξ, якщо подія

$$A = \{\ \exists \lim \xi_n = \xi\} = \cap_{k \ge 1} \cup_{m \ge 1} \cap_{n \ge m} \{|\xi_n - \xi| \le 1/k\}$$

має одиничну ймовірність: $\mathbf{P}(A) = 1$.

Теорема (про критерій збіжності майже напевне). *Послідовність* (ξ_n) *збігається майже напевне до величини* ξ *тоді й тільки тоді, коли*

$$\sup_{k\ge n}|\xi_k-\xi| \to^P 0,\ n\to\infty.$$

Доведення. Позначимо $A_{km}=\cap_{n\ge m}\{|\xi_n-\xi|\le 1/k\}$.

Оскільки ці події монотонно не зростають за k, і $\cup_{m\ge 1}A_{km}\downarrow A$ при $k\to\infty$, то за властивістю *неперервності ймовірності* $\mathbf{P}(A)=1$ тоді й тільки тоді, коли $\mathbf{P}(\cup_{m\ge 1}A_{km})=1$ для кожного k. Події A_{km} монотонно не спадають за m, тому остання рівність еквівалентна співвідношенню: $\lim_{m\to\infty}\mathbf{P}(A_{km})=1$ при кожному k. За означенням збіжності *за ймовірністю* останнє твердження еквівалентне умові теореми, оскільки

$$\mathbf{P}(A_{km})=\mathbf{P}(\sup_{n\ge m}|\xi_n-\xi|\le 1/k)=1-\mathbf{P}(\sup_{n\ge m}|\xi_n-\xi|>1/k)\ \boxdot$$

1.20.4. Збіжність та фундаментальність

Означення. *Послідовність випадкових величин* $(\xi_n, n\ge 1)$ *назвемо*
(а) фундаментальною майже напевне, *якщо*
$\mathbf{P}(\lim_{n,m\to\infty}|\xi_n-\xi_m|=0)=1,$
(б) фундаментальною за ймовірністю, *якщо*
$\lim_{n,m\to\infty}\mathbf{P}(|\xi_n-\xi_m|\ge\varepsilon)=0,\ \forall\varepsilon>0,$
(в) фундаментальною у середньому, *якщо*
$\lim_{n,m\to\infty}\mathbf{E}\,|\xi_n-\xi_m|=0.$

Теорема (про еквівалентність збіжності та фундаментальності). *Послідовність випадкових величин* $(\xi_n,\ n\ge 1)$ *збігається (а) майже напевне, (б) за ймовірністю, (в) у середньому тоді й тільки тоді, коли вона фундаментальна у відповідному розумінні.*

Доведення

(а) Оскільки множини тих ω, для яких послідовність $\xi_n(\omega)$ збігається, чи є фундаментальною, однакові, то їх імовірності одиничні одночасно.

(б) *Необхідність.* Нехай $\xi_n\to^P\xi$. Тоді

$$\mathbf{P}(|\xi_n-\xi_m|\ge\varepsilon)\le\mathbf{P}(|\xi_n-\xi|\ge\varepsilon/2)+\mathbf{P}(|\xi_m-\xi|\ge\varepsilon/2)\to 0,\ n,m\to\infty.$$

Достатність. З означення *фундаментальності за ймовірністю* побудуємо зростаючу послідовність $n_k\uparrow\infty$ таку, що

$$\mathbf{P}\left(\left|\xi_{n_{k+1}}-\xi_{n_k}\right|\ge 2^{-k}\right)\le 2^{-k}.$$

Тоді за *лемою Бореля-Кантеллі* $\mathbf{P}\left(\overline{\lim}_{k\to\infty}\left\{\left|\xi_{n_{k+1}}-\xi_{n_k}\right|\ge 2^{-k}\right\}\right)=0$, отже, ряд $\xi\equiv\xi_{n_1}+\sum_{k\ge 1}\left(\xi_{n_{k+1}}-\xi_{n_k}\right)$ збігається *майже напевне*, оскільки з імовірністю одиниця мажорується рядом з 2^{-k}, починаючи з деякого номера. Тому має місце збіжність частинних сум

$$\xi_{n_i} = \xi_{n_1} + \sum\nolimits_{k=1}^{i-1} \left(\xi_{n_{k+1}} - \xi_{n_k} \right) \to^{P1} \xi, \ i \to \infty.$$

З теореми *про співвідношення між різними видами збіжності* виводимо, що $\xi_{n_i} \to^P \xi$. Звідси

$$\mathbf{P}(|\xi_n - \xi| \geq \varepsilon) \leq \mathbf{P}(\left|\xi_n - \xi_{n_i}\right| \geq \varepsilon/2) + \mathbf{P}(\left|\xi_{n_i} - \xi\right| \geq \varepsilon/2) \leq \delta,$$

якщо спочатку обрати номер i, починаючи з якого другий доданок менший за $\delta/2$, а потім визначити з умови фундаментальності номери n, n_i так, щоб і перший не перевищував $\delta/2$.

(в) *Необхідність* випливає з нерівності

$$\mathbf{E}\,|\xi_n - \xi_m| \leq \mathbf{E}\,|\xi_n - \xi| + \mathbf{E}\,|\xi_n - \xi|\,.$$

Достатність. За *загальною нерівністю Чебишева* з фундаментальності у середньому виводимо фундаментальність за ймовірністю:

$$\mathbf{P}(|\xi_n - \xi_m| \geq \varepsilon) \leq \mathbf{E}\,|\xi_n - \xi_m|\,/\varepsilon \to 0, \ n, m \to \infty.$$

Тому $\xi_{n_i} \to^P \xi$. для деякої підпослідовності $n_i \to \infty$, як доведено у (б). За нерівністю Фату

$$\mathbf{E}\,|\xi_n - \xi| = \mathbf{E}\underline{\lim}_{i\to\infty} \left|\xi_n - \xi_{n_i}\right| \leq \underline{\lim}_{i\to\infty}\mathbf{E} \left|\xi_n - \xi_{n_i}\right| \leq \delta,$$

починаючи з деякого номера, за умовою фундаментальності $\boxdot$

Вправи

(1, теорема Рісса) Довести, що з $\xi_n \to^P \xi$ випливає збіжність $\xi_{n_k} \to^{P1} \xi$ для деякої підпослідовності $n_k \to \infty$.

(2) Збіжність $\xi_n \to^P \xi$ має місце тоді й тільки тоді, коли кожна підпослідовність $(\xi_{n_k}, \ k \geq 1)$ містить підпідпослідовність, що збігається до ξ м.н.

(3) Побудувати приклад послідовності випадкових величин $(\xi_n), \xi$ таких, що $\overline{\lim}_{n\to\infty}\xi_n = \infty$ і $\underline{\lim}_{n\to\infty}\xi_n = -\infty$ м.н., та одночасно $\xi_n \to^P \xi$.

(4) У дискретному ймовірнісному просторі збіжність майже напевне еквівалентна збіжності за ймовірністю.

(5, теорема Єгорова) Якщо $\xi_n \to^P \xi, \ n \to \infty$, то для кожного $\varepsilon > 0$ знайдеться подія A_ε така, що $\mathbf{P}(A_\varepsilon) \geq 1 - \varepsilon$ та збіжність $\xi_n(\omega) \to \xi(\omega)$ при $n \to \infty$ є рівномірною за $\omega \in A_\varepsilon$.

(6) Величини $(\xi_n, n \geq 1)$ невід'ємні та $\xi_n \to^P \xi, n \to \infty$, і $\mathbf{E}\xi_n \to \mathbf{E}\xi$. Довести, що $\mathbf{E}\,|\xi_n - \xi| \to 0, n \to \infty$. Узагальнити це твердження для збіжності у просторі $L_p, p > 1$.

(7) Випадкові величини $(\xi_n,\ n \geq 1)$ незалежні однаково розподілені, причому $\mathbf{P}(|\xi_1| > c) > 0$ для всіх c. Довести, що існує така послідовність сталих c_n, що $c_n\xi_n \to^P 0, n \to \infty$, та $c_n\xi_n \nrightarrow^{P1} 0$.

(8) Якщо $\xi_n \to^{L_2} \xi$, та $\eta_n \to^{L_2} \eta$, то $\mathbf{E}\xi_n\eta_n \to \mathbf{E}\xi\eta, n \to \infty$.

(9) Якщо $(\xi_n - \xi)^2 \to^P 0$, то $\xi_n^2 \to^P \xi^2, n \to \infty$.

(10) Послідовність $(\xi_n,\ n \geq 1)$ обмежена за ймовірністю, $\eta_n \to^P 0, n \to \infty$, а g – дійсна неперервна функція. Довести, що $g(\xi_n + \eta_n) - g(\xi_n) \to^P 0, n \to \infty$.

(11) Випадкові величини $(\xi_n,\ n \geq 1)$ незалежні, центровані: $\mathbf{E}\xi_n = 0$ та квадратично інтегровні: $\mathbf{E}\xi_n^2 < \infty$. Визначимо $\eta_n = n^{-1}\sum_{k=1}^n \xi_k$. (а) Довести, що $\eta_n \to^{P1} 0, n \to \infty$, тоді й тільки тоді, коли $\eta_{2^n} \to^{P1} 0, n \to \infty$. (б) У (а) збіжність м.н. можна замінити на середньоквадратичну збіжність.

1.21. Закон великих чисел

Означення. *Для послідовності випадкових величин* $(\xi_n,\ n \geq 1)$ *виконується* закон великих чисел*, якщо частинні суми* $S_n = \xi_1 + ... + \ \xi_n$ *задовольняють граничне співвідношення*

$$(S_n - \mathbf{E}S_n)\ /\ n \to^P 0, \text{ при } n \to \infty.$$

У випадку *однаково розподілених* випадкових величин $\mathbf{E}S_n = n\mathbf{E}\xi_1$, тому твердження закону великих чисел еквівалентне збіжності *за ймовірністю* середніх арифметичних

$$S_n\ /\ n \to^P \mathbf{E}\xi_1, \text{ при } n \to \infty.$$

1.21.1. Теорема Чебишева про закон великих чисел

Теорема (теорема Чебишева про закон великих чисел).
Нехай випадкові величини $(\xi_n,\ n \geq 1)$ *попарно незалежні і*

$$\mathbf{D}(\xi_n) = o(n),\ n \to \infty.$$

Тоді для послідовності $(\xi_n,\ n \geq 1)$ *справджується закон великих чисел.*

Доведення. З умови теореми випливає більш сильне твердження – збіжність центрованих та нормованих сум S_n до нуля *у середньому квадратичному*. Дійсно, за теоремою *про дисперсію суми незалежних величин* та за теоремою Штольця про границю відношення послідовностей при $n \to \infty$:

$$\lim \mathbf{E}((S_n - \mathbf{E}S_n)\ /\ n)^2 = \lim \mathbf{D}S_n\ /\ n^2 =$$
$$\lim \sum_{k=1}^n \mathbf{D}\xi_k\ /\ n^2 = \lim \mathbf{D}\xi_n/(n^2 - (n-1)^2) = \lim \mathbf{D}\xi_n/(2n - 1) = 0.$$

Завершує доведення теореми застосування імплікації $L_2 \Longrightarrow \mathbf{P}$, що базується на *нерівності Чебишева для дисперсій*

$$\mathbf{P}(|S_n - \mathbf{E}S_n| \geq n\varepsilon) \leq \mathbf{D}S_n \;/\; n^2\varepsilon^2 \to 0 \;\boxdot$$

Зауваження. Остання умова теореми на дисперсії доданків виконується, зокрема, якщо дисперсії $\mathbf{D}\xi_n$ обмежені. Наприклад, для *однаково розподілених* величин дисперсії не залежать від n, і закон великих чисел виконується.

Вправи

(1) Нехай $(\xi_n,\ n \geq 1)$ незалежні додатні випадкові величини. Знайти умови збіжності за ймовірністю величин $\sqrt[n]{\prod_{k=1}^{n} \xi_k}$.

(2) Для $g \in C[0,1]$ обчислити $\lim_{n\to\infty} \int_0^1 .. \int_0^1 g\left(\sqrt[n]{x_1..x_n}\right) dx_1..dx_n$.

(3) Величини $(\xi_n,\ n \geq 1)$ незалежні та $\mathbf{P}(\xi_n = \pm n^\alpha) = 1/2$. Для яких α має місце закон великих чисел ?

(4 – теорема Бернштейна про закон великих чисел) Нехай $(\xi_n, n \geq 1)$ – така послідовність випадкових величин, що $\sup \mathbf{D}\xi_n < \infty$ та коефіцієнти кореляції задовольняють умову $\lim_{|n|\to\infty} \sup_{k\geq 1} \rho(\xi_{n+k}, \xi_k) = 0$. Тоді (ξ_n) задовольняє закон великих чисел.

(5 – теорема Хінчина) Довести, що попарно незалежні однаково розподілені інтегровні випадкові величини задовольняють закон великих чисел.

(6) На прикладі переконатися, що в теоремі Чебишева про закон великих чисел умову на дисперсії не можна замінити на умову $\sup_{n\geq 1} \mathbf{E}\,|\xi_n| < \infty$, навіть для незалежних величин.

(7) Випадкові величини $(\varepsilon_n,\ n \geq 0)$ незалежні однаково розподілені та квадратично інтегровні. Довести, що для величин $\xi_n = \sum_{k=0}^{m} c_k \varepsilon_{n-k}$ виконується закон великих чисел, де c_k – довільні сталі.

(8, теорема Феллера) Нехай $(\xi_n, n \geq 1)$ незалежні однаково розподілені величини. Для того, щоб $(\xi_1 + ... + \xi_n - m_n)/n \to^P 0, n \to \infty$, з числовою послідовністю m_n, необхідно і достатньо, щоб $\mathbf{P}(|\xi_1| > x) = o(1/x), x \to \infty$.

Розглянемо деякі застосування ЗВЧ.

1.21.2. Теорема Бернуллі

Теорема (теорема Бернуллі про асимптотику відносної частоти). *Нехай ν_n – кількість успіхів у перших n випробуваннях нескінченної послідовності випробувань Бернуллі з імовірністю успіху p. Тоді*

$$\nu_n \;/\; n \to^P p,\ n \to \infty.$$

Доведення виводимо з подання $\nu_n = \sum_{k=1}^{n} \chi_k$, де доданки χ_k– індикатори успіхів – незалежні і однаково розподілені випадкові величини,

$$\mathbf{E}\chi_k = p, \quad \mathbf{D}(\nu_n/n) = p(1-p)/n \to 0 \boxdot$$

1.21.3. Поліноми Бернштейна

На початку 20 ст. С.Н. Бернштейн запропонував конструктивне доведення знаменитої (але не конструктивної) теореми Вейєрштраса про наближення неперервної функції поліномами. Це доведення використовує основні ідеї доведення *закону великих чисел*.

Означення. Поліномом Бернштейна *порядку* n *для дійсної функції* f *називається поліном*

$$B_n(f,x) = \sum_{k=0}^{n} f\,(k/n)\ C_n^k\ x^k(1-x)^{n-k},\ x \in [0,1].$$

Теорема (теорема Бернштейна про рівномірне наближення неперервної функції). *Якщо* $f \in C[0,1]$, *то* $B_n(f,x) \to f(x)$ *при* $n \to \infty$ *рівномірно за* $x \in [0,1]$.

Доведення. Позначимо через $\nu_n(x)$ – кількість успіхів у послідовності з n *випробувань Бернуллі* з імовірністю успіху x. Тоді $\nu_n(x)$ має за означенням *біноміальний розподіл* з параметрами n, x, і за теоремою *про обчислення математичного сподівання функції від дискретної величини*

$$B_n(f,x) = \mathbf{E}f(\nu_n(x)/n).$$

Визначимо модуль неперервності та верхню межу

$$w_\delta(f) = \sup\ _{|x-y|\le\delta}\ |f(x) - f(y)|\,, \quad L(f) = \sup_{x\in[0,1]} |f(x)|\,.$$

Використаємо ідеї доведень пунктів (а) і (б) теореми *про властивості збіжності за ймовірністю*:

$$|f\,(\nu_n(x)/n) - f(x)| \le w_\delta(f) + 2\,L(f)\ \mathbb{1}\ _{\{|\nu_n(x)/n - x|>\delta\}},$$

звідки знаходимо оцінку:

$$|B_n(f,x) - f(x)| \le \mathbf{E}\,|f\,(\nu_n(x)/n) - f(x)| \le$$
$$w_\delta(f) + 2\,L(f)\ \mathbf{P}\,(|\nu_n(x)/n - x| > \delta) \le w_\delta(f) + 2\,L(f)\ \mathbf{D}\,(\nu_n(x)/n)\,/\delta^2 =$$
$$w_\delta(f) + 2\,L(f)\ x(1-x)/n\delta^2 \le w_\delta(f) + 2\,L(f)\ /\ 4n\delta^2 \le 2\,\varepsilon,$$

рівномірно за x, починаючи з деякого n. Тут сталу $\delta > 0$ обрано для наперед заданого довільного $\varepsilon > 0$ так, щоб $w_\delta(f) \le \varepsilon$ $\boxdot$

Зауваження. Швидкість збіжності в теоремі Бернштейна можна уточнювати залежно від додаткових властивостей функції f. Наприклад, для

$f \in C_1[0,1]$ маємо $w_\delta(f) \le D\delta$. Обираючи $\delta = 1/\sqrt[3]{n}$, дістанемо

$$|B_n(f,x) - f(x)| \le D\delta + E \ / \ n\delta^2 = O(1/\sqrt[3]{n}), n \to \infty,$$

рівномірно за $x \in [0,1]$.

1.21.4. Метод Монте-Карло

Закон великих чисел можна застосувати до іншого розділу прикладної математики – до наближеного обчислення інтегралів. В основі методу Монте-Карло лежить *закон великих чисел* та теорема *про обчислення математичного сподівання функції від випадкової величини*:

$$I \equiv \int_0^1 f(x)dx = \mathbf{E}f(\alpha),$$

де α – *рівномірно розподілена* на $[0,1]$ випадкова величина.

Теорема (про наближення інтегралу методом Монте-Карло). *Нехай функція f інтегровна в квадраті: $f \in L_2[0,1]$, та $(\alpha_k, k \ge 1)$ – послідовність незалежних рівномірно розподілених на $[0,1]$ випадкових величин. Визначимо величини $Y_k \equiv f(\alpha_k)$. Тоді має місце збіжність*

$$I_n \equiv (Y_1 + ... + \ Y_n) \ / \ n \to^P I, \ n \to \infty.$$

Доведення. Випадкові величини $(Y_k, \ k = \overline{1,n})$ є *незалежними в сукупності*, *однаково розподіленими* за теоремою *про перетворення незалежних величин*, та мають скінченні математичні сподівання $\mathbf{E}f(\alpha_k) = I$ і дисперсії $s^2 = \int_0^1 \ f^2(x)dx - I^2$. Тому для них виконується *закон великих чисел*. Більше того, можна оцінити швидкість збіжності в середньому квадратичному $\mathbf{E}(I - I_n)^2 = \mathbf{D}(I_n) = s^2/n = O(1/n)$ $\boxdot$

На відміну від схем наближеного обчислення інтегралів із регулярно розміщеними вузлами метод Монте-Карло є збіжним і для розривних функцій.

Вправи

(1) Довести, що в теоремі Бернуллі $\mathbf{P}(|\nu_n/n - p| \ge \varepsilon) \le 2\exp(-\alpha\varepsilon^2 n)$ для всіх $n \ge 1, \varepsilon > 0$ та деякого $\alpha > 0$.

(2) Для $f \in C_1([0,1])$ довести рівномірну збіжність $dB_n(f,x)/dx \to f'(x)$.

(3) Степінь поліному $B_n(f,x)$ не більший за степінь поліному f.

(4) Узагальнити теорему Бернштейна на функції з класу $C([0,1]^n)$.

(5) Позначимо $I_n = \int_0^1 ... \int_0^1 g((x_1 + ... + x_n)/n) dx_1 ... dx_n$ для даної функції $g \in C[0,1]$. (а) Довести, що $\lim_{n\to\infty} I_n = I \equiv g(1/2)$. (б) У припущенні, що $g \in C_2[0,1]$, обчислити $\lim_{n\to\infty} n(I_n - I \ )$.

(6) Довести, що для $f \in C_b(\mathbb{R}_+)$ має місце збіжність

$\exp(-nx)\sum_{k=0}^{\infty} f\,(k/n)\,(nx)^k/\; k! \to f(x),\;\; n \to \infty,$

рівномірно на кожному обмеженому інтервалі. Вказівка: ліва частина дорівнює $\mathbf{E}f((\xi_1 + ... + \xi_n)/n)$, де ξ_k незалежні та мають розподіл Пуассона $\Pi(x)$.

(7) Довести, що для $f \in C_b(\mathbb{R}_+)$ мають місце збіжності

$\sum_{k\geq 0} f(x + k/n)(cn)^k \exp(-cn)/k! \to f(x + c),\;\; n \to \infty,\;\; \forall x \in \mathbb{R}_+,$

$\sum_{k\geq 0} f(k/(n + 1))C^k_{n+k}x^k(1 + x)^{-n-k-1} \to f(x),\;\; n \to \infty,\;\; \forall x \in \mathbb{R}_+.$

(8) Перетворенням Лапласа неперервної обмеженої функції f на $\mathbb{R}_+$ називається $\varphi(p) \equiv \int_0^\infty \exp(-pt)f(t)dt$ для $p \in \mathbb{R}_+$. Довести формулу обернення Уідлера: $f(s) = \lim_{n\to\infty} \frac{(-1)^{n-1}}{(n-1)!}\left(\frac{n}{s}\right)^n \varphi^{(n-1)}(\frac{n}{s})$, де $\varphi^{(k)}(p)$ – k-та похідна функції φ. *Вказівка*: величина під знаком границі дорівнює $\mathbf{E}f(s(\xi_1 + ... + \xi_n)/n)$, де величини ξ_k незалежні, та $\xi_k \simeq Exp(1)$.

(9) Величини $(\xi_n, n \geq 1)$ незалежні, однаково розподілені, $\mathbf{E}\xi_1 = \mu$, $\mathbf{D}\xi_1 = \sigma^2$, $\mathbf{P}(\xi_1 = 0) = 0$, та $S^{(i)}_n = \sum_{k=1}^n \xi^i_k$. Довести, що $S^{(1)}_n/S^{(2)}_n \to^P \mu/(\mu^2 + \sigma^2)$.

1.22. Збіжність з імовірністю 1 та її властивості

Послідовність випадкових величин $(\xi_n\;, n \geq 1)$ збігається *з імовірністю 1* до величини ξ, якщо одиничну ймовірність має подія

$$A = \{\; \exists \lim_{n\to\infty} \xi_n = \xi\} = \cap_{k\geq 1}\; \underline{\lim}_{n\to\infty}\{|\xi_n - \xi| \leq 1/k\} = \overline{\cup_{k\geq 1}\overline{\lim}_{n\to\infty}\{|\xi_n - \xi| > 1/k\}}$$

Збіжність майже напевне позначається як $\xi_n \to^{P1} \xi$.

Тісний зв'язок між границями випадкових величин та послідовностей випадкових подій відображає також таке твердження.

Зауваження (про верхню границю величин та подій). *Для кожного ε мають місце включення*

$$\{\overline{\lim}_{n\to\infty}\xi_n > \varepsilon\} \subset \overline{\lim}_{n\to\infty}\{\xi_n > \varepsilon\},\quad \overline{\lim}_{n\to\infty}\{\xi_n \geq \varepsilon\} \subset \{\overline{\lim}_{n\to\infty}\xi_n \geq \varepsilon\}.$$

Дійсно, якщо верхня границя числової послідовності більша за ε, то ця послідовність нескінченно часто перевищуватиме рівень ε.

З урахуванням зауваження важливе значення для доведення збіжності майже напевне має наступне твердження.

1.22.1. Лема Бореля – Кантеллі

Теорема (лема Бореля – Кантеллі). *Нехай* $(A_n,\ n \geq 1)$ *– довільна послідовність випадкових подій.*

(а) Якщо ряд $\sum_{n\geq 1} \mathbf{P}(A_n)$ *збігається, то* $\mathbf{P}(\overline{\lim} A_n) = 0$.

(б) Якщо події $(A_n,\ n \geq 1)$ *незалежні в сукупності і ряд* $\sum_{n\geq 1} \mathbf{P}(A_n)$ *розбігається, то* $\mathbf{P}(\overline{\lim} A_n) = 1$.

Доведення. Позначимо $B_m = \cup_{n\geq m} A_n$. Тоді $B_m \downarrow \cap_{\ m\geq 1} B_m = \overline{\lim} A_n$. За властивістю *неперервності ймовірності* $\mathbf{P}(\overline{\lim} A_n) = \lim_{m\to\infty} \mathbf{P}(B_m)$.

(а) За означенням B_m та властивістю *напівадитивності* ймовірності $\mathbf{P}(B_m) \leq \sum_{n\geq m} \mathbf{P}(A_n) \to 0, m \to \infty$, як залишок збіжного ряду.

(б) Для даного $n > 1$ розглянемо клас подій, що не залежать від A_n:

$$\Lambda(A_n) \equiv \{A \in \mathfrak{F}\ : \mathbf{P}(A_n \cap A) = \mathbf{P}(A_n) P(A)\}.$$

За *формулою включення-виключення*

$$\mathbf{P}(A_n \cap \cup_{i=1}^{n-1} A_i) = \mathbf{P}(\cup_{i=1}^{n-1}(A_i \cap A_n)) =$$
$$\sum_{k=1}^{n-1}(-1)^{k-1} \sum_{i_1<\ldots<i_k<n} \mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_k} \cap A_n) =$$
$$\sum_{k=1}^{n-1}(-1)^{k-1} \sum_{i_1<\ldots<i_k<n} \mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) \mathbf{P}(A_n) = \mathbf{P}(\cup_{i=1}^{n-1} A_i) \mathbf{P}(A_n).$$

Отже, $\cup_{i=1}^{n-1} A_i \in \Lambda(A_n)$. Тому за теоремою *про перетворення незалежних подій* $\overline{\cup_{i=1}^{n-1} A_i} \in \Lambda(A_n)$ і $\cap_{i=1}^{n-1} \overline{A_i} \in \Lambda(\overline{A_n})$. Звідси за індукцією виводимо, що $\mathbf{P}\left(\bigcap_{n=1}^{m} \overline{A_n}\right) = \prod_{n=1}^{m} \mathbf{P}(\overline{A_n})$.

Обчислимо з урахуванням останнього твердження та властивості *неперервності ймовірності* :

$$\mathbf{P}(B_m) = \lim_{N\to\infty} \mathbf{P}\left(\bigcup_{n=m}^{N} A_n\right) = 1 - \lim_{N\to\infty} \mathbf{P}\left(\bigcap_{n=m}^{N} \overline{A_n}\right) =$$
$$1 - \lim_{N\to\infty} \prod_{n=m}^{N} \mathbf{P}(\overline{A_n}) = 1\ - \lim_{N\to\infty} \prod_{n=m}^{N}(1 - \mathbf{P}(A_n)) \geq$$
$$1 - \lim_{N\to\infty} \exp\left(-\sum_{n=m}^{N} \mathbf{P}(A_n)\right) = 1 - \exp\left(-\sum_{n=m}^{\infty} \mathbf{P}(A_n)\right) = 1,$$

оскільки справедлива елементарна нерівність $1 - x \leq \exp(-x)$, а залишок розбіжного ряду розбігається $\boxdot$

Вправи

(1) Якщо ряд $\sum_{n\geq 1} \mathbf{P}(A_n \cap A)$: (а) збігається для деякої події A, то $\mathbf{P}(\overline{\lim} A_n) \leq 1 - \mathbf{P}(A)$, (б) розбігається для всіх A з $\mathbf{P}(A) > 0$, то $\mathbf{P}(\overline{\lim} A_n) = 1$.

(2) Побудувати приклад залежних подій A_n таких, що $\sum_{n\geq 1} \mathbf{P}(A_n) = \infty$ та $\mathbf{P}(\overline{\lim} A_n) = 0$.

(3) Події (A_n) незалежні у сукупності та невироджені: $\mathbf{P}(A_n) = p_n \in (0,1)$. (а) Довести, що $\mathbf{P}(\overline{\lim}_{n\to\infty} A_n) = 1$ тоді й тільки тоді, коли $\mathbf{P}(\cup_{n\ge 1} A_n) = 1$. (б) Остання умова еквівалентна $\sum_{n\ge 1} p_n = \infty$. (в) За таким припущенням знайти розподіл величини $\nu(\omega) = \inf(n : \omega \in A_n) < \infty$.

(4) Довести нерівність $\mathbf{P}(\cup_{k\le n} A_k) \sum_{i,j\le n} \mathbf{P}(A_i \cap A_j) \ge (\sum_{k\le n} \mathbf{P}(A_k))^2$. Вивести звідси, що твердження пункту (б) леми Бореля-Кантеллі виконується вже за припущення попарної незалежності подій A_n.

(5) Довести, що простір класів м.н. еквівалентних випадкових величин зі збіжністю м.н. не може бути метризований.

(6) Випадкові величини $(\xi_n, n \ge 1)$ незалежні та $\xi_n \to^P \xi, n \to \infty$. Довести, що $\xi = c$ м.н. для деякої сталої c.

(7) Величини $(\xi_n, n \ge 1)$ незалежні однаково розподілені, $\mathbf{E}\xi_1^+ = \mathbf{E}\xi_1^- \le \infty$. Довести, що $\mathbf{P}(\xi_n = o(n),\ n \to \infty) = 1$ тоді й тільки тоді, коли $\mathbf{E}\,|\xi_1| < \infty$.

(8) Випадкові величини $(\xi_n, n \ge 1)$ незалежні однаково розподілені і $\mathbf{E}\xi_1 = 0$. (а) Нехай $\mathbf{E}\,|\xi_1|^\alpha < \infty$ при $\alpha > 0$, тоді $\mathbf{P}(|\xi_n| = o(n^{1/\alpha}),\ n \to \infty) = 1$. (б) Якщо $\mathbf{E}\exp(\alpha\xi_1) < \infty$ для $\alpha > 0$, то $\mathbf{P}(\xi_n = o(\ln n),\ n \to \infty) = 1$.

(9) Випадкові величини $(\xi_n,\ n \ge 0)$ незалежні однаково експоненційно розподілені з параметром $\lambda = 1$. Довести, що (а) $\overline{\lim}_{n\to\infty} \xi_n / \ln n = 1$ м.н., та (б) $\underline{\lim}_{n\to\infty} \ln \xi_n / \ln n = -1$ м.н.

(10) Випадкові величини $(\xi_n, n \ge 1)$ незалежні та рівномірно розподілені на відрізку $[0,1]$. Довести, що послідовність ξ_n м.н. всюди щільна на $[0,1]$.

1.22.2. Нерівність Колмогорова

Нерівність Колмогорова є засобом доведення *збіжності майже напевне*, так само, як нерівність Чебишева використовується для встановлення збіжності *за ймовірністю*.

Теорема (нерівність Колмогорова). *Нехай $(\xi_k, k = \overline{1,n})$ – незалежні у сукупності випадкові величини такі, що $\mathbf{E}\xi_k = 0$, $\mathbf{E}\xi_k^2 < \infty$, а $S_k = \xi_1 + ... + \xi_k$. Тоді для кожного $\varepsilon > 0$ виконується нерівність*

$$\mathbf{P}\left(\max_{1\le k\le n} |S_k| \ge \varepsilon\right) \le \mathbf{D}S_n \,/\, \varepsilon^2.$$

Доведення. Розглянемо випадкові події

$$A = \{\max_{1\le k\le n} |S_k| \ge \varepsilon\}, \quad A_k = \{|S_1| < \varepsilon, ..., |S_{k-1}| < \varepsilon, |S_k| \ge \varepsilon\}.$$

Тоді $A_i \cap A_j = \varnothing$, $i \ne j$, та $A = \cup_{k=1}^n A_k$, звідки $1\!\!1_A = \sum_{k=1}^n 1\!\!1_{A_k}$. Відзначимо, що для всіх елементарних подій справедлива нерівність $|S_k|\,1\!\!1_{A_k} \ge \varepsilon 1\!\!1_{A_k}$. Тому за *монотонністю* математичного сподівання

$$\mathbf{D}S_n = \mathbf{E}S_n^2 \ge \mathbf{E}S_n^2 1\!\!1_A = \sum_{k=1}^n \mathbf{E}S_n^2 1\!\!1_{A_k} =$$

$$\sum\nolimits_{k=1}^{n}\mathbf{E}(S_k+S_n-S_k)^2\ 1\!\!\mathrm{I}_{A_k}=\sum\nolimits_{k=1}^{n}\mathbf{E}S_k^2 1\!\!\mathrm{I}_{A_k}+$$
$$\sum\nolimits_{k=1}^{n}\mathbf{E}(S_n-S_k)^2\ 1\!\!\mathrm{I}_{A_k}+2\sum\nolimits_{k=1}^{n}\mathbf{E}S_k 1\!\!\mathrm{I}_{A_k}\ (S_n-S_k)\geq$$
$$\sum\nolimits_{k=1}^{n}\mathbf{E}\varepsilon^2 1\!\!\mathrm{I}_{A_k}+0+2\sum\nolimits_{k=1}^{n}\mathbf{E}S_k 1\!\!\mathrm{I}_{A_k}\ \mathbf{E}(S_n-S_k)=$$
$$\varepsilon^2\sum\nolimits_{k=1}^{n}\mathbf{P}(A_k)=\varepsilon^2\mathbf{P}(A),$$

де використані невід'ємність другого доданку, та незалежність множників $S_k 1\!\!\mathrm{I}_{A_k}$ і $S_n - S_k$. Дійсно, перший з них є функцією від перших k доданків $\xi_1, ..., \xi_k$, а другий дорівнює $\xi_{k+1} + ... +\ \xi_n$ і містить лише решту доданків, тому незалежність випливає з теореми *про векторні перетворення незалежних величин*. Крім того, другий множник є центрованим за умовою: $\mathbf{E}(S_n - S_k) = \mathbf{E}S_n - \mathbf{E}S_k = 0$, отже, математичне сподівання добутку, що дорівнює добуткові за теоремою *про математичне сподівання добутку незалежних величин*, є нульовим ⊡

Вправи

(1) Нехай g – неперервна неспадна опукла донизу функція, а послідовність (ξ_k) задовольняє умови теореми про нерівність Колмогорова. Тоді ліва частина цієї нерівності не перевищує $\mathbf{E}g(|S_n|)/g(\varepsilon)$.

(2) Випадкові величини $(\xi_k, k = \overline{1,n})$ незалежні, $\mathbf{P}(\xi_k = \pm 1) = 1/2$, $S_n = \xi_1 + ... + \xi_n$. Довести, що: (a) $\mathbf{E}\exp(tS_k) \leq \exp(kt^2/2)$ при $t \in \mathbb{R}$, (б) $\mathbf{P}(\max_{1\leq k\leq n}|S_k| \geq nt) \leq 2\exp(-nt^2/2)$ для $t > 0$.

(3) Випадкові величини $(\xi_k, k = \overline{1,n})$ незалежні, $\mathbf{E}\xi_k = 0$, $\mathbf{E}\xi_k^2 < \infty$, а $S_k = \xi_1 + ... + \xi_k$. Для довільних сталих $0 < a_1 \leq a_2 \leq ... \leq a_n$ довести нерівність $\mathbf{P}(|S_k| \leq a_k, \forall k = \overline{1,n}) \geq 1 - \sum_{k=1}^{n}\mathbf{E}\xi_k^2/a_k^2$.

1.23. Посилений закон великих чисел

Означення. *Для послідовності випадкових величин* $(\xi_n,\ n \geq 1)$ *виконується* посилений закон великих чисел*, якщо послідовність їх частинних сум* $S_n \equiv \xi_1 + ... + \xi_n$ *задовольняє граничне співвідношення*

$$(S_n - \mathbf{E}S_n)\ /\ n\ \ \to^{P1} 0,\ \text{ при } n \to \infty.$$

Як і для *закону великих чисел*, у випадку *однаково розподілених* випадкових величин це співвідношення еквівалентне збіжності середніх арифметичних

$$S_n\ /\ n \to^{P1} \mathbf{E}\xi_1,\ \text{ при } n \to \infty.$$

1.23.1. Теорема Колмогорова про посилений закон великих чисел

Теорема (теорема Колмогорова про посилений закон великих чисел). *Нехай* $(\xi_n\ , n \geq 1)$ *– послідовність незалежних у сукупності випадкових величин така, що* $\mathbf{D}\xi_n < \infty$ *та*

$$\sum\nolimits_{n\geq 1} \mathbf{D}\xi_n \ / \ n^2 < \infty.$$

Тоді для цієї послідовності має місце посилений закон великих чисел.

Доведення. Не обмежуючи загальності, вважатимемо, що $\mathbf{E}\xi_n = 0$. Дійсно, в загальному випадку досить розглянути послідовність незалежних величин $\xi_n' = \xi_n - \mathbf{E}\xi_n$ із тими ж дисперсіями. Очевидно, що твердження *посиленого закону великих чисел* для ξ_n і ξ_n' збігаються.

Для кожного k знайдеться єдиний номер n такий, що $2^{n-1} \leq k < 2^n$, причому $k \to \infty$ тоді й тільки тоді, коли $n \to \infty$. Розглянемо випадкові величини $\zeta_n = \max_{1\leq\ k\ \leq 2^n} |S_k|$. При такому виборі для всіх k справедлива нерівність $|S_k/k| \ \leq \zeta_n/2^{n-1}$. Тому для збіжності $S_k \ / \ k \to^{P1} 0, k \to \infty$, достатньо, щоб $\zeta_n \ / \ 2^n \to^{P1} 0,\ n \to \infty$.

Для заданого $\varepsilon > 0$ розглянемо випадкові події $A_n(\varepsilon) = \{\zeta_n/2^n \geq \varepsilon\}$.

За *нерівністю Колмогорова*

$$\sum\nolimits_{n\geq 1} \mathbf{P}(A_n(\varepsilon)) = \sum\nolimits_{n\geq 1} \mathbf{P}\left(\max\nolimits_{1\leq k\leq 2^n} |S_k| \geq \varepsilon 2^n\right) \leq \sum\nolimits_{n\geq 1} \mathbf{D}S_{2^n} \ /2^{2n}\varepsilon^2 =$$
$$\sum\nolimits_{n\geq 1} \sum\nolimits_{k\leq 2^n} \mathbf{D}\xi_k/2^{2n}\varepsilon^2 = \varepsilon^{-2} \sum\nolimits_{k\geq 1} \mathbf{D}\xi_k \sum\nolimits_{n:\ 2^n\geq k} 2^{-2n} \leq$$
$$2\,\varepsilon^{-2} \sum\nolimits_{k\geq 1} \mathbf{D}\xi_k/k^2 < \infty,$$

де враховано теорему *про дисперсію суми незалежних величин.* Отже, за *лемою Бореля – Кантеллі,* (а), $\mathbf{P}(\overline{\lim} A_n(\varepsilon)) = 0$, для кожного $\varepsilon > 0$. Згідно з зауваженням *про верхню границю величин та подій* має місце включення $\{\overline{\lim}(\zeta_n/2^n) > \varepsilon\} \subset \overline{\lim} A_n(\varepsilon)$, тому $\mathbf{P}(\{\overline{\lim}(\zeta_n/2^n) > \varepsilon\}) = 0$, при всіх $\varepsilon > 0$, звідки $\overline{\lim}_{n\to\infty}(\zeta_n/2^n) \leq \varepsilon$ *м.н.* для кожного $\varepsilon > 0$.

Тому остання верхня границя нульова, $\zeta_n/2^n \to^{P1} 0, n \to \infty$, та виконується *посилений закон великих чисел*, як показано вище $\boxdot$

1.23.2. Теорема Бореля

Теорема (теорема Бореля про асимптотику частоти успіху). *Нехай* ν_n *– кількість успіхів у перших* n *випробуваннях нескінченної послідовності випробувань Бернуллі з імовірністю успіху* p. *Тоді*

$$\nu_n \ / \ n \to^{P1} p,\ n \to \infty.$$

Доведення випливає з подання $\nu_n = \sum_{k=1}^n \chi_k$, де χ_k – індикатори успіхів, *незалежні в сукупності* та *однаково розподілені*, та $\mathbf{E}\chi_k = p$ $\boxdot$

Вправи

(1) Знайти необхідну та достатню умову посиленого закону великих чисел для незалежних випадкових величин $(\xi_n, n \geq 1)$ таких, що (а) $\xi_n \simeq \Pi(\lambda_n)$, (б) $\xi_n \simeq Exp(\lambda_n)$, (в) $\xi_n \simeq N(0, \sigma_n^2)$, (г) $\xi_n \simeq B(n, p_n)$.

(2) Випадкові величини $(\xi_n, n \geq 3)$ незалежні, $\mathbf{P}(\xi_n = \pm n) = 1/n \ln n$, $\mathbf{P}(\xi_n = 0) = 1 - 2/n \ln n$, $S_n = 0 + \xi_3 + ... + \xi_n$. Довести, що $S_n/n \to^P 0, n \to \infty$, та $\mathbf{P}(S_n/n \to 0) = 0$.

(3) Нехай $\xi_n \to 0, n \to \infty$, м.н. Довести, що існує борелєва додатна функція g така, що $\sum_{n \geq 1} g(\xi_n) < \infty$ м.н. Це твердження не виконується, якщо збіжність м.н. замінити на збіжність у середньому.

(4) Для випадкових величин (ξ_n) ряд $\sum_{n \geq 1} \xi_n / n$ збігається м.н. і $\mathbf{E}\xi_n = 0$. Довести, що ці величини задовольняють посилений закон великих чисел.

(5) Величини (ξ_n) незалежні, задовольняють умови теореми Колмогорова про посилений закон великих чисел та $\mathbf{E}\xi_n = 0$. Довести, що знайдуться незалежні величини (η_n) такі, що: (а) $\mathbf{P}(\eta_n = \xi_n$ починаючи з деякого $n) = 1$, (б) для (η_n) має місце посилений закон великих чисел, (в) $\sum_{n \geq 1} \mathbf{D}\eta_n / n^2 = \infty$. Тому умова теореми Колмогорова про посилений закон великих чисел не є необхідною.

(6) Випадкові події (A_n) незалежні у сукупності, $S_n = \sum_{k=1}^n 1\!\!1_{A_k}$. Довести, що $S_n / \mathbf{E}S_n \to^{P1} 1, \ n \to \infty$.

(7) Випадкова величина ν_n збігається з кількістю успіхів у n випробуваннях Бернуллі з імовірністю успіху p. Довести при $x \in (0,1)$ таку тотожність

$\lim_{n\to\infty} n^{-1} \ln \mathbf{P}(\nu_n \geq nx) = -x \ln(x/p) - (1-x) \ln((1-x)/(1-p))$.

(8, теорема Чернова) Випадкові величини $(\xi_n, n \geq 1)$ незалежні однаково розподілені та прості, $\mathbf{E}\xi_1 < 0, \mathbf{P}(\xi_1 < 0) < 1$. Позначимо $\rho = \inf_{t \in \mathbb{R}} \exp(t\xi_1)$. Довести, що $\rho \in (0,1)$ та $\lim_{n\to\infty} n^{-1} \ln \mathbf{P}(\xi_1 + ... + \xi_n \geq 0) = \ln \rho$. *Вказівка*: послідовність під знаком логарифму напівмультиплікативна.

1.23.3. Критерій Колмогорова для однаково розподілених величин

Для однаково розподілених незалежних величин існує критерій ПЗВЧ.

Теорема (критерій Колмогорова посиленого закону великих чисел). *Нехай* $(\xi_n, \ n \geq 1)$ *– послідовність незалежних в сукупності однаково розподілених випадкових величин, а* $S_n = \xi_1 + ... + \ \xi_n$.

Для того, щоб існувала стала a *така, що* $S_n \ / \ n \to^{P1} a, \ n \to \infty$, *необхідно і достатньо, щоб* $\mathbf{E}\,|\xi_1| < \infty$. *У цьому випадку* $a = \mathbf{E}\xi_1$.

Доведення

Достатність. Позначимо $\xi_n' = \xi_n 1\!\!1_{\{|\xi_n| \le n\}}$, $\xi_n'' = \xi_n - \xi_n'$.

Нехай S_n', $S_n'' = S_n - S_n'$ означають суми відповідних величин. Доведемо, що три доданки в правій частині тотожності

$$(S_n - \mathbf{E}S_n)/n = (S_n' - \mathbf{E}S_n')/n + S_n'' \,/\, n - \mathbf{E}S_n'' \,/\, n \equiv a_n + b_n - c_n$$

збігаються майже напевне до нуля.

(a) Випадкові величини $(\xi_n' \ , n \ge 1)$ незалежні за теоремою *про перетворення незалежних величин*. Доведемо, що вони задовольняють умови *теореми Колмогорова про посилений закон великих чисел*. Для цього обчислимо з урахуванням однакової розподіленості величин $(\xi_n,\ n \ge 1)$

$$\sum\nolimits_{n\ge1} \mathbf{D}\xi_n'/n^2 \le \sum\nolimits_{n\ge1} \mathbf{E}(\xi_n')^2/n^2 = \sum\nolimits_{n\ge1} \mathbf{E}\xi_n^2 \ 1\!\!1_{\{|\xi_n| \le n\}} \,/n^2 =$$

$$\sum\nolimits_{n\ge1} n^{-2}\mathbf{E}\xi_1^2 \ 1\!\!1_{\{|\xi_1| \le n\}} = \sum\nolimits_{n\ge1} n^{-2} \sum\nolimits_{k=1}^n \mathbf{E}\xi_1^2 \ 1\!\!1_{\{k-1< |\xi_1| \le k\}} =$$

$$\sum\nolimits_{k\ge1} \mathbf{E}\xi_1^2 \ 1\!\!1_{\{k-1< |\xi_1| \le k\}} \sum\nolimits_{n\ge k} n^{-2} \le \sum\nolimits_{k\ge1} \mathbf{E}\xi_1^2 \ 1\!\!1_{\{k-1< |\xi_1| \le k\}} \ 2k^{-1} \le$$

$$2 \sum\nolimits_{k\ge1} \mathbf{E}\,|\xi_1| \ \ 1\!\!1_{\{k-1< |\xi_1| \le k\}} \le 2 \ \ \mathbf{E}\,|\xi_1| < \infty,$$

де використані нерівності $\xi_1^2 \,/\, k \le \ |\xi_1|$ на *попарно несумісних* випадкових подіях $\{k-1 < |\xi_1| \le k\}$. Отже, за вказаною теоремою Колмогорова має місце збіжність $(S_n' - \mathbf{E}S_n')/n \equiv a_n \to 0, n \to \infty$, *майже напевне*.

(b) Далі, знайдемо

$$\sum\nolimits_{n\ge1} \mathbf{P}(\xi_n'' \ne 0) = \sum\nolimits_{n\ge1} \mathbf{P}(|\xi_n| > n) = \sum\nolimits_{n\ge1} \mathbf{P}(|\xi_1| > n) =$$

$$\sum\nolimits_{n\ge1} \sum\nolimits_{k\ge n} \mathbf{E} 1\!\!1_{\{k<|\xi_1|\le k+1\}} =$$

$$\sum\nolimits_{k\ge1} \mathbf{E}\ k \ 1\!\!1_{\{k<|\xi_1|\le k+1\}} \le \sum\nolimits_{k\ge1} \mathbf{E}\,|\xi_1| \ \ 1\!\!1_{\{k<|\xi_1|\le k+1\}} \ \le \mathbf{E}\,|\xi_1| < \infty.$$

За *лемою Бореля – Кантеллі*, (a), $\mathbf{P}(\overline{\lim}\{\xi_n'' \ne 0\}) = 0$, звідки виводимо $\mathbf{P}(\underline{\lim}\{\xi_n'' = 0\}) = 1$, отже, $S_n'' \,/\, n \equiv b_n \to 0, n \to \infty$, *майже напевне*.

(c) Нарешті, згідно з теоремою Штольця та за умовою *однакової розподіленості*

$$\lim \mathbf{E}S_n'' \,/\, n = \lim \mathbf{E}\xi_n''/1 = \lim \mathbf{E}\xi_n 1\!\!1_{\{|\xi_n|>n\}} =$$

$$\lim \mathbf{E}\xi_1 1\!\!1_{\{|\xi_1|>n\}} = \mathbf{E} \lim \xi_1 1\!\!1_{\{|\xi_1|>n\}} = \mathbf{E}0 = 0$$

внаслідок *теореми Лебега про мажоровану збіжність* та за припущенням $\mathbf{E}\,|\xi_1| < \infty$. Отже, $c_n \to 0, n \to \infty$. Достатність доведена.

Необхідність. Зауважимо, що при $n \to \infty$

$$\lim \frac{\xi_n}{n} = \lim \left(\frac{S_n}{n}\right) - \lim \left(\frac{n-1}{n}\frac{S_{n-1}}{n-1}\right) = a - a = 0 \ \text{ майже напевне.}$$

Розглянемо незалежні події $A_n = \{|\xi_n| > n\}$. Унаслідок означення отримуємо $\overline{\lim} A_n \subset \{\overline{\lim}(\xi_n/n) \neq 0\}$, та $\mathbf{P}(\overline{\lim} A_n) = 0$ і за *лемою Бореля – Кантеллі,*(б), наступний ряд має збігатися:

$$\infty > \sum_{n\geq1} \mathbf{P}(A_n) = \sum_{n\geq1} \mathbf{P}(|\xi_n| > n) = \sum_{n\geq1} \mathbf{P}(|\xi_1| > n) =$$
$$\sum_{n\geq1} \sum_{k\geq n} \mathbf{P}(k < |\xi_1| \leq k+1) =$$
$$\sum_{k\geq1} \mathbf{E}\; k\; 1\!\!\mathrm{I}_{\{k<|\xi_1|\leq k+1\}} \geq \sum_{k\geq1} \mathbf{E}(|\xi_1| - 1)\; 1\!\!\mathrm{I}_{\{k<|\xi_1|\leq k+1\}} \geq \mathbf{E}\,|\xi_1| - 2,$$

де використані *однакова розподіленість* випадкових величин ξ_n, та очевидна нерівність $k\; 1\!\!\mathrm{I}_{\{k<|\xi_1|\leq k+1\}} \geq (|\xi_1| - 1)\; 1\!\!\mathrm{I}_{\{k<|\xi_1|\leq k+1\}}$, що виконується для всіх ω. Отже, $\mathbf{E}\,|\xi_1| < \infty$ і внаслідок (а) отримуємо $a = \mathbf{E}\xi_1$ $\boxdot$

Вправи

(1, Хаусдорф) Нехай $(\xi_n,\ n \geq 1)$ послідовність незалежних однаково розподілених випадкових величин, $\mathbf{E}\xi_1 = \mu,\ \mathbf{E}\xi_1^2 < \infty,\ \ S_n = \xi_1 + ... +\ \xi_n$. Тоді $S_n - n\mu = o(n^{1/2+\delta}), n \to \infty$, майже напевне, для кожного $\delta > 0$.

(2) Нехай $(\xi_n,\ n \geq 1)$ послідовність незалежних однаково розподілених випадкових величин, $\mathbf{E}\xi_1^+ = \mathbf{E}\xi_1^-$. Для цієї послідовності виконується закон великих чисел: $(\xi_1 + ... + \xi_n)/n \to^P 0$ тоді й тільки тоді, коли $\mathbf{P}(|\xi_1| > n) = o(1)$ та $\mathbf{E}\xi_1 1\!\!\mathrm{I}_{\{|\xi_1|\leq n\}} = o(1), n \to \infty$.

(3) Нехай $(\xi_n,\ n \geq 1)$ послідовність незалежних однаково розподілених величин з $\mathbf{E}\xi_1 = \infty$, $S_n = \xi_1 + ... +\ \xi_n$. Довести, що (а) $S_n/n \to \infty,\ n \to \infty$, м.н. (б) $\overline{\lim}_{n\to\infty} |S_n/n - c_n| = \infty$ м.н. для довільної послідовності $c_n \in \mathbb{R}$.

(4) Випадкові величини $(\xi_n,\ n \geq 1)$ незалежні однаково розподілені, а $\alpha \in (0, 2)$. Довести, що послідовність $n^{-1/\alpha} \sum_{k=1}^n \xi_k$ збігається м.н. тоді й тільки тоді, коли $\mathbf{E}\,|\xi_1|^\alpha < \infty$ та у випадку, коли $\alpha \geq 1$, $\mathbf{E}\xi_1 = 0$. При $\alpha = 1$ границя дорівнює $\mathbf{E}\xi_1$ та 0 у інших випадках.

(5) Випадкові величини $(\xi_n,\ n \geq 1)$ незалежні, однаково розподілені та інтегровні, а $S_n = \xi_1 + ... +\ \xi_n$. Довести, що $\mathbf{E}\,|S_n/n - \mathbf{E}\xi_1| \to 0, n \to \infty$.

1.24. Збіжність функцій в основному

Означення. *Визначимо клас* узагальнених функцій розподілу

$$\mathfrak{M}_{01} = \{G : \mathbb{R} \to [0, 1],\ G \uparrow,\ G(x - 0) = G(x),\ \forall x \in \mathbb{R}\}.$$

Теорема (про характеризацію функцій розподілу в класі узагальнених). *Функція* $G \in \mathfrak{M}_{01}$ *є функцією розподілу тоді й тільки тоді, коли* $0 = G(-\infty)$, $G(\infty) = 1$.

Доведення очевидне: для неспадної функції умова $G(\mathbb{R}) \subset [0,1]$ еквівалентна $0 \le G(-\infty)$, $G(\infty) \le 1$, звідки заміною нерівностей на рівності отримуємо.умову *нормованості* функції розподілу $\boxdot$

Означення. *Нехай* G, $(G_n, n \ge 1) \subset \mathfrak{M}_{01}$ *– узагальнені функції розподілу. Визначимо, що має місце збіжність* **в основному:** $G_n \to^O G$ *при* $n \to \infty$, *якщо* $G_n(x) \to G(x)$ *для кожної точки* x, *що є точкою неперервності правої частини, тобто* $G(x+0) = G(x)$.

Вправа. Множина точок неперервності функції $G \in \mathfrak{M}_{01}$ всюди щільна в $\mathbb{R}$, оскільки її доповнення не більш ніж зліченне.

Теорема (про збіжність в основному та на щільній множині). *Нехай* $D \subset \mathbb{R}$ *всюди щільна множина.*

(а) Якщо G, $(G_n,\ n \ge 1) \subset \mathfrak{M}_{01}$ *деякі узагальнені функції розподілу та* $G_n(x) \to G(x)$ *для всіх* $x \in D$, *то* $G_n \to^O G,\ n \to \infty$.

(б) Якщо $(G_n,\ n \ge 1) \subset \mathfrak{M}_{01}$ *деякі узагальнені функції розподілу і для всіх* $x \in D$ *існує* $\lim_{n\to\infty} G_n(x)$, *то для деякої узагальненої функції розподілу* $G \in \mathfrak{M}_{01}$ *має місце збіжність в основному* $G_n \to^O G,\ n \to \infty$.

Доведення

(а) Нехай x – точка неперервності $G(x)$. Оскільки множина D щільна, то існують монотонні послідовності $x'_m \uparrow x$, $x''_m \downarrow x$, x'_m, $x''_m \in D$. Тоді за монотонністю G_n

$$G(x'_m) = \lim_{n\to\infty} G_n(x'_m) \le \underline{\lim}_{n\to\infty} G_n(x) \le$$
$$\overline{\lim}_{n\to\infty} G_n(x) \le \lim_{n\to\infty} G_n(x''_m) = G(x''_m).$$

Перейдемо в цих нерівностях до границі $x'_m \uparrow x$, $x''_m \downarrow x$:

$$G(x) \le \underline{\lim}_{n\to\infty} G_n(x) \le \overline{\lim}_{n\to\infty} G_n(x) \le G(x),$$

що і доводить (а).

(б) Визначимо при $y \in D$ та $x \in \mathbb{R}$ функції

$$H(y) = \lim G_n(y),\ G(x) = \sup(H(y),\ y < x,\ y \in D).$$

З означення верхньої межі та з монотонності H виводимо, що для всіх $y < x < z,\ \ y, z \in D$:

$$H(y) \le G(x) \le \sup(H(z),\ y < x,\ y \in D) = H(z).$$

Функція $G(x)$ є *узагальненою функцією розподілу*, оскільки вона є неспадною, набуває значень із $[0,1]$ та неперервна зліва. Дійсно, монотонність G та включення множини значень до $[0,1]$ є очевидним наслідком відповідних властивостей функції H, а останні справедливі, оскільки зберігаються при обчисленні поточкової границі в означенні H. Для перевірки неперервності зліва знайдемо за означенням верхньої межі $y_n \in D,\ y_n \uparrow x$, такі, що $H(y_n) \uparrow G(x)$. Тоді з $H(y_n) \le G((y_n + x)/2) \le G(x)$ отримуємо $G(x) \le G(x - 0) \le G(x)$ і $G(x - 0) = G(x)$. Отже, $G \in \mathfrak{M}_{01}$.

Нехай x – точка неперервності $G(x)$, а $x'_m,\ x''_m \in D,\ x'_m < x < x''_m$. Тоді виконуються нерівності

$$G(x'_m - 1/m) \le H(x'_m) = \lim_{n\to\infty} G_n(x'_m) \le \underline{\lim}_{n\to\infty} G_n(x) \le$$

$$\overline{\lim}_{n\to\infty} G_n(x) \le \lim_{n\to\infty} G_n(x''_m) = H(x''_m) \le G(x''_m + 1/m).$$

Оберемо в цих нерівностях $x'_m \uparrow x,\ x''_m \downarrow x,\ m \to \infty$:

$$G(x) \le \underline{\lim}_{n\to\infty} G_n(x) \le \overline{\lim}_{n\to\infty} G_n(x) \le G(x),$$

звідки $\lim G_n(x) = G(x)$ $\boxdot$

Вправи

(1) Границя в основному не визначена однозначно, однак границею збіжної в основному послідовності з класу $\mathfrak{M}_{01}$ завжди є елемент $\mathfrak{M}_{01}$.

(2) Знайти функції розподілу, що збігаються в основному, та не поточково.

(3) Довести, що функція

$L(F, G) = \inf(\varepsilon > 0 : F(x - \varepsilon) - \varepsilon \le G(x) \le F(x + \varepsilon) + \varepsilon,\ \forall x \in \mathbb{R})$

є метрикою на просторі функцій розподілу, а збіжність $L(F_n, F) \to 0$ для функцій розподілу F_n, F еквівалентна збіжності в основному $F_n \to^O F$.

(4) Довести, що зі збіжності функцій розподілу $F_n \to^O F$ до неперервної границі F випливає рівномірна за x збіжність. Чи є суттєвою неперервність F ?

(5) Якщо $\xi_n \le \zeta_n \le \eta_n$ м.н. та $\xi_n \to^O \zeta,\ \eta_n \to^O \zeta,\ n \to \infty$, то $\zeta_n \to^O \zeta$.

(6) Нехай F_n, F – функції розподілу і $F_n \to^O F$ Довести, що існує ймовірнісний простір $(\Omega,\ \mathfrak{F},\ \mathbf{P})$ та випадкові величини ξ_n, ξ на ньому з функціями розподілу F_n, F відповідно такі, що $\xi_n \to^{P1} \xi,\ n \to \infty$.

1.25. Слабка збіжність функцій розподілу і випадкових величин

Означення. *Нехай $\xi, (\xi_n, n \ge 1)$ – випадкові величини з функціями розподілу $F, (F_n,\ n \ge 1)$ відповідно. Має місце* слабка збіжність:

$$\xi_n \to^W \xi, \quad F_n \to^W F,\ n \to \infty,$$

якщо для кожної неперервної обмеженої функції g (тобто $\forall g \in C_b(\mathbb{R})$) збігаються математичні сподівання

$$\mathbf{E}g(\xi_n) \to \mathbf{E}g(\xi),\ n \to \infty,$$

або ж збігаються інтеграли

$$\int_{-\infty}^{\infty} g(x)dF_n(x) \to \int_{-\infty}^{\infty} g(x)dF(x),\ n \to \infty.$$

Очевидно, що обидва означення еквівалентні за теоремою *про обчислення математичного сподівання функції від випадкової величини.*

Надалі позначатимемо через $Cl(B)$ – замикання, $Int(B)$ – внутрішність множини $B \subset \mathbb{R}$.

1.25.1. Однозначність слабкої границі

Теорема (про однозначність слабкої границі).
(а) Якщо F, G – функції розподілу і

$$\int_{-\infty}^{\infty} g(x)\ dF(x) = \int_{-\infty}^{\infty} g(x)\ dG(x)$$

для всіх $g \in C_b(\mathbb{R})$, то $F = G$.
(б) Слабка границя функцій розподілу визначена однозначно.
Доведення
(а) Визначимо відстань від точки до множини

$$\rho(x, A) = \inf{}_{y \in A}\ |x - y|\,.$$

Ця функція невід'ємна та неперервна за x, що є наслідком нерівності трикутника для відстані: $|\rho(x, A) - \rho(y, A)| \le |x - y|$. Тому функція

$$g_m(x, A) \equiv \exp(-m\ \rho(x, A)).$$

неперервна і обмежена за x. За означенням замикання $\rho(x, A) = 0$ тоді й тільки тоді, коли $x \in Cl(A)$. Звідси випливає, що

$$g_m(x, A) \downarrow \mathbb{1}_{Cl(A)}(x),\ m \to \infty.$$

Підставляючи $g = g_m$ у рівність умови (а), з *теореми Лебега про мажоровану збіжність* при $m \to \infty$ дістанемо

$$F(Cl(A)) = \int_{-\infty}^{\infty} \mathbb{1}_{Cl(A)}(x)dF(x) = \lim_{m\to\infty} \int_{-\infty}^{\infty} g_m(x)dF(x) =$$
$$\lim_{m\to\infty} \int_{-\infty}^{\infty} g_m(x)dG(x) = \int_{-\infty}^{\infty} \mathbb{1}_{Cl(A)}(x)dG(x) = G(Cl(A)).$$

Переходячи до доповнень, приходимо до висновку, що F і G збігаються на відкритих множинах, зокрема на інтервалах $(-\infty, x)$. Отже, відповідні функції розподілу однакові.

(б) Якщо $F_n \to^W F$ і $F_n \to^W G$, то, очевидно, виконується умова пункту (а), згідно з яким $F = G$ $\boxdot$

1.25.2. Властивості слабкої збіжності

Слабка збіжність є слабшою за вже відомі види збіжностей.

Теорема (про співвідношення слабкої збіжності та збіжності за ймовірністю). *Якщо* $\xi_n \to^P \xi$, *то* $\xi_n \to^W \xi, n \to \infty$.

Доведення теореми є очевидним наслідком пункту (в) теореми *про властивості збіжності за ймовірністю* та означення *слабкої збіжності* ⊡

Теорема (про властивості слабкої збіжності). *Нехай* $F_n \to^W F$.

(а) Для довільної множини $B \in \mathfrak{B}(\mathbb{R})$ *справедливі нерівності*

$$F(Int(B)) \leq \underline{\lim}_{n\to\infty} F_n(B) \leq \overline{\lim}_{n\to\infty} F_n(B) \leq F(Cl(B)).$$

(б) Якщо $F(\partial B) \equiv F(Cl(B) \setminus Int(B)) = 0$, *то* $\lim_{n\to\infty} F_n(B) = F(B)$.

Доведення

(а) Доведемо праву нерівність. Нехай функції g_m визначені в теоремі *про однозначність слабкої границі.* За *теоремою Лебега про монотонну збіжність*:

$$\overline{\lim}_{n\to\infty} F_n(B) = \overline{\lim}_{n\to\infty} \int_{-\infty}^{\infty} 1\!\!\mathrm{I}_{x\in B}\, dF_n(x) \leq$$

$$\overline{\lim}_{n\to\infty} \int_{-\infty}^{\infty} g_m(x,B) dF_n(x) = \int_{-\infty}^{\infty} g_m(x,B) dF(x) \downarrow F(Cl(B)),\ m \to \infty,$$

де використані співвідношення $g_m(x,B) \downarrow 1\!\!\mathrm{I}_{x\ \in\ Cl(B)} \geq 1\!\!\mathrm{I}_{x\ \in\ B}$ з доведення теореми *про однозначність слабкої границі.*

Для доведення лівої нерівності перейдемо до доповнень:

$$\underline{\lim}\, F_n(B) = 1 - \overline{\lim}\ F_n(\overline{B}) \geq 1 - F(Cl(\overline{B})) = F(\overline{Cl(\overline{B})}) = F(Int(B)).$$

(б) З умови випливає, що $F(Cl(B)) = F(B) = F(Int(B))$, тому з твердження (а) дістанемо

$$F(B) \leq \underline{\lim} F_n(B) \leq \overline{\lim} F_n(B) \leq F(B)$$

що і доводить збіжність $F_n(B)$ до $F(B)$ ⊡

Теорема (про еквівалентність слабкої збіжності та в основному). *Нехай* $(F_n,\ n \geq 1)$ *– послідовність функцій розподілу. Для того, щоб* $F_n \to^W F, n \to \infty$, *необхідно і достатньо, щоб* $F_n \to^O F, n \to \infty$, *і* F *була функцією розподілу.*

Зауваження. З теореми *про збіжність в основному та на щільній множині* випливає, що збіжні в основному функції розподілу збігаються до границі з $\mathfrak{M}_{01}$ (**вправа**). Однак ця границя не обов'язково є *функцією розподілу,* наприклад: $F_n(x) = 1\!\!\mathrm{I}_{n\ <\ x} \to 0, n \to \infty, \forall x$, де $0 \in \mathfrak{M}_{01}$.

Доведення

Необхідність. Нехай $F_n \to^W F$. Для довільної точки x неперервності функції $F(x)$ та для множини $B = (-\infty, x)$ маємо $\partial B = \{x\}$ і $F(\partial B) = F(x+0) - F(x) = 0$. Тому за теоремою *про властивості слабкої збіжності* $F(x) = F(B) = \lim F_n(B) = \lim F_n(x)$, що доводить збіжність в основному.

Достатність. Нехай $F_n \to^O F$ і F є функцією розподілу. Оберемо всюди щільну множину D точок неперервності функції F. З умови *нормованості* F знайдемо для заданого $\varepsilon > 0$ сталі $-c, c \in D$ такі, що $F(-c) + 1 - F(c) \le \varepsilon$.

Розглянемо розбиття $-c = a_0 < a_1 < \ldots < a_m = c$ відрізка $[-c, c]$ точками $a_k \in D$ із діаметром $\delta = \max_k |a_k - a_{k+1}|$.

Нехай $g \in C_b(\mathbb{R})$. Покладемо

$$g_\delta(x) \equiv \sum_{k=0}^{m-1} g(a_k)\, 1\!\!\mathrm{I}_{x \in [a_k, a_{k+1})}, \quad L(g) \equiv \sup |g(x)|\,.$$

З неперервності функції g випливає, що вона рівномірно неперервна на відрізку $[-c, c]$, тому модуль неперервності прямує до нуля:

$$w_{c\delta}(g) \equiv \sup_{|x| \le c}\ |g(x) - g_\delta(x)| \to 0$$

при $\delta \to 0$. Виходячи з попередніх означень, оцінимо

$$\overline{\lim}_{n\to\infty} \left| \int_{-\infty}^{\infty} g(x) dF_n(x) - \int_{-\infty}^{\infty} g(x) dF(x) \right| \le$$

$$\overline{\lim}_{n\to\infty} \left| \int_{|x|>c} g(x) dF_n(x) \right| + \overline{\lim}_{n\to\infty} \left| \int_{|x|>c} g(x) dF(x) \right| +$$

$$\overline{\lim}_{n\to\infty} \int_{|x|\le c} |g(x) - g_\delta(x)|\, dF_n(x) + \overline{\lim}_{n\to\infty} \int_{|x|\le c} |g(x) - g_\delta(x)|\, dF(x) +$$

$$\overline{\lim}_{n\to\infty} \left| \int_{|x|\le c} g_\delta(x) dF_n(x) - \int_{|x|\le c} g_\delta(x) dF(x) \right| \le$$

$$L(g)\overline{\lim}_{n\to\infty}(F_n(-c) + 1 - F_n(c) + F(-c) + 1 - F(c)) + 2\, w_{g,c}(\delta) +$$

$$\overline{\lim}_{n\to\infty} \left| \sum_{k=0}^{m-1} g(a_k)(F_n(a_{k+1}) - F_n(a_k) - F(a_{k+1}) + F(a_k)) \right| \le$$

$$2\, L(g)(F(-c) + 1 - F(c)) + 2\, w_{c\delta}(g) +$$

$$\sum_{k=0}^{m-1}\ |g(a_k)| \quad \overline{\lim}_{n\to\infty} \left(|F_n(a_{k+1}) - F(a_{k+1})| + |F_n(a_k) - F(a_k)|\right) \le$$

$$2\, L(g)\varepsilon + 2\, w_{c\delta}(g) + 0,$$

оскільки за умови збіжності *в основному* $F_n(a_k) \to F(a_k)$ у кожній точці $a_k \in D$. Праву частину нерівності можна зробити вибором ε і δ як завгодно малою. Отже, $\int g dF_n \to \int g dF$ при $g \in C_b(\mathbb{R})$ та $F_n \to^W F, n \to \infty$ $\boxdot$

Вправи

(1) Нехай F_n функція розподілу дискретної величини, що рівномірно розподілена на множині $\{0, 1/n, 2/n..., (n-1)/n\}$. Знайти слабку границю F_n.

(2) Випадкові величини ξ_n, ξ – цілозначні. Довести, що $\xi_n \to^W \xi$ тоді й тільки тоді, коли $\mathbf{P}(\xi_n = k) \to \mathbf{P}(\xi = k), n \to \infty$, для всіх $k \in \mathbb{Z}$.

(3) Довести, що для сталої c збіжність $\xi_n \to^W c$ має місце тоді й тільки тоді, коли $\xi_n \to^P c$, $n \to \infty$.

(4) Навести приклад послідовності $F_n \to^W F$ такої, що $F_n(B) \to F(B)$ не для всіх $B \in \mathfrak{B}(\mathbb{R})$, що вказує на суттєвість умови $F(\partial B) = 0$.

(5) Довести, що кожне з тверджень теореми про властивості слабкої збіжності еквівалентне збіжності $F_n \to^W F, n \to \infty$.

(6) Нехай $F_n \to^W F, n \to \infty$, а функція g обмежена та неперервна майже скрізь за мірою, що породжена F. Довести, що $\int g dF_n \to \int g dF, n \to \infty$.

(7, теорема Слуцького) Якщо $\xi_n \to^W \xi$, $\eta_n \to^W c$, то $\xi_n + \eta_n \to^W \xi + c$ та $\xi_n \eta_n \to^W c\xi, n \to \infty$.

(8) Нехай $\xi_n \to^W \xi, n \to \infty$. Довести, що $\mathbf{E}\,|\xi| \le \underline{\lim}_{n\to\infty} \mathbf{E}\,|\xi_n|$.

(9) Довести, що слабка збіжність $\xi_n \to^W \xi, n \to \infty$, має місце тоді й тільки тоді, коли $\overline{\lim}_{n\to\infty} \mathbf{E} g(\xi_n) \le \mathbf{E} g(\xi)$ (а) для всіх $g \in C_b(\mathbb{R})$, (б) для всіх $g \in Lip_b(\mathbb{R}) \equiv \{g : \sup_{x \ne y} |g(x) - g(y)| \,/\, |x - y| < \infty\} \cap C_b(\mathbb{R})$.

(10) Нехай $\xi_n \to^W \xi_0, n \to \infty$, а $\mu_n \equiv \mathrm{med}\ \xi_n$ – медіана величини ξ_n, що є розв'язком рівняння $\mathbf{P}(\xi_n < \mu_n) = 1/2$. За умови однозначної визначеності медіан довести, що $\mu_n \to \mu_0, n \to \infty$.

(11) Нехай $\xi_n \to^W \xi, n \to \infty$, та виконується одна з умов: (а) для деякого $\delta > 0$ $\sup_{n\ge1} \mathbf{E}\,|\xi_n|^{1+\delta} < \infty$, (б) $\lim_{N\to\infty} \sup_{n\ge1} \int_N^\infty \mathbf{P}(|\xi_n| \ge x) dx = 0$. Довести, що $\lim \mathbf{E}\xi_n = \mathbf{E}\xi$.

(12) Нехай $\xi_n \to^W \xi, n \to \infty$, а послідовність випадкових величин $\tau_n \in \mathbb{N}$ така, що $\tau_n / n \to^P \alpha \in (0, 1)$. Довести, що $\xi_{\tau_n} \to^W \xi, n \to \infty$.

1.26. Слабка компактність

Означення. *Клас $\mathfrak{M} \subset \mathfrak{M}_{01}$ узагальнених функцій розподілу є* компактним в основному, *якщо будь-яка послідовність $(G_n, n \ge 1) \subset \mathfrak{M}$ містить збіжну в основному підпослідовність:*

$$\exists (G_{n_k},\ k \ge 1) \subset (G_n,\ n \ge 1), \quad G_{n_k} \to^O G,\ n_k \to \infty,$$

для деякої узагальненої функції розподілу $G \in \mathfrak{M}_{01}$.

1.26.1. Теорема Хеллі про компактність в основному

Теорема (теорема Хеллі про компактність в основному). *Клас усіх узагальнених функцій розподілу $\mathfrak{M}_{01}$ є компактним в основному.*

Доведення. Нехай $(G_n,\ n \geq 1) \subset \mathfrak{M}_{01}$. Позначимо через $K_0 = \mathbb{N}$ множину всіх натуральних чисел, $D = \{x_1, ..., x_n, ...\}$ злічену щільну множину на числовій осі.

Розглянемо числову послідовність $\{G_n(x_1),\ n \in K_0\} \subset [0, 1]$. Оскільки $[0, 1]$ – компакт, то знайдуться нескінченна підпослідовність $K_1 \subset K_0$ та число g_1 такі, що $G_n(x_1) \to g_1$ при $n \to \infty$, $n \in K_1$.

Аналогічно, розглянемо послідовність $\{G_n(x_2),\ n \in K_1\} \subset [0, 1]$ і побудуємо підпослідовність $K_2 \subset K_1 \cap [2, \infty)$ так, щоб $G_n(x_2) \to g_2$ при $n \to \infty$, $n \in K_2$, для деякого числа g_2.

Продовжуючи за індукцією, знайдемо вкладені нескінченні підпослідовності $K_j \subset K_{j-1} \cap [j, \infty)$ та числа g_j такі, що $G_n(x_j) \to g_j$ при $n \to \infty$, $n \in K_j$. Зауважимо, що одночасно при кожному $i < j$ має місце збіжність $G_n(x_i) \to g_i$ при $n \to \infty$, $n \in K_j$, оскільки $K_j \subset K_i$, а довільна підпослідовність збіжної послідовності збігається до тієї ж границі.

За методом Кантора визначимо діагональну послідовність

$$K = \{\min K_j,\ j \geq 1\}.$$

Оскільки за побудовою $\min K_j \geq j$, то послідовність K нескінченна.

Крім того, $K \subset K_j$ за винятком перших j елементів. Дійсно, для всіх $i \geq j$ справедливі включення $\min K_i \in K_i \subset K_j$.

Оскільки будь-яка підпослідовність збіжної послідовності збігається до тієї ж границі, то $G_n(x_j) \to g_j$ при $n \to \infty$, $n \in K$, вже для кожного j.

Застосувавши теорему *про збіжність в основному та на щільній множині*, пункт (б), приходимо до висновку, що для деякої узагальненої функції розподілу G має місце збіжність $G_n \to^O G$ при $n \to \infty, n \in K$ ⊡

1.26.2. Теорема Прохорова про слабку компактність

Означення. *Клас $\digamma$ функцій розподілу називається* слабко компактним, *якщо будь-яка послідовність $(F_n, n \geq 1) \subset \digamma$ містить слабко збіжну підпослідовність, тобто для деякої функції розподілу F:*

$$\exists (F_{n_k},\ k \geq 1) \subset (F_n,\ n \geq 1), \quad F_{n_k} \to^W F,\ n_k \to \infty.$$

Зауважимо, що замкненість множини $\digamma$ не передбачається, тобто включення $F \in \digamma$ може не виконуватись.

Символом $F(\overline{[-c,c)}) = F(-c) + 1 - F(c)$ будемо позначати значення *міри Лебега – Стілтьєса* F від доповнення інтервалу $[-c, c)$.

Теорема (теорема Прохорова про критерій слабкої компактності). *Клас* $\mathcal{F}$ *функцій розподілу слабко компактний тоді і тільки тоді, коли*

$$\lim_{c\to\infty} \sup_{F\in\mathcal{F}} F(\overline{[-c,c)}) = 0, \text{ або } \lim_{c\to\infty} \inf_{F\in\mathcal{F}} F([-c,c)) = 1.$$

Зауваження. Відомому російському математику Ю.В. Прохорову належить доведення більш загального критерію слабкої компактності класів імовірнісних *розподілів* на повному метричному просторі. Наведене твердження для $\mathbb{R}$ було відоме раніше.

Доведення. Еквівалентність двох наведених умов є очевидним наслідком формули про *ймовірність доповнення* $F(\overline{[-c,c)}) = 1 - F([-c,c))$.

Достатність. Нехай $\mathfrak{M}_{01}$ – визначений вище клас *узагальнених функцій розподілу*. Оскільки $\mathcal{F} \subset \mathfrak{M}_{01}$, то за *теоремою Хеллі про компактність в основному* для кожної послідовності $(F_n, n \geq 1) \subset \mathcal{F}$ знайдемо *узагальнену функцію розподілу* $G \in \mathfrak{M}_{01}$ і підпослідовність функцій розподілу $(F_{n_k}, k \geq 1) \subset (F_n, n \geq 1)$, такі, що $F_{n_k} \to^O G$. Нехай D – щільна множина точок неперервності G. Якщо $-c, c \in D$, то за означенням збіжності *в основному*

$$G(-c) + 1 - G(c) = \lim_{k\to\infty} (F_{n_k}(-c) + 1 - F_{n_k}(c)) =$$

$$\lim_{k\to\infty} F_{n_k}(\overline{[-c,c)}) \leq \sup_{F\in\mathcal{F}} F(\overline{[-c,c)}) \to 0, \ c \to \infty.$$

Отже, функція G – *нормована*: $G(-\infty) = 0$, $G(\infty) = 1$. Оскільки $G \in \mathfrak{M}_{01}$, то G є *функцією розподілу* внаслідок теореми *про характеризацію функцій розподілу в класі узагальнених*. Тому за теоремою *про еквівалентність слабкої збіжності та в основному* з $F_{n_k} \to^O G$ випливає збіжність $F_{n_k} \to^W G$. Отже, клас $\mathcal{F}$ є *слабко компактним*.

Необхідність. Припустимо, що умова Прохорова не виконана, тобто відповідна границя $\varepsilon > 0$. Оскільки функція $\varepsilon(c) \equiv \sup_{F\in\mathcal{F}} F(\overline{[-c,c)})$ не зростає за c, то вона не менша за свою границю при $c \to \infty$, тобто $\varepsilon(c) \geq \varepsilon$ при всіх c. За означенням верхньої межі в $\varepsilon(c)$ для кожного $n \geq 1$ знайдеться $F_n \in \mathcal{F}$, така, що $F_n(\overline{[-n,n)}) \geq \varepsilon/2$. За означенням *слабкої компактності* $\mathcal{F}$ оберемо $(F_{n_k}, k \geq 1) \subset (F_n, n \geq 1)$ так, щоб $F_{n_k} \to^W F$, де F – деяка *функція розподілу*. Нехай $-c, c$ – точки неперервності F. Оскільки за теоремою *про еквівалентність слабкої збіжності та в основному* $F_{n_k} \to^O F$, то, починаючи з $n_k > c$, маємо $\overline{[-c,c)} \supset \overline{[-n_k,n_k)}$ і за означенням збіжності *в основному*

$$F(-c) + 1 - F(c) = \lim_{k\to\infty}(F_{n_k}(-c) + 1 - F_{n_k}(c)) =$$

$$\lim_{k\to\infty} F_{n_k}(\overline{[-c,c)}) \geq \underline{\lim}_{k\to\infty} F_{n_k}(\overline{[-n_k,n_k)}) \geq \varepsilon/2 > 0.$$

Спрямовуючи тут $c \to \infty$, отримуємо суперечність з умовою *нормованості* функції розподілу $F(-\infty) + 1 - F(\infty) \geq \varepsilon/2 > 0$ ⊡

Лема (про верхні межу та границю монотонної послідовності). *Нехай числова послідовність* $\varepsilon_n(c)$ *така, що*

(а) $\varepsilon_n(c) \downarrow 0$ *при* $c \to \infty$ *для кожного* n, *та*

(б) $\overline{\lim}_{n\to\infty} \varepsilon_n(c) \to 0$ *при* $c \to \infty$.

Тоді $\sup_{n\geq 1} \varepsilon_n(c) \to 0$ *при* $c \to \infty$.

Доведення. Припустимо, що твердження леми не виконується. Оскільки функція $\sup_{n\geq 1} \varepsilon_n(c)$ не зростає за c, то всі її значення не менші за деяке $\delta > 0$ і для кожного m знайдеться n_m таке, що $\varepsilon_{n_m}(m) \geq \delta/2$. Послідовність n_m не може бути обмеженою, інакше $n_m = k$ для нескінченної кількості номерів m, і $\varepsilon_k(m) \geq \delta/2$ для цих номерів, що суперечить умові (а). Тоді для кожного $c > 0$, починаючи з $m > c$ внаслідок монотонності (а) маємо $\varepsilon_n(c) \geq \varepsilon_n(m)$, і

$$\overline{\lim}_{n\to\infty}\varepsilon_n(c) \geq \overline{\lim}_{m\to\infty}\varepsilon_{n_m}(c) \geq \overline{\lim}_{m\to\infty}\varepsilon_{n_m}(m) \geq \delta/2.$$

Ця нерівність суперечить умові (б) ⊡

Теорема (про критерій слабкої компактності послідовності). *Послідовність* $(F_n,\ n \geq 1)$ *функцій розподілу є слабко компактною тоді й тільки тоді, коли*

$$\lim_{c\to\infty}\ \overline{\lim}_{n\to\infty}\ F_n(\overline{[-c,c)}) = 0.$$

Доведення. *Необхідність* очевидна, оскільки верхня границя не перевищує верхню межу, а остання прямує до нуля за *теоремою Прохорова про критерій слабкої компактності*.

Для доведення *достатності* застосуємо лему *про верхні межу та границю монотонної послідовності* до послідовності $\varepsilon_n(c) \equiv F_n(\overline{[-c,c)})$. Умова (а) леми випливає з *нормованості* функцій розподілу F_n, умова (б) збігається з умовою теореми, а твердження леми є умовою Прохорова слабкої компактності. Тому шукане випливає з теореми Прохорова ⊡

Вправи

(1) Послідовність нормальних розподілів $N(\mu_n, \sigma_n^2), n \geq 1$, слабко компактна тоді й тільки тоді, коли $\sup_{n\geq 1} |\mu_n| < \infty$, $\sup_{n\geq 1} \sigma_n^2 < \infty$.

(2) Клас $\mathcal{F}$ функцій розподілу є слабко компактним тоді й тільки тоді, коли знайдеться невід'ємна борелєва функція V така, що $V(x) \to \infty$ при $|x| \to \infty$ і $\sup_{F\in\mathcal{F}} \int_{\mathbb{R}} V(x)dF(x) < \infty$.

1.27. Характеристичні функції

Перетворення Фур'є широко застосовуються в математичній фізиці, теорії диференціальних рівнянь та інших розділах математики. У теорії ймовірностей це поняття відоме як характеристична функція.

Означення. *Нехай ξ – випадкова величина, а F – її функція розподілу.* Характеристичною функцією *величини ξ та функції розподілу F називається така комплекснозначна функція дійсної змінної $t \in \mathbb{R}$:*

$$\varphi_\xi(t) \equiv \mathbf{E}\exp(it\xi) = \int_{-\infty}^{\infty} \exp(itx)\, dF(x) \equiv \varphi_F(t).$$

Рівність посередині випливає з теореми *про обчислення математичного сподівання функції від випадкової величини.* В останньому означенні та надалі *математичне сподівання* (інтеграл Лебега за мірою $\mathbf{P}$) від комплекснозначних величин визначається за лінійністю: $\mathbf{E}(\xi_1 + i\xi_2) = \mathbf{E}\xi_1 + i\mathbf{E}\xi_2$.

Вправа. Для комплекснозначних величин $\overline{\mathbf{E}\xi} = \mathbf{E}\overline{\xi}$, $|\mathbf{E}\xi| \le \mathbf{E}|\xi|$.

1.27.1. Однозначність відповідності

Теорема (про однозначну відповідність між функціями розподілу та характеристичними функціями). *Відповідність між функціями розподілу та характеристичними функціями є взаємно однозначною.*

Доведення. Припустимо, що функції розподілу F, G мають однакові характеристичні функції: $\varphi_F(t) = \varphi_G(t)$ при всіх $t \in \mathbb{R}$. Розглянемо клас

$$\mathfrak{G} = \left\{ g \in C_b(\mathbb{R}) : \int_{-\infty}^{\infty} g(x)dF(x) = \int_{-\infty}^{\infty} g(x)dG(x) \right\}.$$

За припущенням, кожна гармоніка $\exp(itx) \in \mathfrak{G}$ при $t \in \mathbb{R}$. Клас $\mathfrak{G}$ є лінійним, тому він містить всі комплексні тригонометричні поліноми вигляду $g(x) = \sum c_k \exp(it_k x)$. Крім того, за *теоремою Лебега про мажоровану збіжність* клас $\mathfrak{G}$ замкнений відносно обмеженої поточкової збіжності.

Застосовуючи теорему Стоуна – Вейєрштраса про наближення неперервної обмеженої функції поліномами, доведемо, що клас $\mathfrak{G}$ містить усі фінітні неперервні функції. Дійсно, нехай функція $g \in C_b(\mathbb{R})$, причому $g(x) = 0$ при $|x| > c - 1$. Оберемо тригонометричний поліном g_ε такий, що $|g(x) - g_\varepsilon(x)| \le \varepsilon$ при $|x| \le c$ та $g_\varepsilon(-c) = g_\varepsilon(c)$. Після періодичного продовження його на $\mathbb{R}$ з періодом $2c$ отримуємо функцію $g_\varepsilon \in C_b(\mathbb{R})$. Вибором c різницю між інтегралами:

$$\left|\int gdF - \int gdG\right| = \left|\int_{|x|\le c} gdF - \int_{|x|\le c} gdG\right| \le$$

$$\left|\int_{|x|\leq c}(g-g_\varepsilon)dF\right| + \left|\int_{|x|\leq c}(g-g_\varepsilon)dG\right| +$$
$$\left|\int g_\varepsilon dF - \int g_\varepsilon dG\right| + \left|\int_{|x|>c} g_\varepsilon dF - \int_{|x|>c} g_\varepsilon dG\right| \leq$$
$$2\,\varepsilon + 0 + (F(\overline{[-c,c]}) + G(\overline{[-c,c]})(\sup|g| + \varepsilon)$$

можна зробити як завгодно малою. Тому $g \in \mathfrak{G}$.

Оскільки кожна неперервна обмежена функція є обмеженою поточковою границею фінітних неперервних обмежених функцій, то $\mathfrak{G} = C_b(\mathbb{R})$. Тому з теореми *про однозначність слабкої границі*, (а), маємо $F = G$ $\boxdot$

Вправи

(1) Якщо ξ – цілозначна випадкова величина, то її характеристична функція $\varphi(t)$ має період 2π і $\mathbf{P}(\xi = n) = (2\pi)^{-1}\int_{-\pi}^{\pi}\exp(-itn)\varphi(t)dt$.

(2) Нехай випадкова величина ξ інтегровна в будь-якій степені, причому: $\mathbf{E}\xi^n = (n+k)!/k!$ для деякого k та всіх $n \geq 0$. Довести, що розподіл ξ визначається однозначно та збігається з розподілом Ерланга.

(3) Довести, що розподіл інтегровної випадкової величини ξ однозначно визначається функцією $\mathbf{E}\left|\xi - t\right|, t \in \mathbb{R}$.

1.27.2. Властивості характеристичної функції

Теорема (про основні властивості характеристичної функції). *Нехай φ – характеристична функція. Вона має такі властивості:*

(а – нормованість) $\varphi(0) = 1$ і $|\varphi(t)| \leq 1,\ \forall t \in \mathbb{R}$,

(б – ермітовість) $\varphi(-t) = \overline{\varphi(t)}$,

(в – неперервність) $\varphi(t)$ неперервна в нулі і рівномірно неперервна,

(г – невід'ємна визначеність) для довільних дійсних $t_1, ..., t_n$ та комплексних $c_1, ..., c_n$ справедлива нерівність

$$\sum\nolimits_{k,\ j\ =1}^{n} c_k\overline{c_j}\ \varphi(t_k - t_j) \geq 0.$$

Зауваження. Теорема Бохнера-Хінчина стверджує, що будь-яка комплекснозначна функція дійсної змінної, що задовольняє умови (а)-(г), є характеристичною для певної функції розподілу.

Доведення теореми.

(а) $\varphi(0) = \int_{-\infty}^{\infty} dF(x) = F(\mathbb{R}) = 1$ з умови *нормованості* функції розподілу, $|\varphi(t)| = \left|\int_{-\infty}^{\infty}\exp(itx)dF(x)\right| \leq \int_{-\infty}^{\infty}\left|\exp(itx)\right| dF(x) = 1$.

(б) $\varphi(-t) = \int_{-\infty}^{\infty}\exp(-itx)dF(x) = \int_{-\infty}^{\infty}\overline{\exp(itx)}dF(x) =$
$\overline{\int_{-\infty}^{\infty}\exp(itx)dF(x)}$.

(в) $|\varphi(t)-\varphi(s)| \le \int_{-\infty}^{\infty} |\exp(itx)-\exp(isx)|\, dF(x) =$
$\int_{-\infty}^{\infty} |\exp(i(t-s)x)-1|\, dF(x) \to 0,\ t-s\to 0,$

за *теоремою Лебега про мажоровану збіжність*, оскільки $|\exp(ihx)-1| \le 2$ та $|\exp(ihx)-1| \to 0$ при $h\to 0$ для кожного x.

(г) Сума з умови дорівнює

$$\int_{-\infty}^{\infty} \sum_{k,\ j\ =1}^{n} c_k \overline{c_j} \exp(i(t_k-t_j)x) dF(x) =$$
$$\int_{-\infty}^{\infty} \left|\sum_{k=1}^{n} c_k \exp(it_k x)\right|^2 dF(x) \ge 0 \ \boxdot$$

Вправи

(1) Довести, що з невід'ємної визначеності (г) випливає ермітовість (б).

(2) Знайти характеристичну функцію для випадкової величини з щільністю: (а) $(T-|x|)^+/T^2$, (б) $(1-\cos(Tx))/\pi T x^2$.

(3) Нехай $\varphi(t)$ – характеристична функція така, що $\varphi(t)=0$ при $|t|>a$. Вивести з теореми Бохнера-Хінчина, що періодичне продовження з періодом $2a$ функції $\varphi(t)$ є характеристичною функцією.

(4) Довести, що функція з періодом $2T$, що дорівнює $(T-|x|)^+/T$ при $|x|\le T$, є характеристичною.

(5) Якщо випадкова величина ξ має щільність, то $\left|\varphi_\xi(t)\right| < 1\ \forall t\ne 0$.

(6) Випадкова величина ξ має щільність та $\mathbf{E}\xi=0$, $\mathbf{E}\xi^2<\infty$. Довести, що знайдеться $\delta>0$ таке, що $\sup_{|t|\ge u} \left|\varphi_\xi(t)\right| \le \exp(-\delta u^2)$ для всіх $u\in(0,1)$.

(7) Якщо випадкова величина ξ не дорівнює м.н. сталій, то $\left|\varphi_\xi(t)\right| < 1$ у деякому виколотому околі нуля.

(8) Рівність $\left|\varphi_\xi(t)\right| = 1$ для деякого $t\ne 0$ виконується тоді й тільки тоді, коли $\mathbf{P}(\xi\in a+b\mathbb{Z})=1$ для деяких $a,b\in\mathbb{R}$.

(9) Існують незалежні величини ξ_1,ξ_2,ξ_3 такі, що ξ_2 та ξ_3 мають різні розподіли, однак $\xi_1+\xi_2$ і $\xi_1+\xi_3$ – однаково розподілені.

(10) Для довільних розподілу $(p_k, k=\overline{1,n})$ та додатних $(a_k, k=\overline{1,n})$ опуклий полігон $\sum_{k=1}^{n} p_k(1-|t|\, a_k^{-1})^+$ є характеристичною функцією.

(11 – теорема Пойа) Нехай φ – дійсна неперервна парна функція на $\mathbb{R}$, така, що $\varphi(0)=1, \varphi(\infty)=0$, причому графік φ на $\mathbb{R}_+$ є опуклим донизу. Тоді φ є характеристичною функцією. *Вказівка*: довести, що $\varphi(t)$ можна зобразити у вигляді інтегралу $\int_0^\infty (1-|t|/s)^+\mu(ds)$ для деякої міри μ.

(12) Графік кусково-лінійної функції ψ утворений точками $(n,\varphi(n)), n\in\mathbb{Z}$, де φ – характеристична функція. Довести, що ψ – характеристична функція.

(13) За означенням випадкова величина ξ симетрична, якщо ξ та $-\xi$ мають однакові функції розподілу. Довести, що симетричність еквівалентна дійснозначності характеристичної функції.

Теорема (про властивості характеристичної функції).

(а) $\varphi_{a+b\xi}(t) = \exp(ita)\ \varphi_\xi(bt)$.

(б) Якщо ξ, η *незалежні, то* $\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\ \varphi_\eta(t)$. *Аналогічно, для незалежних у сукупності величин* $(\xi_k, k = \overline{1,n})$ *з сумою* $S_n = \xi_1 + ... + \xi_n$

$$\varphi_{S_n}(t) = \prod\nolimits_{k=1}^n \varphi_{\xi_k}(t).$$

(в) Якщо ξ *інтегровна, то* $\varphi_\xi(t) = 1 + it\ \mathbf{E}\xi + o(t)$, $t \to 0$.

(г) Якщо ξ *квадратично інтегровна, то*

$$\varphi_\xi(t) = 1 + it\ \mathbf{E}\xi - t^2\ \mathbf{E}\xi^2/2 + o(t^2),\ t \to 0.$$

(д) За інтегровності ξ *або квадратичної інтегровності відповідно*

$$\mathbf{E}\xi = -i\varphi'_\xi(0), \quad \mathbf{E}\xi^2 = -\varphi''_\xi(0).$$

(е) Якщо $\xi \simeq N(\mu, \sigma^2)$ *нормальна випадкова величина, то*

$$\varphi_\xi(t) = \exp(it\mu - \sigma^2 t^2/2).$$

(ж) Якщо $\xi \simeq \Pi(\lambda)$ *має розподіл Пуассона з параметром* λ, *то*

$$\varphi_\xi(t) = \exp(\lambda(\exp(it) - 1)).$$

(з) Якщо $\xi = c$ *– стала, то* $\varphi_\xi(t) = \exp(itc)$.

Доведення

(а) За мультиплікативністю експоненти та *однорідністю* математичного сподівання

$$\varphi_{a+b\xi}(t) \ = \mathbf{E}\exp(it(a + b\xi)) \ =$$

$$\mathbf{E}\exp(ita)\exp(itb\xi) \ = \exp(ita)\mathbf{E}\exp(itb\xi) \ = \exp(ita)\varphi_\xi(bt).$$

(б) За теоремою *про перетворення незалежних величин* випадкові величини $\exp(it\xi)$ і $\exp(it\eta)$ також незалежні. Тому за теоремою *про математичне сподівання добутку незалежних величин*

$$\varphi_{\xi+\eta}(t) = \mathbf{E}\exp(it(\xi + \eta)) = \mathbf{E}\exp(it\xi)\exp(it\eta) =$$
$$\mathbf{E}\exp(it\xi)\ \mathbf{E}\exp(it\eta) = \varphi_\xi(t)\varphi_\eta(t).$$

Далі, за теоремою *про перетворення незалежних величин* у зображенні $S_n = S_{n-1} + \xi_n$ доданки є незалежними. Тому $\varphi_{S_n}(t) = \varphi_{S_{n-1}}(t)\varphi_{\xi_n}(t)$ за доведеним вище. Звідси за індукцією виводимо друге твердження (б).

(в) З елементарної нерівності $|\exp(itx) - 1 - itx| \le 2\,|tx|$, $\forall tx \in \mathbb{R}$, випливає, що до інтегралу

$$(\varphi_\xi(t) - 1 - it\ \mathbf{E}\xi)t^{-1} = \int_{-\infty}^{\infty}(\exp(itx) - 1 - itx)t^{-1}\ dF(x)$$

можна застосувати *теорему Лебега про мажоровану збіжність* при $t \to 0$, тому що $\int_{-\infty}^{\infty} 2\,|x|\,dF(x) = 2\mathbf{E}\,|\xi| < \infty$. Оскільки підінтегральна функція в правій частині поточково прямує до нуля, то права частина прямує до нуля при $t \to 0$, звідки дістанемо співвідношення (в).

(г) Аналогічно застосовуємо до інтегралу

$$\int_{-\infty}^{\infty} (\exp(itx) - 1 - itx + t^2x^2/2)\, t^{-2} dF(x) =$$
$$(\varphi_\xi(t) - 1 - it\,\mathbf{E}\xi + t^2\mathbf{E}\xi^2/2)\ /\ t^2$$

нерівність $|\exp(itx) - 1 - itx + t^2x^2/2| \ \le |tx|^2$.

(д) Випливає з формули розкладу в ряд Тейлора в околі нуля диференційовної функції $\varphi_\xi(t)$ та тверджень (в) чи (г).

(е) Якщо $\zeta \simeq N(0,1)$ *стандартна нормальна* величина, то

$$\varphi_\zeta(t) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \exp\left(-x^2/2 + itx\right) dx =$$
$$\exp\left(-t^2/2\right)(2\pi)^{-1/2} \int_{-\infty}^{\infty} \exp(-(x - it)^2/2)dx = \exp\left(-t^2/2\right).$$

У загальному випадку досить застосувати властивість (а) до означення величини $\xi = \mu + \sigma\zeta$ з *нормальним розподілом*.

(ж) За формулою *про обчислення математичного сподівання функції від дискретної величини*

$$\varphi_\xi(t) = \sum\nolimits_{n \ge 0} \exp(itn)\lambda^n \exp(-\lambda)/n! \ = \exp(\lambda(\exp(it) - 1)) \ \boxdot$$

Вправи

(1) Випадкова величина ξ з розподілом $\mathbf{P}(\xi = \pm 2n) = C/(n^2 \ln n)$ не інтегровна, однак її характеристична функція диференційовна в нулі.

(2) Якщо характеристична функція $\varphi_\xi(t)$ випадкової величини ξ має похідну парного порядку $2n$, то випадкова величина ξ^{2n} інтегровна.

(3) Якщо $\mathbf{E}\,|\xi|^\alpha < \infty$, $\alpha \in (0,1)$, то φ_ξ задовольняє умову Гельдера рівня α.

(4) Довести, що: (а) характеристична функція величини зі щільністю Коші $(\pi(1 + x^2))^{-1}$ дорівнює $\exp(-|t|)$, (б) для незалежних однаково розподілених величин $\xi_1, ..., \xi_n$ зі щільністю Коші нормована сума $(\xi_1 + ... + \xi_n)/n$ має таку саму щільність, (в) послідовність незалежних випадкових величин із щільністю Коші не задовольняє закон великих чисел.

(5) Характеристична функція рівномірного розподілу на симетричному відрізку $[-a, a]$ дорівнює $(\sin at)/at$.

(6) Якщо величина ξ має гама-розподіл $\Gamma(\lambda, \alpha)$, то $\varphi_\xi(t) = (1 - it/\lambda)^{-\alpha}$.

(7) Нехай $\varphi(t)$ – характеристична функція стандартного нормального розподілу. (а) Інтегруванням за частинами вивести диференціальне рівняння $\varphi'(t) = -t\varphi(t)$, (б) отримати звідси формулу для $\varphi(t)$.

(8) Функція $\exp(-t^4)$ не є характеристичною.

(9) Для незалежності випадкових величин ξ, η необхідно і достатньо, щоб $\mathbf{E}\exp(it\xi + is\eta) = \mathbf{E}\exp(it\xi)\mathbf{E}\exp(is\eta)$ при всіх $s, t \in \mathbb{R}$.

(10) Знайти залежні величини ξ, η такі, що $\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\ \varphi_\eta(t), \forall t \in \mathbb{R}$.

(11) Випадкові величини $(\xi_k, k = \overline{1,n})$ незалежні у сукупності тоді й тільки тоді, коли величини $\xi_1 + ... + \xi_k$ та ξ_{k+1} незалежні для всіх $k = \overline{1, n-1}$.

(12) Для характеристичної функції φ довести, що $\operatorname{Re}\varphi(2t) \geq 4\operatorname{Re}\varphi(t) - 3$.

(13) Якщо випадкова величина ξ обмежена, то функція $t^{-2}\ln\left|\varphi_\xi(t)\right|$ обмежена та відділена від нуля у деякому околі нуля.

(14) Якщо $\varphi(t), t \in \mathbb{R}$, – характеристична функція, то (а) $\varphi^n(t)$, (б) $|\varphi(t)|^2$, (в) $t^{-1}\int_0^t \varphi(s)ds$ – також характеристичні функції.

(15) Довести, що існують різні характеристичні функції $\varphi_k(t), k = 1, 2$, такі, що $\varphi_1^2(t) = \varphi_2^2(t)$ для всіх $t \in \mathbb{R}$.

(16, С.Н.Бернштейн) Якщо величини ξ, η незалежні, однаково розподілені і квадратично інтегровні, причому $\xi + \eta$ та $\xi - \eta$ також незалежні, то ξ, η – нормальні випадкові величини.

(17) Випадкові величини ξ, η незалежні, однаково розподілені і з нульовим середнім та одиничною дисперсією. Якщо розподіл величини $(\xi + \eta)/\sqrt{2}$ збігається з розподілом ξ, то $\xi \simeq N(0, 1)$.

1.27.3. Інші інтегральні перетворення

Одночасно з *характеристичними функціями* у теорії ймовірностей використовують інші види перетворень.

Означення. Твірною функцією моментів *випадкової величини ξ називається така функція змінної t* :

$$m_\xi(t) \equiv \mathbf{E}\exp(t\xi)$$

у припущенні, що величина під знаком сподівання інтегровна для всіх t з деякого околу нуля.

Твірна функція моментів розкладається у ряд Тейлора

$$m_\xi(t) = \sum\nolimits_{n\geq 0} t^n \mathbf{E}\xi^n/n!.$$

Логарифмічне перетворення $c_\xi(t) \equiv \ln m_\xi(t)$ називається твірною функцією кумулянт. Коефіцієнти σ_n розкладу у ряд Тейлора

$$c_\xi(t) = \sum\nolimits_{n\ge 0} t^n \sigma_n / n!$$

називаються *кумулянтами* (*семіінваріантами*) величини ξ. Можна довести (**вправа**), що $\sigma_1 = \mathbf{E}\xi$, $\sigma_2 = \mathbf{D}\xi$, $\sigma_3 = \mathbf{E}(\xi - \mathbf{E}\xi)^3$.

Означення. Перетворенням Лапласа *випадкової величини ξ називається така функція дійсної змінної t* :

$$l_\xi(t) \equiv \mathbf{E}\exp(-t\xi).$$

Очевидно, що для невід'ємних величин ξ перетворення Лапласа визначено коректно та є аналітичною функцією при $\mathrm{Re}\, t > 0$.

Зауважимо, що твердження теореми *про основні властивості характеристичних функцій*, як і інші властивості характеристичних функцій, з необхідними модифікаціями виконуються як для твірних функцій моментів, так і для перетворень Лапласа.

1.27.4. Генератриси випадкових величин

Для цілочисельних величин використовують метод генератрис, які є дискретними аналогами характеристичних функцій.

Означення. *Нехай ν – невід'ємна цілозначна випадкова величина з розподілом $p_n = \mathbf{P}(\nu = n)$, $n = 0, 1, \ldots$* Генератрисою випадкової величини *ν називається генератриса послідовності p_n* :

$$\varphi_\nu(z) \equiv \sum\nolimits_{n\ge 0} z^n p_n = \mathbf{E}z^\nu.$$

Остання формула є наслідком теореми *про обчислення математичного сподівання функції від дискретної величини.*

Теорема (про властивості генератрис).

(а) Ряд $\varphi_\nu(z)$ абсолютно збігається при $|z| \le 1$ та однозначно визначає розподіл p_n.

(б) $\varphi_\nu(1) = 1$.

(в) $|\varphi_\nu(z)| \le 1$ при $|z| \le 1$.

(г) $\varphi'_\nu(1) = \mathbf{E}\nu$.

(д) Якщо величини ξ, η – незалежні, то $\varphi_{\xi+\eta}(z) = \varphi_\xi(z)\varphi_\eta(z)$.

(е) Якщо величини $(\xi_n, n \ge 1)$ та ν – незалежні в сукупності, а величина ζ є сумою випадкового числа ν випадкових доданків ξ_n вигляду

$$\zeta = \sum\nolimits_{k=1}^{\nu} \xi_k = \sum\nolimits_{k\ge 1} \xi_k \, \mathbb{1}_{\{k \le \nu\}} \ ,$$

то її генератриса є суперпозицією генератрис:

$$\varphi_\zeta(z) = \varphi_\nu(\varphi_\xi(z)).$$

Доведення

(а) Випливає зі збіжності ряду $\sum p_n = 1$, однозначність є наслідком формули обертання

$$p_n = (2\pi)^{-1} \int_{-\pi}^{\pi} \exp(-int)\varphi_\xi(\exp(it))dt.$$

Остання виводиться з ортогональності гармонік $\exp(int)$ у просторі інтегровних з квадратом функцій $L_2[-\pi, \pi]$:

$$\int_{-\pi}^{\pi} \exp(-int)\varphi_\xi(\exp(it))dt = \sum_{m \geq 0} p_n \int_{-\pi}^{\pi} \exp(-int + imt)dt = 2\pi p_n.$$

(б), (в) очевидні.

(г) При $z \in (0, 1)$ похідну можна внести під знак суми ряду, оскільки ряд із похідних збігається абсолютно:

$$\varphi'_\nu(z) = \sum_{n \geq 0} nz^{n-1} p_n = \mathbf{E}\nu z^{\nu-1}.$$

Спрямовуючи тут $0 \leq z \uparrow 1$, за *теоремою Лебега про монотонну збіжність* дістанемо рівність (г).

(д) Випливає з теореми *про перетворення незалежних величин* z^ξ, z^η, та теореми *про математичне сподівання добутку незалежних величин.*

(е) За означенням

$$\varphi_\zeta(z) = \mathbf{E}z^\zeta = \sum_{k \geq 0} \mathbf{E}z^\zeta 1\!\!\mathrm{I}_{\{\nu = k\}} = \sum_{k \geq 0} \mathbf{E}z^{\xi_1 + \ldots + \xi_\nu} 1\!\!\mathrm{I}_{\{\nu = k\}} =$$
$$\sum_{k \geq 0} \mathbf{E}z^{\xi_1 + \ldots + \xi_k} 1\!\!\mathrm{I}_{\{\nu = k\}} = \sum_{k \geq 0} \mathbf{E}z^{\xi_1 + \ldots + \xi_k}\, \mathbf{E}1\!\!\mathrm{I}_{\{\nu = k\}} =$$
$$\sum_{k \geq 0} (\mathbf{E}z^{\xi_1})^k\, \mathbf{P}(\nu = k) = \sum_{k \geq 0} (\varphi_\xi(z))^k\, \mathbf{P}(\nu = k) =$$
$$\sum_{k \geq 0} u^k\, \mathbf{P}(\nu = k) \mid_{u = \varphi_\xi(z)} = \varphi_\nu(\varphi_\xi(z)),$$

де використано незалежність суми $\xi_1 + \ldots + \xi_k$ і події $\{\nu = k\}$ та попередній пункт, внаслідок якого генератриса суми *незалежних у сукупності, однаково розподілених* величин дорівнює відповідній степені генератриси доданку.

Зауваження. Твердження (е) залишається справедливим для дійснозначних величин ξ_n, якщо тільки коректно визначено перетворення $\varphi_\xi(z) = \mathbf{E}z^{\xi_1}$, зокрема для дійсних ξ_1 та при $z = \exp(it)$, $t \in \mathbb{R}$. Дійсно, характеристична функція суми незалежних величин також дорівнює добуткові характеристичних функцій доданків $\boxdot$

Вправи

(1) Довести методом генератрис, що (а) сума незалежних величин із розподілами Пуассона, (б) слабка границя величин із розподілами Пуассона знову має розподіл Пуассона.

(2) Випадкова величина η набуває цілих невід'ємних значень, а величини $(\delta_k, k \geq 0)$ не залежать від η, і $\mathbf{P}(\delta_k = 0) = 1 - \mathbf{P}(\delta_k = 1) \in (0,1)$. Розглянемо суми $\eta_1 = \sum_{k=0}^{\eta} \delta_k$, $\eta_0 = \sum_{k=0}^{\eta}(1 - \delta_k)$. Довести, що (а) величини η_0 та η_1 незалежні тоді й тільки тоді, коли η має розподіл Пуассона, (б) η_i мають розподіл Пуассона.

(3) Довести, що кількість успіхів у випадковому числі випробувань Бернуллі, яке не залежить від результатів окремих випробувань та має розподіл Пуассона, теж має пуассонівський розподіл. Знайти параметр цього розподілу.

(4) Довести для невід'ємної величини $\xi \leq \infty$ такі тотожності
$\lim_{t\to\infty} \mathbf{E}\exp(-t\xi) = \mathbf{P}(\xi = 0)$, $\lim_{t\downarrow 0} \mathbf{E}\exp(-t\xi) = \mathbf{P}(\xi < \infty)$.

(5) В умовах пункту (е) теореми про властивості генератрис довести, що $\mathbf{E}\zeta = \mathbf{E}\nu\mathbf{E}\xi_1$ та $\mathbf{D}\zeta = \mathbf{E}\nu\mathbf{D}\xi_1 + (\mathbf{E}\xi_1^2)\mathbf{D}\nu$.

1.27.5. Формула обертання для характеристичної функції

Теорема (про формулу обертання для характеристичної функції). *Нехай функція розподілу F має характеристичну функцію φ_F.*

(а) Для всіх $a < b$, що є точками неперервності F, справедлива тотожність

$$F(b) - F(a) = \frac{1}{2\pi} \lim_{\varepsilon\downarrow 0} \int_a^b dx \int_{-\infty}^{\infty} \exp(-itx - \varepsilon^2 t^2/2)\, \varphi_F(t)\, dt.$$

(б) Якщо функція φ_F абсолютно інтегровна на $\mathbb{R}$, то функція розподілу F має щільність f, що дорівнює

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\, \varphi_F(t)\, dt.$$

Зауваження. Абсолютна збіжність кратного інтегралу в (а) при $\varepsilon > 0$ обумовлена обмеженістю характеристичної функції. Така збіжність при $\varepsilon = 0$ може порушуватись.

Доведення

(а) Нехай випадкова величина ξ має функцію розподілу F, а величина $\zeta \simeq N(0, \varepsilon^{-2})$ не залежить від ξ. За *теоремою Фубіні про кратний та* *повторні інтеграли* отримуємо **тотожність Парсеваля** для $t \in \mathbb{R}$:

$$\mathbf{E}\exp(i\zeta(\xi - t)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(ix(y-t))dF_\zeta(x)dF(y) =$$

$$\int_{-\infty}^{\infty}\varphi_\zeta(y-t)dF(y) = \int_{-\infty}^{\infty}\exp(-itx)f_\zeta(x)\varphi_F(x)dx.$$

Нехай величина $\eta \simeq N(0,\varepsilon^2)$ не залежить від ξ. Тоді за теоремою *про властивості характеристичної функції*, (е), щільності та характеристичні функції величин η та ζ пов'язані рівняннями

$$f_\eta(y) = (2\pi)^{-1/2}\varepsilon^{-1}\varphi_\zeta(y), \quad \varphi_\eta(x) = (2\pi)^{1/2}\varepsilon^{-1}f_\zeta(x) = \exp(-\varepsilon^2x^2/2).$$

Множенням *тотожності Парсеваля* на $(2\pi)^{-1/2}\varepsilon^{-1}$, з використанням парності функції f_η прийдемо при всіх $t \in \mathbb{R}$ до тотожності

$$g_\varepsilon(t) \equiv \int_{-\infty}^{\infty} f_\eta(t-y)dF(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-itx - \varepsilon^2x^2/2)\varphi_F(x)dx.$$

За теоремою *про функцію розподілу суми незалежних величин* ліва частина збігається зі щільністю суми незалежних величин $\xi + \eta$. Тому

$$\mathbf{P}(\xi + \eta \in [a,b)) = \int_a^b g_\varepsilon(t)dt.$$

Оскільки $\mathbf{D}\eta = \varepsilon^2$, то за *нерівністю Чебишева для дисперсій* $\eta \to^P 0$ при $\varepsilon \to 0$, звідки $\xi + \eta \to^P \xi$. Отже, за теоремою *про співвідношення слабкої збіжності та збіжності за ймовірністю* $\xi + \eta \to^W \xi$, а за теоремою *про еквівалентність слабкої збіжності та в основному* $\xi + \eta \to^O \xi$. Згідно з означенням збіжності *в основному*

$$\mathbf{P}(\xi \in [a,b)) = \lim{}_{\varepsilon\downarrow 0}\mathbf{P}(\xi + \eta \in [a,b)) = \lim{}_{\varepsilon\downarrow 0}\int_a^b g_\varepsilon(t)dt,$$

для всіх точок неперервності функції розподілу F, що і доводить (а).

(б) За умови інтегровності границю у формулі (а) за *теоремою Лебега про мажоровану збіжність* можна внести під знак кратного інтегралу, оскільки підінтегральний вираз не перевищує за модулем функції $|\varphi_F(t)|$, яка інтегровна на $[a,b] \times \mathbb{R}$ за x,t. Таким переходом виводимо, що

$$F(b) - F(a) = \int_a^b f(t)dt,$$

де функція f визначена у формулюванні твердження (б). Спрямовуючи тут $a \to -\infty$, за означенням *щільності* розподілу робимо висновок, що функція розподілу F має щільність f $\boxdot$

Вправи

(1) Нехай ξ – випадкова величина з характеристичною функцією φ_ξ, а величини ζ_T не залежать від ξ і мають щільності $(1 - \cos Tx)/\ (\pi Tx^2)$, $x \in \mathbb{R}$.

(а) Обчислити характеристичні функції величин $\zeta_T : \varphi_{\zeta_T}(t) = (1 - |t|/T)\mathbb{1}_{|t|<T}$. (б) Вивести звідси слабку збіжність $\xi + \zeta_T \to^W \xi$, $T \to \infty$. (в) Довести формулу обертання для точок неперервності функції розподілу ξ :

$$F_\xi(b) - F_\xi(a) = \frac{1}{2\pi}\lim_{T\to\infty}\int_a^b dx \int_{-T}^T \exp(-itx)(1 - \tfrac{|t|}{T})\ \varphi_\xi(t)\ dt.$$

(2) Нехай ξ – випадкова величина з характеристичною функцією φ_ξ. Довести таку формулу обертання для точок неперервності функції розподілу ξ :

$$F_\xi(b) - F_\xi(a) = \frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^T (\exp(-ita) - \exp(-itb))(it)^{-1}\varphi_\xi(t)dt.$$

(3) Нехай ξ – випадкова величина з характеристичною функцією φ_ξ, а $\varepsilon > 0$. Довести, що $F_\xi([-2\varepsilon, 2\varepsilon]) \ge \varepsilon \left|\int_{-1/\varepsilon}^{1/\varepsilon} \varphi_\xi(t)dt\right| - 1$.

(4) Нехай φ – характеристична функція для функції розподілу F з множиною точок розриву $\{a_n\}$. Довести, що

$$\lim_{T\to\infty}(2T)^{-1}\int_{|t|\le T}\exp(-itx)\varphi(t)dt = F(x+0) - F(x),$$

$$\lim_{T\to\infty}(2T)^{-1}\int_{|t|\le T}|\varphi(t)|^2\,dt = \sum_{n\ge 1}(F(a_n+0) - F(a_n))^2.$$

(5, формула Пуассона) Нехай характеристична функція φ абсолютно інтегровна на $\mathbb{R}$, а f – відповідна щільність розподілу. Довести при $t, x \in \mathbb{R}$ тотожність $\sum_{k=-\infty}^\infty \varphi(x + 2kt) = \pi t^{-1}\sum_{n=-\infty}^\infty f(\pi n t^{-1})\exp(i\pi n t^{-1}x)$, за умови, що ряд у лівій частині збігається до неперервної функції.

(6) Характеристична функція φ випадкової величини ξ задовольняє асимптотичне зображення $\varphi(s) = 1 + O(|s|^\alpha), s \to 0$, для деякого $\alpha \in (0, 2]$. Довести, що $\mathbf{P}(|\xi| \ge x) = O(x^{-\alpha}), x \to \infty$.

(7) Випадкові величини $(\xi_k, k = \overline{1,n})$ незалежні та рівномірно розподілені на відрізку $(-1, 1)$. Довести, що сума $S_n = \xi_1 + \xi_2 + ... + \xi_n$ має щільність $\pi^{-1}\int_0^\infty (t^{-1}\sin t)^n \cos(tx)dt$.

(8) Нехай ξ – довільна випадкова величина з характеристичною функцією φ_ξ. Довести, що $\mathbf{E}\,|\xi| = \pi^{-1}\int_{-\infty}^\infty (1 - \mathrm{Re}\,\varphi_\xi(t))t^{-2}dt$.

1.27.6. Теорема Леві

Теорема (теорема Леві про критерій слабкої збіжності). *Нехай* $(F_n,\ n \ge 1)$ *– функції розподілу з характеристичними функціями* φ_n.

Слабка збіжність $F_n \to^W F$, $n \to \infty$, *до деякої функції розподілу* F *має місце тоді і тільки тоді, коли характеристичні функції* $\varphi_n(t)$ *збігаються при всіх* $t \in \mathbb{R}$ *до функції* $\varphi(t)$, *яка неперервна в нулі. У цьому випадку* φ *є характеристичною функцією для* F.

Доведення

(а) *Необхідність* випливає з означення *слабкої збіжності*, тому що гармоніки $\exp(itx) \in C_b(\mathbb{R})$ є неперервними обмеженими для всіх $t \in \mathbb{R}$.

(б) *Достатність.* Нехай $\varphi_n(t) \to \varphi(t),\ n \to \infty$, для всіх t, де функція $\varphi(t)$ неперервна в нулі.

Доведемо спочатку, що послідовність $(F_n,\ n \ge 1)$ є *слабко компактною.* Нехай $a > 0,\ b = 2/a$. Враховуючи означення характеристичної функції, абсолютну збіжність кратного інтегралу та *теорему Фубіні про кратний та повторні інтеграли*, обчислимо

$$\frac{1}{2a}\int_{-a}^{a}(1-\varphi_n(t))dt = \frac{1}{2a}\int_{-a}^{a}\left(\int_{-\infty}^{\infty}(1-\exp(itx))dF_n(x)\right)dt =$$

$$\frac{1}{2a}\int_{-\infty}^{\infty}\left(\int_{-a}^{a}(1-\exp(itx))dt\right)dF_n(x) =$$

$$\int_{-\infty}^{\infty}\left(1-\frac{\sin ax}{ax}\right)\,dF_n(x) \ge \int_{|x|\,\ge\,b}\left(1-\frac{\sin ax}{ax}\right)\,dF_n(x) \ge$$

$$\int_{|x|\,\ge\,b}\frac{1}{2}\,dF_n(x) \ge \frac{1}{2}\,F_n(\overline{[-b,b)}).$$

Справедливість цих нерівностей випливає, по-перше, з невід'ємності

$$1-\frac{\sin ax}{ax} \ge 0,$$

оскільки $|\sin y| \le |y|$, а по-друге, з оцінки

$$\left|\frac{\sin ax}{ax}\right| \le \frac{1}{|ax|} \le \frac{1}{2},\ \forall\,|x| \ge b = 2/a.$$

З нерівності для F_n та *теореми Лебега про мажоровану збіжність* маємо

$$\overline{\lim}_{n\to\infty}F_n(\overline{[-b,b)}) \le \frac{2}{2a}\overline{\lim}_{n\to\infty}\int_{-a}^{a}(1-\varphi_n(t))dt =$$

$$\frac{1}{a}\int_{-a}^{a}(1-\varphi(t))dt \to 0,\ a \to 0,\ b \to \infty,$$

оскільки функція $\varphi(t)$ неперервна в нулі та $\varphi(0) = \lim_{n\to\infty}\varphi_n(0) = 1$.

Отже, виконана умова теореми *про критерій слабкої компактності послідовності* функцій розподілу, і послідовність $(F_n,\ n \ge 1)$ є *слабко компактною.*

За означенням слабкої компактності (F_n) знайдемо підпослідовність $(F_{n_k},\ k \ge 1) \subset (F_n,\ n \ge 1)$, і функцію розподілу F такі, що $F_{n_k} \to^W F$, $k \to \infty$. Нехай φ_F – характеристична функція F. Тоді внаслідок вже доведеного твердження необхідності $\varphi_{n_k} \to \varphi_F$. Отже, $\varphi = \lim \varphi_n = \lim \varphi_{n_k} = \varphi_F$ є характеристичною функцією функції розподілу F.

Доведемо, що $F_n \to^W F$. Припустимо, що ця збіжність не має місця. Тоді за означенням *слабкої збіжності* знайдуться $\varepsilon > 0$, функція $g \in C_b(\mathbb{R})$

та нескінченна підпослідовність n_m такі, що

$$\left|\int_{-\infty}^{\infty} g(x)dF_{n_m}(x) - \int_{-\infty}^{\infty} g(x)dF(x)\right| \geq \varepsilon, \ \forall m.$$

Оскільки $(F_{n_m}, \ m \geq 1) \subset (F_n, \ n \geq 1)$ є підпослідовністю *слабко компактної* множини, то знайдеться слабко збіжна підпідпослідовність

$$(F_{n_{m_k}}, \ k \geq 1) \subset (F_{n_m}, \ m \geq 1), \ F_{n_{m_k}} \to^W G,$$

де G – деяка функція розподілу. Оскільки $\varphi_{n_{m_k}} \to \varphi_G$, то $\varphi_G = \varphi = \varphi_F$. Отже, $G = F$ за теоремою *про однозначність відповідності між функціями розподілу та характеристичними функціями*. Але в цьому випадку

$$\int_{-\infty}^{\infty} g(x)dF(x) = \int_{-\infty}^{\infty} g(x)dG(x) = \lim_{k\to\infty} \int_{-\infty}^{\infty} g(x)dF_{n_{m_k}}(x),$$

що суперечить нерівності, яка отримана вище: відстань між правою та лівою частинами повинна бути не меншою за $\varepsilon > 0$.

Отже, від супротивного $F_n \to^W F, n \to \infty$ $\boxdot$

Вправи

(1) Нехай величини ξ_n, ξ мають характеристичні функції φ_n, φ, причому $\int_a^b \varphi_n(t)dt \to \int_a^b \varphi(t)dt, n \to \infty$, для всіх $a < b$. Довести, що $\xi_n \to^W \xi$.

(2) Довести теорему Леві після заміни характеристичних функцій на перетворення Лапласа за припущення невід'ємності випадкових величин.

(3) Перевірити суттєвість умови неперервності в нулі в теореми Леві – побудувати функції розподілу F_n, що не збігаються слабко та мають характеристичні функції φ_n такі, що границя $\lim_{n\to\infty} \varphi_n(t)$ існує для всіх $t \in \mathbb{R}$.

(4) Випадкові величини $(\xi_n, \ n \geq 1)$ та ξ такі, що $|\xi_n| \leq c, |\xi| \leq c$ для сталої c, та $\mathbf{E}\xi_n^k \to \mathbf{E}\xi^k, n \to \infty$, при всіх $k \geq 1$. Довести, що $\xi_n \to^W \xi$.

(5) Випадкові величини $(\xi_n, n \geq 1)$ незалежні однаково розподілені на відрізку $[0, 1)$, мають обмежені щільності і $|\mathbf{E}\exp(2\pi i k\xi_1)| < 1$ для всіх $k \in \mathbb{N}\backslash\{0\}$. Довести, що $S_n = (\xi_1 + .. + \xi_n)(\text{mod } 1) \to^W \chi$, де величина $\chi \simeq U(0, 1)$.

(6) Випадкові величини ξ_n мають характеристичні функції φ_n. Довести, що $\xi_n \to^W 0$ тоді й тільки тоді, коли $\varphi_n(t) \to 1$ для t з деякого околу нуля.

(7) Функції розподілу F_n мають характеристичні функції φ_n, і для кожного $t \in \mathbb{R}$ існує границя $\lim_{n\to\infty} \varphi_n(t) = \varphi(t)$. Довести, що існує узагальнена функція розподілу $F \in \mathfrak{M}_{01}$ така, що $F_n \to^O F, n \to \infty$. Довести еквівалентність таких тверджень: (а) існує функція розподілу F така, що $F_n \to^W F, n \to \infty$, (б) послідовність (F_n) слабко компактна, (в) φ – характеристична функція, (г) φ – неперервна, (д) φ – неперервна в нулі.

(8) Сім'я функцій розподілу (F_α) слабко компактна тоді й тільки тоді, коли $\sup_\alpha \sup_{|t|\leq\varepsilon} \left|1 - \varphi_{F_\alpha}(t)\right| \to 0, \varepsilon \to 0$.

(9) Випадкові величини $(\xi_n,\ n \geq 1)$ незалежні, однаково розподілені з характеристичною функцією φ. Довести, що збіжність $(\xi_1 + ... + \xi_n)/n \to^P a$ до деякої сталої a виконується тоді й тільки тоді, коли існує похідна $\varphi'(0)$.

(10, теорема Хінчина) Випадкові величини $(\xi_n,\ n \geq 1)$ незалежні, однаково розподілені та інтегровні. Довести, що $(\xi_1 + ... + \xi_n)/n \to^P \mathbf{E}\xi_1$. *Вказівка*: досить довести слабку збіжність.

(11) Величини $(\xi_n, n \geq 1)$ незалежні, $\mathbf{P}(\xi_n = 0) = 1/2 = \mathbf{P}(\xi_n = 1)$. Довести, що: (а) сума ряду $\sum_{n\geq 1} 2^{-n}\xi_n$ рівномірно розподілена на $[0, 1]$, (б) величина $\xi = 2\sum_{n\geq 1} 3^{-n}\xi_n$ має (сингулярний) розподіл Кантора. Обчислити відповідну характеристичну функцію.

(12) Сформулювати теоретико-ймовірнісну інтерпретацію такої тотожності: $(\sin t)/t = \prod_{n\geq 1} \cos(2^{-n}t)$.

(13) Нехай $\varphi(t)$ – характеристична функція для функції розподілу $F(x)$. (а) Тоді функція $\varphi_\alpha(t) \equiv (2\varphi(t) - \varphi(t + \alpha) - \varphi(t - \alpha))/(2 - \varphi(\alpha) - \varphi(-\alpha))$ також є характеристичною при $\alpha \in \mathbb{R}$. (б) За умови абсолютної неперервності F довести, що $\varphi_\alpha(t) \to \varphi(t), \alpha \to \infty$, але відповідні щільності не збігаються.

Теорема (про рівномірну збіжність характеристичних функцій). *Нехай φ і $\varphi_n, n \geq 1$, – характеристичні функції та $\varphi_n(t) \to \varphi(t), n \to \infty$, при всіх t. Тоді функції $\varphi_n(t)$ є неперервними за t рівномірно по t і n, а збіжність $\varphi_n(t) \to \varphi(t)$ є рівномірною за t на кожному обмеженому інтервалі.*

Доведення. Нехай F_n – функція розподілу, що відповідає φ_n. За *теоремою Леві про критерій слабкої збіжності* послідовність F_n слабко збігається, а тому є *слабко компактною*. Отже, за *теоремою Прохорова про критерій слабкої компактності* $\sup_n F_n(\overline{[-c, c)}) \to 0,\ c \to \infty$. Звідси

$$|\varphi_n(t) - \varphi_n(s)| \ \leq \int_{|x|>c} |e^{itx} - e^{isx}|\, dF_n(x) + \int_{|x|\leq c} |e^{itx} - e^{isx}|\, dF_n(x) \leq$$
$$2\ \sup_n \int_{|x|>c} dF_n(x) + \sup_{|x|\leq c} \left|e^{i(t-s)x} - 1\right| \int_{|x|\leq c} dF_n(x) \leq$$
$$2\ \sup_n F_n(\overline{[-c, c)}) +\ c\,|t - s| \to 0,\ |t - s| \to 0,$$

тобто функції $\varphi_n(t)$ є рівномірно за n і t неперервними.

Якщо вказана в теоремі збіжність не є рівномірною за $t \in [-T, T]$, то знайдуться $\delta > 0$ і $t_n \in [-T, T]$ такі, що $|\varphi_n(t_n) - \varphi(t_n)| \geq \delta$ для всіх n. Переходячи до підпослідовності, можемо вважати, що $t_n \to t_0$. Отже,

$$\delta \leq \overline{\lim}_{n\to\infty} |\varphi_n(t_n) - \varphi(t_n)| \leq \overline{\lim}_{n\to\infty} |\varphi_n(t_n) - \varphi_n(t_0)| +$$
$$\overline{\lim}_{n\to\infty} |\varphi_n(t_0) - \varphi(t_0)| + \overline{\lim}_{n\to\infty} |\varphi(t_n) - \varphi(t_0)| = 0,$$

де враховано рівномірну неперервність φ_n, що доведена вище. Отримана суперечність свідчить про справедливість твердження теореми $\boxdot$

Теорема (про добуток слабко збіжної послідовності зі збіжною). *Нехай* $(\xi_n,\ n \geq 1)$ *послідовність випадкових величин така, що* $\xi_n \to^W \xi$, $n \to \infty$, *а числова послідовність* $a_n \to a, n \to \infty$. *Тоді*

$$a_n \xi_n \to^W a\xi,\ n \to \infty.$$

Доведення. Позначимо через φ_n – характеристичну функцію ξ_n.

Зі збіжності $\varphi_n(t) \to \varphi(t), n \to \infty$, та з теореми *про рівномірну збіжність характеристичних функцій* виводимо, що вказана збіжність є рівномірною за t на обмеженому інтервалі. Оскільки послідовність $a_n t \to at$ є обмеженою, то внаслідок рівномірності збіжності $\varphi_n(a_n t) \to \varphi(at)$ для всіх t, що за *теоремою Леві про критерій слабкої збіжності* доводить шукане твердження $\boxdot$

Вправи

(1) Довести, що в останній теоремі $a_n + \xi_n \to^W a + \xi, n \to \infty$.

(2) Довести, що в попередніх вправі та теоремі послідовність сталих $a_n \to a$ можна замінити на послідовність випадкових величин $\eta_n \to^P a$.

1.27.7. Сумісна характеристична функція та слабка збіжність випадкових векторів

Означення. Сумісною характеристичною функцією *випадкового вектора* $\xi = (\xi_1, ..., \xi_d)$ *називається така функція від* $t = (t_1, ..., t_d) \in \mathbb{R}^d$:

$$\varphi_\xi(t) = \mathbf{E}\exp(it'\xi) = \mathbf{E}\exp\left(\sum_{k=1}^{d} it_k\xi_k\right).$$

Сумісна характеристична функція має такі ж властивості, що і звичайна *характеристична функція*. Зокрема, відповідність між сумісними функціями розподілу та характеристичними функціями є взаємно однозначною.

Означення. *Випадкові вектори* $\xi_n \in \mathbb{R}^d$ *слабко збігаються:* $\xi_n \to^W \xi$, $n \to \infty$, *якщо* $\mathbf{E}g(\xi_n) \to \mathbf{E}g(\xi)$ *для всіх* $g \in C_b(\mathbb{R}^d)$.

Теорема (теорема Леві про критерій слабкої збіжності випадкових векторів). *Послідовність* d*-вимірних випадкових векторів слабко збігається:* $\xi_n \to^W \xi, n \to \infty$, *тоді й тільки тоді, коли поточково збігаються відповідні сумісні характеристичні функції:* $\varphi_n(t) \to \varphi(t)$, $n \to \infty, \forall t \in \mathbb{R}^d$, *до границі* $\varphi(t)$, *що неперервна в нулі.*

Доведення повністю аналогічне доведенню для скалярного випадку.

Теорема (теорема Крамера-Волда про критерій слабкої збіжності випадкових векторів). *Послідовність d-вимірних випадкових векторів слабко збігається: $\xi_n \to^W \xi, n \to \infty$, тоді й тільки тоді, коли для довільного $t \in \mathbb{R}^d$ слабко збігається лінійна форма $t'\xi_n \to^W t'\xi, n \to \infty$.*

Доведення. Нехай φ_n, φ – сумісні характеристичні функції векторів ξ_n, ξ. За *теоремою Леві про критерій слабкої збіжності* збіжність при кожному $t \in \mathbb{R}^d$ лінійних форм: $t'\xi_n \to^W t'\xi$, $n \to \infty$, еквівалентна збіжності характеристичних функцій $\mathbf{E}\exp(is\ t'\xi_n) \to \mathbf{E}\exp(is\ t'\xi)$ для всіх $s \in \mathbb{R}$. Остання лише формою запису відрізняється від збіжності $\varphi_n(st) \to \varphi(st)$, $\forall s \in \mathbb{R}$. Нарешті, за *теоремою Леві про критерій слабкої збіжності випадкових векторів* $\xi_n \to^W \xi$ тоді й тільки тоді, коли $\varphi_n(st) \to \varphi(st)$, $\forall s \in \mathbb{R}, \forall t \in \mathbb{R}^d$, оскільки множина $\{st,\ s \in \mathbb{R}, t \in \mathbb{R}^d\} = \mathbb{R}^d$ $\boxdot$

Вправи

(1) Нехай $\xi_n \to^W \xi \in \mathbb{R}^d$, а функція $g \in C(\mathbb{R}^d)$. Довести, що має місце збіжність $g(\xi_n) \to^W g(\xi), n \to \infty$.

(2) Довести, що (а) сумісна характеристична функція нормального випадкового вектора $\xi \simeq N_n(m, V)$ дорівнює $\varphi_\xi(t) = \exp(it'm - t'Vt/2)$, (б) ξ є узагальненим нормальним вектором тоді й тільки тоді, коли остання рівність має місце для деяких $m \in \mathbb{R}^n$ і симетричної невід'ємно визначеної матриці V.

(3) Вектор $\xi = (\xi_1, ..., \xi_n)$ є нормальним випадковим вектором тоді й тільки тоді, коли для кожного $t \in \mathbb{R}^n \backslash \{0\}$ величина $t'\xi$ має нормальний розподіл.

(4) Нехай послідовності випадкових величин $(\xi_n), (\eta_n)$ такі, що ξ_n, η_n незалежні при кожному n, $\xi_n \to^W \xi$, $\eta_n \to^W \eta$, причому ξ та η незалежні. Довести слабку збіжність векторів $(\xi_n, \eta_n) \to^W (\xi, \eta)$.

(5) Нехай φ, ψ, γ – характеристичні функції випадкових величин $\xi.\eta, \zeta$. Довести, що функція $\varphi(t_1)\psi(t_2)\gamma(t_1 + t_2)$ є сумісною характеристичною функцією двовимірного вектора. Побудувати цей вектор.

(6) Координати вектора $\xi = (\xi_1, ..., \xi_n)$ незалежні тоді й тільки тоді, коли їх сумісна характеристична функція є добутком: $\varphi_\xi(t) = \prod_{k=1}^n \varphi_{\xi_k}(t_k), \forall t \in \mathbb{R}^d$.

(7) Довести рівність $\mathbf{E}\xi\eta = \mathbf{E}\xi\mathbf{E}\eta$ для комплекснозначних незалежних інтегровних величин ξ, η.

1.28. Класична центральна гранична теорема

У центральній граничній теоремі доводиться, що центрована та нормована сума незалежних величин слабко збігається до нормальної величини для практично довільних *розподілів* окремих доданків. Цей клас теорем є підставою для широкого застосування *нормального розподілу* в статистиці, принаймні у випадку великої кількості спостережень..

Теорема (класична центральна гранична теорема). *Нехай* (ξ_n) – *послідовність незалежних у сукупності однаково розподілених випадкових величин зі скінченними середніми та дисперсіями:* $\mu = \mathbf{E}\xi_1$, $\sigma^2 = \mathbf{D}\xi_1 > 0$, *а* $S_n = \xi_1 + ... + \ \xi_n, n \geq 1$. *Тоді*

$$(S_n - \mathbf{E}S_n)/\sqrt{\mathbf{D}S_n} = (S_n - n\mu)/\sigma\sqrt{n} \to^W \zeta \simeq N(0,1),\ n \to \infty.$$

Наслідок. *Оскільки збіжність в основному випливає зі слабкої збіжності, а нормальна функція розподілу неперервна, то для всіх* $x \in \mathbb{R}$ *має місце збіжність функцій розподілу*

$$\mathbf{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < x\right) \to \Phi(x) \equiv \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\ y^2\right) dy,\ n \to \infty.$$

Зокрема, звідси при довільному розподілі окремих доданків за правилом трьох сигма для досить великих n *випливає наближена рівність*

$$\mathbf{P}(|S_n - n\mu| < 3\sigma\sqrt{n}) \approx 0.997.$$

Доведення. Рівності $\mathbf{E}S_n = n\mu$, $\mathbf{D}S_n = \sigma^2 n$ є наслідками незалежності та однакової розподіленості доданків, за *лінійністю* математичного сподівання та за теоремою *про дисперсію суми незалежних величин.*

Нехай φ – характеристична функція ξ_1. Тоді за теоремою *про властивості характеристичної функції*, пункти (а) та (б):

$$\varphi_n(t) \equiv \mathbf{E}\exp(it((S_n - n\mu)/\sigma\sqrt{n})) =$$

$$\exp(-it\ n\ \mu/\sigma\sqrt{n})\ \varphi_{S_n}(t/\sigma\sqrt{n}) = \exp(-it\ n\ \mu/\sigma\sqrt{n})\ \varphi^n(t/\sigma\sqrt{n}).$$

Використаємо розклад у ряд Тейлора логарифмічної функції та твердження (г) теореми *про властивості характеристичної функції*:

$$\ln\varphi(s) = \ln(1 + (\varphi(s) - 1)) = (\varphi(s) - 1) - (\varphi(s) - 1)^2/2 + o(s^2) =$$

$$is\mu - s^2\mu_2/2 + s^2\mu^2/2 + o(s^2) = is\mu - s^2\sigma^2/2 + o(s^2),\ s \to 0.$$

Звідси отримаємо

$$\ln\varphi_n(t) = -it\ n\ \mu\ /\ \sigma\sqrt{n}+$$

$$n\left(it\ \mu/\sigma\sqrt{n} - \sigma^2 t^2\ /\ 2\sigma^2 n + o(1/n)\right) = -t^2/2\ + o(1),\ n \to \infty.$$

Отже, $\varphi_n(t) \to \exp(-t^2/2) = \varphi_\zeta(t)$. Застосування *теореми Леві про критерій слабкої збіжності* довершує доведення $\boxdot$

Вправи

(1) Довести класичну центральну граничну теорему без вживання комплекснозначної логарифмічної функції за нерівністю $|z^n - \zeta^n| \le n\,|z - \zeta|$.

(2) Випадкова величина $\zeta_{\lambda\alpha}$ має гама-розподіл $\Gamma(\lambda, \alpha)$.Довести слабку збіжність $(\lambda\zeta_{\lambda\alpha} - \alpha)/\sqrt{\alpha} \to^W \zeta \simeq N(0,1), \alpha \to \infty$.

(3) Випадкова величина ζ_λ має розподіл Пуассона $\Pi(\lambda)$. (а) Довести, що $(\zeta_\lambda - \lambda)/\sqrt{\lambda} \to^W \zeta \simeq N(0,1), \lambda \to \infty$. (б) Перевірити, що вказане наближення можна вважати прийнятним вже при $\lambda \ge 5$. (в) Вивести звідси тотожність $\lim_{n\to\infty} \exp(-n) \sum_{k=0}^n n^k/k! = 1/2$.

(4) Випадкові величини $(\xi_n,\ n \ge 1)$ незалежні, однаково розподілені, та $a_n^{-1}(\xi_1 + ... + \xi_n) - b_n \to^W \eta, n \to \infty$, для деяких $a_n > 0, b_n$, причому $\eta \ne 0$ м.н. Довести, що $a_n \to \infty, a_n/a_{n-1} \to 1, n \to \infty$.

(5) Випадкові величини $(\xi_n,\ n \ge 1)$ незалежні, однаково рівномірно розподілені на інтервалі $[0,1]$ (а) Знайти таке перетворення добутків $\prod_{k=1}^n \xi_k$, яке б слабко збігалося до стандартної нормальної випадкової величини $N(0,1)$. (б) Обчислити граничний при $n \to \infty$ розподіл величин $n \min_{1\le k\le n} \xi_k$.

(6) Величини $(\xi_n, n \ge 1)$ незалежні і мають щільність Коші $(\pi(1 + x^2))^{-1}$. Знайти граничний розподіл для максимумів $n^{-1} \max_{1\le k\le n} \xi_k$.

(7) Величини $(\xi_n, n \ge 1)$ незалежні однаково розподілені, $\mathbf{E}\xi_1 = 0$, $\mathbf{D}\xi_1 = 1$. Довести, що $\max_{1\le k\le n} |\xi_k| /\sqrt{n} \to^W 0$.

(8) Випадкові величини $(\xi_n, n \ge 1)$ незалежні, $\mathbf{P}(\xi_n = \pm 1) = 1/2$. Довести, що $\max_{1\le k\le n}(\xi_1 + ... + \xi_k)/\sqrt{n} \to^W |\zeta|, n \to \infty$, де $\zeta \simeq N(0,1)$.

(9) Випадкові величини $(\xi_n, n \ge 1)$ незалежні, однаково розподілені, та мають неперервну парну додатну у околі нуля щільність. Довести слабку збіжність $n^{-1} \sum_{k=1}^n \xi_k^{-1} \to^W \kappa,\ n \to \infty$, де κ має розподіл Коші.

(10) Випадкові вектори $(\xi_{nk}, k = \overline{1, n+1})$ рівномірно розподілені на одиничній сфері у $\mathbb{R}^{n+1}$. Довести, що $\sqrt{n}\xi_{n1} \to^W \zeta, n \to \infty,\ \zeta \simeq N(0,1)$.

(11) Випадкові величини $(\xi_{nk},\ k = \overline{1,n},\ n \ge 1)$ незалежні при кожному n, $\xi_{nk} \in \{0,1\}$, а ймовірності $p_{nk} \equiv \mathbf{P}(\xi_{nk} = 1)$ такі, що $\max_{k\le n} p_{nk} \to 0$. Довести, що збіжність $\sum_{k=1}^n p_{nk} \to \lambda, n \to \infty$, необхідна і достатня для того, щоб $\mathbf{P}\left(\sum_{k=1}^n \xi_{nk} = m\right) \to \exp(-\lambda)\lambda^m/m!$, $\forall m \in \mathbb{Z}_+$.

(12) Величини $(\xi_n, n \ge 1)$ задовольняють умови класичної центральної граничної теореми, а функція $g \in C_2(\mathbb{R})$ така, що $g'(\mu) = 0$. Довести такі збіжності: (а) $n\left(g(S_n/n) - g(\mu)\right) \to^W g''(\mu)\sigma^2\zeta^2/2, n \to \infty$, де величина $\zeta \simeq N(0,1)$, (б) $S_n(n - S_n)/n - n/4 \to^W -\sigma^2\zeta^2/2, n \to \infty$, при $\mu = 1/2$.

1.29. Загальна гранична теорема для стандартних серій

1.29.1. Стандартні послідовності серій

Означення. Стандартною послідовністю серій *називається подвійна послідовність випадкових величин* $(\zeta_{nk},\ k = \overline{1, k_n},\ n \geq 1)$, *що задовольняє такі умови:*

(1 – незалежність) випадкові величини в кожній серії $(\zeta_{nk},\ k = \overline{1, k_n})$ *незалежні в сукупності,*

(2 – центрованість) $\mathbf{E}\zeta_{nk} = 0,\ \forall k,\ \forall n,$

(3 – нормованість) $\exists\ \mathbf{D}\zeta_{nk} = \sigma_{nk}^2$ *і* $\sum_{k=1}^{k_n} \sigma_{nk}^2 = 1.$

(4 – рівномірна малість) $\max_{1 \leq k \leq k_n} \sigma_{nk}^2 \to 0$ *при* $n \to \infty.$

Приклад. Якщо величини $(\xi_n, n \geq 1)$ задовольняють умови *класичної центральної граничної теореми*, то послідовність

$$\zeta_{nk} \equiv (\xi_k - \mu) \ / \ \sigma\sqrt{n},\ k = \overline{1, n},$$

є *стандартною послідовністю серій* (**Вправа**).

Нехай $(\zeta_{nk},\ k = \overline{1, k_n},\ n \geq 1)$ – стандартна послідовність серій. Введемо такі позначення:

$$\zeta_n = \sum\nolimits_{k=1}^{k_n} \zeta_{nk},\ \varphi_n(t) = \mathbf{E}\exp(it\zeta_n),$$

$$F_{nk}(x) = \mathbf{P}(\zeta_{nk} < x),\ \varphi_{nk}(t) = \mathbf{E}\exp(it\zeta_{nk}).$$

За *адитивністю* математичного сподівання, теоремою *про дисперсію суми незалежних величин* та теоремою *про властивості характеристичної функції*, пункт (б),

$$\mathbf{E}\zeta_n = 0,\ \mathbf{D}\zeta_n = 1,\ \varphi_n(t) = \prod\nolimits_{k=1}^{k_n} \varphi_{nk}(t).$$

1.29.2. Допоміжні леми

Доведення загальної граничної теореми базується на таких лемах.

Лема 1. *Для стандартної послідовності серій виконується граничне співвідношення*

$$\ln \varphi_n(t) - \sum\nolimits_{k=1}^{k_n} (\varphi_{nk}(t) - 1) \to 0,\ n \to \infty.$$

Доведення. З елементарної нерівності $|\exp(ix) - 1 - ix| \leq x^2/2$ та з умови центрованості $\mathbf{E}\zeta_{nk} = 0$ означення стандартних серій дістаємо

$$|\varphi_{nk}(t) - 1| = |\mathbf{E}(\exp(it\zeta_{nk}) - 1 - it\zeta_{nk})| \leq t^2\sigma_{nk}^2 \leq t^2 \max_k \sigma_{nk}^2 \to 0, n \to \infty.$$

Зокрема, $|\varphi_{nk}(t)| \geq 1/2$, починаючи з деякого n при всіх k, отже коректно визначена логарифмічна функція на комплексній площині як від $\varphi_{nk}(t)$, так і від $\varphi_n(t)$. З урахуванням нерівності $|\ln(1+z) - z| \leq |z|^2$ для всіх $z \in \mathbb{C}$, та попередньої нерівності оцінимо

$$\left|\ln \varphi_n(t) - \sum_{k=1}^{k_n} (\varphi_{nk}(t) - 1)\right| =$$

$$\left|\sum_{k=1}^{k_n} (\ln \varphi_{nk}(t) - \varphi_{nk}(t) + 1)\right| \leq \sum_{k=1}^{k_n} |\varphi_{nk}(t) - 1|^2 \leq$$

$$\sum_{k=1}^{k_n} t^4\sigma_{nk}^4 \leq t^4 \max_k \sigma_{nk}^2 \sum_{k=1}^{k_n} \sigma_{nk}^2 = t^4 \max_k \sigma_{nk}^2 \to 0,\ n \to \infty,$$

де враховане також означення стандартних серій $\boxdot$

Лема 2. (а) *При кожному n функції*

$$G_n(x) \equiv \sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2 1\!\!\mathrm{I}_{\{\zeta_{nk}\ <\ x\}} = \sum_{k=1}^{k_n} \int_{(-\infty,\, x)} y^2\, dF_{nk}(y)$$

є функціями розподілу.

(б) *Для всіх обмежених борелєвих функцій g має місце тотожність*

$$\sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2\, g(\zeta_{nk}) = \int_{-\infty}^{\infty} g(y)\, dG_n(y).$$

Доведення

(а) Монотонність G_n очевидна. *Нормованість* випливає з *теореми Лебега про монотонну збіжність* і нормованості (3) *стандартної послідовності серій*:

$$\lim_{x\to\infty} G_n(x) = \textstyle\sum_{k=1}^{k_n} \lim_{x\to\infty} \mathbf{E}\zeta_{nk}^2 1\!\!\mathrm{I}_{\{\zeta_{nk}\ <\ x\}} =$$

$$\textstyle\sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2 = \sum_{k=1}^{k_n} \mathbf{D}\zeta_{nk} = \sum_{k=1}^{k_n} \sigma_{nk}^2 = 1.$$

Неперервність G_n зліва також виводиться з *теореми Лебега про монотонну збіжність*, з урахуванням неперервності зліва функції $1\!\!\mathrm{I}_{\{\xi\ <\ y\}}$:

$$\lim_{y\uparrow x} G_n(y) = \sum_{k=1}^{k_n} \lim_{y\uparrow x}\ \mathbf{E}\zeta_{nk}^2 1\!\!\mathrm{I}_{\{\zeta_{nk}\ <\ y\}} = \sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2 1\!\!\mathrm{I}_{\{\zeta_{nk}\ <\ x\}} = G_n(x).$$

(б) Позначимо через $\mathfrak{G}$ клас борелєвих функцій g, для яких виконується тотожність твердження (б) леми:

$$\mathfrak{G} = \left\{g : \sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2\, g(\zeta_{nk}) = \int_{-\infty}^{\infty} g(y)\, dG_n(y)\right\}.$$

З означення G_n випливає, що $g(y) = 1\!\!\mathrm{I}_{y\,<\,x} \in \mathfrak{G}$ при довільному $x \in \mathbb{R}$. З лінійності $\mathfrak{G}$ виводимо, що $1\!\!\mathrm{I}_A \in \mathfrak{G}$ для кожної множини A з алгебри $\mathfrak{A}(\mathbb{R})$. Крім того, за *теоремою Лебега про монотонну збіжність* та *теоремою Лебега про мажоровану збіжність* для *математичного сподівання* та для *інтеграла Лебега – Стілтьєса* клас $\mathfrak{G}$ замкнений відносно монотонної та мажорованої збіжності. Тому, зокрема, клас $\{B \in \mathfrak{B}(\mathbb{R}) : 1\!\!\mathrm{I}_B \in \mathfrak{G}\}$ замкнений відносно *монотонної збіжності* множин, звідки зі включень $1\!\!\mathrm{I}_A \in \mathfrak{G}$, $\forall A \in \mathfrak{A}(\mathbb{R})$ за теоремою *про монотонний клас* виводимо, що $1\!\!\mathrm{I}_B \in \mathfrak{G}$, $\forall B \in \mathfrak{B}(\mathbb{R})$. За лінійністю $\mathfrak{G}$ містить всі *прості* функції, а за *теоремою Лебега про мажоровану збіжність* – всі обмежені *борелєві* $\boxdot$

Лема 3. *Визначимо при кожному $t \in \mathbb{R}$ функцію*

$$f_t(x) = \begin{cases} \ (\exp(itx) - 1 - itx)\ /\ x^2, & x \neq 0, \\ -\ t^2/2, & x = 0. \end{cases}$$

(а) $f_t(\cdot) \in C_b(\mathbb{R})$, $\forall t \in \mathbb{R}$.

(б) Для всіх $t \in \mathbb{R}$ виконується тотожність

$$\sum\nolimits_{k=1}^{k_n} (\varphi_{nk}(t) - 1) = \int_{-\infty}^{\infty} f_t(x)\, dG_n(x),$$

де функції розподілу G_n визначені в лемі 2.

Доведення

(а) Включення $f_t \in C_b(\mathbb{R})$ очевидне, оскільки значення $f_t(0)$ є продовженням означення f_t за неперервністю в точці $x = 0$.

(б) Застосуємо тотожність леми 2 до функції $g(x) = f_t(x)$:

$$\int_{-\infty}^{\infty} f_t(x)\, dG_n(x) = \sum\nolimits_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2\, f_t(\zeta_{nk}) =$$

$$\sum\nolimits_{k=1}^{k_n} \mathbf{E}(\exp(it\zeta_{nk}) - 1 - it\zeta_{nk}) = \sum\nolimits_{k=1}^{k_n} (\varphi_{nk}(t) - 1),$$

оскільки за означенням $x^2 f_t(x) = \exp(itx) - 1 - itx$, і за властивостями *стандартної послідовності серій* $\mathbf{E}\zeta_{nk} = 0$ $\boxdot$

1.29.3. Гранична теорема для стандартних серій

Теорема (загальна гранична теорема для стандартних серій). *Нехай $(\zeta_{nk},\ k = \overline{1, k_n},\ n \geq 1)$ – стандартна послідовність серій випадкових величин. Припустимо, що послідовність функцій розподілу*

$$G_n(x) \equiv \sum\nolimits_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2 1\!\!\mathrm{I}_{\{\zeta_{nk}\,<\,x\}}$$

слабко збігається: $G_n \to^W G$, до деякої функції розподілу G.

Тоді суми у серіях слабко збігаються: $\zeta_n \equiv \sum_{k=1}^{k_n} \zeta_{nk} \to^W \zeta$, *при* $n \to \infty$, *до випадкової величини* ζ *із характеристичною функцією*

$$\varphi_\zeta(t) = \exp\left(\int_{-\infty}^{\infty} f_t(x)\, dG(x)\right),$$

де функція $f_t(x) \equiv (\exp(itx) - 1 - itx) \;/\; x^2$.

Доведення. Функція G_n є *функцією розподілу* за лемою 2.

Оскільки $f_t \in C_b(\mathbb{R})$ за лемою 3, то зі збіжності $G_n \to^W G$ та з лем 3, 1 виводимо, що

$$\ln\varphi_n(t) - \int_{-\infty}^{\infty} f_t(x)\, dG(x) = \ln\varphi_n(t) - \sum_{k=1}^{k_n} (\varphi_{nk}(t) - 1) +$$

$$+ \int_{-\infty}^{\infty} f_t(x)\, dG_n(x) - \int_{-\infty}^{\infty} f_t(x)\, dG(x) \to 0,\ n \to \infty,\quad \forall t \in \mathbb{R}.$$

Отже, $\varphi_n(t) \to \varphi_\zeta(t), n \to \infty, \forall t \in \mathbb{R}$. Функція $\varphi_\zeta(t)$ неперервна в нулі, тому що інтеграл під знаком експоненти в означенні φ_ζ неперервний в нулі за t за *теоремою Лебега про мажоровану збіжність*, оскільки $|f_t(x)| \le 1$ при $|t| \le 1$ та $f_t(x) \to f_0(x) \equiv 0,\ t \to 0$, для всіх x. Тому за *теоремою Леві про критерій слабкої збіжності* $\varphi_\zeta(t)$ є характеристичною функцією і $\zeta_n \to^W \zeta, n \to \infty$ $\boxdot$

1.30. Центральні граничні теореми для послідовностей серій

1.30.1. Теорема Ліндеберга для стандартних серій

Теорема (центральна гранична теорема Ліндеберга для стандартних серій). *Нехай* $(\zeta_{nk},\ k = \overline{1, k_n},\ n \ge 1)$ – *стандартна послідовність серій випадкових величин, що задовольняє* ***умову Ліндеберга***:

$$L_n(\varepsilon) \equiv \sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2\ \mathbb{1}_{\{|\zeta_{nk}| \,\ge\, \varepsilon\}} \to 0,\ n \to \infty,\ \forall\ \varepsilon > 0.$$

Тоді має місце збіжність

$$\zeta_n \equiv \sum_{k=1}^{k_n} \zeta_{nk} \to^W \zeta \simeq N(0,1).$$

Зауваження. Умова рівномірної малості (г) в означенні *стандартної послідовності серій* випливає з умови Ліндеберга, оскільки

$$\max\nolimits_{k \le k_n} \mathbf{E}\zeta_{nk}^2 = \max\nolimits_{k \le k_n} \left(\mathbf{E}\zeta_{nk}^2 (\mathbb{1}_{\{|\zeta_{nk}| < \varepsilon\}} + \zeta_{nk}^2 \mathbb{1}_{\{|\zeta_{nk}| \ge \varepsilon\}})\right) \le \varepsilon^2 + L_n(\varepsilon).$$

Зауваження. В.Феллер довів, що *умова Ліндеберга* є необхідною для асимптотичної нормальності сум у стандартній послідовності серій.

Доведення теореми. Досить оцінити в умовах *загальної граничної теореми для стандартних серій* при кожному $\varepsilon > 0$

$$G_n(-\varepsilon) + 1 - G_n(\varepsilon) = \sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2 \, 1\!\mathrm{I}_{\{\zeta_{nk} < -\varepsilon\}} +$$

$$\sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2 - \sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2 \, 1\!\mathrm{I}_{\{\zeta_{nk} < \varepsilon\}} = \sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2 \, 1\!\mathrm{I}_{\{\zeta_{nk} < -\varepsilon\} \cup \{\zeta_{nk} \geq \varepsilon\}} \leq$$

$$\sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2 \, 1\!\mathrm{I}_{\{|\zeta_{nk}| \geq \varepsilon\}} = L_n(\varepsilon) \to 0, n \to \infty,$$

отже, $G_n(-\varepsilon) \to 0$, $G_n(\varepsilon) \to 1, n \to \infty$, $\forall \varepsilon > 0$ і $G_n \to^O G, n \to \infty$, де $G(x) = 1\!\mathrm{I}_{0 < x}$ є *функцією розподілу*.

Тому за теоремою *про еквівалентність слабкої збіжності та в основному* $G_n \to^W G, n \to \infty$. Отже, згідно з *загальною граничною теоремою для стандартних серій* $\zeta_n \to^W \zeta, n \to \infty$, де гранична характеристична функція дорівнює

$$\varphi_\zeta(t) = \exp\left(\int_{-\infty}^{\infty} f_t(x) \, dG(x)\right) = \exp(f_t(0)) = \exp(-t^2/2)$$

і є характеристичною функцією *стандартного нормального розподілу*. Звідси за теоремою *про однозначну відповідність між функціями розподілу та характеристичними функціями* $\zeta \simeq N(0,1)$ $\boxdot$

1.30.2. Теорема Ляпунова для стандартних серій

Теорема (центральна гранична теорема Ляпунова для стандартних серій). *Нехай* $(\zeta_{nk}, \; k = \overline{1, k_n}, \; n \geq 1)$ – *стандартна послідовність серій випадкових величин, що задовольняє* ***умову Ляпунова****:*

$$\exists \delta > 0 : \; L_n^{2+\delta} \equiv \sum_{k=1}^{k_n} \mathbf{E} \left|\zeta_{nk}\right|^{2+\delta} \to 0, \; n \to \infty.$$

Тоді має місце збіжність

$$\zeta_n \equiv \sum_{k=1}^{k_n} \zeta_{nk} \to^W \zeta \simeq N(0,1).$$

Доведення. Оцінимо при всіх $\varepsilon > 0$ показник Ліндеберга

$$L_n(\varepsilon) = \sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2 \, 1\!\mathrm{I}_{\{|\zeta_{nk}| \geq \varepsilon\}} \leq$$

$$\sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2 \left(\left|\zeta_{nk}\right| / \varepsilon\right)^\delta \, 1\!\mathrm{I}_{\{|\zeta_{nk}| \geq \varepsilon\}} \leq \varepsilon^{-\delta} L_n^{2+\delta} \to 0, n \to \infty,$$

оскільки $\left|\zeta_{nk}\right| / \varepsilon \geq 1$ на множині $\{\left|\zeta_{nk}\right| \geq \varepsilon\}$. Отже, з умови Ляпунова випливає *умова Ліндеберга*, а отже, і асимптотична нормальність $\boxdot$

1.30.3. Теорема Ліндеберга для загальних серій

Означення. Загальною послідовністю серій *називається подвійна послідовність випадкових величин* $(\xi_{nk},\ k=\overline{1,k_n},\ n\geq 1)$ *така, що:*
(1) величини в кожній серії $(\xi_{nk},\ k=\overline{1,k_n})$ *незалежні в сукупності,*
(2) існують $\mathbf{E}\xi_{nk}=\mu_{nk}$, $\mu_n\equiv\sum_{k=1}^{k_n}\mu_{nk}$,
(3) існують $\mathbf{D}\xi_{nk}=\sigma_{nk}^2$, $\sigma_n^2\equiv\sum_{k=1}^{k_n}\sigma_{nk}^2$.

Теорема (центральна гранична теорема Ліндеберга для загальних серій). *Нехай* $(\xi_{nk},\ k=\overline{1,k_n},\ n\geq 1)$ *– загальна послідовність серій випадкових величин, що задовольняє* ***умову Ліндеберга****:*

$$L_n(\varepsilon)\equiv\sigma_n^{-2}\sum\nolimits_{k=1}^{k_n}\mathbf{E}\left|\xi_{nk}-\mu_{nk}\right|^2\mathbb{I}_{\{|\xi_{nk}-\mu_{nk}|\,\geq\,\varepsilon\sigma_n\}}\to 0,\ n\to\infty,\forall\varepsilon>0.$$

Тоді суми $\xi_n\equiv\sum_{k=1}^{k_n}\xi_{nk}$ *після центрування та нормування є асимптотично нормальними:*

$$(\xi_n-\mu_n)\ /\ \sigma_n\to^{W}\zeta\simeq N(0,1),\ n\to\infty.$$

Доведення. Розглянемо подвійну послідовність випадкових величин

$$\zeta_{nk}\equiv(\xi_{nk}-\mu_{nk})\ /\ \sigma_n.$$

Тоді ζ_{nk} є *стандартною послідовністю серій*, оскільки за означенням

$$\mathbf{E}\zeta_{nk}=(\mathbf{E}\xi_{nk}-\mu_{nk})/\sigma_n=0,\ \ \sum\nolimits_{k=1}^{k_n}\mathbf{D}\zeta_{nk}=\sum\nolimits_{k=1}^{k_n}\sigma_{nk}^2/\sigma_n^2=1.$$

Як зазначено у зауваженні до *центральної граничної теореми Ліндеберга для стандартних серій,* умова (г) рівномірної малості в означенні стандартної послідовності серій випливає з умови Ліндеберга:

$$\overline{\lim}_{n\to\infty}\max\nolimits_{1\leq k\leq k_n}\mathbf{D}\zeta_{nk}\ \leq\varepsilon^2+\overline{\lim}_{n\to\infty}L_n(\varepsilon)=\varepsilon^2\to 0,\ \varepsilon\to 0.$$

Далі, показники в *умовах Ліндеберга* для ζ_{nk} та ξ_{nk} збігаються, а величина $\zeta_n=\sum_{k=1}^{k_n}\zeta_{nk}=(\xi_n-\mu_n)\ /\ \sigma_n$. Тому твердження теореми випливає з граничної теореми Ліндеберга для стандартних серій ⊡

Теорема (центральна гранична теорема Ліндеберга для однорідних серій). *Якщо величини в кожній серії* $(\xi_{nk},\ k=\overline{1,k_n})$ *однаково розподілені та*

$$\lim_{n\to\infty}\ \sigma_{n1}^{-2}\ \mathbf{E}\left|\xi_{n1}-\mu_{n1}\right|^2\mathbb{I}_{\left\{|\xi_{n1}-\mu_{n1}|\,\geq\,\varepsilon\sqrt{n}\ \sigma_{n1}\right\}}=0,\ \forall\varepsilon>0,$$

то виконується умова Ліндеберга і має місце асимптотична нормальність сум.

Доведення. Всі k_n доданків в правій частині визначення показника $L_n(\varepsilon)$ з *умови Ліндеберга* та в сумі для дисперсії σ_n^2 є однаковими, тому вираз у лівій частині останньої умови збігається з $L_n(\varepsilon)$ $\boxdot$

1.30.4. Теорема Ляпунова для загальних серій

Теорема (центральна гранична теорема Ляпунова для загальних серій). *Нехай* $(\xi_{nk},\ k = \overline{1,k_n},\ n \geq 1)$ *– загальна послідовність серій випадкових величин, що задовольняє* ***умову Ляпунова***:

$$\exists \delta > 0: \ L_n^{2+\delta} \equiv \sigma_n^{-2-\delta} \sum\nolimits_{k=1}^{k_n} \mathbf{E}\, |\xi_{nk} - \mu_{nk}|^{2+\delta} \to 0,\ n \to \infty.$$

Тоді суми $\xi_n \equiv \sum_{k=1}^{k_n} \xi_{nk}$ *є асимптотично нормальними:*

$$(\xi_n - \mu_n) \ / \ \sigma_n \to^W \zeta \simeq N(0,1),\ n \to \infty.$$

Доведення, як і в *центральній граничній теоремі Ліндеберга для загальних серій*, спирається на стандартні серії $\zeta_{nk} = (\xi_{nk} - \mu_{nk})/\sigma_n$ і нерівність $L_n(\varepsilon) \leq \varepsilon^{-\delta} L_n^{2+\delta}$ між показниками Ліндеберга та Ляпунова $\boxdot$

Наслідок. *Умова Ляпунова виконується, коли величини в серіях* $(\xi_{nk}, k = \overline{1,k_n})$ *однаково розподілені, а моменти:* $\mathbf{E}\,|\xi_{nk}|^{2+\delta}$ *– обмежені.*

Дійсно, у такому випадку всі доданки в сумі $L_n^{2+\delta}$ однакові та обмежені, тому сума зростає як k_n, у той час як знаменник має порядок $k_n^{2+\delta}$.

Вправи

(1) Випадкові величини $(\xi_k, k \geq 1)$ незалежні у сукупності. За яких умов існують додатні сталі $b_n \to \infty$ такі, що $b_n^{-1} \sum_{k=1}^n \xi_k \to^W \zeta \simeq N(0,1)$, якщо: (а) $\mathbf{P}(\xi_k = \pm a_k) = 1/2$, (б) величини ξ_k рівномірно розподілені на $[-a_k, a_k]$?

(2) Випадкові величини $(\xi_k, k \geq 1)$ незалежні та $\mathbf{P}(\xi_k = \pm k^\alpha) = k^{-\beta}/2$, $\mathbf{P}(\xi_k = 0) = 1 - k^{-\beta}$, де $2\alpha > \beta - 1$. Довести, що для утвореної ними загальної послідовності серій $(\xi_k, k = \overline{1,n},\ n \geq 1)$ умова Ліндеберга виконується тоді й тільки тоді, коли $\beta \in [0,1)$.

(3) Випадкові величини $(\xi_k, k \geq 1)$ незалежні та $\mathbf{E}\xi_k = 0$, $|\xi_k| \leq c_k$, причому має місце зображення $c_n = o(\mathbf{D}S_n),\ n \to \infty$, де $S_n = \xi_1 + ... + \xi_n$. Довести, що $S_n/\sqrt{\mathbf{D}S_n} \to^W \zeta \simeq N(0,1), n \to \infty$.

(4) Випадкові величини $(\xi_k, k \geq 1)$ незалежні, $\mathbf{P}(\xi_k = \pm 1) = (1 - k^{-2})/2$, $\mathbf{P}(\xi_k = \pm k) = k^{-2}/2$. Довести, що $S_n/\sqrt{n} \to^W \zeta \simeq N(0,1)$, при $n \to \infty$, однак $\mathbf{D}S_n/n \to 2,\ n \to \infty$.

(5) Випадкові величини $(\xi_k, k \geq 1)$ незалежні, $\mathbf{P}(\xi_k = \pm 1) = (1 - k^{-1})/2$, $\mathbf{P}(\xi_k = \pm\sqrt{k}) = 1/2k$. Довести, що не існує таких сталих σ_n, що має місце збіжність $S_n/\sigma_n \to^W \zeta \simeq N(0,1), n \to \infty$.

(6) Випадкові величини $(\xi_k, k \geq 1)$ незалежні, $\mathbf{P}(\xi_k = 0) = 1 - k^{-1}$, та $\mathbf{P}(\xi_k = 1) = k^{-1}$. Довести, що $(S_n - \ln n)/\sqrt{\ln n} \to^W \zeta \simeq N(0,1), n \to \infty$.

(7) Випадкові величини $(\xi_k, k \geq 1)$ незалежні, та мають гама-розподіли: $\xi_k \simeq \Gamma(1, \alpha_k)$. Загальна послідовність серій $(\xi_k, k = \overline{1,n},\ n \geq 1)$ задовольняє умову Ліндеберга, якщо $\left(\sum_{k=1}^n \alpha_k^2\right)\left(\sum_{k=1}^n \alpha_k\right)^{-1} \to 0,\ n \to \infty$.

(8) Випадкові величини у кожній серії $(\xi_{nk}, k = \overline{1,k_n})$ незалежні та невід'ємні. Довести, що збіжність $\sum_{k=1}^{k_n} \xi_{nk} \to^P 0, n \to \infty$, має місце тоді й тільки тоді, коли для кожного $\varepsilon > 0$ виконуються умови: $\sum_{k=1}^{k_n} \mathbf{P}(\xi_{nk} > \varepsilon) \to 0$ та $\sum_{k=1}^{k_n} \mathbf{E}\xi_{nk} 1\!\!\mathrm{I}_{\{\xi_{nk} \leq \varepsilon\}} \to 0$ при $n \to \infty$.

(9) Розглянемо послідовність серій $(\xi_{nk},\ k = \overline{1,n},\ n \geq 1)$ випадкових величин із розподілом Пуассона з параметром $1/n$ всередині n-ої серії. Доведіть, що виконані всі умови теореми Ляпунова, за винятком умови Ляпунова, причому центральна гранична теорема також не виконується.

(10) Випадкові величини $(\xi_n, n \geq 1)$ незалежні і однаково симетрично розподілені: $\xi_i \simeq -\xi_j$,.причому $x^2\mathbf{P}(|\xi_1| > x) = o(\mathbf{E}(\xi_1^2 1\!\!\mathrm{I}_{\{|\xi_1| \leq x\}})), x \to \infty$. Тоді знайдуться $\sigma_n \in \mathbb{R}$ такі, що $(\xi_1 + ... + \xi_n)/\sigma_n \to^W N(0,1), n \to \infty$.

1.30.5. Теорема Пуассона для стандартних серій

Граничним розподілом центрованих та нормованих сум незалежних величин може бути не тільки нормальний розподіл.

Теорема (гранична теорема Пуассона для стандартних серій). *Нехай $(\zeta_{nk},\ k = \overline{1,k_n},\ n \geq 1)$ – стандартна послідовність серій випадкових величин, що при деякому $a > 0$ задовольняє умову*

$$P_n(\varepsilon) \equiv \sum_{k=1}^{k_n} \mathbf{E}\zeta_{nk}^2\ 1\!\!\mathrm{I}_{\{|\zeta_{nk} - a| \,\geq\, \varepsilon\}} \to 0,\ n \to \infty,\ \forall \varepsilon > 0.$$

Тоді $\zeta_n \equiv \sum_{k=1}^{k_n} \zeta_{nk} \to^W \zeta$, де гранична характеристична функція

$$\varphi_\zeta(t) = \exp((\exp(ita) - 1 - ita)\ /\ a^2)$$

є характеристичною функцією випадкової величини ζ із розподілом Пуассона з параметром $\lambda = 1/a^2$ і множиною значень $\{na - 1/a,\ n \in \mathbb{Z}_+\}$.

Доведення. Як і в *центральній граничній теоремі Ліндеберга для стандартних серій*, доводимо, що $G_n \to^O G, n \to \infty$, де $G(x) = 1\!\!\mathrm{I}_{a\,<\,x}$.

Звідси за *загальною граничною теоремою для стандартних серій*

$$\lim_{n\to\infty} \varphi_{\zeta_n}(t) = \exp\left(\int_{-\infty}^{\infty} f_t(x)\ dG(x)\right) = \exp(f_t(a)) = \varphi_\zeta(t),$$

де пуассонів розподіл границі ζ визначається з розкладу у ряд Тейлора експоненти від $f_t(a) = (\exp(ita) - 1 - ita)\ /\ a^2$:

$$\varphi_\zeta(t) = \sum\nolimits_{n\geq 0} \frac{\exp(it(an - 1/a))\ a^{-2n}}{n!} \exp(-a^{-2})\ \boxdot$$

Вправа. Вивести з даної теореми граничну теорему Пуассона для кількості успіхів у схемі серій випробувань Бернуллі.

1.30.6. Теорема Реньї з теорії надійності

На відміну від центральної граничної теореми, в теорії надійності розглядають граничні теореми для сум випадкового числа незалежних величин. Такі суми можна інтерпретувати як час до відмови високонадійної системи, яка періодично відновлюється. При певних припущеннях щодо розподілу числа доданків (кількості відновлень до відмови) виявляється, що граничний розподіл сум є показниковим і визначається виключно математичними сподіваннями окремих доданків.

Теорема (теорема Реньї). *Нехай величини* $(\xi_n,\ n \geq 1)$ *незалежні в сукупності невід'ємні однаково розподілені,* $\mu = \mathbf{E}\xi_1 < \infty$*, і не залежать від випадкової величини* ν*, що має геометричний розподіл із параметром* $\theta : \mathbf{P}(\nu = n) = \theta(1-\theta)^{n-1}, n \geq 1.$

Позначимо $\zeta_\nu \equiv \sum_{k=1}^{\nu} \xi_k$*. Тоді для всіх* $x > 0$

$$\mathbf{P}(\theta\ \zeta_\nu\ /\mu \geq x) \to \exp(-x),\ \ \theta \to 0.$$

Зауваження. Якщо інтерпретувати ймовірність успіху θ – як імовірність відмови за період до відновлення, ν – як кількість відновлень системи до виходу з ладу, а ξ_k – як інтервал між послідовними відновленнями, то сума ζ_ν дорівнює часу до першої відмови системи.

Доведення. Позначимо через $\varphi(t)$ – *характеристичну функцію* величини ξ_k. За теоремою *про властивості характеристичних функцій*

$$\varphi(t) = 1 + it\mu + o(t),\ \ t \to 0.$$

Генератриса величини ν дорівнює $\psi_\nu(z) = \theta z/(1 - z(1-\theta))$. Тому з теореми *про властивості генератрис*, пункт (е), дістанемо при $z = \exp(it)$

$$\mathbf{E}\exp(it\zeta_\nu) = \psi_\nu(\mathbf{E}\exp(it\xi_1)) = \theta\varphi(t)/(1 - \varphi(t)(1-\theta)).$$

Підставимо в цю формулу $t = s\theta/\mu \to 0,\ \theta \to 0$, та використаємо розклад $\varphi(t)$ в околі нуля

$$\lim\nolimits_{\theta\to 0} \mathbf{E}\exp(is\ \zeta_\nu\theta/\mu) = \lim\nolimits_{\theta\to 0} \theta\varphi(s\theta/\mu)/(1 - \varphi(s\theta/\mu)(1-\theta)) =$$

$$\lim_{\theta\to 0}\theta\ (1+O(\theta))\ /\ (1-(1+i\mu\ s\theta/\mu)(1-\theta)+o(\theta)) =$$
$$\lim_{\theta\to 0}\theta\ /\ (\theta - is\theta + o(\theta)) = 1/(1-is).$$

Гранична функція неперервна в нулі та є *характеристичною функцією* величини τ з *показниковим розподілом* та одиничним параметром. За *теоремою Леві про критерій слабкої збіжності* $\zeta_\nu\theta/\mu \to^W \tau, \theta \to 0$ ⊡

Приклад. Плановий ремонт мережі зв'язку виконується щорічно через 350-380 днів після попереднього ремонту з однаковими ймовірностями для дня з цього проміжку. Відремонтована мережа виходить із ладу на інтервалі між ремонтами з імовірністю 0.01. Яка ймовірність безвідмовної роботи за 10 років (періодів планових ремонтів) ?

За *теоремою Реньї*

$\mathbf{P}(\zeta_\nu > 10 * 365) = \mathbf{P}(0.01\frac{\zeta_\nu}{(350+380)/2} > 0.10) \approx \exp(-0.10) \simeq 0.904.$

Вправа. Нехай величини $(\xi_n, n \geq 1)$ незалежні та однаково розподілені, $\mathbf{E}\xi_1 = 0$, $\mathbf{D}\xi_1 = 1$, а ν така сама, як і в теоремі Реньї та не залежить від (ξ_n). Довести, що $\sqrt{\theta}\zeta_\nu \to^W \eta, \theta \to 0$, де $\varphi_\eta(t) = (1 + t^2/2)^{-1}$.

1.30.7. Центральна гранична теорема для випадкових векторів

Теорема (центральна гранична теорема для випадкових векторів). *Нехай* $(\eta_{nk},\ k = \overline{1,k_n},\ n \geq 1)$ *– послідовність серій випадкових векторів у* $\mathbb{R}^d$ *така, що:*

(1) вектори $(\eta_{nk},\ k = \overline{1,k_n})$ *незалежні в сукупності в кожній серії,*

(2) існують середні $m_{nk} = \mathbf{E}\eta_{nk}$ *і коваріаційні матриці* $V_{nk} = Cov(\eta_{nk})$,

(3) сумарні показники $m_n = \sum_{k=1}^{k_n} m_{nk}$, $V_n = \sum_{k=1}^{k_n} V_{nk}$ *мають границі:* $m_n \to m$, $V_n \to V$ *при* $n \to \infty$.

(4) Якщо виконана умова Ліндеберга:

$$L_n^d(\varepsilon) \equiv \sum\nolimits_{k=1}^{k_n} \mathbf{E}\|\eta_{nk} - m_{nk}\|^2 \mathbb{I}_{\{\|\eta_{nk}-m_{nk}\| \,\geq\, \varepsilon\}} \to 0,\ n \to \infty, \forall\varepsilon > 0,$$

то суми $\eta_n \equiv \sum_{k=1}^{k_n} \eta_{nk}$ *слабко збігаються до нормального вектора:*

$$\eta_n \to^W \xi \simeq N_d(m, V),\ n \to \infty.$$

Доведення. Нехай $t \in \mathbb{R}^d$ фіксоване. Доведемо збіжність лінійної форми $t'\eta_n \to^W t'\xi$. Зауважимо, що за теоремою *про лінійні перетворення нормальних векторів*

$$t'\xi \simeq N(t'm,\ t'Vt).$$

Розглянемо послідовності випадкових величин

$$\xi_{nk} = t'\eta_{nk},\ \xi_n = \sum\nolimits_{k=1}^{k_n} \xi_{nk} = t'\eta_n.$$

Тоді $\mathbf{E}\xi_n = t'm_n \to t'm$ та за теоремою *про дисперсію лінійної форми від випадкового вектора* $\mathbf{D}\xi_{nk} = t'V_{nk}t$. Позначимо

$$\sigma_n^2 = t'V_n t,\ \ \sigma^2 = t'Vt.$$

За умовою

$$\mathbf{D}\xi_n = \sum\nolimits_{k=1}^{k_n} \mathbf{D}\xi_{nk} = \sum\nolimits_{k=1}^{k_n} t'V_{nk}t = t'V_n t = \sigma_n^2,\ \ \sigma_n^2 \to \sigma^2,\ n \to \infty.$$

Припустимо, що $\sigma = 0$. Тоді $\mathbf{D}\xi_n \to 0$ і $\xi_n - \mathbf{E}\xi_n \to^{L2} 0, n \to \infty$. Тому $\xi_n \to^P t'm$. Оскільки за теоремою *про співвідношення між різними видами збіжності* зі збіжності *за ймовірністю* випливає *слабка збіжність*, то $\xi_n = t'\eta_n \to^W t'm = t'\xi \simeq N(t'm,\ 0)$, що доводить збіжність лінійної форми $t'\eta_n$ за умови зробленого припущення.

Отже, можна вважати, що $\sigma > 0$. Тоді $\sigma_n^2 \ge \delta^2 > 0$, починаючи з деякого номера. З умов (1) – (3) теореми виводимо, що послідовність (ξ_{nk}) є *загальною послідовністю серій.*

Показник з *умови Ліндеберга* в *центральній граничній теоремі Ліндеберга для загальних серій* (ξ_{nk}) має вигляд

$$L_n(\varepsilon) = \sigma_n^{-2} \sum\nolimits_{k=1}^{k_n} \mathbf{E}\,|\xi_{nk} - \mathbf{E}\xi_{nk}|^2\ \ 1\!\!\mathrm{I}_{\{|\xi_{nk}-\mathbf{E}\xi_{nk}|\,\ge\,\varepsilon\sigma_n\}} \le$$

$$\sigma_n^{-2} \sum\nolimits_{k=1}^{k_n} \mathbf{E}\|\eta_{nk} - m_{nk}\|^2\|t\|^2\ \ 1\!\!\mathrm{I}_{\{\|\eta_{nk}-m_{nk}\|\ \ \|t\|\,\ge\,\varepsilon\sigma_n\}} \le$$

$$\|t\|^2\delta^{-2} \sum\nolimits_{k=1}^{k_n} \mathbf{E}\|\eta_{nk} - m_{nk}\|^2 1\!\!\mathrm{I}_{\{\|\eta_{nk}-m_{nk}\|\ge\varepsilon\delta/\|t\|\}} =$$

$$\|t\|^2\delta^{-2}L_n^d(\varepsilon\delta/\|t\|) \to 0,\ n \to \infty.$$

Отже, за *центральною граничною теоремою Ліндеберга для загальних серій* $(\xi_n - t'm_n)/\sigma_n \to^W \zeta \simeq N(0,1)$. Тому за теоремою *про добуток слабко збіжної послідовності зі збіжною*

$$\xi_n - t'm_n = \sigma_n\ (\xi_n - t'm_n)/\sigma_n \to^W \sigma\zeta \simeq N(0,\sigma^2).$$

Оскільки $t'm_n \to t'm$, то

$$\xi_n \to^W t'm + \sigma\zeta \simeq N(t'm,\ \sigma^2) = N(t'm,\ t'Vt),\ n \to \infty.$$

Отже, і в даному випадку кожна лінійна форма від сум $\eta_n \equiv \sum_{k=1}^{k_n} \eta_{nk}$ слабко збігається:

$$t'\eta_n \to^W t'\xi \simeq N(t'm,\ t'Vt),\ n \to \infty,\ \forall t \in \mathbb{R}^d.$$

Остаточно зі збіжності $t'\eta_n \to^W t'\xi$ та з *теореми Крамера-Волда про критерій слабкої збіжності випадкових векторів* виводимо слабку збіжність векторів $\eta_n \to^W \xi, n \to \infty$ ⊡

Теорема (класична центральна гранична теорема для випадкових векторів). *Нехай* $(\gamma_n, n \ge 1)$ *послідовність незалежних у сукупності однаково розподілених випадкових векторів у* $\mathbb{R}^d$*,* $\mathbf{E}\gamma_1 = m$, *Cov*$(\gamma_1) = V$, *а* $\eta_n \equiv \frac{1}{\sqrt{n}} \sum_{k=1}^n (\gamma_k - m)$. *Тоді*

$$\eta_n \to^W \zeta \simeq N_d(0, V),\ n \to \infty.$$

Доведення. Покладемо $\eta_{nk} = (\gamma_k - m)/\sqrt{n},\ k = \overline{1, n}$ та застосуємо для цієї послідовності *центральну граничну теорему для випадкових векторів*. Умова незалежності (1) виконується за теоремою *про перетворення незалежних величин*. Величини з умови (2) дорівнюють $m_{nk} = 0,\ V_{nk} = V/n$, тому виконується умова (3).

Нарешті, показник з *умови Ліндеберга* прямує до нуля за *теоремою Лебега про монотонну збіжність*, оскільки з урахуванням однакової розподіленості

$$L_n(\varepsilon) = n\, \mathbf{E}\|\eta_{n1}\|^2\, 1\!\!1_{\{\|\eta_{n1}\| \ge \varepsilon\}} =$$
$$\mathbf{E}\|\gamma_1 - m\|^2\, 1\!\!1_{\{\|\gamma_1 - m\| \ge \varepsilon\sqrt{n}\}} \downarrow 0,\ n \to \infty.$$

Отже, слабка збіжність випливає з *центральної граничної теореми для випадкових векторів* ⊡

Розділ 2

Імовірність 2

Вступ

За своїм змістом даний розділ є вступом до курсу теорії випадкових процесів, що призначений для студентів III курсу спеціальностей математика, статистика, та був започаткований І.І.Гіхманом. Матеріал курсу містить одночасно як нові фундаментальні поняття теорії ймовірностей, а саме: випадкових елементів, незалежних систем, умовного сподівання та інших, так і уточнення граничних теорем – теорему про три ряди, закон повторного логарифма, розклад Еджуорта, нерівності Беррі-Ессеєна. Наведені приклади випадкових блукань, ланцюгів Маркова та строго стаціонарних послідовностей.

Розділ містить матеріал семестрового курсу додаткових розділів теорії ймовірностей та випадкових процесів, що розрахований на 34 години лекцій та 17 годин самостійної роботи. Останні параграфи розділу – щодо процесів Пуассона та вінерівського – викладаються в рамках загального курсу теорії ймовірностей.

За браком часу такі теми, як нескінченно подільні розподіли, граничні задачі для випадкових блукань, гіллясті процеси, мартингальні послідовності, викладаються у рамках курсу теорії випадкових процесів.

2.1. Імовірнісні простори випадкових послідовностей і функцій

Послідовність випадкових величин $(\xi_n, n \geq 1)$ набуває значень у просторі дійсних послідовностей

$$\mathbb{R}^\infty \equiv \{x = (x_n, n \geq 1),\ x_n \in \mathbb{R}\}.$$

2.1.1. Борелєва сигма-алгебра множин послідовностей

Означення. *Борелєвою сигма-алгеброю підмножин простору $\mathbb{R}^\infty$ називається сигма-алгебра, що породжена координатними функціями:*

$$\mathfrak{B}(\mathbb{R}^\infty) \equiv \sigma[\{x : x_n < a\}\,,\ a \in \mathbb{R}, n \geq 1].$$

За означенням борелєвими є також скінченні перетини:

$$\bigcap_{k=1}^{n} \{x \in \mathbb{R}^\infty : x_k < a_k\} = \{x \in \mathbb{R}^\infty : (x_1, ...x_n) \in \Pi_a\} \in \mathfrak{B}(\mathbb{R}^\infty),$$

$$\forall a = (a_1, ...a_n) \in \mathbb{R}^n, \text{ де } \quad \Pi_a = \prod_{k=1}^{n} (-\infty, a_k).$$

З останнього зображення приходимо до такого означення.

Означення. Циліндричною множиною *розмірності n із основою $B_n \in \mathfrak{B}(\mathbb{R}^n)$ називається така підмножина $\mathbb{R}^\infty$:*

$$J_n(B_n) \equiv \{x \in \mathbb{R}^\infty : (x_1, ...x_n) \in B_n\}.$$

Означення. *Сигма-алгеброю циліндричних множин J_n розмірності n та* класом циліндричних множин J у просторі $\mathbb{R}^\infty$ *називаються такі класи множин:*

$$J_n \equiv \{J_n(B_n),\ B_n \in \mathfrak{B}(\mathbb{R}^n)\}, \quad J \equiv \cup_{n \geq 1} J_n.$$

Очевидно, що клас J_n ізоморфний $\mathfrak{B}(\mathbb{R}^n)$ та є сигма-алгеброю.

Зауважимо, що для всіх $n \geq 1$, $B_n \in \mathfrak{B}(\mathbb{R}^n)$ за означенням *циліндричних множин* справедлива тотожність узгодженості

$$J_{n+1}(B_n \times \mathbb{R}) = J_n(B_n).$$

Тому, зокрема, мають місце включення $J_n \subset J_{n+1}$. Оскільки класи J_n є алгебрами, то J – також алгебра внаслідок наведеного включення.

Оператор $J_n()$ витримує вкладені різниці та зліченні об'єднання:

$$J_n(B_n \backslash C_n) = J_n(B_n) \backslash J_n(C_n), \qquad J_n(\cup_{k \geq 1} B_n^k) = \cup_{k \geq 1} J_n(B_n^k).$$

Крім того, $J_n(\Pi_a) \in \mathfrak{B}(\mathbb{R}^\infty)$, як показано вище. Тому $J_n(A_n) \in \mathfrak{B}(\mathbb{R}^\infty)$ для всіх $A_n \in \mathfrak{A}(\mathbb{R}^n)$, та $J_n(B_n) \in \mathfrak{B}(\mathbb{R}^\infty)$ для всіх $B_n \in \mathfrak{B}(\mathbb{R}^n)$, оскільки клас $\{B_n : J_n(B_n) \in \mathfrak{B}(\mathbb{R}^\infty)\}$ є сигма-алгеброю та містить $\mathfrak{A}(\mathbb{R}^n)$. Отже, справедливе еквівалентне означення:

$$\mathfrak{B}(\mathbb{R}^\infty) = \sigma[J_n(B_n),\ B_n \in \mathfrak{B}(\mathbb{R}^n),\ n \geq 1] = \sigma[\cup_{n \geq 1} J_n] = \sigma[J].$$

Вправи

(1) Клас J не є сигма-алгеброю.

(2) Множини $\{x \in \mathbb{R}^\infty : \sup x_n < a\}$, $\{x \in \mathbb{R}^\infty : \exists \lim x_n < a\}$ – борелєві.

(3) Визначимо на $\mathbb{R}^\infty$ метрику $\rho(x, y) = \sum_{n \geq 1} 2^{-n} \min(|x_n - y_n|\,, 1)$, що визначає покоординатну збіжність. Тоді $\mathfrak{B}(\mathbb{R}^\infty)$ породжується відкритими кулями.

2.1.2. Побудова ймовірності на множинах послідовностей

Нехай $\mathbf{P}$ – імовірнісна міра на $\mathfrak{B}(\mathbb{R}^\infty)$. Тоді з властивостей оператора $J_n()$ з означення *циліндричної множини* випливає, що функція множин

$$\mathbf{P}_n(B_n) \equiv \mathbf{P}(J_n(B_n)),\ \forall B_n \in \mathfrak{B}(\mathbb{R}^n),$$

є *ймовірністю* при кожному $n \geq 1$. Ця ймовірність $\mathbf{P}_n$ називається **скінченновимірним розподілом** порядку n для міри $\mathbf{P}$.

З *тотожності узгодженості* виводимо, що скінченновимірні розподіли задовольняють **умову узгодженості:**

$$\mathbf{P}_{n+1}(B_n \times \mathbb{R}) = \mathbf{P}_n(B_n),\quad \forall B_n \in \mathfrak{B}(\mathbb{R}^n),\ \forall n \geq 1,$$

оскільки обидві частини даної рівності є значеннями міри $\mathbf{P}$ від тієї самої циліндричної підмножини $\mathbb{R}^\infty$.

Теорема (теорема Колмогорова про побудову ймовірності на множинах послідовностей). *Нехай* $(\mathbf{P}_n : \mathfrak{B}(\mathbb{R}^n) \to [0,1],\ n \geq 1)$ *– послідовність імовірнісних мір, що задовольняють умову узгодженості. Існує єдина ймовірність* $\mathbf{P}$ *на* $\mathfrak{B}(\mathbb{R}^\infty)$ *така, що для кожного* $n \geq 1$ *міра* $\mathbf{P}_n$ *є її скінченновимірним розподілом порядку* n*, тобто*

$$\mathbf{P}_n(B_n) = \mathbf{P}(J_n(B_n)),\ \forall B_n \in \mathfrak{B}(\mathbb{R}^n).$$

Доведення. Визначимо таку міру $\mathbf{P}$ на *класі циліндричних множин* J :

$$\mathbf{P}(J_n(B_n)) \equiv \mathbf{P}_n(B_n),\ B_n \in \mathfrak{B}(\mathbb{R}^n).$$

Це визначення є коректним внаслідок *умови узгодженості*, та задає нормовану, невід'ємну і адитивну функцію на алгебрі J. Останнє твердження виводиться за допомогою зведення всіх циліндричних множин у скінченому об'єднанні до єдиного порядку n.

Розглянемо такі підкласи *класу циліндричних множин* J:

$$J_n^a = \{J_n(A_n), A_n \in \mathfrak{A}(\mathbb{R}^n)\} \subset J_n, n \geq 1.$$

Очевидно, що J_n^a є алгеброю, яка ізоморфна $\mathfrak{A}(\mathbb{R}^n)$ та породжує сигма-алгебру J_n, а об'єднання $J^a \equiv \cup_{n\geq 1} J_n^a$ також є алгеброю. Тому справедливе включення $\sigma[J^a] \supset \cup_{n\geq 1} J_n = J$ та $\sigma[J^a] = \sigma[J] = \mathfrak{B}(\mathbb{R}^\infty)$, де остання рівність встановлена вище.

Оскільки $\mathfrak{B}(\mathbb{R}^\infty) = \sigma[J^a]$, то твердження теореми випливає з *теореми Каратеодорі про продовження міри*, якщо довести *неперервність в нулі* міри $\mathbf{P}$ на алгебрі J^a.

Для доведення цієї неперервності від супротивного припустимо, що існує послідовність множин $\widehat{B}_n = J_n(B_n) \in J^a$, $B_n \in \mathfrak{A}(\mathbb{R}^n)$, таких, що $\widehat{B}_n \downarrow \varnothing$, однак $\mathbf{P}(\widehat{B}_n) \geq \delta > 0$.

Оскільки множина $B_n \in \mathfrak{A}(\mathbb{R}^n)$ є скінченним об'єднанням паралелепіпедів, то за властивостями неперервності та нормованості *сумісної функції розподілу* для даного $\delta > 0$ існує компактна множина (об'єднання вписаних компактних паралелепіпедів) $C_n \subset B_n$ така, що $\mathbf{P}_n(B_n \backslash C_n) \leq \delta 2^{-n-1}$.

Позначимо $\widehat{C}_n = J_n(C_n)$. Тоді $\widehat{C}_n \subset \widehat{B}_n$, $\cap_{n\geq 1}\widehat{C}_n \subset \cap_{n\geq 1}\widehat{B}_n = \varnothing$.

Розглянемо незростаючу послідовність $\widehat{D}_n = \cap_{k=1}^n \widehat{C}_k$. За *тотожністю узгодженості* $\widehat{D}_n = J_n(D_n)$, де $D_n = \cap_{k=1}^n (C_k \times \mathbb{R}^{n-k}) \subset C_n$ є компактом як замкнена підмножина компакту. Оскільки послідовність $\widehat{B}_n$ не спадає, то справедливі співвідношення

$$\mathbf{P}(\widehat{B}_n \backslash \widehat{D}_n) = \mathbf{P}\left(\bigcup_{k=1}^n (\widehat{B}_n \backslash \widehat{C}_k)\right) \leq \mathbf{P}\left(\bigcup_{k=1}^n (\widehat{B}_k \backslash \widehat{C}_k)\right) \leq$$

$$\sum_{k=1}^n \mathbf{P}(\widehat{B}_k \backslash \widehat{C}_k) = \sum_{k=1}^n \mathbf{P}_k(B_k \backslash C_k) \leq \sum_{k=1}^n \delta 2^{-k-1} < \delta/2.$$

Тому $\mathbf{P}(\widehat{D}_n) = \mathbf{P}(\widehat{B}_n) - \mathbf{P}(\widehat{B}_n \backslash \widehat{D}_n) \geq \delta/2 > 0$. Отже, $\widehat{D}_n \neq \varnothing$ при кожному n. Оберемо довільні точки $x(n) \in \widehat{D}_n$.

Оскільки $\widehat{D}_n \subset \widehat{D}_1 = J_1(D_1)$, то послідовність відповідних координат $\{x_1(n), n \geq 1\}$ міститься у компакті D_1. За означенням компактності оберемо нескінченну підпослідовність номерів $K_1 \subset \mathbb{N}$ так, щоб $x_1(n) \to y_1$ для деякого y_1 при $n \to \infty, n \in K_1$. Тоді $y_1 \in D_1$. Далі, з включення $(x_1(n), x_2(n)) \in D_2$ знайдемо підпослідовність $K_2 \subset K_1$ таку, щоб $(x_1(n), x_2(n)) \to (y_1, y_2)$, $n \to \infty, n \in K_3$, для деякого $(y_1, y_2) \in D_2$. Перша координата останнього вектора дійсно збігається з y_1, оскільки $\{x_1(n), n \in K_2\} \subset \{x_1(n), n \in K_1\}$. Продовжуючи за індукцією, для кожного $m \geq 1$ знайдемо граничну точку $(y_1, ..., y_m) \in D_m$.

Тоді з означення циліндру $\widehat{D}_m$ виводимо включення для точки $y = (y_n, n \geq 1) \in J_m(D_m) = \widehat{D}_m$ при всіх m. Отже,.$y \in \cap_{m\geq 1}\widehat{D}_m$, що суперечить доведеній вище несумісності $\cap_{n\geq 1}\widehat{D}_n = \varnothing$ $\boxdot$

Теорема (теорема Колмогорова про побудову ймовірності на множинах послідовностей через сумісні функції розподілу). *Нехай послідовність сумісних функцій розподілу* $(F_n : \mathbb{R}^n \to [0,1], n \geq 1)$ *задовольняє умову узгодженості:*

$$F_{n+1}(x_1, ..., x_n, \infty) = F_n(x_1, ..., x_n), \ \forall x_k \in \mathbb{R}, \ \forall n \geq 1.$$

Тоді існує єдина ймовірність $\mathbf{P}$ *на* $\mathfrak{B}(\mathbb{R}^\infty)$ *така, що при всіх* $n \geq 1$

$$\mathbf{P}(\{x \in \mathbb{R}^\infty : x_1 < a_1, ..., x_n < a_n\}) = F_n(a_1, ..., a_n), \quad \forall a_k \in \mathbb{R}.$$

Доведення. Визначимо на $\mathfrak{B}(\mathbb{R}^n)$ як $\mathbf{P}_n$ *міру Лебега – Стілтьєса, що породжена сумісною функцією розподілу* F_n. Розглянемо клас множин

$$\aleph_n = \{B_n \in \mathfrak{B}(\mathbb{R}^n) : \mathbf{P}_{n+1}(B_n \times \mathbb{R}) = \mathbf{P}_n(B_n)\}.$$

З умови узгодженості на функції розподілу F_n виводимо, що всі *кути* $\Pi_a = \Pi_{k=1}^n(-\infty, a_k)$ належать класу $\aleph_n$. Оскільки $\mathbf{P}_n$ та $\mathbf{P}_{n+1}$ є ймовірностями, то цей клас є сигма-алгеброю. Тому $\aleph_n = \mathfrak{B}(\mathbb{R}^n)$. Отже, послідовність мір $\mathbf{P}_n$ задовольняє *умову узгодженості* і з *теореми Колмогорова про побудову ймовірності на множинах послідовностей* виводимо твердження теореми $\boxdot$

Вправи

(1, продакт-міра) Нехай задано послідовність довільних функцій розподілу $(G_n, n \geq 1)$. Тоді існує єдина ймовірність $\mathbf{P}$ на $\mathfrak{B}(\mathbb{R}^\infty)$ така, що

$\mathbf{P}(\{x \in \mathbb{R}^\infty : x_1 < a_1, ..., x_n < a_n\}) = \prod_{k=1}^n G_k(a_k),\ \forall a_k \in \mathbb{R},\ \forall n \geq 1.$

(2, гауссівська послідовність) Нехай $a \in \mathbb{R}^\infty$, а $V = (V_{ij}, i, j \geq 1)$ симетрична додатно визначена матриця. Позначимо вектор $a^n = (a_1, ..., a_n)$, та матрицю $V_n = (V_{ij}, 1 \leq i, j \leq n)$. Тоді сумісні функції розподілу F_n зі щільностями

$$(2\pi)^{-n/2} |\det V_n|^{-1/2} \exp(-(x - a^n)' V_n^{-1}(x - a^n)),\ x \in \mathbb{R}^n,$$

задовольняють умову узгодженості.

(3, дискретний простір) Нехай E – дискретна множина. (а) Якщо розподіли $(p_n(x_1, ..., x_n), x_k \in E, n \geq 1)$ такі, що $\sum_{x \in E} p_{n+1}(x_1, .., x_n, x) = p_n(x_1, ..., x_n)$, то існує єдина ймовірність $\mathbf{P}$ на $\mathfrak{B}(E^\infty)$, для якої при всіх $n \geq 1$ та $a_k \in E$ має місце рівність $\mathbf{P}(\{x \in E^\infty : x_1 = a_1, ...x_n = a_n\}) = p_n(a_1, ..., a_n)$. (б) Якщо $(q(x), x \in E)$ та рядки матриці $(p(x, y), x, y \in E)$ є розподілами ймовірностей, то послідовність $p_n(x_1, ..., x_n) = q(x_1)p(x_1, x_2)...p(x_{n-1}, x_n)$ задовольняє умови (а) для кожного розподілу q.

(4) Нехай $C(\mathbb{Q})$ – простір неперервних функцій на множині раціональних чисел $\mathbb{Q}$, з борелєвою сигма-алгеброю $\mathfrak{B}(\mathbb{R}^\mathbb{Q})$. Функція Q задана на циліндрах $C_F = \{x : x_t \in B_t, t \in F\}$ при скінчених $F \subset \mathbb{Q}$ та довільних $B_t \in \mathfrak{B}(\mathbb{R})$ рівністю $Q(C_F) = \prod_{t \in F} G(B_t)$, де G – імовірнісна міра Лебега – Стілтьєса на $\mathfrak{B}(\mathbb{R})$. Довести, що Q продовжується до адитивної імовірності на алгебру циліндрів, що породжена множинами C_F, однак ця ймовірність не є σ-адитивною.

(5, простір дійсних функцій) Нехай T – довільна множина. Розглянемо простір $\mathbb{R}^T = \{x : T \to \mathbb{R}\}$. Нехай $\mathfrak{B}(\mathbb{R}^T) \equiv \sigma[\{x : x(t) < a\}, a \in \mathbb{R}, t \in T]$. Довести, що $\mathfrak{B}(\mathbb{R}^T) = \cup_{\tau \subset T} \mathfrak{B}(\mathbb{R}^\tau)$, де об'єднання береться по зліченним підмножинам τ множини T, а сигма-алгебра $\mathfrak{B}(\mathbb{R}^\tau)$ визначається так само, як і $\mathfrak{B}(\mathbb{R}^\infty)$. У випадку, коли $T = [0, 1]$, довести, що $S_a \equiv \{x : \sup_{t \in [0,1]} x(t) < a\} \notin \mathfrak{B}(\mathbb{R}^T)$. Одночасно перетин $S_a \cap C([0, 1])$ вже належить $\mathfrak{B}(\mathbb{R}^T)$.

2.2. Випадкові елементи

Випадковий елемент є узагальненням випадкового вектора.

2.2.1. Означення, породжені сигма-алгебра та розподіл

Означення. *Нехай* $(C, \mathfrak{C})$ *– довільний вимірний простір.* C*-значним* випадковим елементом *на просторі* $(\Omega, \mathfrak{F}, \mathbf{P})$ *називається довільна вимірна функція* $\zeta : \Omega \to C$*, тобто така, що* $\zeta^{(-1)}(B) \in \mathfrak{F},\ \forall B \in \mathfrak{C}$.

C*-значний випадковий елемент називається:*

(а) при $(C, \mathfrak{C}) = (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ *– випадковою величиною,*

(б) при $(C, \mathfrak{C}) = (\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$ *– випадковим вектором,*

(в) при $(C, \mathfrak{C}) = (\mathbb{R}^\infty, \mathfrak{B}(\mathbb{R}^\infty))$ – випадковою послідовністю.

Теорема (про критерій вимірності функції). *Нехай* σ*-алгебра* $\mathfrak{C}$ *породжується класом* $\mathfrak{D} : \mathfrak{C} = \sigma[\mathfrak{D}]$*. Для того, щоб функція* $\zeta : \Omega \to C$ *була випадковим елементом, необхідно і достатньо, щоб виконувались включення:* $\zeta^{(-1)}(B) \in \mathfrak{F},\ \forall B \in \mathfrak{D}$.

Доведення. *Необхідність* очевидна.

Достатність випливає з того, що клас $\{B \in \mathfrak{C} : \zeta^{(-1)}(B) \in \mathfrak{F}\}$ містить $\mathfrak{D}$ та є сигма-алгеброю, оскільки $\mathfrak{F}$ – сигма-алгебра $\boxdot$

Для довільної випадкової послідовності $\zeta = (\zeta_n, n \geq 1)$ її n-та координата $\zeta_n = \pi_n(\zeta)$ є випадковою величиною за теоремою *про вимірність суперпозиції*, оскільки координатне відображення $\pi_n : \mathbb{R}^\infty \to \mathbb{R}$ є вимірною функцією. Навпаки, послідовність $\zeta = (\zeta_n, n \geq 1)$, що складена з випадкових величин, є випадковою послідовністю. Дійсно, прообрази *циліндричних множин* з основами у вигляді вимірних прямокутників є випадковими подіями:

$$\{\zeta \in J_n(B_1 \times ... \times B_n)\} = \bigcap\nolimits_{k=1}^{n} \{\zeta_k \in B_k\} \in \mathfrak{F}.$$

Оскільки циліндричні множини породжують борелєву сигма-алгебру $\mathfrak{B}(\mathbb{R}^\infty)$, то сформульоване твердження є наслідком теореми *про критерій вимірності функції.*

Означення. Породженою сигма-алгеброю *для випадкового елемента* ζ *називається клас випадкових подій* $\sigma[\zeta] \equiv \{\zeta^{(-1)}(B),\ B \in \mathfrak{C}\}$.

Отже, функція $\zeta : \Omega \to C$ є *випадковим елементом* тоді й тільки тоді, коли $\sigma[\zeta] \subset \mathfrak{F}$.

Означення. *Випадковий елемент* ζ *вимірний відносно сигма-алгебри* $\mathfrak{D} \subset \mathfrak{F}$, *якщо* $\sigma[\zeta] \subset \mathfrak{D}$, *тобто коли* $\zeta^{(-1)}(B) \in \mathfrak{D},\ \forall B \in \mathfrak{C}$.

Означення. *Нехай* $(\zeta_\gamma, \gamma \in \Gamma)$ *– довільна сім'я випадкових елементів. Породженою ними сигма-алгеброю називається сигма-алгебра*

$$\sigma[\zeta_\gamma, \gamma \in \Gamma] \equiv \sigma[\cup_{\gamma \in \Gamma} \sigma[\zeta_\gamma]].$$

Означення. Розподілом випадкового елемента ζ *називається така функція на вимірних множинах* $B \in \mathfrak{C}$:

$$P_\zeta(B) \equiv \mathbf{P}(\zeta \in B) = \mathbf{P}(\zeta^{(-1)}(B)).$$

За властивостями прообразу $\zeta^{(-1)}$ ця функція є *ймовірністю* на $\mathfrak{C}$.

Теорема (про канонічну побудову випадкового елемента). *Нехай* F *– імовірність на сигма-алгебрі* $\mathfrak{C}$ *підмножин простору* C. *Тоді існує ймовірнісний простір* $(\Omega, \mathfrak{F}, \mathbf{P})$ *та випадковий елемент* ζ *на ньому, що має розподіл* $P_\zeta(B) = F(B), \forall B \in \mathfrak{C}$.

Доведення. Визначимо ймовірнісний простір $(\Omega, \mathfrak{F}, \mathbf{P}) \equiv (C, \mathfrak{C}, F)$ та $\zeta(\omega) = \omega$. Тоді $P_\zeta(B) = F(\{\omega : \zeta(\omega) \in B\}) = F(\{\omega : \omega \in B\}) = F(B)$ $\boxdot$

Приклад. Нехай $(G_n, n \geq 1)$ – послідовність довільних функцій розподілу. Тоді знайдеться ймовірнісний простір $(\Omega, \mathfrak{F}, \mathbf{P})$ та послідовність незалежних у сукупності випадкових величин $(\xi_n, n \geq 1)$ на ньому таких, що ξ_n має функцію розподілу G_n.

Для доведення визначимо простір $(\Omega, \mathfrak{F}) = (\mathbb{R}^\infty, \mathfrak{B}(\mathbb{R}^\infty))$ та побудуємо на $\mathfrak{F}$ продакт-міру $\mathbf{P}$ за *теоремою Колмогорова про побудову ймовірності на множинах послідовностей через сумісні функції розподілу* так, щоб

$$\mathbf{P}(\{x \in \mathbb{R}^\infty : x_1 < a_1, ..., x_n < a_n\}) = \prod\nolimits_{k=1}^{n} G_k(a_k), \quad \forall a_k \in \mathbb{R}, \forall n \geq 1.$$

Це можливо, оскільки права частина є послідовністю *сумісних функцій розподілу*, що задовольняє *умову узгодженості*. Якщо покласти $\xi(\omega) = \omega$ та $\xi_n(\omega) = \omega_n$, то за означенням $\xi = (\xi_n, n \geq 1)$ і

$$\mathbf{P}(\xi_1 < a_1, ..., \xi_n < a_n) = \mathbf{P}(\xi \in \{x \in \mathbb{R}^\infty : x_1 < a_1, ..., x_n < a_n\}) =$$
$$\mathbf{P}(\{x \in \mathbb{R}^\infty : x_1 < a_1, ..., x_n < a_n\}) = \prod\nolimits_{k=1}^{n} G_k(a_k) \ \boxdot$$

2.2.2. Вимірність відносно породженої сигма-алгебри

Теорема (про вимірність відносно сигма-алгебри, що породжена величиною). *Нехай* ξ, η *– випадкові величини. Для того, щоб величина* η *була вимірною відносно породженої сигма-алгебри* $\sigma[\xi]$, *необхідно* *і достатньо, щоб для деякої невипадкової борелєвої функції g виконувалась тотожність $\eta(\omega) = g(\xi(\omega))$, $\forall \omega \in \Omega$.*

Доведення

Достатність. За означенням прообразу відображення $\{g(\xi) \in B\} = \{\xi \in g^{(-1)}(B)\} \in \mathfrak{F}$ для всіх $B \in \mathfrak{B}(\mathbb{R})$, оскільки функція g – борелєва.

Необхідність. Нехай випадкова величина η – проста, тобто має місце зображення $\eta = \sum_{k=1}^n c_k 1\!\!\mathrm{I}_{A_k}$, де $A_k = \{\eta = c_k\} \in \sigma[\xi]$. З означення *породженої сигма-алгебри* випливає, що $A_k = \{\xi \in B_k\}$ для деякої множини $B_k \in \mathfrak{B}(\mathbb{R})$. Розглянемо борелєву функцію $g(x) = \sum_{k=1}^n c_k 1\!\!\mathrm{I}_{x \in B_k}$. Тоді $g(\xi) = \sum_{k=1}^n c_k 1\!\!\mathrm{I}_{\{\xi \in B_k\}} = \sum_{k=1}^n c_k 1\!\!\mathrm{I}_{A_k} = \eta$, що доводить за зробленого припущення потрібне зображення.

Нехай η – довільна $\sigma[\xi]$-вимірна випадкова величина. За теоремою *про апроксимацію випадкових величин простими* знайдеться послідовність простих випадкових величин η_n, що вимірні відносно сигма-алгебри $\sigma[\xi]$ і $\eta_n \to \eta$, $n \to \infty$, для всіх ω. За доведеним вище при кожному n існує борелєва функція g_n така, що $\eta_n = g_n(\xi)$. Визначимо множину

$$B = \{x : \exists \lim_{n\to\infty} g_n(x)\} = \cap_{k\ge 1} \cup_{N\ge 1} \cap_{n,m\ge N} \{x : |g_n(x) - g_m(x)| < 1/k\}$$

та функцію $g(x) = \lim_{n\to\infty} g_n(x) 1\!\!\mathrm{I}_{x\in B}$. Остання границя існує за означенням B. З наведеного зображення виводимо, що $B \in \mathfrak{B}(\mathbb{R})$, а тому функція $g(x)$ є борелєвою. Оскільки $g_n(\xi) = \eta_n \to \eta$, то за означенням B має місце включення $\xi(\omega) \in B$ при всіх ω. Однак $g_n(\xi) \to g(\xi)$ за означенням g. Тому $\eta = \lim_{n\to\infty} \eta_n = \lim_{n\to\infty} g_n(\xi) = g(\xi)$ $\boxdot$

Наслідок (про критерій вимірності відносно розбиття). *Нехай $\{H_n, n \ge 1\}$ – розбиття простору Ω, а $\mathfrak{C} = \sigma[H_n, n \ge 1]$. Випадкова величина η є вимірною відносно сигма-алгебри $\mathfrak{C}$ тоді й тільки тоді, коли при всіх ω та для деяких дійсних c_n має місце рівність $\eta(\omega) = \sum_{n\ge 1} c_n 1\!\!\mathrm{I}_{\omega \in H_n}$.*

Доведення. Визначимо $\xi = \sum_{n\ge 1} n 1\!\!\mathrm{I}_{\omega\in H_n}$. Тоді $\sigma[\xi] = \mathfrak{C}$ та з попередньої теореми виводимо, що $\eta = g(\xi) = \sum_{n\ge 1} g(n) 1\!\!\mathrm{I}_{\omega \in H_n}$ $\boxdot$

Вправи

(1) Нехай $S_n = \xi_1 + ... + \xi_n$ для деяких (ξ_k). Тоді $\sigma[\xi_1, ..., \xi_n] = \sigma[S_1, ..., S_n]$.

(2) Функції розподілу F, G такі, що $F(x) \le G(x)$ для всіх $x \in \mathbb{R}$. Довести, що існує ймовірнісний простір $(\Omega, \mathfrak{F}, \mathbf{P})$ та випадкові величини ξ, η на ньому з функціями розподілу F, G відповідно такі, що $\xi \ge \eta$ при всіх ω.

(3) Нехай $\xi, (\xi_n, n \ge 1)$ випадкові величини такі, що ξ вимірна відносно породженої сигма-алгебри $\sigma[\xi_n, n \ge 1]$. Тоді знайдеться така борелєва функція $g : \mathbb{R}^\infty \to \mathbb{R}$, що $\xi = g((\xi_n, n \ge 1))$ для всіх ω.

(4) Функції розподілу слабко збігаються: $F_n \to^W F_0, n \to \infty$. Довести, що знайдеться ймовірнісний простір $(\Omega, \mathfrak{F}, \mathbf{P})$ та випадкові величини $\xi_n, n \geq 0$, на ньому такі, що ξ_n має функцію розподілу F_n та $\xi_n \to^{P1} \xi_0, n \to \infty$.

(5) Послідовності $(\xi_n), (\eta_n)$ мають однакові скінченно-вимірні розподіли: $F_{(\xi_1,...,\xi_n)}(x_1,...,x_n) = F_{(\eta_1,...,\eta_n)}(x_1,...,x_n), \forall x_k \in \mathbb{R}, \forall n \geq 1$, та $\xi_n \to^P \xi$, $n \to \infty$. Довести, що $\eta_n \to^P \eta$ для деякої випадкової величини $\eta \simeq \xi$.

2.3. Незалежні класи подій та випадкових величин

Нехай $(\Omega, \mathfrak{F}, \mathbf{P})$ – фіксований *імовірнісний простір*. Позначимо через Θ довільну параметричну множину.

2.3.1. Незалежні класи подій

Означення. *Множина $(\mathfrak{C}_\theta, \theta \in \Theta)$ з $\mathfrak{C}_\theta \subset \mathfrak{F}$ називається системою* незалежних класів подій, *якщо для всіх $n \geq 1$, для всіх різних $\theta_k \in \Theta$, та при довільному виборі $A_k \in \mathfrak{C}_{\theta_k},\ k = \overline{1,n}$, ці події незалежні:*

$$\mathbf{P}\left(\bigcap\nolimits_{k=1}^{n} A_k\right) = \prod\nolimits_{k=1}^{n} \mathbf{P}(A_k).$$

Теорема (про критерій незалежності класів подій). *Множина $(\mathfrak{C}_\theta, \theta \in \Theta) \subset \Theta^{\mathfrak{F}}$ є системою незалежних класів подій тоді й тільки тоді, коли будь-яка скінченна підмножина $(\mathfrak{C}_{\theta_k}, k = \overline{1,n}) \subset (\mathfrak{C}_\theta, \theta \in \Theta)$ є системою незалежних класів подій.*

Доведення очевидне, оскільки умова у означенні стосується лише скінченної кількості представників з класів $\mathfrak{C}_\theta$ $\boxdot$

Означення. *Множина $\mathfrak{C} \subset \mathfrak{F}$ називається π-класом подій, якщо для всіх $A_k \in \mathfrak{C}$ виконується включення $A_1 \cap A_2 \in \mathfrak{C}$.*

Означення. *Множина $\mathfrak{D} \subset \mathfrak{F}$ називається d-класом, якщо:*

(а) $\Omega \in \mathfrak{C}$,

(б) з $A_k \in \mathfrak{D}$ та $A_2 \subset A_1$ виплива\u0454, що $A_1 \backslash A_2 \in \mathfrak{D}$,

(в) якщо $A_k \in \mathfrak{D}$ попарно несумісні, то $\cup_{k=1}^{\infty} A_k \in \mathfrak{D}$.

Теорема (про п-клас та d-клас). *Нехай d-клас $\mathfrak{D} \subset \mathfrak{F}$ містить π-клас: $\mathfrak{C} \subset \mathfrak{D}$. Тоді $\sigma[\mathfrak{C}] \subset \mathfrak{D}$.*

Доведення. Позначимо через $d[\mathfrak{C}]$ найменший d-клас, що містить клас $\mathfrak{C}$ (тобто перетин всіх таких d-класів). Доведемо, що твердження теореми виконується вже для класу $\mathfrak{D} = d[\mathfrak{C}]$, тоді воно буде справедливе для довільного d-класу, що містить $\mathfrak{C}$.

Нехай $\widetilde{\mathfrak{C}} = \{B \in \mathfrak{D} : A \cap B \in \mathfrak{D},\ \forall A \in \mathfrak{C}\}$. Тоді $\mathfrak{C} \subset \widetilde{\mathfrak{C}}$, оскільки $\mathfrak{C}$ є π-класом та $\mathfrak{C} \subset \mathfrak{D}$. Крім того, $\widetilde{\mathfrak{C}}$ є d-класом, оскільки d-операції сумісні з операцією перетину, а $\mathfrak{D}$ є d-класом. Тому $\mathfrak{D} = d[\mathfrak{C}] \subset \widetilde{\mathfrak{C}}$. Отже, $A \cap B \in \mathfrak{D}$, $\forall A \in \mathfrak{C}, \forall B \in \mathfrak{D}$.

Нехай тепер $\widehat{\mathfrak{C}} = \{A \in \mathfrak{D} : A \cap B \in \mathfrak{D},\ \forall B \in \mathfrak{D}\}$. Як і вище, доводимо, що $\widehat{\mathfrak{C}}$ є d-класом. Крім того, за попереднім твердженням $\mathfrak{C} \subset \widehat{\mathfrak{C}}$. Отже, $A \cap B \in \mathfrak{D}$, $\forall A \in \mathfrak{D}, \forall B \in \mathfrak{D}$.

Тому клас $\mathfrak{D} = d[\mathfrak{C}]$ є π-класом, та одночасно і d-класом. Отже, він замкнений відносно доповнень $\overline{A} = \Omega \backslash A$ та зліченних об'єднань, оскільки $\cup_{n\ge1} A_n = \cup_{n\ge1}(A_n \backslash \cap_{k=1}^n A_k)$. Звідси робимо висновок, що $\mathfrak{D}$ є сигма-алгеброю, тому $\sigma[\mathfrak{C}] \subset \mathfrak{D}$, що і доводить теорему $\boxdot$

Вправа. Нехай $\mathfrak{C} \subset \mathfrak{F}$ – $\pi-$клас подій, а $\mathfrak{H}$ – простір $\mathfrak{F}$-вимірних дійсних функцій, що має такі властивості: (а) $1\!\!1_A \in \mathfrak{H}$ при $A \in \mathfrak{C}$, (б) $f + g \in \mathfrak{H}$ та $cf \in \mathfrak{H}$ для всіх $f, g \in \mathfrak{H}$, $c \in \mathbb{R}$, (в) з $g_n \in \mathfrak{H}$ та $0 \le g_n \uparrow g, n \to \infty$, випливає включення $g \in \mathfrak{H}$. Довести, що простір $\mathfrak{H}$ містить всі $\sigma[\mathfrak{C}]$-вимірні обмежені функції. Вивести звідси теорему про вимірність відносно σ-алгебри, що породжена величиною.

2.3.2. Перетворення незалежних класів подій

Теорема (про незалежні π-класи подій). *Нехай* $(\mathfrak{C}_\theta, \theta \in \Theta)$ – *незалежні π-класи подій. Тоді породжені сигма-алгебри* $(\sigma[\mathfrak{C}_\theta], \theta \in \Theta)$ *також є незалежними класами подій.*

Доведення. За теоремою *про критерій незалежності класів подій* можна вважати, що множина Θ – скінченна. Оскільки природа параметричної множини неістотна, припускатимемо, що $\Theta = \{1, ..., n\}$.

Для кожного $m \in [1, n]$ визначимо такий клас множин:

$$\mathfrak{B}^*_m = \{B_m \in \sigma[\mathfrak{C}_m] : B_m \text{ та } (\cap_{k<m} B_k) \cap (\cap_{k>m} A_k) \text{ незалежні},$$

$$\forall B_k \in \sigma[\mathfrak{C}_k],\ \forall A_k \in \mathfrak{C}_k\}.$$

(1) Клас $\mathfrak{B}^*_1$ має такі властивості.

(1а) $\mathfrak{C}_1 \in \mathfrak{B}^*_1$ за умовою незалежності $\mathfrak{C}_k$.

(1б) За теоремою *про перетворення незалежних подій* клас $\mathfrak{B}^*_1$ містить множину Ω та замкнений відносно вкладеної різниці та зліченного об'єднання множин, які не перетинаються.

(1в) З (1а), (1б) та з означення d-класу виводимо, що $\mathfrak{B}_1^*$ містить d-клас $d[\mathfrak{C}_1]$, що породжується класом $\mathfrak{C}_1$. За теоремою *про п-клас та d-клас* $\sigma[\mathfrak{C}_1] \subset d[\mathfrak{C}_1]$, тому $\mathfrak{B}_1^* = \sigma[\mathfrak{C}_1]$.

(2) Розглянемо клас $\mathfrak{B}_2^*$.

(2а) З отриманого твердження (1в) та з незалежності $\mathfrak{C}_k$ виводимо, що $\mathfrak{C}_2 \in \mathfrak{B}_2^*$, оскільки за означенням $\mathfrak{B}_1^*$

$$\mathbf{P}\left(B_1 \cap \left(\bigcap_{k>1} A_k\right)\right) = \mathbf{P}(B_1)\mathbf{P}\left(\bigcap_{k>1} A_k\right) = \mathbf{P}(B_1)\prod_{k=2}^n \mathbf{P}(A_k) =$$
$$\mathbf{P}\left(B_1 \cap \left(\bigcap_{k>2} A_k\right)\right)\mathbf{P}(A_2),\ \forall B_1 \in \sigma[\mathfrak{C}_1],\ \forall A_k \in \mathfrak{C}_k,$$

де остання рівність випливає з попередніх, якщо в них покласти $A_2 = \Omega$.

(2б) Як і у (1б), доводимо, що клас $\mathfrak{B}_2^*$ замкнений відносно d-операцій.

(2в) Тому $\mathfrak{B}_2^* = d[\mathfrak{C}_2] = \sigma[\mathfrak{C}_2]$, тобто

$$\mathbf{P}\left(B_1 \cap B_2 \cap \left(\bigcap_{k>2} A_k\right)\right) = \mathbf{P}(B_2)\mathbf{P}\left(B_1 \cap \left(\bigcap_{k>2} A_k\right) =\right.$$
$$\mathbf{P}(B_1)\mathbf{P}(B_2)\prod_{k=3}^n \mathbf{P}(A_k),$$

де остання рівність є наслідком твердження (2в) при $B_2 = \Omega$.

Повторюючи ці міркування, за індукцією доводимо, що події B_m і $\cap_{k<m} B_k$ незалежні для всіх $m \le n$ та $B_k \in \sigma[\mathfrak{C}_k]$. Тому

$$\mathbf{P}\left(B_1 \cap B_2 ... \cap B_n\right) = \mathbf{P}\left(B_1 ... \cap B_{n-1}\right)\mathbf{P}\left(B_n\right) = ... = \prod_{k=1}^n \mathbf{P}(B_k)\ \boxdot$$

Теорема (про об'єднання незалежних п-класів). *Нехай незалежні* π*-класи подій* $(\mathfrak{C}_\theta, \theta \in \Theta)$ *мають параметричну множину* Θ, *що утворена розбиттям* $\Theta = \cup_{\gamma\in\Gamma}\Theta_\gamma,\ \Theta_\gamma \cap \Theta_\beta = \varnothing, \gamma \ne \beta$. *Тоді породжені сигма-алгебри* $(\sigma[\cup_{\theta\in\Theta_\gamma}\mathfrak{C}_\theta], \gamma \in \Gamma)$ *також є незалежними класами подій.*

Доведення. Враховуючи теорему *про незалежні п-класи подій*, можна вважати, що класи $\mathfrak{C}_\theta$ є сигма-алгебрами, адже породжені сигма-алгебри $\sigma[\mathfrak{C}_\theta]$ – незалежні, причому $\sigma[\cup_{\theta\in\Theta_\gamma}\mathfrak{C}_\theta] = \sigma[\cup_{\theta\in\Theta_\gamma}\sigma[\mathfrak{C}_\theta]]$.

Розглянемо для фіксованого $\gamma \in \Gamma$ такий клас подій:

$$\Psi_\gamma = \{A_{\theta_1} \cap A_{\theta_2} ... \cap A_{\theta_n},\ \forall A_{\theta_k} \in \mathfrak{C}_{\theta_k}, \forall \theta_k \in \Theta_\gamma : \theta_k \ne \theta_j\ \forall k \ne j,\ n \ge 1\}.$$

Цей клас є π-класом, оскільки перетин двох подій з Ψ_γ перегрупуванням можна зобразити у вигляді перетину подій з різними індексами θ. Далі, за умовою незалежності $\mathfrak{C}_\theta$ класи $(\Psi_\gamma, \gamma \in \Gamma)$ є *незалежними класами подій*:

$$\mathbf{P}\left(\bigcap_{i=1}^m \bigcap_{k=1}^{n_i} A_{\theta_{ik}}\right) = \prod_{i=1}^m \prod_{k=1}^{n_i} \mathbf{P}(A_{\theta_{ik}}) = \prod_{i=1}^m \mathbf{P}\left(\bigcap_{k=1}^{n_i} A_{\theta_{ik}}\right),\ \theta_{ik} \in \Psi_{\gamma_i},$$

де індекси θ_{ik} є різними для кожного фіксованого i за означенням класів Ψ_γ, а при різних i – через попарну несумісність індексних множин Θ_γ.

Отже, за теоремою *про незалежні π-класи подій* класи $(\sigma[\Psi_\gamma], \gamma \in \Gamma)$ є *незалежними класами подій.*

Нарешті, зі включень $\Psi_\gamma \subset \sigma[\cup_{\theta\in\Theta_\gamma}\mathfrak{C}_\theta]$ та $\cup_{\theta\in\Theta_\gamma}\mathfrak{C}_\theta \subset \Psi_\gamma$ виводимо, що $\sigma[\Psi_\gamma] = \sigma[\cup_{\theta\in\Theta_\gamma}\mathfrak{C}_\theta]$, що доводить незалежність останніх класів подій ⊡

Вправи

(1) Якщо $(\mathfrak{C}_\theta, \theta \in \Theta)$ – незалежні класи подій, а $\Upsilon \subset \Theta$, то $(\mathfrak{C}_\theta, \theta \in \Upsilon)$ також є незалежними класами подій.

(2) Сигма-алгебри $(\mathfrak{F}_k, k = \overline{1,n})$ незалежні у сукупності тоді й тільки тоді, коли при кожному $k > 1$ незалежні сигма-алгебри $(\mathfrak{F}_k, \sigma[\cup_{i<k}\mathfrak{F}_i])$.

(3) Вказати незалежні класи $\mathfrak{C}_1, \mathfrak{C}_2$ такі, що $\sigma[\mathfrak{C}_1]$, $\sigma[\mathfrak{C}_2]$ не є незалежними.

(4) Нехай $\mathfrak{F}_1, \mathfrak{F}_2, \mathfrak{F}_3 \subset \mathfrak{F}$ під-сигма-алгебри випадкових подій такі, що $\mathfrak{F}_1 \subset \sigma[\mathfrak{F}_2 \cup \mathfrak{F}_3]$ та $\mathfrak{F}_2 \subset \sigma[\mathfrak{F}_1 \cup \mathfrak{F}_3]$, причому сигма-алгебри $\mathfrak{F}_3$ та $\sigma[\mathfrak{F}_1 \cup \mathfrak{F}_2]$ незалежні. Довести, що сигма-алгебри $\mathfrak{F}_1, \mathfrak{F}_2$ нерозрізнимі: для кожної події $A_1 \in \mathfrak{F}_1$ знайдеться $A_2 \in \mathfrak{F}_2$ така, що $\mathbf{P}(A_1 \Delta A_2) = 0$.

2.3.3. Закон 0 та 1 Колмогорова

Означення. *Нехай* $\mathfrak{F}_n \subset \mathfrak{F}, n \geq 1,$ *– послідовність під-сигма-алгебр. Їх* залишковою сигма-алгеброю *називається сигма-алгебра подій*

$$\overline{\lim}_{n\to\infty}\mathfrak{F}_n \equiv \cap_{n\geq 1}\sigma[\cup_{k\geq n}\mathfrak{F}_k].$$

Теорема (про закон нуля та одиниці Колмогорова). *Нехай* $(\mathfrak{F}_n)$ *– послідовність незалежних сигма-алгебр, а* $\Delta_\infty \equiv \overline{\lim}_{n\to\infty}\mathfrak{F}_n$ *– породжена залишкова сигма-алгебра. Тоді всі події* $A \in \Delta_\infty$ *– вироджені:* $\mathbf{P}(A) \in \{0, 1\}$.

Доведення. Позначимо $\Delta_n = \sigma[\cup_{k\geq n}\mathfrak{F}_k]$ та $\Lambda_n = \sigma[\cup_{k\leq n}\mathfrak{F}_k]$. За означенням $\Delta_n \downarrow \Delta_\infty, n \to \infty$.

Нехай $A \in \Delta_\infty$. Тоді $A \in \Delta_n$ для всіх $n \geq 1$. За теоремою *про об'єднання незалежних π-класів* класи Δ_n та Λ_{n-1} незалежні, тому клас $\{A\}$ не залежить від Λ_{n-1} для всіх $n > 1$. Отже, клас $\{A\}$ не залежить від $\cup_{n>1}\Lambda_{n-1} \equiv \Lambda_\infty$. Оскільки останній клас є алгеброю і, зокрема, π-класом, то за теоремою *про незалежні π-класи подій,* сигма-алгебра $\sigma[\{A\}]$ не залежить від сигма-алгебри $\sigma[\Lambda_\infty] \supset \sigma[\cup_{n\geq 1}\mathfrak{F}_n] = \Delta_1$. Як показано вище, $A \in \Delta_1$. Тому за означенням *незалежних класів подій* подія $A \in \sigma[\{A\}]$ не залежить від події $A \in \Delta_1 : \mathbf{P}(A) = \mathbf{P}(A \cap A) = \mathbf{P}(A)\mathbf{P}(A)$, отже $\mathbf{P}(A) \in \{0, 1\}$ ⊡

2.3.4. Незалежні класи випадкових векторів

Означення. *Випадкові вектори з класу* $(\xi_\theta, \theta \in \Theta)$ *незалежні у сукупності, якщо породжені сигма-алгебри* $\{\sigma[\xi_\theta], \theta \in \Theta\}$ *є незалежними класами подій, тобто для всіх* $n \geq 1$, *та для всіх різних* $\theta_k \in \Theta$ *і довільних* $A_k \in \sigma[\xi_{\theta_k}], k = \overline{1, n}$,

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

Надалі позначатимемо

$$\sigma[\xi_\theta, \theta \in \Theta] \equiv \sigma[\cup_{\ \theta\in\Theta}\ \sigma[\xi_\theta]].$$

Теорема (про критерій незалежності класів векторів). *Випадкові вектори з класу* $(\xi_\theta, \theta \in \Theta)$ *незалежні у сукупності тоді й тільки тоді, коли незалежними є класи подій* $(\{\xi_\theta < x\}_{x\in\mathbb{R}^d}, \theta \in \Theta)$.

Доведення. Клас подій $\{\xi_\theta < x\}_{x\in\mathbb{R}^d}$ є π-класом та породжує сигма-алгебру $\sigma[\xi_\theta]$ за теоремою *про критерій вимірності функції*. Тому твердження теореми випливає з теореми *про незалежні п-класи подій* $\boxdot$

Теорема (про незалежність згрупованих класів векторів). *Нехай випадкові вектори з класу* $(\xi_\theta, \theta \in \Theta)$ *незалежні у сукупності і має місце розбиття* $\Theta = \cup_{\gamma\in\Gamma}\Theta_\gamma,\ \Theta_\gamma \cap \Theta_\beta = \varnothing, \gamma \neq \beta$. *Тоді породжені сигма-алгебри* $(\sigma[\xi_\theta, \theta \in \Theta_\gamma],\ \gamma \in \Gamma)$ *незалежні у сукупності.*

Доведення є очевидним наслідком теореми *про об'єднання незалежних п-класів* та означення сигма-алгебри $\sigma[\xi_\theta, \theta \in \Theta_\gamma]$ $\boxdot$

Наслідок. *Якщо випадкові вектори з класу* $(\xi_\theta, \theta \in \Theta)$ *незалежні у сукупності, а функції* $g_\gamma : \mathbb{R}^{\Theta_\gamma} \to \mathbb{R}^d$ *– борелєві, то випадкові величини з суперпозицій* $(g_\gamma((\xi_\theta, \theta \in \Theta_\gamma)), \gamma \in \Gamma)$ *є незалежними у сукупності.*

Наслідок. *Якщо випадкові величини* $(\xi_n, n \geq 1)$ *незалежні у сукупності, то ймовірності наступних подій дорівнюють* 0 *або* 1:

$$\{\exists \lim_{n\to\infty} \xi_n < a\}, \quad \left\{\sum\nolimits_{n\geq 1} |\xi_n| < \infty\right\},$$

$$\{\overline{\lim}_{n\to\infty}(\xi_1 + ... + \xi_n)/n < b\},\ \{\exists \lim_{n\to\infty}(\xi_1 + ... + \xi_n)/n < c\}.$$

Доведення випливає з теореми *про закон нуля та одиниці Колмогорова,* оскільки наведені випадкові події належать *залишковій сигма-алгебрі* $\overline{\lim}_{n\to\infty}\sigma[\xi_n]$ $\boxdot$

Вправи

(1) Нехай $\mathfrak{F}_n = \{\{x \in \mathbb{R}^\infty : x_n \in B\}, B \in \mathfrak{B}[\mathbb{R}]\} \subset \mathfrak{B}[\mathbb{R}^\infty]$. Довести, що залишкова σ-алгебра $\overline{\lim}_{n\to\infty}\mathfrak{F}_n = \cap_{n\geq 1}(\mathbb{R}^n \times \mathfrak{B}[\mathbb{R}^\infty])$.

(2) Випадкові величин $(\xi_n, n \geq 1)$ незалежні у сукупності. Довести, що радіус збіжності ряду $\sum_{n\geq 1} \xi_n$ є м.н. сталою.

(3) Навести приклад послідовності випадкових величин (ξ_n) та залишкової події $A \in \overline{\lim}_{n\to\infty}\sigma[\xi_n]$ такої, що $\mathbf{P}(A) = 1/2$.

(4) Випадкові величин $(\xi_n, n \ge 1)$ незалежні у сукупності, а послідовність $(n_k, k \ge 1)$ строго зростає. Довести, що величини $(g_k(\xi_{n_k}, ..., \xi_{n_{k+1}-1}),\ k \ge 1)$ незалежні для довільних борелєвих функції g_k.

(5) Випадкові вектори $X = (\xi_1, ..., \xi_n)$ та $Y = (\eta_1, ..., \eta_m)$ незалежні, і кожний з них складається з незалежних випадкових величин. Довести, що (а) вектор $(\xi_1, ..., \xi_n, \eta_1, ..., \eta_m)$ містить незалежні у сукупності випадкові величини, (б) обернене твердження також має місце.

2.3.5. Закон 0 та 1 Хьюітта-Севіджа

Нехай $(\xi_n, n \ge 1)$ – послідовність незалежних однаково розподілених випадкових величин, а $S_n = \xi_1 + ... + \xi_n$. Тоді подія $\{S_n = 0$ нескінченно часто$\}$ не належить *залишковій сигма-алгебрі*, однак також є виродженою. Для доведення останнього твердження розглянемо такі визначення.

Означення. *Бієкція* $\pi : \mathbb{N} \to \mathbb{N}$ *є* скінченою перестановкою, *якщо* $\pi_n = n$ *для всіх* $n \ge N$ *та деякого* N.

Для довільної послідовності $x = (x_n, n \ge 1) \in \mathbb{R}^\infty$ та перестановки π позначимо борелєве відображення $\pi(x) = (x_{\pi_n}, n \ge 1) : \mathbb{R}^\infty \to \mathbb{R}^\infty$.

Означення. *Сигма-алгеброю* переставних множин *з* $\mathbb{R}^\infty$ *назвемо клас*

$$\mathfrak{X} = \{B \in \mathfrak{B}(\mathbb{R}^\infty) : B = \pi^{(-1)}(B),\ \forall \text{скінчених перестановок } \pi\}.$$

Для послідовності випадкових величин $(\xi_n, n \ge 1)$ *розглянемо породжену нею сигма-алгебру* переставних подій*:*

$$\mathfrak{X}[\xi] = \{A = \{\xi \in B\},\ B \in \mathfrak{X}\}.$$

Приклад. Підмножини $\mathbb{R}^\infty$ вигляду $B_1 = \{x = (x_n) : \sum_{n\ge1}|x_n| < \infty\}$, $B_2 = \{x = (x_n) : x_1 + ... + x_n = 0$ для нескінченої множини номерів $n\}$, $B_3 = \{x = (x_n) : \exists \lim x_n\}$ належать $\mathfrak{X}$. Зауважимо, що на відміну від інших, множина B_2 не належить *залишковій сигма-алгебрі* $\cap_{n\ge1}(\mathbb{R}^n \times \mathfrak{B}(\mathbb{R}^\infty))$, тому до події $\{\xi \in B_2\}$ з довільною послідовністю $\xi = (\xi_n)$ незалежних випадкових величин не можна застосувати теорему *про закон нуля та одиниці Колмогорова*.

Теорема (про закон нуля та одиниці Хьюітта – Севіджа). *Нехай* $\xi = (\xi_n)$ *– послідовність незалежних однаково розподілених випадкових величин. Тоді* σ*-алгебра* $\mathfrak{X}[\xi]$ *вироджена:* $\mathbf{P}(A) \in \{0, 1\}$ *для всіх* $A \in \mathfrak{X}[\xi]$.

Доведення. Позначимо $P_\xi(B) = \mathbf{P}(\xi \in B)$ розподіл випадкового елемента ξ на $\mathfrak{B}(\mathbb{R}^\infty)$. Нехай $A \in \mathfrak{X}[\xi]$, тобто $A = \{\xi \in B\}$ для деякої $B \in \mathfrak{X}$.

Оскільки борелєва сигма-алгебра $\mathfrak{B}(\mathbb{R}^\infty)$ породжується алгеброю *циліндричних множин* $J \equiv \cup_{n\ge 1}\{J_n(B_n),\ B_n \in \mathfrak{B}(\mathbb{R}^n)\}$, то знайдеться послідовність множин $B_n \in \mathfrak{B}(\mathbb{R}^n)$ така, що $P_\xi(B\Delta J_n(B_n)) \to 0, n \to \infty$.

Позначимо $A_n = \{\xi \in J_n(B_n)\} = \{(\xi_1, ..., \xi_n) \in B_n\}$. Тоді $\mathbf{P}(A\Delta A_n) = P_\xi(B\Delta J_n(B_n)) \to 0, n \to \infty$.

Визначимо перестановку $\pi_n = (n+1, ..., 2n, 1, ..., n, 2n+1, 2n+2, ...)$, що переставляє перші n номерів з наступними n номерами. З незалежності та однакової розподіленості величин послідовності $\xi = (\xi_n, n \ge 1)$ внаслідок *теореми Колмогорова про побудову ймовірності на множинах послідовностей через сумісні функції розподілу* виводимо, що $P_\xi(B) = P_{\pi_n(\xi)}(B)$ для всіх $B \in \mathfrak{B}(\mathbb{R}^\infty)$, оскільки послідовність $\pi_n(\xi)$ також складається з незалежних однаково розподілених величин з таким же розподілом, і має ті самі скінченно-вимірні розподіли, що і ξ. Тому

$$\mathbf{P}(A\Delta A_n) = P_\xi(B\Delta J_n(B_n)) = P_{\pi_n(\xi)}(B\Delta J_n(B_n)) =$$
$$\mathbf{P}(\{\pi_n(\xi) \in B\}\Delta\{\pi_n(\xi) \in J_n(B_n)\}) =$$
$$\mathbf{P}(\{\xi \in \pi_n^{-1}(B)\}\Delta\{(\xi_{n+1}, ..., \xi_{2n}) \in B_n\}) = \mathbf{P}(A\Delta A_n'),$$

де згідно з включенням $B \in \mathfrak{X}$ враховано тотожність $\pi_n^{-1}(B) = B$, і за означенням подія A_n' дорівнює $\{(\xi_{n+1}, ..., \xi_{2n}) \in B_n\}$.

Отже, $\mathbf{P}(A\Delta(A_n \cap A_n')) \le\ \mathbf{P}(A\Delta A_n) + \mathbf{P}(A\Delta A_n') \to 0, n \to \infty$. Звідси з нерівності $|\mathbf{P}(A_1) - \mathbf{P}(A_2)| \le \mathbf{P}(A_1\Delta A_2)$ виводимо, що $\mathbf{P}(A_n) \to \mathbf{P}(A)$, $\mathbf{P}(A_n') \to \mathbf{P}(A)$, $\mathbf{P}(A_n \cap A_n') \to \mathbf{P}(A), n \to \infty$. Оскільки події A_n, A_n' незалежні, то $\mathbf{P}(A_n \cap A_n') = \mathbf{P}(A_n)\mathbf{P}(A_n') \to \mathbf{P}^2(A)$. Тому $\mathbf{P}^2(A) = \mathbf{P}(A)$ $\boxdot$

Вправи

(1) Випадкові величини (ξ_n) незалежні у сукупності. (а) Довести, що величини $\overline{\lim}_{n\to\infty}\xi_n$, $\underline{\lim}_{n\to\infty}\xi_n$ є сталими м.н. (б) Якщо $S_n = \xi_1 + ... + \xi_n$, і $0 < b_n \to \infty$, то величини $\overline{\lim}_{n\to\infty}S_n/b_n, \underline{\lim}_{n\to\infty}S_n/b_n$ також є сталими м.н.

(2) Випадкові величини (ξ_n) незалежні, $\mathbf{P}(\xi_n = 1) = p = 1 - \mathbf{P}(\xi_n = -1)$, а $S_n = \xi_1 + ... + \xi_n$ – блукання Бернуллі. Тоді $\mathbf{P}(\overline{\lim}_{n\to\infty}\{S_n = 0\}) \in \{0, 1\}$. Ця ймовірність нульова тоді й тільки тоді, коли $p \ne 1/2$.

(3) Випадкові величини (ξ_n) незалежні та однаково розподілені, причому $\mathbf{E}\xi_1 = 0 < \mathbf{E}\xi_1^2 < \infty$. Довести, що $\overline{\lim}_{n\to\infty}n^{-1/2}\,|\xi_1 + ... + \xi_n| = \infty$ м.н.

2.4. Рівномірна інтегровність

Властивість рівномірної інтегровності є більш адекватним аналогом умови мажорованої збіжності для здійснення граничного переходу під знаком математичного сподівання.

Означення. *Послідовність інтегровних величин* $(\xi_n,\ n \geq 1)$ *називається* рівномірно інтегровною*, якщо*

$$\sup_{n\geq 1} \mathbf{E}\,|\xi_n|\, 1\!\!1_{\{|\xi_n|\geq c\}} \to 0,\ c \to \infty.$$

Існування границі тут випливає з монотонності за c лівої частини.

Зауваження. Для сталої послідовності $\xi_n \equiv \xi$ рівномірна інтегровність є наслідком інтегровності за *теоремою Лебега про мажоровану збіжність*, оскільки $|\xi|\, 1\!\!1_{\{|\xi|\geq c\}} \to 0,\ c \to \infty$, та $|\xi|\, 1\!\!1_{\{|\xi|\geq c\}} \leq |\xi|$.

Теорема (про умови рівномірної інтегровності).

(1) З умови мажорованої інтегровності: $|\xi_n| \leq \eta$, *де величина* η *– інтегровна, випливає рівномірна інтегровність.*

(2) Послідовність інтегровних величин (ξ_n) *рівномірно інтегровна тоді й тільки тоді, коли*

$$\lim_{c\to\infty} \overline{\lim}_{n\to\infty} \mathbf{E}\,|\xi_n|\, 1\!\!1_{\{|\xi_n|\geq c\}} = 0.$$

(3) Якщо послідовність (ξ_n) *рівномірно інтегровна, і* $|\zeta_n| \leq |\xi_n|$, *то послідовність* (ζ_n) *також рівномірно інтегровна.*

(4) Якщо послідовності $(\xi_n), (\zeta_n)$ *рівномірно інтегровні, то їх сума* $(\xi_n + \zeta_n)$ *також рівномірно інтегровна.*

(5) Якщо послідовність (ξ_n) *рівномірно інтегровна, а* (ζ_n) *– обмежена, то їх добуток* $(\xi_n \zeta_n)$ *– рівномірно інтегровний.*

(6) Для рівномірної інтегровності (ξ_n) *необхідно і достатньо, щоб для деякої невід'ємної борелєвої функції* $V(x), x \geq 0, \lim_{x\to\infty} V(x)/x = \infty$ *та* $\sup_{n\geq 1} \mathbf{E}V(|\xi_n|) < \infty$.

Доведення. Розглянемо неспадну за x функцію $\psi_c(x) = x 1\!\!1_{x\geq c}$ для $x \geq 0$, та відмітимо, що умова рівномірної інтегровності еквівалентна такій умові: $\sup_{n\geq 1} \mathbf{E}\psi_c(|\xi_n|) \to 0, c \to \infty$.

(1) За означенням $\mathbf{E}\psi_c(|\xi_n|) \leq \mathbf{E}\psi_c(\eta) \to 0$ при $c \to \infty$ рівномірно за n за *теоремою Лебега про мажоровану збіжність*, оскільки $\psi_c(\eta) \to 0$ при $c \to \infty$ та $\psi_c(\eta) \leq \eta$.

(2) Необхідність наведеної умови очевидна. Оскільки величини інтегровні за умовою, то для кожного n внаслідок твердження (1) існує монотонна границя $\mathbf{E}\,|\xi_n|\, 1\!\!1_{\{|\xi_n|\geq c\}} \downarrow 0, c \to \infty$. Тому дане твердження є наслідком леми *про верхні межу та границю монотонної послідовності* $\varepsilon_n(c) \equiv \mathbf{E}\,|\xi_n|\, 1\!\!1_{\{|\xi_n|\geq c\}}$.

(3) З монотонності $\psi_c(x)$ виводимо, що $\mathbf{E}\psi_c(|\zeta_n|) \leq \mathbf{E}\psi_c(|\xi_n|) \to 0$ при $c \to \infty$ рівномірно за n.

(4) Твердження є наслідком означення рівномірної інтегровності та нерівності $\psi_c(x+y) \leq 2\psi_{c/2}(x) + 2\psi_{c/2}(y)$.

(5) Якщо $|\zeta_n| \le a$, то $\mathbf{E}\psi_c(|\xi_n\zeta_n|) \le \mathbf{E}\psi_c(|\xi_n a|) = a\mathbf{E}\psi_{c/a}(|\xi_n|) \to 0$ при $c \to \infty$ рівномірно за n.

(6) Для доведення *достатності* позначимо $\varepsilon(c) = \sup_{x \ge c} x/V(x)$ та зауважимо, що $\varepsilon(c) \to 0, c \to \infty$, і $\psi_c(x) \le \varepsilon(c)V(x)$ для всіх $x \ge 0$. Звідси $\mathbf{E}\psi_c(|\xi_n|) \le \varepsilon(c)\mathbf{E}V(|\xi_n|) \le \varepsilon(c)\sup_{n\ge1}\mathbf{E}V(|\xi_n|) \to 0$ при $c \to \infty$ рівномірно за n. *Необхідність* доводиться вибором $V(x) = x\left(\sup_{n\ge1}\mathbf{E}\psi_x(|\xi_n|)\right)^{-1/2}$ ⊡

Теорема (про граничний перехід для рівномірно інтегрованої послідовності). *Нехай послідовність (ξ_n) є рівномірно інтегровною.*

(а –нерівності Фату) Тоді

$$\mathbf{E}\,\underline{\lim}_{n\to\infty}\xi_n \le \underline{\lim}_{n\to\infty}\mathbf{E}\xi_n \le \overline{\lim}_{n\to\infty}\mathbf{E}\xi_n \le \mathbf{E}\,\overline{\lim}_{n\to\infty}\xi_n.$$

(б –граничний перехід) Якщо додатково $\xi_n \to \xi,\ n \to \infty$, то величина ξ є інтегровною і має місце збіжність

$$\mathbf{E}\xi_n \to \mathbf{E}\xi, \quad \mathbf{E}\,|\xi_n - \xi| \to 0,\ n \to \infty.$$

Доведення.

(а) Позначимо $\zeta_n^c = \xi_n \mathbb{1}_{\{\xi_n \ge c\}}$, $\xi_n^c = \xi_n - \zeta_n^c = \xi_n \mathbb{1}_{\{\xi_n < c\}}$. За означенням $\xi_n^c \le c, \xi_n^c \le \xi_n$. Звідси за лівою *нерівністю Фату теореми Лебега про мажоровану збіжність* отримуємо праву нерівність:

$$\mathbf{E}\overline{\lim}_{n\to\infty}\xi_n \ge \mathbf{E}\overline{\lim}_{n\to\infty}\xi_n^c = c - \mathbf{E}\underline{\lim}_{n\to\infty}(c - \xi_n^c) \ge$$

$$c - \underline{\lim}_{n\to\infty}\mathbf{E}(c - \xi_n^c) = \overline{\lim}_{n\to\infty}\mathbf{E}\xi_n^c \ge \overline{\lim}_{n\to\infty}\mathbf{E}\xi_n - \sup_{n\ge1}\mathbf{E}\zeta_n^c,$$

де завдяки умові *рівномірної інтегровності* останній доданок можна зробити як завгодно малим вибором сталої c.

Для доведення лівої нерівності досить застосувати праву нерівність до послідовності $(-\xi_n)$, яка також є рівномірно інтегровною.

(б) З $\xi_n \to^{\mathbf{P}1} \xi$ випливає, що $|\xi_n| \to^{\mathbf{P}1} |\xi|, n \to \infty$, де послідовність $|\xi_n|$ *рівномірно інтегровна*. Тому послідовність $\mathbf{E}\,|\xi_n|$ обмежена:

$$\mathbf{E}\,|\xi_n| \le c + \sup_{n\ge1}\mathbf{E}\,|\xi_n|\,\mathbb{1}_{\{|\xi_n|\ge c\}} < \infty.$$

Отже, з лівої *нерівності Фату* для невід'ємних величин виводимо інтегровність: $\mathbf{E}\,|\xi| \le \underline{\lim}_{n\to\infty}\mathbf{E}\,|\xi_n| < \infty$.

Зі збіжності $\lim_{n\to\infty}\xi_n = \xi$ виводимо, що всі нерівності в (а) є рівностями, а тому існує $\lim_{n\to\infty}\mathbf{E}\xi_n = \mathbf{E}\xi$

Нарешті, застосувавши останнє твердження до рівномірно інтегровної послідовності $|\xi_n - \xi|$, що збігається до нуля, отримаємо збіжність *у середньому* $\mathbf{E}\,|\xi_n - \xi| \to 0$ ⊡

Теорема (про необхідність рівномірної інтегровності). *Нехай послідовність невід'ємних інтегровних величин* (ξ_n) *поточково збігається* $\xi_n \to \xi$, $n \to \infty$, *і* $\mathbf{E}\xi < \infty$. *Для того, щоб* $\lim_{n\to\infty} \mathbf{E}\xi_n = \mathbf{E}\xi$, *необхідно і достатньо, щоб послідовність* (ξ_n) *була рівномірно інтегровною.*

Доведення. *Достатність* доведено у теоремі *про граничний перехід для рівномірно інтегрованої послідовності.*

Для доведення *необхідності* позначимо $C = \{c : \mathbf{P}(\xi = c) = 0\}$. Оскільки доповнення C не більш ніж зліченне, то множина C – щільна. Функція $x 1\!\!1_{x<c}$ неперервна у точці $\xi(\omega)$ при майже всіх $\omega \in \Omega$, та обмежена сталою c, тому за *теоремою Лебега про мажоровану збіжність*

$$\lim_{n\to\infty} \mathbf{E}\xi_n 1\!\!1_{\{\xi_n<c\}} = \mathbf{E}\lim_{n\to\infty} \xi_n 1\!\!1_{\{\xi_n<c\}} = \mathbf{E}\xi 1\!\!1_{\{\xi<c\}}.$$

Звідси за теоремою *про умови рівномірної інтегровності:*

$$\overline{\lim}_{n\to\infty}\mathbf{E}\xi_n 1\!\!1_{\{\xi_n\ge c\}} =$$

$$\overline{\lim}_{n\to\infty} \left(\mathbf{E}\xi_n - \mathbf{E}\xi - \mathbf{E}\xi_n 1\!\!1_{\{\xi_n<c\}} + \mathbf{E}\xi 1\!\!1_{\{\xi<c\}} + \mathbf{E}\xi 1\!\!1_{\{\xi\ge c\}}\right) =$$
$$0 + 0 + \mathbf{E}\xi 1\!\!1_{\{\xi\ge c\}} \to 0,\ c \to \infty.$$

Унаслідок цієї ж теореми послідовність (ξ_n) – рівномірно інтегровна ⊡

Вправи

(1) Послідовність $(\xi_n^+,\ n \ge 1)$ рівномірно інтегровна. Довести нерівність Фату: $\overline{\lim}_{n\to\infty}\mathbf{E}\xi_n \le \mathbf{E}\overline{\lim}_{n\to\infty}\xi_n$.

(2) Імовірнісний простір $(\Omega, \mathfrak{F}, \mathbf{P})$ побудований на відрізку $[0,1]$ з борелєвою сигма-алгеброю та мірою Лебега. Нехай $\xi_n(\omega) = n 1\!\!1_{\omega\in(1/(n+1),1/n]}$. Тоді $\xi_n \to 0$, $n \to \infty$, та $\mathbf{E}\xi_n \to 0$, однак ця послідовність не є інтегровано мажорованою, оскільки $\mathbf{E}\sup_{n\ge1} \xi_n = \infty$, та є рівномірно інтегровною.

(3) Невід'ємна випадкова величина ζ невід'ємна та інтегровна: $\mathbf{E}\zeta < \infty$. Тоді знайдеться невід'ємна борелєва функція $V(x), x \ge 0$, така, що $\mathbf{E}V(\zeta) < \infty$ та $\lim_{x\to\infty} V(x)/x = \infty$.

(4) Послідовність випадкових величин $(\xi_n, n \ge 1)$ рівномірно інтегровна. Довести, що $\mathbf{E}\max_{1\le k\le n} |\xi_k| = o(n), n \to \infty$.

(5) Випадкові величини $(\xi_n, n \ge 1)$ незалежні однаково розподілені та інтегровні: $\mathbf{E}|\xi_1| < \infty$. Довести, що послідовність $(\xi_1 + ... + \xi_n)/n$ рівномірно інтегровна. Вивести звідси, що в *критерії Колмогорова посиленого закону великих чисел* має місце також збіжність у середньому.

(6) Випадкові величини $(\xi_n, n \ge 1)$ незалежні і $\mathbf{E}\xi_n = 0$, $\mathbf{E}\xi_n^2 = 1$. Довести, що послідовність $((\xi_1 + ... + \xi_n)/n, n \ge 1)$ рівномірно інтегровна.

(7) Довести теорему про граничний перехід для рівномірно інтегровної послідовності за умови її збіжності лише за ймовірністю.

(8) Випадкові величини (ξ_n) інтегровні у степені $p>0$ та $\xi_n \to^P \xi_0$, $n\to\infty$. Довести, що наступні твердження еквівалентні: (a) $\mathbf{E}\,|\xi_n-\xi_0|^p \to 0, n\to\infty$, (б) $\mathbf{E}\,|\xi_n|^p \to \mathbf{E}\,|\xi_0|^p\,, n\to\infty$, (в) послідовність $(|\xi_n|^p)$ рівномірно інтегровна.

(9) Послідовність $(\xi_n, n\ge 1)$ рівномірно інтегровна і $\mathbf{E}\xi_n = a$, а величини $(\eta_n, n\ge 1)$ обмежені: $|\eta_n|\le b$ та $\eta_n \to^P c\in\mathbb{R}$, при $n\to\infty$. Довести збіжність $\mathbf{E}\xi_n\eta_n \to ac, n\to\infty$.

(10) Послідовність $(\xi_n, n\ge 1)$ рівномірно інтегровна. Довести, що для довільної послідовності випадкових подій з $A_m\downarrow\varnothing$, $m\to\infty$, випливає співвідношення $\lim_{m\to\infty}\sup_{n\ge1}\mathbf{E}\,|\xi_n|\,1\!\!\mathrm{I}_{A_m}=0$.

(11) Послідовність $(\xi_n, n\ge 1)$ рівномірно інтегровна тоді й тільки тоді, коли функція $\sup_{n\ge1}\mathbf{E}\,|\xi_n|\,1\!\!\mathrm{I}_A, A\in\mathfrak{F}$, обмежена і абсолютно неперервна відносно ймовірності $\mathbf{P}$.

2.5. Умовне математичне сподівання

Поняття *умовної ймовірності* $\mathbf{P}(A\mid B)$ за умови, що відбулася випадкова подія B, використовується для описання умовного *стохастичного експерименту*, у якому постулюється, що подія B вже відбулася. Однак визначення такого поняття суттєво спирається на припущення про додатність ймовірності умови $\mathbf{P}(B)$. Водночас низка прикладів вказує, що умовну ймовірність доцільно визначити і у випадках, коли вказане припущення не виконується. Наприклад, у складному стохастичному експерименті, де ймовірність успіху у схемі випробувань Бернуллі визначається результатом вибору випадкової точки на інтервалі $[0,1]$, значення відповідних умовних імовірностей очевидні, однак не мають формальних визначень у рамках елементарної теорії ймовірностей. Для подолання вказаної суперечності використовується поняття умовного математичного сподівання відносно системи випадкових подій.

2.5.1. Умовне сподівання відносно розбиття

Означення. *Нехай ξ – інтегровна випадкова величина, а B – випадкова подія з додатною ймовірністю:* $\mathbf{P}(B)>0$. *Назвемо умовним математичним сподіванням ξ за умови B число*

$$\mathbf{E}(\xi\mid B)\equiv \mathbf{E}(\xi 1\!\!\mathrm{I}_B)/\mathbf{P}(B).$$

Нехай $\Delta=\{H_1,...,H_n\}$ скінченне *розбиття* простору Ω, тобто $\Omega = \cup H_i$, $H_i\cap H_j=\varnothing, i\ne j$, $\mathbf{P}(H_i)>0$, а ξ – інтегровна випадкова величина.

Тоді для кожного $i = \overline{1,n}$ визначено умовне сподівання $\mathbf{E}(\xi \mid H_i)$. Наступне означення дає можливість одночасної фіксації всіх цих сподівань.

Означення. *Умовним математичним сподіванням інтегровної випадкової величини ξ відносно скінченого розбиття $\Delta = \{H_1, ..., H_n\}$ називається випадкова величина*

$$\mathbf{E}(\xi \mid \Delta) \equiv \zeta = \sum\nolimits_{k=1}^{n} \mathbf{E}(\xi \mid H_k) \mathbb{1}_{\omega \in H_k}.$$

Теорема (про характеризацію умовного сподівання відносно розбиття). *Випадкова величина ζ збігається з умовним сподіванням величини ξ відносно розбиття $\Delta = \{H_1, ..., H_n\}$ тоді й тільки тоді, коли вона задовольняє такі умови.*

(а – вимірність) Величина ζ вимірна відносно сигма-алгебри $\sigma[\Delta]$, що породжена розбиттям Δ.

(б – баланс) Для всіх $A \in \sigma[\Delta]$ справедлива рівність $\mathbf{E}\xi \mathbb{1}_A = \mathbf{E}\zeta \mathbb{1}_A$.

Доведення. Зауважимо, що внаслідок скінченності Δ має місце зображення $\sigma[\Delta] = \{\cup_{i \in I} H_i,\ I \subset \overline{1,n}\}$. Позначимо $m_k = \mathbf{E}(\xi \mid H_k)$.

Необхідність. Умова (а) вимірності випливає з означення умовного сподівання: $\{\zeta < x\} = \cup_{k: m_k < x:} H_k \in \sigma[\Delta]$. Нехай $A \in \sigma[\Delta]$. Тоді $A = \cup_{i \in I} H_i$, для деякої множини $I \subset \overline{1,n}$ і за означенням

$$\mathbf{E}\zeta \mathbb{1}_A = \mathbf{E}\zeta \sum\nolimits_{i \in I} \mathbb{1}_{H_i} = \sum\nolimits_{i \in I} \mathbf{E}\zeta \mathbb{1}_{H_i} = \sum\nolimits_{i \in I} m_i \mathbf{P}(H_i) =$$
$$\sum\nolimits_{i \in I} \mathbf{E}(\xi \mid H_i)\mathbf{P}(H_i) = \sum\nolimits_{i \in I} \mathbf{E}(\xi \mathbb{1}_{H_i}) = \mathbf{E}(\xi \mathbb{1}_A),$$

що доводить умову балансу (б).

Достатність. З умови вимірності (а) та наслідку *про критерій вимірності відносно розбиття* випливає, що величина ζ має вигляд $\zeta = \sum_{k=1}^{n} c_k \mathbb{1}_{\omega \in H_k}$ для деяких дійсних c_k. Оберемо в умові балансу (б) $A = H_k$. Тоді $\mathbf{E}\xi \mathbb{1}_{H_k} = \mathbf{E}\zeta \mathbb{1}_{H_k} = \mathbf{E}(c_k \mathbb{1}_{\omega \in H_k}) = c_k \mathbf{P}(H_k)$, звідки $c_k = \mathbf{E}(\xi \mid H_k) = m_k$, отже за означенням $\zeta = \mathbf{E}(\xi \mid \Delta)$ $\boxdot$

Вправи

(1) Довести, що умовне математичне сподівання за умови події $\mathbf{E}(\xi \mid B)$ має всі властивості математичного сподівання.

(2) Нехай $|\Omega| = n$. Позначимо через $d(n)$ кількість різних розбиттів множини Ω. Довести, що $d(n) = \sum_{k=0}^{n-1} C_{n-1}^{k} d(k)$, де $d(0) = 1$.

(3) Довести що для кожної обмеженої $\sigma[\Delta]$-вимірної величини η справедлива тотожність $\mathbf{E}(\xi\eta \mid \Delta) = \eta \mathbf{E}(\xi \mid \Delta)$.

2.5.2. Умовне математичне сподівання відносно сигма-алгебри

Для загального означення використаємо умови *вимірності та балансу* з теореми *про характеризацію умовного сподівання відносно розбиття.*

Означення. Умовним математичним сподіванням *випадкової величини ξ відносно під-сигма-алгебри $\mathfrak{C} \subset \mathfrak{F}$ називається випадкова величина ζ, що позначається через $\mathbf{E}(\xi \mid \mathfrak{C})$, та задовольняє умови:*

(а – вимірність) ζ є вимірною відносно сигма-алгебри $\mathfrak{C}$,

(б – баланс) $\mathbf{E}\xi 1\!\!1_A = \mathbf{E}\zeta 1\!\!1_A$ для всіх $A \in \mathfrak{C}$.

Для визначеності умовного сподівання припускається, що математичні сподівання у (б) визначено коректно в узагальненому розумінні.

Означення. *Випадкова величина ζ* узагальнено інтегровна*, якщо:*

$$\min(E\zeta^+, E\zeta^-) < \infty.$$

Лема (про знаковизначеність випадкової величини). *Нехай величина ζ вимірна відносно $\mathfrak{C}$ і узагальнено інтегровна.*

(а) Якщо $\mathbf{E}\zeta 1\!\!1_A \ge 0$ для всіх $A \in \mathfrak{C}$, то $\zeta \ge 0$ майже напевне.

(б) У твердженні (а) всі нерівності можна замінити на рівності.

Доведення. За умови узагальненої інтегровності всі величини $\zeta 1\!\!1_A$ так само узагальнено інтегровні.

(а) Подія $A = \{\zeta < 0\} \in \mathfrak{C}$, причому величина $\zeta 1\!\!1_A$ недодатна, і має нульове математичне сподівання, оскільки останнє невід'ємне за умовою та недодатне за властивістю *невід'ємності* математичного сподівання. Отже, з властивості *додатності* математичного сподівання виводимо, що $\zeta 1\!\!1_A = 0$ майже напевне. Тому $\mathbf{P}(\zeta \ge 0) = 1$.

(б) Застосуємо твердження (а) одночасно для величин ζ та $-\zeta$ $\boxdot$

Теорема (про існування та єдиність умовного математичного сподівання). *Нехай випадкова величина ξ : (1) інтегровна, або (2) невід'ємна, а сигма-алгебра $\mathfrak{C} \subset \mathfrak{F}$.*

Тоді умовне математичне сподівання $\zeta \equiv \mathbf{E}(\xi \mid \mathfrak{C})$ існує як узагальнена випадкова величина зі значеннями у $\mathbb{R} \cup \{\infty\}$, визначено однозначно м.н. та у припущенні (1) – інтегровне, а за умови (2) – невід'ємне м.н.

Зауваження. Умовне сподівання невід'ємної величини ξ визначене завжди, однак може набувати значення ∞. Тому для знакозмінної величини ξ можна визначити поняття узагальненого умовного сподівання за лінійністю: $\mathbf{E}(\xi \mid \mathfrak{C}) \equiv \mathbf{E}(\xi^+ \mid \mathfrak{C}) - \mathbf{E}(\xi^- \mid \mathfrak{C})$, якщо тільки виконується умова $\min(\mathbf{E}(\xi^+ \mid \mathfrak{C}), \mathbf{E}(\xi^- \mid \mathfrak{C})) < \infty$ м.н.

Доведення

(1) Нехай величина ξ інтегровна. Тоді функція $Q(A) = \mathbf{E}\xi 1\!\!\mathrm{I}_A$ є скінченним зарядом на сигма-алгебрі $\mathfrak{C}$. Цей заряд *абсолютно неперервний відносно міри* $\mathbf{P}$ на $\mathfrak{C}$, оскільки з $\mathbf{P}(A) = 0$ випливає $\xi 1\!\!\mathrm{I}_A = 0$ м.н. та $\mathbf{E}\xi 1\!\!\mathrm{I}_A = 0$ за властивістю *центрованості* з теореми *про властивості математичного сподівання.*

Тому за *теоремою Радона-Нікодима* внаслідок абсолютної неперервності існує щільність f заряду Q відносно міри $\mathbf{P}$ на $\mathfrak{C}$, тобто така $\mathfrak{C}$-вимірна інтегровна за $\mathbf{P}$ функція, що $Q(A) = \int_A f(\omega)\mathbf{P}(d\omega),\ \forall A \in \mathfrak{C}$.

Функція f є шуканою, оскільки за означенням випадкова величина $\zeta(\omega) \equiv f(\omega)$ вимірна відносно сигма-алгебри $\mathfrak{C}$ та

$$\mathbf{E}\xi 1\!\!\mathrm{I}_A \equiv Q(A) = \int_A f(\omega)\mathbf{P}(d\omega) \equiv \mathbf{E}\zeta 1\!\!\mathrm{I}_A,\ \forall A \in \mathfrak{C}.$$

Одночасно ζ є інтегровною, як зазначено вище.

(2) Нехай $\xi \geq 0$. Розглянемо інтегровні величини $\xi_n = \min(\xi, n)$ та позначимо $\zeta_n = \mathbf{E}(\xi_n \mid \mathfrak{C})$. Оскільки послідовність (ξ_n) – неспадна, то за умовою балансу $\mathbf{E}((\zeta_{n+1} - \zeta_n)1\!\!\mathrm{I}_A) = \mathbf{E}((\xi_{n+1} - \xi_n)1\!\!\mathrm{I}_A) \geq 0$ для всіх $A \in \mathfrak{C}$. Тому $\zeta_{n+1} - \zeta_n \geq 0$ *м.н.* за лемою *про знаковизначеність випадкової величини*, (а). Отже, послідовність (ζ_n) майже напевне неспадна та невід'ємна.. Позначимо $\zeta = \overline{\lim}_{n\to\infty}\zeta_n$. Тоді випадкова величина ζ вимірна відносно сигма-алгебри $\mathfrak{C}$ і м.н. збігається з $\lim_{n\to\infty}\zeta_n$. З умови балансу для умовних сподівань ζ_n виводимо, що $\mathbf{E}(\zeta_n 1\!\!\mathrm{I}_A) = \mathbf{E}(\xi_n 1\!\!\mathrm{I}_A)$ для всіх $A \in \mathfrak{C}$. З *теореми Лебега про монотонну збіжність* граничним переходом $n \to \infty$ та з останньої рівності виводимо умову балансу для ζ. Невід'ємність ζ м.н. є наслідком невід'ємності ξ за лемою *про знаковизначеність випадкової величини*, (а), оскільки $\mathbf{E}\zeta 1\!\!\mathrm{I}_A = \mathbf{E}\xi 1\!\!\mathrm{I}_A \geq 0, \forall A \in \mathfrak{C}$..

Для доведення *єдиності* припустимо, що величини $\zeta_1, \zeta_2 \in \mathbb{R} \cup \{\infty\}$ одночасно задовольняють умови (а),(б).

Розглянемо випадкові події $C_{nk} = \{\zeta_k \leq n\} \in \mathfrak{C}$ згідно (а). З (б) виводимо, що $\mathbf{E}\zeta_2 1\!\!\mathrm{I}_{B\cap C_{n1}} = \mathbf{E}\zeta_1 1\!\!\mathrm{I}_{B\cap C_{n1}}, \forall B \in \mathfrak{C}$, де величина $\zeta_1 1\!\!\mathrm{I}_{B\cap C_{n1}} \leq n$. Тому $\mathbf{E}\left(\zeta_2 - \zeta_1\right) 1\!\!\mathrm{I}_{B\cap C_{n1}} = 0$ і $\left(\zeta_2 - \zeta_1\right) 1\!\!\mathrm{I}_{C_{n1}} = 0$ м.н. за лемою *про знаковизначеність випадкової величини.* Оскільки $C_{nk} \uparrow C_k \equiv \{\zeta_k < \infty\}, n \to \infty$, то з останньої рівності отримуємо $\left(\zeta_2 - \zeta_1\right) 1\!\!\mathrm{I}_{C_1} = 0$ м.н. Аналогічно виводимо, що $\left(\zeta_2 - \zeta_1\right) 1\!\!\mathrm{I}_{C_2} = 0$ м.н. Нарешті, $\zeta_2 = \zeta_1 = \infty$ при $\omega \in \overline{C_1} \cap \overline{C_2}$ за означенням $\boxdot$

Теорема (про еквівалентне означення умовного сподівання). *Нехай величина ξ інтегровна або невід'ємна. Випадкова величина ζ збігається м.н. з умовним математичним сподіванням $\mathbf{E}(\xi \mid \mathfrak{C})$ тоді й тільки* *тоді, коли виконуються умови*

(а) величина ζ вимірна відносно сигма-алгебри $\mathfrak{C}$,

(в) $\mathbf{E}(\xi\eta) = \mathbf{E}(\zeta\eta)$ для всіх обмежених $\mathfrak{C}$-вимірних величин η.

Доведення. *Достатність.* Оскільки величина $\eta = 1\!\!\mathrm{I}_A$ є обмеженою $\mathfrak{C}$-вимірною величиною при $A \in \mathfrak{C}$, то з умови (в) випливає умова (б) балансу. Отже, з (а),(в) та з теореми *про існування та єдиність умовного математичного сподівання* виводимо, що $\zeta = \mathbf{E}(\xi \mid \mathfrak{C})$ м.н.

Необхідність. Нехай $\zeta = \mathbf{E}(\xi \mid \mathfrak{C})$ м.н. Тоді з (б) робимо висновок, що (в) має місце для $\eta = 1\!\!\mathrm{I}_A, A \in \mathfrak{C}$. Враховуючи лінійність математичного сподівання, виводимо (в) для простих η.

(1) Припустимо, що ξ – інтегровна. Тоді ζ також є інтегровною за теоремою *про існування та єдиність умовного математичного сподівання*. Для довільної обмеженої $\mathfrak{C}$-вимірної випадкової величини η за теоремою *про апроксимацію випадкових величин простими* знайдемо прості величини $\eta_n \to \eta$ такі, що $|\eta_n| \le |\eta|$. Зокрема, величини η_n обмежені рівномірно за n. Тому в рівностях $\mathbf{E}\xi\eta_n = \mathbf{E}\zeta\eta_n$ можна перейти до границі під знаком математичних сподівань за *теоремою Лебега про мажоровану збіжність*, що і доводить (в).

(2) Нехай ξ – невід'ємна. Для довільної невід'ємної $\mathfrak{C}$-вимірної обмеженої величини η побудуємо прості величини такі, що $0 \le \eta_n \uparrow \eta$. Тоді до рівностей $\mathbf{E}\xi\eta_n = \mathbf{E}\zeta\eta_n$ можна застосувати *теорему Лебега про монотонну збіжність*, звідки виводимо (в) для невід'ємних η. На знакозмінні η рівність (в) поширюється за лінійністю $\boxdot$

2.5.3. Властивості умовного математичного сподівання

Зауважимо, що умовне математичне сподівання визначене з точністю до рівності *майже напевне*. Тому його властивості також виконуються *м.н.*

Теорема (про загальні властивості умовного сподівання). *Нехай сигма-алгебра $\mathfrak{C} \subset \mathfrak{F}$. Тоді за припущення існування відповідних умовних сподівань виконуються такі твердження.*

(1) $\mathbf{E}(c \mid \mathfrak{C}) = c$ м.н. $\forall c \in \mathbb{R}$.

(2) Якщо $\xi \ge 0$, то $\mathbf{E}(\xi \mid \mathfrak{C}) \ge 0$ м.н.

(3) $\mathbf{E}(c\xi \mid \mathfrak{C}) = c\mathbf{E}(\xi \mid \mathfrak{C})$ м.н.

(4) $\mathbf{E}(\xi_1 + \xi_2 \mid \mathfrak{C}) = \mathbf{E}(\xi_1 \mid \mathfrak{C}) + \mathbf{E}(\xi_2 \mid \mathfrak{C})$ м.н.

(5) Якщо $\xi_1 \le \xi_2$, то $\mathbf{E}(\xi_1 \mid \mathfrak{C}) \le \mathbf{E}(\xi_2 \mid \mathfrak{C})$ м.н.

(6) $|\mathbf{E}(\xi \mid \mathfrak{C})| \le \mathbf{E}(|\xi| \mid \mathfrak{C})$ м.н.

Доведення. Всі наведені рівності ґрунтуються на теоремі *про існування та єдиність умовного математичного сподівання*, де єдиність гарантується з точністю до рівності майже напевне. Отже, наведені тотожності є прямими наслідками умов (а),(б) в означенні відповідних умовних сподівань, а також теореми *про властивості математичного сподівання.*

(1) Виконання умов вимірності та балансу очевидне.

(2) Невід'ємність вже доведена в теоремі *про існування та єдиність умовного математичного сподівання.*

(3) Позначимо $\zeta = \mathbf{E}(\xi \mid \mathfrak{C})$. Тоді (а) $c\zeta$ є $\mathfrak{C}$-вимірною величиною, і (б) за *однорідністю* математичного сподівання

$$\mathbf{E}(c\xi 1\!\!1_A) = c\mathbf{E}(\xi 1\!\!1_A) = c\mathbf{E}(\zeta 1\!\!1_A) = \mathbf{E}(c\zeta 1\!\!1_A), \forall A \in \mathfrak{C},$$

що доводить умову балансу для $c\zeta$ в означенні $\mathbf{E}(c\xi \mid \mathfrak{C})$.

(4) Нехай $\zeta_i = \mathbf{E}(\xi_i \mid \mathfrak{C}), i = 1, 2$. Тоді (а) сума $\zeta_1 + \zeta_2$ є $\mathfrak{C}$-вимірною величиною, і (б) за *адитивністю* математичного сподівання

$$\mathbf{E}((\xi_1 + \xi_2)1\!\!1_A) = \mathbf{E}\xi_1 1\!\!1_A + \mathbf{E}\xi_2 1\!\!1_A = \mathbf{E}\zeta_1 1\!\!1_A + \mathbf{E}\zeta_2 1\!\!1_A =$$
$$\mathbf{E}((\zeta_1 + \zeta_2)1\!\!1_A), \forall A \in \mathfrak{C},$$

що доводить умову балансу для $\zeta_1 + \zeta_2$ в означенні $\mathbf{E}(\xi_1 + \xi_2 \mid \mathfrak{C})$.

(5) У випадку інтегровних ξ_k виводиться з (2),(3),(4), якщо (2) застосувати до різниці $\xi_2 - \xi_1 \geq 0$. У припущенні $\xi_k \geq 0$ маємо $\mathbf{E}\zeta_1 1\!\!1_A = \mathbf{E}\xi_1 1\!\!1_A \leq \mathbf{E}\xi_2 1\!\!1_A = \mathbf{E}\zeta_2 1\!\!1_A$ для всіх $A \in \mathfrak{C}$. Вибором $A = \{\zeta_2 < \infty\}$ звідси виводимо включення $\{\zeta_1 < \infty\} \subset \{\zeta_2 < \infty\}$, внаслідок підстановки $A = A \cap \{\zeta_1 < \infty\}$ та леми *про знаковизначеність випадкової величини* остаточно отримуємо нерівність $\zeta_1 1\!\!1_{\{\zeta_1 < \infty\}} \leq \zeta_2 1\!\!1_{\{\zeta_1 < \infty\}}$.

(6) Є наслідком (3) та (5) для пари нерівностей $-|\xi| \leq \xi \leq |\xi|$ ⊡

Теорема (про властивості умовного сподівання). *Нехай випадкова величина ξ інтегровна або невід'ємна, а σ-алгебри $\mathfrak{C}, \mathfrak{C}_k \subset \mathfrak{F}$. Тоді*

(1) $\mathbf{E}(\xi \mid \mathfrak{F}_*) = \mathbf{E}\xi$ *м.н., де* $\mathfrak{F}_* = \{\varnothing, \Omega\}$.

(2) Якщо величина ξ вимірна відносно $\mathfrak{C}$, то $\mathbf{E}(\xi \mid \mathfrak{C}) = \xi$ *м.н.*

(3) $\mathbf{E}(\mathbf{E}(\xi \mid \mathfrak{C})) = \mathbf{E}\xi$.

(4) Якщо $\mathfrak{C}_2 \subset \mathfrak{C}_1$, то $\mathbf{E}(\mathbf{E}(\xi \mid \mathfrak{C}_2) \mid \mathfrak{C}_1) = \mathbf{E}(\xi \mid \mathfrak{C}_2)$ м.н.

(5) Якщо $\mathfrak{C}_2 \subset \mathfrak{C}_1$, то $\mathbf{E}(\mathbf{E}(\xi \mid \mathfrak{C}_1) \mid \mathfrak{C}_2) = \mathbf{E}(\xi \mid \mathfrak{C}_2)$ м.н.

(6) Якщо $\sigma[\xi]$ та $\mathfrak{C}$ незалежні, то $\mathbf{E}(\xi \mid \mathfrak{C}) = \mathbf{E}\xi$ м.н.

(7) Якщо випадкові величини ξ та $\xi\eta$ невід'ємні або інтегровні, а величина η є $\mathfrak{C}$-вимірною, то $\mathbf{E}(\xi\eta \mid \mathfrak{C}) = \eta\mathbf{E}(\xi \mid \mathfrak{C})$ м.н.

(8) Нехай функція g – опукла донизу, величини ξ і $g(\xi)$ інтегровні. Тоді має місце нерівність Йенсена: $g(\mathbf{E}(\xi \mid \mathfrak{C})) \leq \mathbf{E}(g(\xi) \mid \mathfrak{C})$ м.н.

(9) Якщо подія $A \in \mathfrak{C}$ *і* $\xi_1 1\!\!\mathrm{I}_A = \xi_2 1\!\!\mathrm{I}_A$ *м.н., то* $\mathbf{E}(\xi_1 \mid \mathfrak{C}) = \mathbf{E}(\xi_2 \mid \mathfrak{C})$ *м.н. при* $\omega \in A$.

Доведення. Як і в попередній теоремі, доведення зводиться до перевірки умов (а),(б) в означенні відповідних умовних сподівань та застосуванні теореми *про існування та єдиність умовного математичного сподівання*.

(1) Вимірними відносно сигма-алгебри $\{\varnothing, \Omega\}$ є сталі (невипадкові) величини, зокрема, $\mathbf{E}\xi$. Для останньої виконується і умова балансу.

(2) Умова (а) $\mathfrak{C}$-вимірності ξ виконана, а умова балансу (б) очевидна.

(3) Досить обрати в умові балансу (б) визначення умовного сподівання $\zeta = \mathbf{E}(\xi \mid \mathfrak{C})$ випадкову подію $A = \Omega$.

(4) Позначимо $\zeta_2 = \mathbf{E}(\xi \mid \mathfrak{C}_2)$. Оскільки $\mathfrak{C}_2 \subset \mathfrak{C}_1$, то величина ζ_2 є $\mathfrak{C}_1$-вимірною. Отже, за властивістю (2) $\mathbf{E}(\zeta_2 \mid \mathfrak{C}_1) = \zeta_2$ м.н.

(5) Позначимо $\zeta_i = \mathbf{E}(\xi \mid \mathfrak{C}_i)$. Шукана рівність має вигляд $\mathbf{E}(\zeta_1 \mid \mathfrak{C}_2) = \zeta_2$ м.н. Для доведення цієї рівності зауважимо, що (а) умова вимірності ζ_2 відносно σ-алгебри $\mathfrak{C}_2$ виконується за означенням ζ_2, (б) за умовою балансу $\mathbf{E}(\zeta_1 1\!\!\mathrm{I}_A) = \mathbf{E}(\xi 1\!\!\mathrm{I}_A) = \mathbf{E}(\zeta_2 1\!\!\mathrm{I}_A)$ для всіх $A \in \mathfrak{C}_2 \subset \mathfrak{C}_1$, звідки випливає умова балансу для ζ_2 у визначенні $\mathbf{E}(\zeta_1 \mid \mathfrak{C}_2)$.

(6) Стала $\mathbf{E}\xi$ є: (а) вимірною відносно сигма-алгебри $\mathfrak{C}$, та (б) задовольняє умову балансу, оскільки $\mathbf{E}\xi 1\!\!\mathrm{I}_A = \mathbf{E}\xi \mathbf{E} 1\!\!\mathrm{I}_A = \mathbf{E}((\mathbf{E}\xi) 1\!\!\mathrm{I}_A)$ внаслідок незалежності ξ та $1\!\!\mathrm{I}_A$ при $A \in \mathfrak{C}$.

(7) Позначимо $\zeta = \mathbf{E}(\xi \mid \mathfrak{C})$. Припустимо, що $\xi \geq 0$. Тоді і $\zeta \geq 0$ м.н. Для доведення рівності досить перевірити умови вимірності та балансу для величини $\zeta\eta$ в означенні $\mathbf{E}(\xi\eta \mid \mathfrak{C})$.

(7а) Оскільки ζ є $\mathfrak{C}$-вимірною, то добуток $\zeta\eta$ також $\mathfrak{C}$-вимірний.

(7б) Умова балансу має вигляд: $\mathbf{E}(\xi\eta 1\!\!\mathrm{I}_A) = \mathbf{E}(\zeta\eta 1\!\!\mathrm{I}_A), A \in \mathfrak{C}$. Для її перевірки розглянемо подію $C \in \mathfrak{C}$. Оскільки $C \cap A \in \mathfrak{C}$ при $A \in \mathfrak{C}$, то

$$\mathbf{E}(\xi 1\!\!\mathrm{I}_C 1\!\!\mathrm{I}_A) = \mathbf{E}(\xi 1\!\!\mathrm{I}_{C\cap A}) = \mathbf{E}(\zeta 1\!\!\mathrm{I}_{C\cap A}) = \mathbf{E}(\zeta 1\!\!\mathrm{I}_C 1\!\!\mathrm{I}_A)$$

за умовою балансу для $\mathbf{E}(\xi \mid \mathfrak{C})$. Остання рівність є потрібною умовою балансу для $\eta = 1\!\!\mathrm{I}_C, C \in \mathfrak{C}$. Унаслідок лінійності ця умова виконується для всіх простих $\mathfrak{C}$-вимірних величин η.

Нехай η – довільна невід'ємна $\mathfrak{C}$-вимірна величина. Оберемо за теоремою *про апроксимацію випадкових величин простими* $\mathfrak{C}$-вимірні прості η_n так, щоб $0 \leq \eta_n \uparrow \eta$. Застосуємо до доведеної рівності $\mathbf{E}(\xi\eta_n 1\!\!\mathrm{I}_A) = \mathbf{E}(\zeta\eta_n 1\!\!\mathrm{I}_A)$ *теорему Лебега про монотонну збіжність* при $n \to \infty$. Отримаємо потрібну умову балансу, звідки за теоремою *про існування та єдиність умовного математичного сподівання* випливає перше твердження в (7). Зокрема, з останньої рівності при $A = \Omega$ виводимо інтегровність $\zeta\eta$ при інтегрованій $\xi\eta$.

Для знакозмінних величин з інтегровними $\xi\eta$ та ξ твердження доводиться за лінійністю:

$$\mathbf{E}(\xi\eta \mid \mathfrak{C}) = \mathbf{E}(\xi^+\eta^+ - \xi^-\eta^+ - \xi^+\eta^- + \xi^-\eta^- \mid \mathfrak{C}) =$$

$$\eta^+\mathbf{E}(\xi^+ \mid \mathfrak{C}) - \eta^+\mathbf{E}(\xi^- \mid \mathfrak{C}) - \eta^-\mathbf{E}(\xi^+ \mid \mathfrak{C}) + \eta^-\mathbf{E}(\xi^- \mid \mathfrak{C}) =$$
$$(\eta^+ - \eta^-)\mathbf{E}((\xi^+ - \xi^-) \mid \mathfrak{C}) = \eta\mathbf{E}(\xi \mid \mathfrak{C}),$$

де враховані м.н. скінченність усіх доданків та $\mathfrak{C}$-вимірність величин $\eta^{\pm}$.

(8) Позначимо $\zeta = \mathbf{E}(\xi \mid \mathfrak{C})$. Як і в доведенні *нерівності Йенсена* для математичного сподівання, з опуклості g виводимо (**вправа**) зображення

$$g(x) = \sup(a + bx, (a,b) \in K_g), \quad \forall x \in \mathbb{R},$$

де K_g - множина опорних прямих у щільній множині точок (x_n) :

$$K_g = \{(a_n, b_n) \in \mathbb{R}^2 : a_n + b_n x \le g(x), \forall x, \ a_n + b_n x_n = g(x_n), \ \forall n\}.$$

Для фіксованих $(a,b) \in K_g$ за монотонністю та лінійністю умовного сподівання $\mathbf{E}(g(\xi) \mid \mathfrak{C}) \ge a + b\mathbf{E}(\xi \mid \mathfrak{C})$ м.н., звідки переходом до верхньої межі за (a,b) та з наведеного зображення отримуємо нерівність (8).

(9) Позначимо $\zeta_i = \mathbf{E}(\xi_i \mid \mathfrak{C})1\!\!\mathrm{I}_A$. Ці величини є $\mathfrak{C}$-вимірними. За властивостями (7) та (3) для всіх $C \in \mathfrak{C}$ виконуються рівності

$$\mathbf{E}(\zeta_1 1\!\!\mathrm{I}_C) = \mathbf{E}(\mathbf{E}(\xi_1 \mid \mathfrak{C})1\!\!\mathrm{I}_{A\cap C}) = \mathbf{E}(\mathbf{E}(\xi_1 1\!\!\mathrm{I}_{A\cap C} \mid \mathfrak{C})) = \mathbf{E}(\xi_1 1\!\!\mathrm{I}_{A\cap C}) =$$

$$\mathbf{E}(\xi_2 1\!\!\mathrm{I}_{A\cap C}) = \mathbf{E}(\mathbf{E}(\xi_2 1\!\!\mathrm{I}_{A\cap C} \mid \mathfrak{C})) = \mathbf{E}(\mathbf{E}(\xi_2 \mid \mathfrak{C})1\!\!\mathrm{I}_{A\cap C}) = \mathbf{E}(\zeta_2 1\!\!\mathrm{I}_C).$$

Отже, дане твердження випливає з леми *про знаковизначеність випадкової величини,* з якої $\zeta \equiv \zeta_1 - \zeta_2 = 0$ м.н. $\boxdot$

Наступна теорема дає підстави інтерпретувати умовне математичне сподівання квадратично інтегровної величини як її проекцію на лінійний підпростір $\mathfrak{C}$-вимірних величин. У зв'язку з цим пункт (5) теореми *про властивості умовного сподівання* можна розглядати як відому з геометрії теорему про три перпендикуляри.

Теорема (про екстремальну властивість умовного сподівання). *Якщо* $\mathbf{E}\xi^2 < \infty$ *та сигма-алгебра* $\mathfrak{C} \subset \mathfrak{F}$, *то умовне сподівання* $\zeta \equiv \mathbf{E}(\xi \mid \mathfrak{C})$ *інтегроване у квадраті і*

$$\zeta = \arg\min{}_{\eta:\ \mathbf{E}\eta^2<\infty,\ \eta\ \mathfrak{C}\text{-вимірна}}\, \mathbf{E}(\xi - \eta)^2 \text{ м.н.},$$

тобто $\mathbf{E}(\xi-\zeta)^2 \le \mathbf{E}(\xi-\eta)^2$ *для всіх квадратично інтегровних вимірних відносно* $\mathfrak{C}$ *величин* η.

Доведення. За теоремою *про властивості умовного сподівання*, (8), з опуклості функції x^2 випливає нерівність $\zeta^2 \le \mathbf{E}(\xi^2 \mid \mathfrak{C})$ м.н., звідки $\mathbf{E}\zeta^2 \le \mathbf{E}(\mathbf{E}(\xi^2 \mid \mathfrak{C})) = \mathbf{E}\xi^2 < \infty$.

Нехай випадкова величина η інтегровна в квадраті: $\mathbf{E}\eta^2 < \infty$ та $\mathfrak{C}$-вимірна. Тоді за теоремою *про властивості умовного сподівання*, (3)

$$\mathbf{E}(\xi-\eta)^2 = \mathbf{E}(\xi-\zeta+\zeta-\eta)^2 =$$

$$\mathbf{E}(\xi-\zeta)^2 + 2\mathbf{E}(\zeta-\eta)(\xi-\zeta) + \mathbf{E}(\zeta-\eta)^2 =$$
$$\mathbf{E}(\xi-\zeta)^2 + 2\mathbf{E}(\mathbf{E}((\zeta-\eta)(\xi-\zeta) \mid \mathfrak{C})) + \mathbf{E}(\zeta-\eta)^2 =$$
$$\mathbf{E}(\xi-\zeta)^2 + 2\mathbf{E}((\zeta-\eta)\mathbf{E}((\xi-\zeta) \mid \mathfrak{C})) + \mathbf{E}(\zeta-\eta)^2 =$$
$$\mathbf{E}(\xi-\zeta)^2 + \mathbf{E}(\zeta-\eta)^2 \ge \mathbf{E}(\xi-\zeta)^2,$$

де використане твердження (7) цієї теореми. Рівність $\mathbf{E}((\xi-\zeta) \mid \mathfrak{C}) = 0$ м.н. є наслідком означення ζ та теореми *про властивості умовного сподівання*. Інтегровність добутку $(\zeta-\eta)(\xi-\zeta)$ та $\xi-\zeta$ виводиться з квадратичної інтегровності величин $\zeta-\eta$, $\xi-\zeta$ застосуванням *нерівності Коші*.

З доведеної нерівності виводимо, що найменше значення $\mathbf{E}(\xi-\eta)^2$ у вказаному класі η набувається саме при $\eta = \zeta$ $\boxdot$

Теорема (про граничний перехід під знаком умовного сподівання). *Нехай сигма-алгебра* $\mathfrak{C} \subset \mathfrak{F}$.

(а) Якщо $0 \le \xi_n \uparrow \xi, n \to \infty$, *то* $\mathbf{E}(\xi_n \mid \mathfrak{C}) \uparrow \mathbf{E}(\xi \mid \mathfrak{C})$ *м.н.*

(б) Якщо $\xi_n \to \xi$ *поточково при* $n \to \infty$ *та* $|\xi_n| \le \eta$, *де* $\mathbf{E}\eta < \infty$, *то* $\mathbf{E}(\xi_n \mid \mathfrak{C}) \to \mathbf{E}(\xi \mid \mathfrak{C})$ *м.н. при* $n \to \infty$.

Доведення. Позначимо $\zeta_n = \mathbf{E}(\xi_n \mid \mathfrak{C})$.

(а) За теоремою *про загальні властивості умовного сподівання*, (5), послідовність ζ_n м.н. неспадна. Тому величина $\zeta = \overline{\lim}\zeta_n$ м.н. збігається з монотонною границею $\lim \zeta_n$. Вимірність ζ відносно $\mathfrak{C}$ є наслідком такої вимірності величин ζ_n. Крім того, ζ задовольняє умову балансу в означенні $\mathbf{E}(\xi \mid \mathfrak{C})$, що випливає за *теоремою Лебега про монотонну збіжність* з рівнянь балансу $\mathbf{E}(\xi_n \mathbb{1}_A) = \mathbf{E}(\zeta_n \mathbb{1}_A)$ для $A \in \mathfrak{C}$ при $n \to \infty$. Отже, $\zeta = \mathbf{E}(\xi \mid \mathfrak{C})$ м.н. за теоремою *про існування та єдиність умовного математичного сподівання.*

(б) Позначимо $\xi_n' = \sup_{k \ge n} |\xi_k - \xi|$. Тоді $\xi_n' \downarrow 0$ при $n \to 0$ та $\xi_n' \le 2\eta$. Звідси $0 \le 2\eta - \xi_n' \uparrow 2\eta$ і за твердженням (а) $\mathbf{E}(2\eta - \xi_n' \mid \mathfrak{C}) \uparrow \mathbf{E}(2\eta \mid \mathfrak{C})$ м.н. Отже, з лінійності умовного сподівання та скінченності $\mathbf{E}(2\eta \mid \mathfrak{C})$ отримуємо збіжність $\mathbf{E}(\xi_n' \mid \mathfrak{C}) \downarrow 0$ м.н. Звідси з нерівності (6) теореми *про загальні властивості умовного сподівання* отримуємо твердження (б):

$$|\zeta_n - \mathbf{E}(\xi \mid \mathfrak{C})| = |\mathbf{E}((\xi_n - \xi) \mid \mathfrak{C})| \le \mathbf{E}(|\xi_n - \xi| \mid \mathfrak{C}) \le \mathbf{E}(\xi_n' \mid \mathfrak{C}) \ \boxdot$$

Вправи

(1) У припущенні квадратичної інтегровності ξ вивести теорему про існування та єдиність умовного математичного сподівання з теореми про існування проекції на замкнений підпростір у гільбертовому просторі $L_2(\mathbf{P})$.

(2) У припущенні квадратичної інтегровності ξ вивести умову балансу у означенні умовного сподівання з його екстремальної властивості $\mathbf{E}(\xi-\zeta)^2 \leq \mathbf{E}(\xi-\zeta-\varepsilon\eta)^2$ для всіх $\varepsilon \in \mathbb{R}$ та всіх $\mathfrak{C}$-вимірних квадратично інтегровних η.

(3) Довести, що для $\mathfrak{C}$-вимірної величини ζ та обмеженої борелівської g має місце рівність $\mathbf{E}(g(\xi,\zeta) \mid \mathfrak{C}) = \mathbf{E}(g(\xi,y) \mid \mathfrak{C}) \mid_{y=\zeta}$.

(4) Довести нерівність Йенсена для опуклої функції g лише на підставі нерівності $g(x) \geq g(x_0) + k(x_0)(x-x_0)$.

(5) Розглянемо ймовірнісний простір $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \mathbf{P})$ з імовірністю $\mathbf{P}$, що задається щільністю $f : \mathbf{P}(B) = \int_B f(x)dx$. Величина ξ дорівнює $\xi(\omega) = g(\omega)$. Обчислити умовне сподівання $\mathbf{E}(\xi \mid \mathfrak{C})$, якщо $\mathfrak{C}$ дорівнює:

(а) $\sigma[\{B \in \mathfrak{B}(\mathbb{R}) : B = B+1\}]$, (б) $\sigma[\{B \in \mathfrak{B}(\mathbb{R}) : B = -B\}]$,

(в) $\sigma[\{x : \sin(x) < a\}, a \in \mathbb{R}]$, (г) $\sigma[\{[k2^{-n}, (k+1)2^{-n})\}, k \in \mathbb{Z}, n \geq 1]$.

Тут за означенням $h(B) = \{h(x), x \in B\}$.

(6) Розглянемо ймовірнісний простір $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n), \mathbf{P})$ з імовірністю $\mathbf{P}$, що задається сумісною щільністю $f : \mathbf{P}(B) = \int_B f(x)dx$. Сигма-алгебра переставних множин $\mathfrak{P}(\mathbb{R}^n)$ містить всі борелеві множини B такі, що $\pi(x) \in B$ для всіх $x \in B$ та перестановок π координат векторів $x \in \mathbb{R}^n$. Обчислити $\mathbf{E}(\xi \mid \mathfrak{C})$ для величин вигляду: (а) $\xi(\omega) = \omega_1$, (б) $\xi(\omega) = \omega_1\omega_2$, (в) $\xi(\omega) = \omega_1^2$.

(7) Випадкова величина ξ інтегровна. Довести, що сім'я випадкових величин $(\mathbf{E}(\xi \mid \mathfrak{C}), \mathfrak{C} \subset \mathfrak{F})$ рівномірно інтегровна.

(8) Випадкова величина ξ невід'ємна. Довести, що $\mathbf{E}(\xi \mid \mathfrak{C}) < \infty$ м.н. тоді й тільки тоді, коли міра $Q(A) \equiv \mathbf{E}\xi 1\!\!\mathrm{I}_A$ при $A \in \mathfrak{C}$ є сигма-скінченною.

(9) Нехай випадкові величини (ξ_n) інтегровні мажоровано зверху: $\xi_n \leq \eta$, $\mathbf{E}\,|\eta| < \infty$. Довести, що $\overline{\lim}_{n\to\infty} \mathbf{E}(\xi_n \mid \mathfrak{C}) \leq \mathbf{E}(\overline{\lim}_{n\to\infty} \xi_n \mid \mathfrak{C})$ м.н.

(10) Випадкові величини $\xi, (\xi_n, n \geq 1)$ такі, що $\xi_n \to \xi, n \to \infty$, $|\xi_n| \leq \eta$, $\mathbf{E}\eta < \infty$, а сигма-алгебри $\mathfrak{C}_k$ не зростають за k та $\cap\mathfrak{C}_k = \mathfrak{C}_\infty$ Довести, що $\mathbf{E}(\xi_n \mid \mathfrak{C}_k) \to \mathbf{E}(\xi \mid \mathfrak{C}_\infty), n, k \to \infty$, м.н. та у середньому.

(11) Якщо послідовність$(\xi_n, n \geq 1)$ рівномірно інтегровна та $\xi_n \to \xi, n \to \infty$, то $\mathbf{E}(\xi_n \mid \mathfrak{C}) \to^P \mathbf{E}(\xi \mid \mathfrak{C})$.

(12) Довести нерівність Коші: $(\mathbf{E}(\xi\eta \mid \mathfrak{C}))^2 \leq \mathbf{E}(\xi^2 \mid \mathfrak{C})\mathbf{E}(\eta^2 \mid \mathfrak{C})$ м.н.

(13) Довести, що оператор умовного сподівання є стискаючим у просторі L_p: $\mathbf{E}(|\mathbf{E}(\xi \mid \mathfrak{C})|)^p \leq \mathbf{E}(|\xi|^p)$.

(14) Визначимо **умовну дисперсію**: $\mathbf{D}(\xi \mid \mathfrak{C}) \equiv \mathbf{E}((\xi - \mathbf{E}(\xi \mid \mathfrak{C}))^2 \mid \mathfrak{C})$. Довести тотожність $\mathbf{D}\xi = \mathbf{E}\mathrm{D}(\xi \mid \mathfrak{C}) + \mathbf{D}\mathrm{E}(\xi \mid \mathfrak{C})$.

(15) Випадкова величина ξ та σ-алгебра $\mathfrak{C} \subset \mathfrak{F}$ такі, що $\mathbf{E}(\exp(it\xi) \mid \mathfrak{C}) = \exp(-t^2/2), \forall t \in \mathbb{R}$. Довести, що ξ – стандартна нормальна і не залежить від $\mathfrak{C}$.

2.5.4. Умовне математичне сподівання відносно випадкових величин

Означення. *Нехай* $(\eta_\gamma, \gamma \in \Gamma)$ *довільна система випадкових величин. Умовним математичним сподіванням величини* ξ *відносно цієї системи називається випадкова величина, що дорівнює умовному сподіванню відносно породженої сигма-алгебри:*

$$\mathbf{E}(\xi \mid \eta_\gamma, \gamma \in \Gamma) \equiv \mathbf{E}(\xi \mid \sigma[\eta_\gamma, \gamma \in \Gamma]).$$

Зокрема, $\mathbf{E}(\xi \mid \eta) \equiv \mathbf{E}(\xi \mid \sigma[\eta])$.

Теорема (про критерій для умовного сподівання відносно величин). *Нехай* ξ *інтегровна або невід'ємна. Випадкова величина* ζ *збігається м.н. з* $\mathbf{E}(\xi \mid \eta)$ *тоді й тільки тоді, коли*

(а) знайдеться борелєва функція m *така, що* $\zeta = m(\eta)$ *м.н. та*

(б) для довільної обмеженої борелєвої функції g *має місце рівність балансу* $\mathbf{E}\xi g(\eta) = \mathbf{E}m(\eta)g(\eta)$.

Зауваження. Аналогічне твердження виконується також для умовного сподівання відносно довільної системи величин $(\eta_\gamma, \gamma \in \Gamma)$.

Доведення. За теоремою *про вимірність відносно сигма-алгебри, що породжена величиною,* випадкова величина ζ задовольняє умову вимірності у визначенні $\mathbf{E}(\xi \mid \sigma[\eta])$ – є вимірною відносно $\sigma[\eta]$, тоді й тільки тоді, коли виконується умова (а) теореми. Так само, умова балансу у тому ж означенні еквівалентна умові (б), оскільки за згаданою теоремою випадкова величина є обмеженою та $\sigma[\eta]$-вимірною тоді й тільки тоді, коли вона має вигляд $g(\eta)$, де функція g – обмежена борелєва $\boxdot$

Означення. *Борелєва функція* $m : \mathbb{R} \to \mathbb{R}$, *для якої згідно з теоремою про критерій для умовного сподівання відносно величин має місце зображення* $\mathbf{E}(\xi \mid \eta) = m(\eta)$ *м.н., називається* функцією регресії *величини* ξ *на* η. *Для цієї функції використовується також позначення* $\mathbf{E}(\xi \mid \eta = y) \equiv m(y)$.

Відмітимо, що формально останнє позначення визначене коректно для майже всіх y навіть у випадку, коли $\mathbf{P}(\eta = y) = 0$ при всіх $y \in \mathbb{R}$.

Всі властивості умовного математичного сподівання відносно σ-алгебри виконуються і для сподівання відносно величин. Наведемо деякі з них.

Теорема (про властивості умовного сподівання відносно величин).

(а) $\mathbf{E}(g(\xi) \mid \xi) = g(\xi)$ *м.н. для борелєвої функції* g.

(б) Для борелєвої функції g *за умови інтегровності величин* ξ *та* $\xi g(\eta)$ *справедлива рівність* $\mathbf{E}(\xi g(\eta) \mid \eta) = g(\eta)\mathbf{E}(\xi \mid \eta)$ *м.н.*

(в) Якщо величини ξ, η *незалежні, а* $f : \mathbb{R}^2 \to \mathbb{R}$ *борелєва функція, для якої величина* $f(\xi, \eta)$ *інтегровна, то*

$$\mathbf{E}(f(\xi, \eta) \mid \eta = y) = \mathbf{E}f(\xi, y)$$

для майже всіх y *за мірою* $\mathbf{P}(\eta \in \cdot)$.

(г) Функція регресії квадратично інтегровної величини ξ *на* η *мінімізує квадратичне відхилення від* ξ *у класі функцій від* η :

$$m = \arg\min_{g:\ \text{борелєва}} \mathbf{E}(\xi - g(\eta))^2.$$

Зауваження. Інші властивості умовного математичного сподівання відносно випадкових величин є наслідками властивостей *умовного математичного сподівання* відносно сигма-алгебри.

Доведення

(а) Випадкова величина $g(\xi)$ вимірна відносно сигма-алгебри $\sigma[\xi]$, тому (а) випливає з теореми *про властивості умовного сподівання*, (2).

(б) Оскільки величина $g(\eta)$ є $\sigma[\eta]$-вимірною, то твердження даного пункту випливає з тієї ж теореми, (7).

(в) Позначимо $m(y) = \mathbf{E}f(\xi, y)$. За теоремою *про критерій для умовного сподівання відносно величин* для доведення рівності (в) досить перевірити умову балансу $\mathbf{E}(f(\xi, \eta)g(\eta)) = \mathbf{E}(m(\eta)g(\eta))$ для всіх обмежених борелєвих g. Остання рівність випливає з теореми *про математичне сподівання функції від незалежних величин*:

$$\mathbf{E}(f(\xi, \eta)g(\eta)) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x, y)g(y)dF_\xi(x) \right) dF_\eta(y) =$$

$$\int_{-\infty}^{\infty} m(y)g(y)dF_\eta(y) = \mathbf{E}m(\eta)g(\eta).$$

(г) Дане твердження є очевидним наслідком теореми *про екстремальні властивості умовного сподівання* з урахуванням теореми *про вимірність відносно сигма-алгебри, що породжена величиною* $\boxdot$

2.5.5. Обчислення умовного сподівання через сумісну щільність

З теореми *про критерій для умовного сподівання відносно величин* випливає, що умовне математичне сподівання однієї випадкової величини відносно іншої визначається їх *сумісною функцією розподілу*.

Теорема (про обчислення умовного сподівання через сумісну щільність). *Нехай випадкові величини* ξ, η *мають сумісну щільність* $f_{\xi,\eta}(x,y)$, *а величина* ξ *інтегровна або невід'ємна. Тоді*

$$\mathbf{E}(\xi \mid \eta = y) = \int_{-\infty}^{\infty} x \ f_{\xi|\eta}(x \mid y)dx,$$

де щільність розподілу $f_{\xi|\eta}$ *дорівнює*

$$f_{\xi|\eta}(x \mid y) \equiv f_{\xi,\eta}(x,y)/f_\eta(y),$$

а функція $f_\eta(y) = \int_{-\infty}^{\infty} f_{\xi,\eta}(x,y)dx$ *є щільністю величини* η.

Зауваження. З останньої рівності виводимо, що $f_{\xi,\eta}(x,y) = 0$ майже всюди для значень y таких, що $f_\eta(y) = 0$. У цьому випадку довизначимо $f_{\xi|\eta}(x \mid y)$ як стандартну нормальну щільність. Тоді $f_{\xi|\eta}(x \mid y)$ є *щільністю* розподілу для всіх $y \in \mathbb{R}$. Ця щільність називається **умовною щільністю** величини ξ за умови $\eta = y$.

Доведення. З теореми *про обчислення ймовірностей, пов'язаних із неперервним вектором,* отримуємо тотожність

$$\mathbf{P}(\eta < y) = \mathbf{P}((\xi,\eta) \in \mathbb{R} \times (-\infty, y)) = \int_{-\infty}^{y} \left(\int_{-\infty}^{\infty} f_{\xi,\eta}(x,u)dx \right) du,$$

звідки виводимо останню рівність теореми.

Позначимо $m(y) = \int_{-\infty}^{\infty} x f_{\xi|\eta}(x \mid y)dx$. Інтеграл тут коректно визначений для майже всіх y внаслідок інтегровності або невід'ємності ξ, оскільки у першому випадку $\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |x| \, f_{\xi,\eta}(x,y)dx \right) dy = \mathbf{E}\,|\xi| < \infty$.

Для перевірки першої рівності теореми використаємо теорему *про критерій для умовного сподівання відносно величин,* за якою досить довести для довільних обмежених борелєвих g тотожність

$$\mathbf{E}(\xi g(\eta)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x g(y) f_{\xi,\eta}(x,y)dxdy =$$

$$\int_{-\infty}^{\infty} g(y) \left(\int_{-\infty}^{\infty} x f_{\xi,\eta}(x,y)dx \right) dy = \int_{-\infty}^{\infty} g(y) \left(\int_{-\infty}^{\infty} x f_{\xi|\eta}(x \mid y)dx \right) f_\eta(y)dy =$$

$$\int_{-\infty}^{\infty} g(y)m(y)f_\eta(y)dy = \mathbf{E}(m(\eta)g(\eta)) \ \boxdot$$

Зауваження. За умови неперервності $f_{\xi,\eta}(x,y)$ має місце рівність

$$\mathbf{E}(\xi \mid \eta = y) = \lim\nolimits_{\varepsilon \downarrow 0} \mathbf{E}(\xi \mid |\eta - y| \le \varepsilon).$$

Дійсно, за означенням умовного математичного сподівання відносно події вираз під знаком границі у правій частині дорівнює

$$\mathbf{E}(\xi \mathbb{1}_{\{|\eta-y| \le \varepsilon\}})/\mathbf{P}(|\eta - y| \le \varepsilon) =$$

$$\textstyle\int_{y-\varepsilon}^{y+\varepsilon} \left(\int_{-\infty}^{\infty} x f_{\xi,\eta}(x,u)dx \right) du \left(\int_{y-\varepsilon}^{y+\varepsilon} f_\eta(u)du \right)^{-1} \to \int_{-\infty}^{\infty} x \ f_{\xi|\eta}(x \mid y)dx, \varepsilon \to 0.$$

Отже, *функція регресії* є природним узагальненням поняття умовного математичного сподівання за умови події додатної ймовірності.

Для прикладу розглянемо задачу обчислення нормальної регресії.

Теорема (про нормальну регресію на площині). *Нехай нормальний вектор на площині* $(\xi_1, \xi_2) \simeq N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ *з середніми* $\mu_i = \mathbf{E}\xi_i$, *дисперсіями* $\sigma_i^2 = \mathbf{D}\xi_i, i = 1, 2,$ *і коефіцієнтом кореляції* ρ. *Тоді*

(а) функція регресії ξ_2 *на* ξ_1 *є лінійною:*

$$m(x) \equiv \mathbf{E}(\xi_2 \mid \xi_1 = x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1),$$

(б) умовна щільність $f_{\xi_2|\xi_1}(y \mid x)$ *є нормальною щільністю розподілу величини* $N(m(x), \sigma_2^2(1 - \rho^2))$.

Зауваження. Позначимо $\varepsilon = \xi_2 - m(\xi_1)$. Унаслідок лінійності функції m за теоремою *про лінійні перетворення нормальних векторів* вектор (ε, ξ_1) є нормальним вектором. Крім того, $\mathbf{E}\varepsilon = 0$ та

$$\mathrm{cov}(\varepsilon, \xi_1) = \mathbf{E}(\xi_2 - m(\xi_1))\xi_1 = \mu_2\mu_1 + \rho\sigma_2\sigma_1 - \mu_2\mu_1 - \rho\frac{\sigma_2}{\sigma_1}\sigma_1^2 = 0.$$

Тому за теоремою *про незалежність координат нормального вектора* випадкові величини ε та ξ_1 незалежні. Отже, для нормального вектора (ξ_1, ξ_2) на площині має місце таке зображення другої координати:

$$\xi_2 = m(\xi_1) + \varepsilon = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(\xi_1 - \mu_1) + \varepsilon,$$

у вигляді суми її функції регресії $m(\xi_1)$ на першу координату та незалежної від цієї координати похибки ε. Вказане рівняння називається рівнянням регресії, причому для нормальної моделі це рівняння є лінійним відносно ξ_1.

Доведення. Сумісна щільність вектора (ξ_1, ξ_2) має вигляд

$$f_{\xi_1,\xi_2}(x, y) = \left(2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}\right)^{-1} \times$$
$$\exp\left(-\tfrac{1}{2(1-\rho^2)}\left[\left(\tfrac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\tfrac{x-\mu_1}{\sigma_1}\tfrac{y-\mu_2}{\sigma_2} + \left(\tfrac{y-\mu_2}{\sigma_2}\right)^2\right]\right).$$

Після ділення цього виразу на щільність величини ξ_1 з нормальним розподілом $N(\mu_1, \sigma_1^2)$ обчислимо

$$f_{\xi_2|\xi_1}(y \mid x) = \left(2\pi\sigma_2^2(1 - \rho^2)\right)^{-1/2} \exp\left(-\left(2\sigma_2^2(1 - \rho^2)\right)^{-1}[y - m(x)]^2\right),$$

що і доводить твердження (б), оскільки остання функція є *нормальною щільністю розподілу* з відповідними параметрами.

Формула (а) є очевидним наслідком теореми *про обчислення умовного сподівання через сумісну щільність* $\boxdot$

Вправи

(1) Довести рівність $\mathbf{E}(\xi \mid \eta = y) = m(y)$ для випадку, коли $\mathbf{P}(\eta = y) > 0$.

(2) Випадкові величини ξ, η незалежні та однаково розподілені. Обчислити умовні сподівання $\mathbf{E}(\xi \mid \xi + \eta)$, $\mathbf{E}(\xi \mid \xi\eta)$.

(3) Величина ξ рівномірно розподілена на $[0, \pi]$. Обчислити $\mathbf{E}(\xi \mid \sin \xi)$.

(4) Випадкова величина ξ має строго додатну щільність f. Обчислити $\mathbf{E}(\xi \mid \sin \xi)$, $\mathbf{E}(\xi \mid \xi^2)$, $\mathbf{E}(\xi \mid |\xi|)$.

(5) Випадкова величина $\xi \ge 0$, сигма-алгебра $\mathfrak{C} \subset \mathfrak{F}$, $\zeta = \mathbf{E}(\xi \mid \mathfrak{C})$, причому $\zeta < \infty$ м.н. Довести, що $\mathbf{P}(\xi > 0, \zeta = 0) = 0$, тобто дріб ξ/ζ визначений м.н. та задовольняє нерівність $\mathbf{E}(\xi/\zeta \mid \mathfrak{C}) \le 1$ м.н. Зокрема, $\mathbf{E}(\xi/\zeta) \le 1 < \infty$.

(6) Випадкові величини $(\xi_n, n \ge 1)$ незалежні, однаково розподілені та інтегровні. Довести, що (а) $\mathbf{E}(\xi_1 \mid S_k, k \ge n) = S_n/n$ м.н., де $S_n = \xi_1 + ... + \xi_n$. (б) $\mathbf{E}(\xi_k \mid S_n) = S_n/n$ м.н при $k \le n$. (в) У припущенні, що $\xi_n \in \mathbb{N}$, вивести звідси рекурентні рівняння $\mathbf{P}(S_n = k) = \frac{n}{k} \sum_{i=1}^{k} i\mathbf{P}(\xi_1 = i)\mathbf{P}(S_{n-1} = k - i)$ для послідовного обчислення розподілів сум S_n.

(7) Випадкові величини $(\xi_n, n \ge 1)$ незалежні, однаково розподілені та мають функцію розподілу F, а $\xi_{(n)} = \max_{1 \le k \le n} \xi_k$. Довести, що:

$$\mathbf{P}(\xi_k < x \mid \xi_{(n)} = t) = (n-1)F(x)/nF(t)1\!\!1_{x \le t} + 1\!\!1_{x > t} \text{ при } k \le n.$$

(8) Узагальнити твердження теореми про лінійну регресію на площині на випадок нормального вектора $(\xi_1, \xi_2) \simeq N_{d_1+d_2}(\mu_1, \mu_2, V_1, V_2, Q_{12})$.

(9) Випадкові величини $(\xi_n, n \ge 1)$ незалежні, а $S_n = \xi_1 + ... + \xi_n$. Довести, що S_k та S_l умовно незалежні за умови S_n при $k < n < l$, тобто $\forall A, B \in \mathfrak{B}[\mathbb{R}]$

$$\mathbf{P}(S_k \in A, S_l \in B \mid S_n = y) = \mathbf{P}(S_k \in A \mid S_n = y)\mathbf{P}(S_l \in B \mid S_n = y).$$

(10) $\mathrm{Cov}(\xi, \eta) = \mathrm{Cov}(\xi, \mathbf{E}(\eta \mid \xi))$ для квадратично інтегровних величин ξ, η.

(11) Нехай $(\xi_1, \xi_2) \simeq N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Тоді (а) частка дисперсії $\mathbf{D}(\xi_1)$, що пояснюється оптимальним прогнозом, дорівнює $\rho^2 = \mathbf{D}(\mathbf{E}(\xi_1 \mid \xi_2))/\mathbf{D}(\xi_1)$, (б) $\mathbf{E}(|(\xi_1 - \mu_1)/\sigma_1| \mid |\xi_2|) = 2(\sqrt{1-\rho^2} + \rho \arcsin \rho)/\pi$.

(12) Нехай величини ξ, η та борелєва функція $g : \mathbb{R}^2 \to \mathbb{R}$ такі, що величина $g(\xi, \eta)$ інтегровна, причому η вимірна відносно сигма-алгебри $\mathfrak{C} \subset \mathfrak{F}$. Довести, що $\mathbf{E}(g(\xi, \eta) \mid \mathfrak{C}) = \mathbf{E}(g(\xi, y) \mid \mathfrak{C}) |_{y=\eta}$.

2.5.6. Умовні ймовірності відносно сигма-алгебри

При побудові теорії ймовірностей первісним було поняття ймовірності, а математичне сподівання визначалося як інтеграл Лебега за заданою ймовірністю. Водночас умовне математичне сподівання визначене автономно, і формально не пов'язане з умовною ймовірністю. Зокрема, це призводить до необхідності для кожної нової випадкової величини повністю заново повторювати обчислення її умовного сподівання відносно тієї ж сигма-алгебри. Для подолання цієї незручності використовується поняття умовної ймовірності відносно сигма-алгебри.

Означення. *Нехай сигма-алгебра* $\mathfrak{C} \subset \mathfrak{F}$, *і* $A \in \mathfrak{F}$. *Умовною ймовірністю події* A *за умови сигма-алгебри* $\mathfrak{C}$ *називається випадкова величина*

$$\mathbf{P}(A \mid \mathfrak{C}) \equiv \mathbf{E}(1\!\!1_A \mid \mathfrak{C}).$$

Теорема (про основні властивості умовних імовірностей відносно сигма-алгебри). *Нехай сигма-алгебра* $\mathfrak{C} \subset \mathfrak{F}$. *Тоді*

(1) $\mathbf{P}(A \mid \mathfrak{C}) \geq 0$ *м.н. для всіх* $A \in \mathfrak{F}$.

(2) $\mathbf{P}(\Omega \mid \mathfrak{C}) = 1$ *м.н.*

(3) Для кожної послідовності попарно несумісних подій $(A_n, n \geq 1)$ $\mathbf{P}(\cup_{n\geq 1} A_n \mid \mathfrak{C}) = \sum_{n\geq 1} \mathbf{P}(A_n \mid \mathfrak{C})$ *м.н.*

Доведення теореми спирається на теорему *про загальні властивості умовного сподівання.*

(1) Є наслідком невід'ємності умовного сподівання від $1\!\!1_A \geq 0$.

(2) Випливає з нормованості умовного сподівання від $1\!\!1_\Omega \equiv 1$.

(3) Виводиться з теореми *про граничний перехід під знаком умовного сподівання*, оскільки $B_N \equiv \cup_{n=1}^N A_n \uparrow B_\infty \equiv \cup_{n=1} A_n$ при $N \to \infty$:

$$\sum_{n\geq 1} \mathbf{P}(A_n \mid \mathfrak{C}) = \lim_{N\to\infty} \sum_{n=1}^N \mathbf{E}(1\!\!1_{A_n} \mid \mathfrak{C}) = \lim_{N\to\infty} \mathbf{E}(1\!\!1_{B_N} \mid \mathfrak{C}) =$$
$$\mathbf{E}(1\!\!1_{B_\infty} \mid \mathfrak{C}) = \mathbf{P}(\cup_{n\geq 1} A_n \mid \mathfrak{C}) \text{ м.н. } \boxdot$$

2.5.7. Регулярні умовні ймовірності

Умовна ймовірність відносно сигма-алгебри є функцією як події $A \in \mathfrak{F}$, так і елементарної події $\omega \in \Omega$. Вона визначена при кожній A для майже всіх ω. Оскільки множин $A \in \mathfrak{F}$ нескінченно багато, то об'єднання відповідних виключних подій нульової ймовірності за різними A може мати додатну ймовірність (та навіть одиничну ймовірність). Тому не можна гарантувати, що ймовірність тієї множини ω, для яких функція $\mathrm{P}(A \mid \mathfrak{C})$ має властивість сигма-адитивності одночасно на всіх множинах A, дорівнює одиниці (більше того, не можна навіть стверджувати, що така множина ω не порожня. **Вправа**: навести відповідний приклад).

Для подолання вказаної обставини використовується певна модифікація поняття умовної ймовірності.

Означення. *Нехай σ-алгебра $\mathfrak{C} \subset \mathfrak{F}$. Функція $P(A,\omega) : \mathfrak{F} \times \Omega \to [0,1]$ називається* регулярною умовною ймовірністю *за умовою $\mathfrak{C}$, якщо:*

(а) для кожної $A \in \mathfrak{F}$ м.н. має місце рівність $P(A,\omega) = \mathbf{E}(1\!\!1_A \mid \mathfrak{C})(\omega)$,

(б) для майже всіх ω функція $P(\cdot,\omega)$ є ймовірністю від $A \in \mathfrak{F}$.

Про корисність введеного поняття свідчить таке твердження.

Теорема (про обчислення умовного сподівання через регулярну ймовірність). *Нехай σ-алгебра $\mathfrak{C} \subset \mathfrak{F}$, і існує регулярна умовна ймовірність $P(A,\omega), A \in \mathfrak{F}, \omega \in \Omega$, за умови $\mathfrak{C}$. Тоді для довільної невід'ємної або інтегровної випадкової величини ξ має місце рівність*

$$\mathbf{E}(\xi \mid \mathfrak{C})(\omega) = \int_\Omega \xi(\varpi) P(d\varpi,\omega) \quad \text{м.н.},$$

де інтеграл у правій частині є інтегралом Лебега від вимірної невід'-ємної або інтегровної функції ξ за ймовірнісною мірою $P(\cdot,\omega)$.

Доведення

(1) Нехай $\xi = 1\!\!1_A, A \in \mathfrak{F}$, індикаторна величина. Тоді права частина дорівнює $P(A,\omega)$ і твердження теореми випливає з означення *регулярної умовної ймовірності*, (а).

(2) Для простих величин ξ рівність теореми виводиться з лінійності м.н. її правої та лівої частини.

(3) Нехай величина ξ невід'ємна. Оберемо прості величини $0 \le \xi_n \uparrow \xi$. Тоді з відповідних рівностей для ξ_n за теоремою *про граничний перехід під знаком умовного сподівання* та за *теоремою Лебега про монотонну збіжність* граничним переходом $n \to \infty$ отримуємо шукане твердження.

(4) Для інтегровної величини ξ теорема доводиться за лінійністю з урахуванням скінченності м.н. $\mathbf{E}(\xi^{\pm} \mid \mathfrak{C})$ та інтегралів за мірою $P(\cdot,\omega)$ $\boxdot$

Теорема (про існування регулярної умовної ймовірності). *Нехай $\mathfrak{F} = \sigma[\xi]$ для деякої випадкової величини ξ, а сигма-алгебра $\mathfrak{C} \subset \mathfrak{F}$. Тоді існує регулярна умовна ймовірність $P(A,\omega)$ за умовою $\mathfrak{C}$.*

Доведення. За означенням *породженої сигма-алгебри* $\sigma[\xi]$ кожна подія $A \in \mathfrak{F} = \sigma[\xi]$ має вигляд $A = \{\xi \in B\}$ для деякої $B \in \mathfrak{B}(\mathbb{R})$. Тому умовну ймовірність $P(\cdot,\omega)$ слід будувати на подіях вигляду $\{\xi \in B\}$.

Нехай $D \subset \mathbb{R}$ зліченна щільна підмножина.

Для кожного $y \in D$ визначимо $H(y,\omega) = \mathbf{E}(1\!\!1_{\{\xi < y\}} \mid \mathfrak{C})(\omega)$.

(1) Розглянемо для $y_1 < y_2,\ y_i \in D$ події

$$O_1(y_1,y_2) = \{\omega : H(y_1,\omega) > H(y_2,\omega)\},\ O_1 = \bigcup_{y_1<y_2,\ y_i \in D} O_1(y_1,y_2).$$

Оскільки $1\!\!1_{\{\xi<y_1\}} \le 1\!\!1_{\{\xi<y_2\}}$, то за теоремою *про загальні властивості умовного сподівання* $\mathbf{P}(O_1(y_1,y_2)) = 0$, і $\mathbf{P}(O_1) = 0$. Зауважимо, що на

доповненні $\overline{O_1}$ функція $H(y,\omega)$ не спадає за y на множині D. Зокрема, при $\omega \in \overline{O_1}$ ця функція має границю зліва та справа $H(y \pm 0, \omega)$ на D.

(2) Визначимо

$$O_2 = \overline{O_1} \cap (\{\omega : H(-\infty, \omega) \neq 0\} \cup \{\omega : H(+\infty, \omega) \neq 1\}).$$

Оскільки $1\!\!\mathrm{I}_{\xi<y} \downarrow 0$ при $y \downarrow -\infty$ та $1\!\!\mathrm{I}_{\xi<y} \uparrow 1$ при $y \uparrow -\infty$, то за теоремою *про граничний перехід під знаком умовного сподівання* $\mathbf{P}(O_2) = 0$, якщо врахувати також, що $\mathbf{P}(\overline{O_1}) = 1$. Зауважимо, що на $\overline{O_2}$ функція $H(y,\omega)$ не спадає на D та *нормована* на D.

(3) Розглянемо події

$$O_3(y) = \overline{O_1} \cap \{\omega : H(y-0, \omega) \neq H(y,\omega)\},\ O_3 = \bigcup\nolimits_{y \in D} O_3(y).$$

Оскільки $1\!\!\mathrm{I}_{\{\xi<u\}} \uparrow 1\!\!\mathrm{I}_{\{\xi<y\}}$ при $u \uparrow y$, то за теоремою *про граничний перехід під знаком умовного сподівання* $\mathbf{P}(O_3(y)) = 0$ та $\mathbf{P}(O_3) = 0$.

Нехай $O = O_1 \cup O_2 \cup O_3$. Тоді $\mathbf{P}(O) = 0$ та для всіх $\omega \in \overline{O}$ функція $H(y,\omega)$ є неспадною, нормованою та неперервною зліва на D.

Крім того, для кожного $y \in D$ за означенням $H(y,\omega) = \mathbf{E}(1\!\!\mathrm{I}_{\{\xi<y\}} \mid \mathfrak{C})$. Зокрема, випадкова величина $H(y,\omega)$ є $\mathfrak{C}$-вимірною та задовольняє умову балансу $\mathbf{E}(1\!\!\mathrm{I}_{\{\xi<y\}} 1\!\!\mathrm{I}_A) = \mathbf{E}(H(y) 1\!\!\mathrm{I}_A), \forall A \in \mathfrak{C}$. Визначимо

$$G(x,\omega) = \sup(H(y,\omega), y < x, y \in D).$$

Так само, як і в теоремі *про збіжність в основному та на щільній множині*, (б), доводимо, що для кожної $\omega \in \overline{O}$ функція $G(x,\omega)$ є *узагальненою функцією розподілу* за аргументом $x \in \mathbb{R}$, а з включення $\omega \in \overline{O_2}$ та теореми *про характеризацію функцій розподілу в класі узагальнених* робимо висновок, що вказана функція є *функцією розподілу*. Далі, верхня межа в означенні G досягається на будь-якій зліченній послідовності $y_n \uparrow x, y_n \in D$, тому $G(x,\omega)$ є $\mathfrak{C}$-вимірною за аргументом ω, як і $H(y,\omega)$. Тоді з рівностей балансу $\mathbf{E}(1\!\!\mathrm{I}_{\{\xi<y_n\}} 1\!\!\mathrm{I}_A) = \mathbf{E}(H(y_n) 1\!\!\mathrm{I}_A), \forall A \in \mathfrak{C}$, та з *теореми Лебега про монотонну збіжність* виводимо при $y_n \uparrow x$, що таку ж умову балансу задовольняє і $G(x,\omega)$. Отже, за означенням *умовного математичного сподівання* рівність $G(x,\omega) = \mathbf{E}(1\!\!\mathrm{I}_{\{\xi<x\}} \mid \mathfrak{C})$ м.н. виконується для кожного $x \in \mathbb{R}$.

Позначимо через $F(B,\omega), B \in \mathfrak{A}[\mathbb{R}]$, *адитивну міру Лебега – Стілтьєса*, що породжена функцією розподілу $G(x,\omega)$ при $\omega \in \overline{O}$, та тим же виразом для інших ω. Оскільки вказаний вираз зводиться до лінійного перетворення, то $F(B,\omega) = \mathbf{E}(1\!\!\mathrm{I}_{\{\xi \in B\}} \mid \mathfrak{C})$ м.н.для всіх $B \in \mathfrak{A}[\mathbb{R}]$, причому ця функція $\mathfrak{C}$-вимірна за ω.

За теоремою *про неперервність в нулі міри Лебега – Стілтьєса* та *теоремою Каратеодорі про продовження міри* для кожної елементарної події $\omega \in \overline{O}$ існує імовірнісна міра $Q(B,\omega), B \in \mathfrak{B}[\mathbb{R}]$, що є продовженням міри $F(B,\omega)$ при $B \in \mathfrak{A}[\mathbb{R}]$. Визначимо $Q(B,\omega) = 1\!\!\mathrm{I}_{\{0\in B\}}$ при $\omega \in O$. Тоді $Q(B,\omega) = F(B,\omega) = \mathbf{E}(1\!\!\mathrm{I}_{\{\xi\in B\}} \mid \mathfrak{C})$ м.н. для кожної множини $B \in \mathfrak{A}[\mathbb{R}]$.

Розглянемо такий клас множин:

$$\mathfrak{D} = \{B \in \mathfrak{B}[\mathbb{R}] : Q(B,\omega) = \mathbf{E}(1\!\!\mathrm{I}_{\{\xi\in B\}} \mid \mathfrak{C}) \text{ м.н.}\}.$$

Як доведено вище, $\mathfrak{A}[\mathbb{R}] \subset \mathfrak{D}$. Крім того, за означенням клас $\mathfrak{D}$ є монотонним, оскільки функція $Q(B,\omega)$ є сигма-адитивною, а тому і неперервною за B, а для умовного сподівання $\mathbf{E}(1\!\!\mathrm{I}_{\{\xi\in B\}} \mid \mathfrak{C})$ м.н. виконується теорема про монотонну збіжність як наслідок теореми *про граничний перехід під знаком умовного.сподівання*. Тому за теоремою *про монотонний клас* $\mathfrak{D} = \mathfrak{B}[\mathbb{R}]$.

Оскільки $Q(B,\omega)$ є ймовірнісною мірою, з урахуванням останньої рівності виводимо, що функція $P(A,\omega) \equiv Q(B,\omega)$ при $A = \{\xi \in B\}$ є шуканою регулярною умовною ймовірністю $\boxdot$

Вправи

(1) Довести такі формули повної ймовірності: (а) $\mathbf{P}(A) = \mathbf{E}\mathrm{P}(A \mid \mathfrak{C})$ для всіх $A \in \mathfrak{F}$, (б) $\mathbf{E}g(\xi) = \int_{\mathbb{R}} \mathbf{E}(g(\xi) \mid \eta = y)dF_\eta(y)$ за умови інтегровності $g(\xi)$.

(2) Довести узагальнену формулу Байєса: для $A \in \mathfrak{F}, H \in \mathfrak{C}$
$\mathbf{P}(H \mid A) = (\mathbf{E}1\!\!\mathrm{I}_H\mathbf{P}(A \mid \mathfrak{C}))\,/\mathbf{E}\mathrm{P}(A \mid \mathfrak{C})$.

(3) Величина ξ має строго додатну щільність f. Обчислити $\mathbf{P}(\xi < x \mid |\xi|)$.

(4) Нехай $(\zeta_k, k = \overline{1,n})$ – стандартний нормальний вектор. Довести, що його умовний розподіл за умови, що фіксовані суми $\sum_{k=1}^n \zeta_k, \sum_{k=1}^n \zeta_k^2$, є рівномірним на відповідній $(n-2)$-вимірній множині.

(5) Сигма-алгебри $\mathfrak{F}_1, \mathfrak{F}_2 \subset \mathfrak{F}$ називаються умовно незалежними за умови $\mathfrak{F}_3 \subset \mathfrak{F}$, якщо $\mathbf{P}(F_1 \cap F_2 \mid \mathfrak{F}_3) = \mathbf{P}(F_1 \mid \mathfrak{F}_3)\mathbf{P}(F_2 \mid \mathfrak{F}_3)$ м.н. для всіх $F_i \in \mathfrak{F}_i$. Довести, що вказана умовна незалежність має місце тоді й тільки тоді, коли $\mathbf{P}(F_1 \mid \sigma[\mathfrak{F}_2 \cup \mathfrak{F}_3]) = \mathbf{P}(F_1 \mid \mathfrak{F}_3)$ м.н. для всіх $F_1 \in \mathfrak{F}_1$.

(6) Для величин ξ, η існує регулярна умовна ймовірність $P(x,B), B \in \mathfrak{B}(\mathbb{R})$, така, що $P(x,B) = \mathbf{P}(\xi \in B \mid \eta = x)$ для майже всіх $x \in \mathbb{R}$ за мірою $\mathbf{P}(\eta \in \cdot)$. Довести, що для довільної борелєвої обмеженої функції g виконується така рівність: $\mathbf{E}(g(\eta,\xi) \mid \eta = x) = \int_{\mathbb{R}} g(x,y)P(x,dy)$ для майже всіх x.

(7) Нехай ξ – випадковий елемент зі значеннями у повному сепарабельному метричному просторі E з борелєвою сигма-алгеброю $\mathfrak{B}(E)$. Довести, що існує регулярна умовна ймовірність $P(B,\omega) = \mathbf{P}(\xi \in B \mid \mathfrak{C})$ м.н.

(8) Випадкові величини $(\xi_k, k \geq 1)$ незалежні, однаково розподілені та мають неперервну функцію розподілу F. Знайти $\mathbf{P}(\xi_1 < x \mid \max_{1\leq k\leq n} \xi_k)$.

(9) Випадкові величини ξ_1, ξ_2 вимірні відносно σ-алгебр $\mathfrak{C}_1, \mathfrak{C}_2$ відповідно, а параметри $p, q, r \in (0, \infty]$ такі, що $p^{-1} + q^{-1} + r^{-1} = 1$. Довести, що

$$|\mathbf{E}\xi_1\xi_2 - \mathbf{E}\xi_1\mathbf{E}\xi_2| \leq 8(\mathbf{E}\,|\xi_1|^p)^{1/p}\,(\mathbf{E}\,|\xi_2|^q)^{1/q}\,(\alpha(\mathfrak{C}_1, \mathfrak{C}_2))^{1/r},$$

де $\alpha(\mathfrak{C}_1, \mathfrak{C}_2) \equiv \sup(|\mathbf{P}(A_1 \cap A_2) - \mathbf{P}(A_1)\mathbf{P}(A_2)|\,,\ A_k \in \mathfrak{C}_k)$.

2.6. Ряди з незалежних величин

Нехай $(\xi_n, n \geq 1)$ – послідовність *незалежних у сукупності* випадкових величин, а $S_n = \xi_1 + ... + \xi_n$ – відповідні частинні суми. Тоді випадкова подія $\{$ряд $\sum \xi_n$ збігається$\} = \{\exists \lim_{n\to\infty} S_n\}$ належить *залишковій сигма-алгебрі* $\Delta_\infty \equiv \overline{\lim}_{n\to\infty}\sigma[\xi_n]$, що породжена незалежними сигма-алгебрами $\sigma[\xi_n]$. За теоремою *про закон нуля та одиниці Колмогорова* $\mathbf{P}(A) \in \{0, 1\}$ для довільної події $A \in \Delta_\infty$. За *теоремою Колмогорова про побудову ймовірності на множинах послідовностей* ймовірність події $A \in \Delta_\infty$ однозначно визначається послідовністю *маргінальних функцій розподілу* $F_{\xi_n}(x), n \geq 1$. А.М.Колмогорову належать також необхідні та достатні умови на ці функції розподілу для справедливості рівності $\mathbf{P}(\exists \lim_{n\to\infty} S_n) = 1$.

2.6.1. Нерівності Колмогорова

Теорема (про нерівності Колмогорова). *Нехай випадкові величини* $(\xi_n, n \geq 1)$ *незалежні у сукупності,* $\mathbf{E}\xi_n = 0$, $\mathbf{E}\xi_n^2 < \infty$, *а* $S_n = \xi_1 + ... + \xi_n$.

(а) Тоді для кожного $\varepsilon > 0$

$$\mathbf{P}(\max_{1\leq k\leq n} |S_k| \geq \varepsilon) \leq \mathbf{D}S_n/\varepsilon^2.$$

(б) Якщо додатково $|\xi_n| \leq c$ *для сталої* c, *то при всіх* $\varepsilon > 0$

$$\mathbf{P}(\max_{1\leq k\leq n} |S_k| \geq \varepsilon) \geq 1 - (c + \varepsilon)^2/\mathbf{D}S_n.$$

Доведення. Твердження (а) є доведена вище *нерівність Колмогорова.*

(б) Розглянемо, як і у згаданій нерівності, випадкові події

$$A = \{\max_{1\leq k\leq n} |S_k| \geq \varepsilon\}, \quad A_k = \{|S_1| < \varepsilon, ..., |S_{k-1}| < \varepsilon, |S_k| \geq \varepsilon\}.$$

Тоді $A_i \cap A_j = \varnothing$, $i \neq j$, та $A = \cup_{k=1}^n A_k$, звідки $1\!\!1_A = \sum_{k=1}^n 1\!\!1_{A_k}$. Зауважимо, що $\mathbf{E}S_n^2 = \mathbf{D}S_n$, оскільки $\mathbf{E}S_n = \sum_{k=1}^n \mathbf{E}\xi_k = 0$.

У доведенні *нерівності Колмогорова* встановлено тотожність

$$\mathbf{E}S_n^2 \mathbb{1}_A = \sum\nolimits_{k=1}^n \mathbf{E}(S_k^2 \mathbb{1}_{A_k}) + \sum\nolimits_{k=1}^n \mathbf{E}(S_n - S_k)^2 \mathbb{1}_{A_k}.$$

Оцінимо

$$\mathbf{E}S_n^2 \mathbb{1}_A = \mathbf{E}S_n^2 - \mathbf{E}S_n^2 \mathbb{1}_{\overline{A}} \geq \mathbf{E}S_n^2 - \varepsilon^2 \mathbf{P}(\overline{A}),$$

де враховано нерівність $S_n^2 \mathbb{1}_{\overline{A}} \leq \varepsilon^2 \mathbb{1}_{\overline{A}}$.

З попередніх тотожності та нерівності виводимо, що

$$\mathbf{E}S_n^2 - \varepsilon^2 \mathbf{P}(\overline{A}) \leq \textstyle\sum_{k=1}^n \mathbf{E}(S_k^2 \mathbb{1}_{A_k}) + \sum_{k=1}^n \mathbf{E}(S_n - S_k)^2 \mathbb{1}_{A_k} \leq$$
$$\textstyle\sum_{k=1}^n \mathbf{E}((c+\varepsilon)^2 \mathbb{1}_{A_k}) + \sum_{k=1}^n \mathbf{E}(S_n - S_k)^2 \ \mathbf{E}\mathbb{1}_{A_k} \leq$$
$$(c+\varepsilon)^2 \textstyle\sum_{k=1}^n \mathbf{P}(A_k) + \sum_{k=1}^n \mathbf{E}S_n^2 \ \mathbf{P}(A_k) = ((c+\varepsilon)^2 + \mathbf{E}S_n^2)\,\mathbf{P}(A).$$

Тут використано також такі чотири співвідношення.

(а) За умови обмеженості ξ_k та внаслідок означення події A_k :

$$S_k^2 \mathbb{1}_{A_k} = (S_{k-1} + \xi_k)^2 \mathbb{1}_{A_k} \leq (\varepsilon + c)^2 \mathbb{1}_{A_k}.$$

(б) За теоремою *про векторні перетворення незалежних величин* величини $S_n - S_k = \xi_{k+1} + ... + \xi_n$ та $\mathbb{1}_{A_k} = g(\xi_1, ..., \xi_k)$ незалежні, тому за теоремою *про математичне сподівання добутку незалежних величин*:

$$\mathbf{E}(S_n - S_k)^2 \mathbb{1}_{A_k} = \mathbf{E}(S_n - S_k)^2 \ \mathbf{E}\mathbb{1}_{A_k}.$$

(в) З урахуванням центрованості сум S_k та теореми *про дисперсію суми незалежних величин*:

$$\mathbf{E}(S_n - S_k)^2 = \mathbf{D}\sum\nolimits_{i=k+1}^n \xi_i = \sum\nolimits_{i=k+1}^n \mathbf{D}\xi_i \leq \sum\nolimits_{i=1}^n \mathbf{D}\xi_i = \mathbf{E}S_n^2.$$

(г) Нарешті, через попарну несумісність A_k :

$$\sum\nolimits_{k=1}^n \mathbf{P}(A_k) = \mathbf{P}(A).$$

Підстановкою $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$ з доведеної нерівності отримуємо

$$\mathbf{P}(A) \geq (\mathbf{E}S_n^2 - \varepsilon^2)/((c+\varepsilon)^2 - \varepsilon^2 + \mathbf{E}S_n^2) =$$
$$1 - (c+\varepsilon)^2/((c+\varepsilon)^2 - \varepsilon^2 + \mathbf{D}S_n) \geq 1 - (c+\varepsilon)^2/\mathbf{D}S_n \ \boxdot$$

Вправи

(1) Величини (ξ_n) незалежні, $\mathbf{E}\xi_n = 0$, $\mathbf{E}\xi_n^2 < \infty$, $S_n = \xi_1 + ... + \xi_n$, а послідовність $a_n > 0$ неспадає. Тоді $\mathbf{P}(\max_{k \leq n} |S_k| / a_k > 1) \leq \sum_{k=1}^n a_k^{-2} \mathbf{E}\xi_k^2$.

(2) Величини (ξ_n) незалежні та $\mathbf{E}\xi_n = 0$, $\sum_{k \geq 1} \mathbf{E}\,|\xi_k| / k < \infty$. Довести: (а) для $\varepsilon > 0$ нерівність $\mathbf{P}(\sum_{k=n}^m |\xi_k| / k > \varepsilon) \leq \varepsilon^{-1} \sum_{k=n}^m \mathbf{E}\,|\xi_k| / k$. (б) збіжність м.н. ряду $\sum_{k=n}^m |\xi_k| / k$ (в) посилений закон великих чисел для (ξ_n).

(3) Випадкові величини $(\xi_n, n \geq 1)$ незалежні у сукупності, $\mathbf{E}\xi_n = 0$, $S_n = \xi_1 + ... + \xi_n$, причому $\mathbf{P}(S_n - S_k \geq a) \geq b > 0$ для всіх $k < n$. Довести нерівність $\mathbf{P}(\max_{1\leq k\leq n} S_k \geq \varepsilon) \leq b^{-1}\mathbf{P}(S_n \geq \varepsilon + a)$.

(4) Величини $(\xi_n, n \geq 1)$ незалежні, $\mathbf{E}\xi_n = 0$, $\mathbf{E}\xi_n^2 < \infty$, $S_n = \xi_1 + ... + \xi_n$. Довести для $\varepsilon > 0$, що $\mathbf{P}(\max_{1\leq k\leq n} S_k \geq \varepsilon) \leq \mathbf{E}S_n^2/(\varepsilon^2 + \mathbf{E}S_n^2)$.

2.6.2. Теореми про один та два ряди

Лема (про критерій фундаментальності майже напевне). *Послідовність випадкових величин $(\xi_n, n \geq 1)$ фундаментальна майже напевне тоді й тільки тоді, коли*

$$\sup_{k\geq 0} \left|\xi_{n+k} - \xi_n\right| \to^P 0,\ n \to \infty.$$

Доведення. Позначимо події $A = \{(\xi_n,\ n \geq 1)$ фундаментальна$\}$ та $A_{kN} = \cap_{n\geq N} \cap_{m\geq N} \{|\xi_n - \xi_m| \leq 1/k\}$. Тоді $A = \cap_{k\geq 1} \cup_{N\geq 1} A_{kN}$.

За означенням події A_{kN} монотонно не зростають за k та $\cup_{N\geq 1} A_{kN} \downarrow A$ при $k \to \infty$. Тому за *неперервністю ймовірності* $\mathbf{P}(A) = 1$ тоді й тільки тоді, коли $\mathbf{P}(\cup_{N\geq 1} A_{kN}) = 1$ для кожного k. Оскільки A_{kN} монотонно не спадають за N, то остання рівність еквівалентна $\lim_{N\to\infty} \mathbf{P}(A_{kN}) = 1$ для кожного k. За означенням збіжності *за ймовірністю* це твердження еквівалентне умові леми, оскільки

$$\mathbf{P}(A_{kN}) = \mathbf{P}(\sup_{n\geq N, m\geq N} |\xi_n - \xi_m| \leq 1/k) =$$
$$1 - \mathbf{P}(\sup_{n\geq N, m\geq N} |\xi_n - \xi_m| > 1/k) \to 1, N \to \infty,$$

тоді й тільки тоді, коли $\sup_{n\geq N, m\geq N} |\xi_n - \xi_m| \to^P 0$, причому

$$\sup_{k\geq 0} \left|\xi_{N+k} - \xi_N\right| \leq \sup_{n\geq N, m\geq N} |\xi_n - \xi_m| \leq 2 \sup_{k\geq 0} \left|\xi_{N+k} - \xi_N\right| \ \boxdot$$

Теорема (про один ряд). *Нехай випадкові величини $(\xi_n, n \geq 1)$ незалежні у сукупності, $\mathbf{E}\xi_n = 0$, $\mathbf{E}\xi_n^2 < \infty$.*

(а) Якщо $\sum_{n\geq 1} \mathbf{E}\xi_n^2 < \infty$, то ряд $\sum_{n\geq 1} \xi_n$ збігається м.н.

(б) Якщо $|\xi_n| \leq c$ і ряд $\sum_{n\geq 1} \xi_n$ збігається м.н., то $\sum_{n\geq 1} \mathbf{E}\xi_n^2 < \infty$.

Доведення. Позначимо $S_n = \xi_1 + ... + \xi_n$.

Збіжність ряду $\sum_{n\geq 1} \xi_n$ еквівалентна фундаментальності послідовності частинних сум S_n. За лемою *про критерій фундаментальності майже напевне* ця фундаментальність має місце тоді й тільки тоді, коли при кожному $\varepsilon > 0$

$$\mathbf{P}(\sup_{k\geq 0} |S_{n+k} - S_n| \geq \varepsilon) \to 0, n \to \infty.$$

За властивістю неперервності ймовірності

$$\mathbf{P}(\sup_{k\geq 0}|S_{n+k}-S_n|\geq\varepsilon)=\lim_{m\to\infty}\mathbf{P}(\max{}_{0\leq k\leq m}|S_{n+k}-S_n|\geq\varepsilon)=$$
$$\lim_{m\to\infty}\mathbf{P}(\max{}_{0\leq k\leq m}|S_{nk}|\geq\varepsilon),$$

де за означенням $S_{nk}\equiv\xi_{n+1}+...+\xi_{n+k}$ є сумами незалежних величин з нульовим математичним сподіванням та скінченною дисперсією.

Отже, ряд $\sum_{n\geq 1}\xi_n$ збігається м.н. тоді й тільки тоді, коли

$$\lim_{n\to\infty}\lim_{m\to\infty}\mathbf{P}(\max{}_{0\leq k\leq m}|S_{nk}|\geq\varepsilon)=0,\quad\forall\varepsilon>0.$$

(а) За теоремою *про нерівності Колмогорова*, (а), та за теоремою *про дисперсію суми незалежних величин*

$$\mathbf{P}(\max_{0\leq k\leq m}|S_{nk}|\geq\varepsilon)\leq\mathbf{D}S_{nm}/\varepsilon^2=\varepsilon^{-2}\mathbf{D}\sum_{k=1}^{m}\xi_{n+k}=$$
$$\varepsilon^{-2}\sum_{k=1}^{m}\mathbf{D}\xi_{n+k}\leq\varepsilon^{-2}\sum_{k\geq 1}\mathbf{D}\xi_{n+k}=\varepsilon^{-2}\sum_{k>n}\mathbf{E}\xi_k^2\to 0,\ n\to\infty,$$

де врахована також збіжність до нуля залишку збіжного ряду. Отже, твердження (а) випливає з наведеного вище критерію.

(б) Припустимо від супротивного розбіжність, тобто: $\sum_{n\geq 1}\mathbf{E}\xi_n^2=\infty$. Тоді $\mathbf{D}S_{nm}=\sum_{k=1}^{m}\mathbf{D}\xi_{n+k}=\sum_{k=n+1}^{n+m}\mathbf{D}\xi_k\to\infty, m\to\infty$ Отже, за теоремою *про нерівності Колмогорова*, (б)

$$\lim_{m\to\infty}\mathbf{P}(\max_{0\leq k\leq m}|S_{nk}|\geq\varepsilon)\geq 1-\lim_{m\to\infty}(c+\varepsilon)^2/\mathbf{D}S_{nm}=1.$$

Ця рівність суперечить збіжності ряду $\sum_{n\geq 1}\xi_n$ внаслідок наведеного вище критерію $\boxdot$

Теорема (про два ряди). *Нехай випадкові величини* $(\xi_n, n\geq 1)$ *незалежні у сукупності і* $\mathbf{E}\xi_n^2<\infty$.

(а) Якщо числові ряди $\sum_{n\geq 1}\mathbf{E}\xi_n$, $\sum_{n\geq 1}\mathbf{D}\xi_n$ *збігаються, то ряд* $\sum_{n\geq 1}\xi_n$ *збігається м.н.*

(б) Якщо $|\xi_n|\leq c$ *і ряд* $\sum_{n\geq 1}\xi_n$ *збігається м.н., то збігаються і ряди* $\sum_{n\geq 1}\mathbf{E}\xi_n$, $\sum_{n\geq 1}\mathbf{D}\xi_n$.

Доведення

(а) Позначимо $\zeta_n=\xi_n-\mathbf{E}\xi_n$. Тоді величини ζ_n незалежні, $\mathbf{E}\zeta_n=0$, $\sum_{n\geq 1}\mathbf{E}\zeta_n^2=\sum_{n\geq 1}\mathbf{D}\xi_n<\infty$. Тому за теоремою *про один ряд*, (а), ряд $\sum_{n\geq 1}\zeta_n$ збігається м.н. Тоді за властивістю суми збіжних рядів та за означенням ζ_n ряд $\sum_{n\geq 1}\xi_n=\sum_{n\geq 1}(\zeta_n+\mathbf{E}\xi_n)$ також збігається м.н.

(б) Розглянемо ймовірнісний простір $(\Omega,\mathfrak{F},\mathbf{P})$, на якому задано послідовність $(\xi_n, n\geq 1)$. Нехай $(\Omega',\mathfrak{F}',\mathbf{P}')$ – копія цього простору, а функції $\xi'_n(\omega')=\xi_n(\omega)$ при $\omega'=\omega$ є копіями функцій ξ_n. Тоді $(\xi'_n, n\geq 1)$ є незалежними величинами та мають такі ж маргінальні та скінченно-вимірні розподіли, що і $(\xi_n, n\geq 1)$. Отже, за *теоремою Колмогорова про побудову* *ймовірності на множині послідовностей* ряд $\sum_{n\geq 1}\xi'_n$ збігається м.н. за ймовірністю $\mathbf{P}'$.

Визначимо *прямий добуток* $(\widetilde{\Omega},\widetilde{\mathfrak{F}},\widetilde{\mathbf{P}}) \equiv (\Omega,\mathfrak{F},\mathbf{P}) \times (\Omega',\mathfrak{F}',\mathbf{P}')$, тобто множина $\widetilde{\Omega} = \Omega\times\Omega' = \{\widetilde{\omega} = (\omega,\omega'), \omega\in\Omega, \omega'\in\Omega'\}$, сигма-алгебра подій $\widetilde{\mathfrak{F}} = \sigma[A\times A', A\in\mathfrak{F}, A'\in\mathfrak{F}']$, та міра $\widetilde{\mathbf{P}}(A\times A') = \mathbf{P}(A)\mathbf{P}'(A')$. Значення $\widetilde{\mathbf{P}}$ на $\widetilde{\mathfrak{F}}$ знаходяться за *теоремою Каратеодорі про продовження міри.*

Розглянемо на ймовірнісному просторі $(\widetilde{\Omega},\widetilde{\mathfrak{F}},\widetilde{\mathbf{P}})$ такі випадкові величини: $\widetilde{\xi}_n(\widetilde{\omega}) = \xi_n(\omega)$, $\widetilde{\xi}'_n(\widetilde{\omega}) = \xi'_n(\omega')$ для $\widetilde{\omega} = (\omega,\omega')$. За означенням

$$\widetilde{\mathbf{P}}((\widetilde{\xi}_1,...,\widetilde{\xi}_n)\in B_1\times...\times B_n,\ (\widetilde{\xi}'_1,...,\widetilde{\xi}'_m)\in C_1\times...\times C_m) =$$

$$\widetilde{\mathbf{P}}(\{(\omega,\omega') : (\xi_1,...,\xi_n)(\omega)\in B_1\times...\times B_n,\ (\xi'_1,...,\xi'_m)(\omega')\in C_1\times...\times C_m\}) =$$

$$\widetilde{\mathbf{P}}(\{\omega : (\xi_1,...,\xi_n)\in B_1\times...\times B_n\}\times\{\omega' :\ (\xi'_1,...,\xi'_m)\in C_1\times...\times C_m\}) =$$

$$\mathbf{P}((\xi_1,...,\xi_n)\in B_1\times...\times B_n)\ \mathbf{P}'((\xi'_1,...,\xi'_m)\in C_1\times...\times C_m) =$$

$$\prod_{i=1}^n \mathbf{P}(\xi_i\in B_i)\ \prod_{j=1}^m \mathbf{P}(\xi_j\in C_j).$$

Отже, випадкові величини $(\widetilde{\xi}_n,\widetilde{\xi}'_m, n\geq 1, m\geq 1)$ незалежні у сукупності, і розподіл $\widetilde{\xi}_n$ та $\widetilde{\xi}'_n$ збігається з розподілом ξ_n. Тому за *теоремою Колмогорова про побудову ймовірності на множині послідовностей*

$$\widetilde{\mathbf{P}}\left(\text{ряд } \sum\nolimits_{n\geq 1}\widetilde{\xi}_n \text{ збігається}\right) = \widetilde{\mathbf{P}}\left(\text{ряд } \sum\nolimits_{n\geq 1}\widetilde{\xi}'_n \text{ збігається}\right) =$$

$$\mathbf{P}\left(\text{ряд } \sum\nolimits_{n\geq 1}\xi_n \text{ збігається}\right) = 1,$$

оскільки ці ймовірності однозначно визначаються послідовністю функцій розподілу величин ξ_n.

За теоремою *про векторні перетворення незалежних величин* випадкові величини $\zeta_n = \widetilde{\xi}_n - \widetilde{\xi}'_n$ на просторі $(\widetilde{\Omega},\widetilde{\mathfrak{F}},\widetilde{\mathbf{P}})$ незалежні. Крім того, внаслідок вказаної вище збіжності розподілів та незалежності $\widetilde{\mathbf{E}}\zeta_n = \widetilde{\mathbf{E}}\widetilde{\xi}_n - \widetilde{\mathbf{E}}\widetilde{\xi}'_n = 0$, $\widetilde{\mathbf{D}}\zeta_n = \widetilde{\mathbf{D}}\widetilde{\xi}_n + \widetilde{\mathbf{D}}\widetilde{\xi}'_n = 2\mathbf{D}\xi_n < \infty$, та $|\zeta_n| \leq \left|\widetilde{\xi}_n - \widetilde{\xi}'_n\right| \leq 2c$. Як доведено вище, $\widetilde{\mathbf{P}}\left(\text{ряд } \sum_{n\geq 1}\zeta_n \text{ збігається}\right) = 1$. Отже, послідовність $(\zeta_n, n\geq 1)$ задовольняє. умови теореми *про один ряд*, (б), і внаслідок цієї теореми збігається ряд: $2\sum_{n\geq 1}\mathbf{D}\xi_n = \sum_{n\geq 1}\widetilde{\mathbf{D}}\zeta_n < \infty$. Тому за цією ж теоремою, (а), ряд із незалежних та центрованих величин $\xi_n - \mathbf{E}\xi_n$ збігається м.н. Нарешті, як сума збіжних рядів ряд $\sum_{n\geq 1}(\xi_n - (\xi_n - \mathbf{E}\xi_n)) = \sum_{n\geq 1}\mathbf{E}\xi_n$ збігається м.н. $\boxdot$

Вправи

(1) Знайти множину тих $\omega \in \Omega$, для яких збігається ряд $\sum_{n\ge1}\xi_n(\omega)$.

(2) Випадкові величини $(\xi_n, n \ge 1)$ незалежні у сукупності, $\mathbf{P}(\xi_n = \pm 1) = 1/2$. Ряд $\sum_{n\ge1} a_n\xi_n$ збігається м.н. тоді й тільки тоді, коли $\sum_{n\ge1} a_n^2 < \infty$.

(3) Нехай $\xi_n \to^P 0, n \to \infty$. Тоді знайдеться числова послідовність c_n така, що $\sum_{n\ge1}|c_n| = \infty$ та ряд $\sum_{n\ge1} c_n\xi_n$ збігається м.н.

(4) Навести приклад (необмеженої) послідовності незалежних квадратично інтегровних центрованих величин $(\xi_n, n \ge 1)$ таких, що ряд $\sum_{n\ge1}\xi_n$ збігається м.н., і одночасно $\sum_{n\ge1}\mathbf{E}\xi_n^2 = \infty$.

(5) Випадкові величини $(\xi_n, n \ge 1)$ такі, що $\sum_{n\ge1}\mathbf{E}\,|\xi_n| < \infty$. Довести, що ряд $\sum_{n\ge1}\xi_n$ збігається м.н.

(6) Випадкові величини $(\xi_n, n \ge 1)$ та числова послідовність ε_n такі, що $\sum_{n\ge1}\varepsilon_n < \infty$ і $\sum_{n\ge1}\mathbf{P}(|\xi_n| \ge \varepsilon_n) < \infty$. Довести, що $\sum_{n\ge1}|\xi_n| < \infty$ м.н.

(7) Величини $(\xi_n, n \ge 1)$ незалежні та квадратично інтегровні. Довести, що: (а) ряд $\sum_{n\ge1}\xi_n$ збігається у середньому квадратичному тоді й тільки тоді, коли збігаються ряди $\sum_{n\ge1}\mathbf{E}\xi_n$, $\sum_{n\ge1}\mathbf{D}\xi_n$. (б) За (а) ряд $\sum_{n\ge1}\xi_n$ збігається м.н.

(8) Величини $(\xi_n, n \ge 1)$ незалежні, нормально розподілені, $\mathbf{E}\xi_n = 0$. Довести, що ряд $\sum_{n\ge1}\xi_n$ збігається м.н. тоді й тільки тоді, коли $\sum_{n\ge1}\mathbf{E}\xi_n^2 < \infty$.

2.6.3. Теорема про три ряди

Теорема (теорема Колмогорова про три ряди). *Нехай випадкові величини $(\xi_n, n \ge 1)$ незалежні у сукупності. Для того, щоб ряд $\sum_{n\ge1}\xi_n$ збігався м.н., достатньо, щоб для деякої сталої $c > 0$, і необхідно, щоб для всіх $c > 0$ збігались три числових ряди:*

$$\text{(а)}\quad \sum_{n\ge1}\mathbf{P}(|\xi_n| > c), \qquad \text{(б)}\quad \sum_{n\ge1}\mathbf{E}\xi_n^c, \qquad \text{(в)}\quad \sum_{n\ge1}\mathbf{D}\xi_n^c,$$

$$\text{де } \xi_n^c \equiv \xi_n 1\!\!\mathrm{I}_{\{|\xi_n|\le c\}}.$$

Доведення. Зауважимо, що випадкові величини ξ_n^c інтегровні у квадраті внаслідок своєї обмеженості: $|\xi_n^c| \le c$.

Достатність. За теоремою *про перетворення незалежних величин* величини ξ_n^c незалежні. Тому внаслідок (б),(в) для величин ξ_n^c виконані умови теореми *про два ряди*, (а). Отже, ряд $\sum_{n\ge1}\xi_n^c$ збігається м.н.

Зазначимо, що $\{\xi_n \ne \xi_n^c\} = \{|\xi_n| > c\}$. Тому зі збіжності ряду (а) та з *леми Бореля-Кантеллі*, (а), виводимо, що $\mathbf{P}(\overline{\lim}_{n\to\infty}\{\xi_n \ne \xi_n^c\}) = 0$. Переходом в останній рівності до доповнення отримуємо $\mathbf{P}(\underline{\lim}_{n\to\infty}\{\xi_n = \xi_n^c\}) = 1$, тобто $\xi_n = \xi_n^c$, починаючи з деякого номера n, м.н. Тоді за ознакою порівняння рядів ряд $\sum_{n\ge1}\xi_n$ збігається м.н.

Необхідність. За властивостями збіжних рядів зі збіжності $\sum_{n\geq 1}\xi_n$ м.н. виводимо, що $\xi_n \to^{P1} 0$. Тому $\mathbf{P}(\overline{\lim}_{n\to\infty}\{|\xi_n| > c\}) = 0$ для довільного $c > 0$. Випадкові події $\{|\xi_n| > c\}$ незалежні у сукупності внаслідок незалежності ξ_n. За *лемою Бореля-Кантеллі*, (б), що застосована до вказаних подій, ряд з умови (а) має збігатися, інакше попередня ймовірність верхньої границі подій дорівнювала б одиниці.

Як доведено вище, зі збіжності ряду (а) випливає, що з імовірністю 1 послідовності ξ_n та ξ_n^c збігаються, починаючи з деякого номера. Тому зі збіжності ряду $\sum_{n\geq 1}\xi_n$ м.н. випливає збіжність $\sum_{n\geq 1}\xi_n^c$ м.н. Оскільки величини ξ_n^c незалежні та $|\xi_n^c| \leq c$ за означенням, то з теореми *про два ряди*, (б), виводимо збіжність рядів (б),(в) даної теореми $\boxdot$

Теорема (узагальнена теорема Колмогорова про посилений закон великих чисел). *Нехай випадкові величини* $(\xi_n, n \geq 1)$ *незалежні у сукупності,* $\mathbf{E}\xi_n^2 < \infty$, $S_n = \xi_1 + ... + \xi_n$, *а числова послідовність* $b_n \uparrow \infty$. *Якщо* $\sum_{n\geq 1}\mathbf{D}\xi_n/b_n^2 < \infty$, *то*

$$(S_n - \mathbf{E}S_n)/b_n \to^{P1} 0,\ n \to \infty.$$

Зауваження. При виборі $b_n = n$ останнє твердження збігається з твердженням *теореми Колмогорова про посилений закон великих чисел.*

Доведення. Позначимо $\zeta_k \equiv (\xi_k - \mathbf{E}\xi_k)/b_k$. Випадкові величини ζ_k незалежні за теоремою *про перетворення незалежних величин*, та задовольняють умови: $\mathbf{E}\zeta_k = 0$, $\sum_{n\geq 1}\mathbf{D}\zeta_n = \sum_{n\geq 1}\mathbf{D}\xi_n/b_n^2 < \infty$. Тому за теоремою *про один ряд*, (а), ряд $\sum_{n\geq 1}\zeta_n$ збігається м.н.

Позначимо $T_n = \sum_{k=1}^n \zeta_k$, $T = \lim_{n\to\infty} T_n$. Унаслідок цих означень заміною індексів підсумовування отримуємо

$$b_n^{-1}(S_n - \mathbf{E}S_n) = b_n^{-1}\sum_{k=1}^n b_k\zeta_k = b_n^{-1}\sum_{k=1}^n b_k(T_k - T_{k-1}) =$$
$$b_n^{-1}\sum_{k=1}^n b_kT_k - b_n^{-1}\sum_{k=1}^{n-1} b_{k+1}T_k =$$
$$T_n - b_n^{-1}\sum_{k=1}^{n-1}(b_{k+1} - b_k)T_k \to T - T = 0,\ n \to \infty,$$

за теоремою Теплиця, оскільки лінійне перетворення збіжної послідовності T_n з матрицею $(b_n^{-1}(b_{k+1} - b_k), k = \overline{1, n-1})$ зберігає границю $\boxdot$

Вправи

(1) Випадкові величини $(\xi_n, n \geq 1)$ незалежні. Довести, що умови збіжності ряду $\sum_{n\geq 1}\xi_n$ (а) м.н., (б) за ймовірністю, (в) слабкої, еквівалентні.

(2) Випадкові величини $\varkappa_n$ незалежні та $\mathbf{P}(\varkappa_n = 0) = \mathbf{P}(\varkappa_n = 1) = 1/2$. (а) Довести, що ряд $\sum_{n\geq 1} 2^{-n}\varkappa_n = \eta$ збігається м.н., а його сума η рівномірно розподілена на $[0, 1]$. (б) Функція розподілу суми ряду $\zeta = \sum_{n\geq 1} 3^{-n}\varkappa_n$ неперервна, але не є абсолютно неперервною.

(3) Випадкові величини (ξ_n) незалежні, інтегровні і $\mathbf{E}\xi_n = 0$. Довести, що ряд $\sum_{n\geq 1} \xi_n$ збігається м.н. тоді й тільки тоді, коли $\sum_{n\geq 1} \mathbf{E}\xi_n^2/(1+|\xi_n|) < \infty$.

(4) Випадкові величини $(\xi_n n \geq 1)$ незалежні та $\mathbf{P}(\xi_n = 0) = 1 - 2^{-2n}$, $\mathbf{P}(\xi_n = n2^n) = \mathbf{P}(\xi_n = -n2^n) = 2^{-2n-1}$. Довести, що $\sum_{n\geq 1} n^{-2}\mathbf{D}\xi_n = \infty$, однак для послідовності (ξ_n) виконується посилений закон великих чисел.

(5) Випадкові величини $(\xi_n n \geq 1)$ незалежні. Знайти у найпростішому вигляді необхідні та достатні умови збіжності майже напевне ряду $\sum_{n\geq 1} \xi_n$, якщо: (а) $\xi_n \simeq Exp(\lambda_n)$, (б) $\xi_n \simeq \Pi(\lambda_n)$, (в) $\xi_n \simeq N(\mu_n, \sigma_n^2)$, (г) $\xi_n \simeq \Gamma(\lambda_n, \alpha_n)$, (д) $\xi_n = c_n \eta_n$, де η_n мають розподіл Коші.

(6) Випадкові величини (ξ_n) незалежні, інтегровні і $\mathbf{E}\xi_n = 0$. Довести, що для справедливості закону великих чисел: $(\xi_1 + ... + \xi_n)/n \to^P 0, n \to \infty$, необхідно і достатньо, щоб при $n \to \infty$: (а) $\sum_{k=1}^n \mathbf{P}(|\xi_k| \geq n) = o(n)$, (б) $\sum_{k=1}^n \mathbf{E}\xi_k^n = o(n)$, (в) $\sum_{k=1}^n \mathbf{D}\xi_k^n = o(n^2)$, де $\xi_k^n = \xi_k \mathbb{I}_{\{|\xi_k|<n\}}$.

2.7. Узагальнення посиленого закону великих чисел

Посилений закон великих чисел для незалежних однаково розподілених величин можна узагальнити за рахунок прискорення швидкості збіжності та відмови від умови незалежності.

2.7.1. Закон повторного логарифму

Розглянемо послідовність незалежних однаково розподілених величин (ξ_n) з нульовим середнім: $\mathbf{E}\xi_1 = 0$ та скінченною дисперсією: $\mathbf{E}\xi_1^2 < \infty$.

Згідно з *посиленим законом великих чисел* $S_n/n \to^{P1} 0, n \to \infty$. Нормуючий знаменник n у цьому співвідношенні можна скоригувати на підставі *узагальненої теореми Колмогорова про посилений закон великих чисел:* унаслідок збіжності ряду $\sum_{n\geq 1} 1/(n \ln^2(n))$ виводимо збіжність м.н. $S_n/\sqrt{n}\ln n \to^{P1} 0$, $n \to \infty$. З іншого боку, з *класичної центральної граничної теореми* випливає, що збіжність $S_n/\sqrt{n} \to^{P1} 0, n \to \infty$, не може виконуватись, оскільки існує ненульова слабка границя. Звідси виникає запитання: який має бути нормуючий знаменник для сум S_n, щоб границя м.н. була б ненульовою. Відповідь є у наступних твердженнях.

Означення. *Числова послідовність $\overline{\varphi}_n$ є верхньою для випадкової послідовності S_n, якщо $\mathbf{P}(\underline{\lim}_{n\to\infty}\{S_n \leq \overline{\varphi}_n\}) = 1$. Послідовність $\underline{\varphi}_n$ є нижньою для S_n, якщо $\mathbf{P}(\overline{\lim}_{n\to\infty}\{S_n \geq \underline{\varphi}_n\}) = 1$.*

Таким чином, асимптотичні властивості сум S_n при $n \to \infty$ відображаються смугою $([\underline{\varphi}_n, \overline{\varphi}_n], n \ge 1)$, що її нижній край ці суми перетинають нескінченно часто, а верхній не перетинають, починаючи з деякого номера, і все це з імовірністю 1.

Теорема (про критерій для верхньої та нижньої послідовності). *Нехай* φ_n *– додатна числова послідовність.*

(а) Послідовність $(1+\varepsilon)\varphi_n$ *є верхньою для* (S_n) *при кожному* $\varepsilon > 0$ *тоді й тільки тоді, коли* $\mathbf{P}(\overline{\lim}_{n\to\infty} S_n/\varphi_n \le 1) = 1$.

(б) Послідовність $(1-\varepsilon)\varphi_n$ *є нижньою для* (S_n) *при всіх* $\varepsilon \in (0,1)$ *тоді й тільки тоді, коли* $\mathbf{P}(\overline{\lim}_{n\to\infty} S_n/\varphi_n \ge 1) = 1$.

Доведення

(а) *Достатність* є наслідком включення

$$\{\overline{\lim}_{n\to\infty} S_n/\varphi_n \le 1\} \subset \underline{\lim}_{n\to\infty}\{S_n \le (1+\varepsilon)\varphi_n\},$$

звідки виводимо, що ймовірність правої частина дорівнює одиниці.

Необхідність випливає зі включення

$$\underline{\lim}_{n\to\infty}\{S_n \le (1+\varepsilon)\varphi_n\} \subset \{\overline{\lim}_{n\to\infty} S_n/\varphi_n \le 1+\varepsilon\},$$

внаслідок якого $\overline{\lim}_{n\to\infty} S_n/\varphi_n \le 1+\varepsilon$ м.н. для кожного $\varepsilon > 0$, звідки отримуємо $\overline{\lim}_{n\to\infty} S_n/\varphi_n \le 1$ м.н.

(б) *Достатність* випливає зі включення

$$\{\overline{\lim}_{n\to\infty} S_n/\varphi_n \ge 1\} \subset \overline{\lim}_{n\to\infty}\{S_n \ge (1-\varepsilon)\varphi_n\},$$

а *необхідність* – зі включення

$$\overline{\lim}_{n\to\infty}\{S_n \ge (1-\varepsilon)\varphi_n\} \subset \{\overline{\lim}_{n\to\infty} S_n/\varphi_n \ge 1-\varepsilon\} \boxdot$$

З останньої теореми можна зробити висновок, що асимптотичні м.н. властивості послідовності сум S_n визначаються значенням верхньої границі $\overline{\lim}_{n\to\infty} S_n/\varphi_n$. Якщо S_n – суми незалежних випадкових величин, а $\varphi_n \to \infty$, то дана границя вимірна відносно залишкової сигма-алгебри і за теоремою *про закон нуля та одиниці Колмогорова* є м.н. сталою.

Наступна теорема дає умову для скінченності цієї сталої.

Теорема (про закон повторного логарифму). *Нехай випадкові величини* $(\xi_n, n \ge 1)$ *незалежні та однаково розподілені,* $\mathbf{E}\xi_1 = 0$, $D\xi_1 = \sigma^2 < \infty$, $S_n = \xi_1 + \ldots + \xi_n$.

Тоді для послідовності $\varphi_n \equiv \sqrt{2\sigma^2 n \ln\ln n}, n \ge 3$, *має місце рівність*

$$\overline{\lim}_{n\to\infty} S_n/\varphi_n = 1 \text{ м.н.}$$

Зауваження. За теоремою *про критерій для верхньої та нижньої послідовності* для кожного $\varepsilon > 0$ послідовність $(1+\varepsilon)\varphi_n$ є верхньою, а $(1-\varepsilon)\varphi_n$ – нижньою для S_n.

Доведення проведемо у припущенні, що величини ξ_n нормально розподілені. Загальний випадок можна звести до сформульованого шляхом використання *класичної центральної граничної теореми*. Крім того, внаслідок однорідності відношення S_n/φ_n можна вважати, що $\sigma = 1$.

Лема 1. *Нехай величини* $(\xi_n, n \geq 1)$ *незалежні та мають симетричні розподіли, тобто* $\xi_n \simeq -\xi_n, n \geq 1$, $S_n = \xi_1 + ... + \xi_n$. *Тоді для всіх* ε

$$\mathbf{P}(\max_{k\leq n} S_k > \varepsilon) \leq 2\mathbf{P}(S_n > \varepsilon).$$

Доведення. Розглянемо події $A_k = \{S_1 \leq \varepsilon, ...S_{k-1} \leq \varepsilon, S_k > \varepsilon\}$. Подію $A = \{\max_{k\leq n} S_k > \varepsilon\}$ можна зобразити у вигляді $A = \cup_{k=1}^n A_k$, де події A_k є *попарно несумісними*. Оскільки $\{S_n > \varepsilon\} \subset A$, то

$$\mathbf{P}(S_n > \varepsilon) = \mathbf{P}(\{S_n > \varepsilon\} \cap A) = \sum_{k=1}^n \mathbf{P}(\{S_n > \varepsilon\} \cap A_k) \geq$$
$$\sum_{k=1}^n \mathbf{P}(\{S_n - S_k \geq 0\} \cap A_k) = \sum_{k=1}^n \mathbf{P}(S_n - S_k \geq 0)\mathbf{P}(A_k) \geq$$
$$\sum_{k=1}^n \tfrac{1}{2}\ \mathbf{P}(A_k) = \tfrac{1}{2}\ \mathbf{P}(A),$$

де враховано також включення $\{S_n > \varepsilon\} \cap A_k \supset \{S_n - S_k \geq 0\} \cap A_k$, незалежність випадкових подій $\{S_n - S_k \geq 0\} = \{\xi_{k+1} + ... + \xi_n \geq 0\}$ і A_k за теоремою *про векторні перетворення незалежних величин*, та симетричність розподілу суми $S_n - S_k = \xi_{k+1} + ... + \xi_n$, внаслідок якої $\mathbf{P}(S_n - S_k \geq 0) \geq \frac{1}{2}$. Вказана симетричність випливає з незалежності та симетричності ξ_k ⊡

Лема 2. *Нехай* $\zeta \simeq N(0,1)$ *– стандартна нормальна величина. Тоді*

$$\mathbf{P}(\zeta > t) \sim (\sqrt{2\pi}t)^{-1} \exp(-t^2/2),\ t \to \infty.$$

Доведення є простим наслідком правила Лопіталя, що застосовується до відношення лівої та правої частин даного співвідношення ⊡

Для **доведення** теореми *про закон повторного логарифму* за теоремою *про критерій для верхньої та нижньої послідовності* досить довести, що послідовність $a\varphi_n$ є верхньою при $a > 1$ і нижньою при $a < 1$.

(a) Доведемо, що послідовність $a\varphi_n$ є верхньою при $a > 1$.

Розглянемо подію $A = \overline{\lim}_{n\to\infty}\{S_n > a\varphi_n\}$. Шукане твердження полягає в тому, що $\mathbf{P}(A) = 0$ для всіх $a > 1$.

Оберемо довільне $a > 1$. Визначимо натуральні $n_0 = 0, n_k = [a^k] + 1$ та випадкові події $A_k = \cup_{n_k < n \leq n_{k+1}}\{S_n > a\varphi_n\}$. Оскільки $\mathbb{N} = \cup_{k\geq 0}(n_k, n_{k+1}]$, де доданки попарно несумісні, то з означення верхньої границі послідовності подій виводимо, що $A = \overline{\lim}_{k\to\infty} A_k$.

Тому для доведення рівності $\mathbf{P}(A) = 0$ за *лемою Бореля-Кантеллі*, (а), досить встановити збіжність ряду: $\sum_{k\geq 0}\mathbf{P}(A_k) < \infty$. Оцінимо, враховуючи монотонність послідовності φ_n та лему 1 (симетричність центрованого нормального розподілу є наслідком парності нормальної щільності):

$$\mathbf{P}(A_k) \leq \mathbf{P}(\cup_{n_k<n\leq n_{k+1}}\{S_n > a\varphi_{n_k}\}) = \mathbf{P}(\max_{n_k<n\leq n_{k+1}} S_n > a\varphi_{n_k}) \leq$$
$$\mathbf{P}(\max_{0<n\leq n_{k+1}} S_n > a\varphi_{n_k}) \leq 2\mathbf{P}(S_{n_{k+1}} > a\varphi_{n_k}) = 2\mathbf{P}(\zeta_k > t_k),$$

де $\zeta_k \equiv S_{n_{k+1}}/\sqrt{n_{k+1}}$, $t_k \equiv a\varphi_{n_k}/\sqrt{n_{k+1}}$. За означенням номерів n_k

$$t_k \equiv a\sqrt{2n_k \ln\ln n_k}/\sqrt{n_{k+1}} \geq$$
$$a\sqrt{2a^k \ln\ln a^k}/\sqrt{a^{k+1}+1} = \sqrt{2a(1+a^{-k-1})^{-1}\ln(k\ln a)},\ \ k > 3.$$

За теоремою *про нормальність суми незалежних нормальних векторів* та теоремою *про інтерпретацію параметрів нормального розподілу* $S_n \simeq N(0,n)$, оскільки $\mathbf{E}S_n = 0$ і $\mathbf{D}S_n = n$ за теоремою *про дисперсію суми незалежних величин*. Тому величина $\zeta_k \simeq N(0,1)$ та внаслідок отриманої нерівності для $\mathbf{P}(A_k)$ і леми 2

$$\mathbf{P}(A_k) \leq 2\mathbf{P}(\zeta_k > t_k) \sim 2(\sqrt{2\pi}t_k)^{-1}\exp(-t_k^2/2) \leq$$
$$b(\ln k)^{-1/2}\exp(-2a(1+a^{-k-1})^{-1}\ln(k\ln a)/2) \sim c(\ln k)^{-1/2}k^{-a},\ \ k\to\infty.$$

Оскільки $a > 1$, то ряд, що складений доданками з правої частини, збігається. Тому з умови порівняння рядів виводимо збіжність ряду, складеного з $\mathbf{P}(A_k)$, що доводить твердження (а), як показано вище.

(б) Доведемо, що послідовність $a\varphi_n$ з $a < 1$ є нижньою. Для цього досить перевірити, що подія $A \equiv \overline{\lim}_{n\to\infty}\{S_n \geq a\varphi_n\}$ справджується *м.н.*

Розглянемо послідовність $(-S_n n \geq 1)$, що утворена сумами стандартних нормальних випадкових величин $(-\xi_k, k = \overline{1,n})$. Визначимо таку подію: $B = \underline{\lim}_{n\to\infty}\{S_n > -2\varphi_n\}$. Оскільки розподіл S_n є симетричним: $-S_n \simeq S_n$, то доповнення $\overline{B} = \overline{\lim}_{n\to\infty}\{-S_n \geq 2\varphi_n\}$ має нульову ймовірність за вже доведеним твердженням (а), отже, $\mathbf{P}(B) = 1$.

Оберемо $b > 1$ та визначимо номери $n_k = [b^k]$. Розглянемо випадкові величини $\eta_k \equiv S_{n_k} - S_{n_{k-1}} = \sum_{n_{k-1}<i\leq n_k}\xi_i$. Інтервали $(n_{k-1}, n_k], k \geq 1$, попарно не перетинаються, тому за теоремою *про векторні перетворення незалежних величин* величини $(\eta_k, k \geq 1)$ незалежні. Як і вище, з теореми *про нормальність суми незалежних нормальних векторів* знаходимо розподіл: $\eta_k \simeq N(0, n_k - n_{k-1})$, оскільки за теоремою *про дисперсію суми незалежних величин* $\mathbf{D}\eta_k = n_k - n_{k-1}$.

Позначимо сталі $c_k \equiv a\varphi_{n_k} + 2\varphi_{n_{k-1}}$ та такі події $C_k = \{\eta_k \geq c_k\}$, $C = \overline{\lim}_{k\to\infty} C_k$. Оскільки величини η_k незалежні, то події C_k також незалежні. Тому до цих подій можна застосувати *лему Бореля-Кантеллі*, (б), за якої $\mathbf{P}(C) = 1$, якщо тільки розбігається ряд $\sum \mathbf{P}(C_k)$. Для доведення цієї розбіжності застосуємо лему 2:

$$\mathbf{P}(C_k) = \mathbf{P}(\eta_k \geq c_k) = \mathbf{P}(\zeta_k \geq t_k) \sim (\sqrt{2\pi} t_k)^{-1} \exp(-t_k^2/2), k \to \infty,$$

де величини $\zeta_k \equiv \eta_k/\sqrt{n_k - n_{k-1}} \simeq N(0,1)$ є *стандартними нормальними,* а права частина останнього співвідношення спадає за t_k, та врахуємо такі співвідношення при $k \to \infty$:

$$t_k \equiv c_k/\sqrt{n_k - n_{k-1}} \leq \sqrt{2 \ln\ln n_k} \left(a\sqrt{n_k} + 2\sqrt{n_{k-1}}\right)/\sqrt{n_k - n_{k-1}} \leq$$
$$\sqrt{2\ln k \ln b} \left(ab^{k/2} + 2b^{(k-1)/2}\right)/b^{(k-1)/2}\sqrt{b - 1 + b^{-k+1}} =$$
$$\rho_b \sqrt{2\ln k + 2\ln\ln b} + O(1), k \to \infty.$$

Сталу $\rho_b \equiv (a\sqrt{b} + 2)/\sqrt{b-1}$ вибором досить великого $b > 1$ можна зробити меншою за одиницю: $\rho_b < 1$, оскільки $\rho_b \to a < 1$ при $b \to \infty$. Отже, внаслідок отриманих вище зображень для $\mathbf{P}(C_k)$ та t_k

$$\textstyle\sum_{k\geq 1} \mathbf{P}(C_k) \geq \sum_{k>1} C(\ln k)^{-1/2} \exp(-\rho_b^2(2\ln k + 2\ln\ln b)/2) =$$
$$\textstyle\sum_{k>1} D(\ln k)^{-1/2} k^{-\rho_b^2} = \infty,$$

оскільки $\rho_b^2 < 1$. Тому, як вказано вище, $\mathbf{P}(C) = 1$.

За означенням B і C, для кожної елементарної події $\omega \in B$, починаючи з деякого номера n, виконується нерівність $S_n > -2\varphi_n$, а для всіх $\omega \in C$ нерівності: $\eta_k \geq c_k = a\varphi_{n_k} + 2\varphi_{n_{k-1}}$ справедливі для нескінченої послідовності номерів (n_k). Якщо $\omega \in B \cap C$, то починаючи з деякого номера k, $S_{n_{k-1}} > -2\varphi_{n_{k-1}}$ звідки $S_{n_k} = \eta_k + S_{n_{k-1}} \geq S_{n_{k-1}} + c_k \geq -2\varphi_{n_{k-1}} + c_k = a\varphi_{n_k}$ для нескінченної кількості номерів n_k. Тому $B \cap C \subset \overline{\lim}_{n\to\infty}\{S_n \geq a\varphi_n\} = A$, і з доведених рівностей $\mathbf{P}(B) = 1, \mathbf{P}(C) = 1$ виводимо $\mathbf{P}(A) = 1$, що і доводить (б) $\boxdot$

Вправи

(1, симетризація) Випадкові величини $(\xi_k, k = \overline{0,2})$ незалежні однаково розподілені, і мають медіану m, тобто $\mathbf{P}(\xi_0 < m) = \mathbf{P}(\xi_0 > m) = 1/2$. Довести при $|t| \geq m$ нерівність $\mathbf{P}(|\xi_0| \geq t) \leq 2\mathbf{P}(|\xi_1 - \xi_2| \geq t - |m|)$.

(2) Величини $(\xi_k, k \geq 1)$ незалежні, $\mathbf{P}(\xi_k = \pm 1) = 1/2$, $S_k = \xi_1 + ... + \xi_k$. Довести, що $\mathbf{P}(\max_{1\leq k\leq n} S_k \geq r) = 2\mathbf{P}(S_n > r) + \mathbf{P}(S_n = r)$ при $r \in \mathbb{Z}_+$.

(3) Випадкові величини $(\xi_n, n \geq 1)$ незалежні та мають стандартний нормальний розподіл. Довести такі зображення: (а) $\xi_n = O(\sqrt{\ln n}), n \to \infty$, м.н., (б) $\mathbf{P}(\xi_n = o(\sqrt{\ln n}), n \to \infty) = 0$, (в) $\mathbf{P}(\overline{\lim}_{n\to\infty}\xi_n/\sqrt{2\ln n} = 1) = 1$.

(4) Випадкові величини $(\xi_n n \geq 1)$ незалежні однаково розподілені, $\mathbf{E}\xi_1 = 0$, $\mathbf{E}\xi_1^2 = 1$, а $S_n = \xi_1 + ... + \xi_n$. Довести, що $\overline{\lim}_{n\to\infty}S_n/\sqrt{n} = \infty$ м.н.

(5) Випадкові величини (ξ_n) незалежні та мають розподіл Пуассона $\Pi(\lambda)$. Довести, що $\overline{\lim}_{n\to\infty}(\xi_n \ln\ln n)/\ln n = 1$ м.н.

(6) Випадкові величини (ξ_n) незалежні та мають показниковий розподіл $Exp(1)$. Довести граничні співвідношення: (а) $\lim_{n\to\infty}\max_{1\leq k\leq n}(\xi_k/\ln k) = 1$ м.н., (б) $\overline{\lim}_{n\to\infty}\xi_n/\ln n \leq 1$ м.н., (в) $\underline{\lim}_{n\to\infty}\max_{\sqrt{n}\leq k\leq n}(\xi_k/2k) \geq 1$ м.н.

(7) Випадкові величини (ξ_n) незалежні однаково розподілені Довести еквівалентність: $\mathbf{E}\sup_{n\geq 1}|\xi_n/n| < \infty$ тоді й тільки тоді, коли $\mathbf{E}\,|\xi_1|\,(\ln|\xi_1|)^+ < \infty$.

(8, теорема Леві) Випадкові величини (ξ_n) незалежні та мають розподіл Бернуллі $\mathbf{P}(\xi_n = \pm 1) = 1/2$. Довести, що послідовність $\varphi_n = 2n\ln\ln n + c\ln\ln\ln n$ є верхньою при $c > 3$, та нижньою при $c \leq 1$.

2.7.2. Строго стаціонарні послідовності

Посилений закон великих чисел можна узагальнити за рахунок послаблення умови незалежності доданків.

Означення. *Послідовність випадкових величин* $(\xi_n, n \geq 1)$ *називається* строго стаціонарною, *якщо для всіх* $n \geq 1$ *та всіх* $k \geq 1$

$$\mathbf{P}((\xi_1, ..., \xi_n) \in B_n) = \mathbf{P}((\xi_{k+1}, ..., \xi_{k+n}) \in B_n),\ \forall B_n \in \mathfrak{B}(\mathbb{R}^n).$$

Зауваження. Унаслідок довільності n для справедливості останньої умови досить перевірити її при $k = 1$. Дійсно, для всіх $k > 1, B_n \in \mathfrak{B}(\mathbb{R}^n)$

$$\mathbf{P}((\xi_{k+1}, ..., \xi_{k+n}) \in B_n) = \mathbf{P}((\xi_2, ..., \xi_{k+1}, ..., \xi_{k+n}) \in \mathbb{R}^{k-1} \times B_n) =$$
$$\mathbf{P}((\xi_1, ..., \xi_k, ..., \xi_{k+n-1}) \in \mathbb{R}^{k-1} \times B_n) = \mathbf{P}((\xi_k, ..., \xi_{k+n-1}) \in B_n),$$

звідки за індукцією отримуємо рівності для всіх $k \geq 1$.

Зауваження. За побудовою *міри Лебега – Стілтьєса, що породжена сумісною функцією розподілу,* та з урахуванням останнього зауваження, послідовність $(\xi_n, n \geq 1)$ є строго стаціонарною тоді й тільки тоді, коли при зсуві номерів на 1 не змінюються *сумісні функції розподілу*:

$$F_{\xi_1,...,\xi_n}(x_1, ..., x_n) = F_{\xi_2,...,\xi_{n+1}}(x_1, ..., x_n),\ \forall x_k \in \mathbb{R},\ \forall n \geq 1.$$

Приклади

(1) Послідовність незалежних однаково розподілених випадкових величин $(\zeta_n, n \geq 1)$ є *строго стаціонарною*..Дійсно, *сумісна функція розподілу*

вектора $(\zeta_{k+1}, ..., \zeta_{k+n})$ дорівнює n-кратному добутку функції розподілу F_{ζ_1} і не залежить від k.

(2) Нехай $(\zeta_n, n \geq 1)$ – послідовність незалежних однаково розподілених величин, а $g : \mathbb{R}^m \to \mathbb{R}$ довільна борелєва функція. Тоді послідовність $\xi_n = g(\zeta_{n+1}, ..., \zeta_{n+m})$ є *строго стаціонарною*. Дійсно, *сумісна функція розподілу* вектора $(\xi_{k+1}, ..., \xi_{k+n})$ визначається сумісною функцією розподілу вектора $(\zeta_{k+2}, ..., \zeta_{k+n+m})$, що не залежить від k згідно з (1).

Вправа. Твердження попереднього прикладу залишається справедливим у випадку, коли $(\zeta_n, n \geq 1)$ – довільна строго стаціонарна послідовність.

Наведені приклади свідчать, що клас строго стаціонарних послідовностей є суттєвим узагальненням класу незалежних однаково розподілених величин, зокрема, містить також послідовності залежних величин.

Оскільки послідовність $\xi \equiv (\xi_n, n \geq 1)$ – *випадковий елемент* у вимірному просторі $(\mathbb{R}^\infty, \mathfrak{B}(\mathbb{R}^\infty))$ то відповідна *породжена сигма-алгебра* має вигляд $\sigma[\xi] = \{\{\xi \in B\}, B \in \mathfrak{B}(\mathbb{R}^\infty)\}$.

Означення. Оператором зсуву *на просторі* $\mathbb{R}^\infty$ *називається відображення* $\theta x = (x_{n+1}, n \geq 1) : \mathbb{R}^\infty \to \mathbb{R}^\infty$ *для точок* $x = (x_n, n \geq 1) \in \mathbb{R}^\infty$.

Теорема (про критерій строгої стаціонарності). *Послідовність випадкових величин* $\xi = (\xi_n)$ *строго стаціонарна тоді й тільки тоді, коли*

$$\mathbf{P}(\xi \in B) = \mathbf{P}(\theta\xi \in B), \ \forall B \in \mathfrak{B}(\mathbb{R}^\infty).$$

Доведення. *Достатність* отримуємо, враховуючи перше зауваження, підстановкою *циліндричних множин* $B = J_n(B_n), \ B_n \in \mathfrak{B}(\mathbb{R}^n)$.

Необхідність. Для циліндричної множини $B = J_n(B_n)$ умова теореми виконується за означенням. Оскільки ж обидві частини наведеної рівності є сигма-адитивними мірами за B, а борелєва сигма-алгебра $\mathfrak{B}(\mathbb{R}^\infty)$ породжується *класом циліндричних множин*, то за *теоремою Каратеодорі про продовження міри* ця рівність виконується на $\mathfrak{B}(\mathbb{R}^\infty)$ $\boxdot$

Означення. *Множина* $B \in \mathfrak{B}(\mathbb{R}^\infty)$ *називається* інваріантною, *якщо* $B = \theta^{(-1)}(B)$. *Клас усіх інваріантних множин позначимо через* $\mathfrak{I}(\mathbb{R}^\infty)$.

З властивостей прообразу $\theta^{(-1)}$ виводимо, що клас інваріантних множин $\mathfrak{I}(\mathbb{R}^\infty)$ є *сигма-алгеброю*.

Приклади. Множини $\{x : \exists \lim_{n\to\infty} x_n\}$, $\{x : \overline{\lim}(x_1 + ... + x_n)/n > 0\}$, $\{x : \sum_{n\geq 1} |x_n| < \infty\}$ є інваріантними.

Означення. *Випадкова подія* $A \in \sigma[\xi]$ *називається* інваріантною подією, *якщо* $A = \{\xi \in B\}$, *де* $B \in \mathfrak{I}(\mathbb{R}^\infty)$. *Клас всіх інваріантних подій з* $\sigma[\xi]$ *є сигма-алгеброю як прообраз* $\mathfrak{I}(\mathbb{R}^\infty)$ *і позначається через* $\mathfrak{I}[\xi]$.

Означення. *Випадкова величина ζ називається* інваріантною величиною*, якщо вона вимірна відносно сигма-алгебри $\mathfrak{I}[\xi]$.*

Лема (про перетворення строго стаціонарної послідовності). *Нехай $\xi = (\xi_n, n \geq 1)$ – строго стаціонарна послідовність, $A \in \mathfrak{I}[\xi]$ – інваріантна подія, а $g : \mathbb{R} \to \mathbb{R}$ – борелєва функція Тоді строго стаціонарною є послідовність суперпозицій*

$$\xi^* = g(\xi)1\!\!1_A \equiv (g(\xi_n)1\!\!1_A, n \geq 1).$$

Доведення. За означенням інваріантних подій $A = \{\xi \in B\}$, де $B \in \mathfrak{I}(\mathbb{R}^\infty)$ – *інваріантна* множина. Тому$\{\xi \in B\} = \{\xi \in \theta^{(-1)}(B)\} = \{\theta\xi \in B\}$.

Звідси $\theta\xi^* \equiv g(\theta\xi)1\!\!1_A = g(\theta\xi)1\!\!1_{\{\xi\in B\}} = g(\theta\xi)1\!\!1_{\{\theta\xi\in B\}}$. Для довільної множини $C \in \mathfrak{B}(\mathbb{R}^\infty)$ позначимо $D = \{x : g(x)1\!\!1_{x\in B} \in C\} \in \mathfrak{B}(\mathbb{R}^\infty)$. Тоді $\{\xi^* \in C\} = \{\xi \in D\}$ і внаслідок строгої стаціонарності ξ

$$\mathbf{P}(\theta\xi^* \in C) = \mathbf{P}(g(\theta\xi)1\!\!1_{\{\theta\xi\in B\}} \in C) = \mathbf{P}(\theta\xi \in D) = \mathbf{P}(\xi \in D) = \mathbf{P}(\xi^* \in C),$$

Шукане випливає з теореми *про критерій строгої стаціонарності* $\boxdot$

Теорема Біркгофа – Хінчина

Наступна теорема є певним узагальненням *критерію Колмогорова посиленого закону великих чисел.*

Теорема (теорема Біркгофа – Хінчина). *Нехай $\xi = (\xi_n, n \geq 1)$ є строго стаціонарною послідовністю та інтегровна: $\mathbf{E}\,|\xi_1| < \infty$. Тоді для послідовних сум $S_n = \xi_1 + ... + \xi_n$ має місце збіжність*

$$S_n \;/\; n \to^{P1} \mathbf{E}(\xi_1 \mid \mathfrak{I}[\xi]), n \to \infty.$$

Доведення спирається на таку теорему.

Теорема (максимальна ергодична теорема Гарсіа). *Нехай виконуються умови теореми Біркгофа – Хінчина. Визначимо випадкову подію $C_n \equiv \{\max_{1\leq k\leq n} S_k > 0\}$. Тоді $\mathbf{E}\xi_1 1\!\!1_{C_n} \geq 0,\ \forall n \geq 1$.*

Доведення теореми Гарсіа.

Позначимо $\mu_n = \max_{0\leq k\leq n} S_k$, де $S_0 = 0$. Тоді $C_n = \{\mu_n > 0\}$, оскільки при виконанні C_n максимум в означенні μ_n не може досягатись при $k = 0$.

Розглянемо одночасно з ξ послідовність $\xi' \equiv \theta\xi = (\xi_{n+1}, n \geq 1)$. Позначимо через S'_n, μ'_n аналогічні до S_n, μ_n величини, що будуються за послідовністю ξ'. Унаслідок *строгої стаціонарності* та теореми *про обчислення ймовірностей, пов'язаних із випадковим вектором*, випадкові вектори $(S_k, k = \overline{1,n})$ та $(S'_k, k = \overline{1,n})$ однаково розподілені. Звідси випливає, що функції розподілу величин μ_n та μ'_n збігаються, зокрема, $\mathbf{E}\mu_n = \mathbf{E}\mu'_n$. Інтегровність всіх розглянутих величин є наслідком умови інтегровності ξ_1 та теореми *про обчислення математичного сподівання функції від випадкової величини*, оскільки за властивістю строгої стаціонарності всі величини ξ_k мають однакові функції розподілу.

За означенням справедливі співвідношення

$$(\xi_1 + \mu'_n)1\!\!1_{C_n} = (\xi_1 + \max_{0\le k\le n} S'_k)\, 1\!\!1_{C_n} = (\max_{0\le k\le n}(\xi_1 + S'_k))1\!\!1_{C_n} =$$
$$(\max_{0\le k\le n} S_{k+1})1\!\!1_{C_n} \ge (\max_{1\le k\le n} S_k)1\!\!1_{C_n} = \mu_n 1\!\!1_{C_n},$$

де остання рівність обґрунтована на початку доведення.

З доведеної нерівності за монотонністю та лінійністю математичного сподівання отримуємо

$$\mathbf{E}\xi_1 1\!\!1_{C_n} \ge \mathbf{E}\mu_n 1\!\!1_{C_n} - \mathbf{E}\mu'_n 1\!\!1_{C_n} = \mathbf{E}(\mu_n - \mu'_n)1\!\!1_{\{\mu_n>0\}} =$$
$$\mathbf{E}(\mu_n - \mu'_n 1\!\!1_{\{\mu_n>0\}}) \ge \mathbf{E}(\mu_n - \mu'_n) = 0,$$

де передостання нерівність є наслідком невід'ємності μ'_n, а остання рівність встановлена вище $\boxdot$

Доведення теореми. Розглянемо при $x \in \mathbb{R}$ події

$$C_x \equiv \{\overline{\lim}_{n\to\infty} S_n/n > x\}.$$

Оскільки верхня границя середніх арифметичних від числової послідовності не змінюються при її зсуві, то $\theta^{(-1)}C_x = C_x$, звідки виводимо *інваріантність події:* $C_x \in \mathfrak{I}[\xi]$.

Нехай A – довільна інваріантна подія: $A \in \mathfrak{I}[\xi]$. Розглянемо послідовність випадкових величин

$$\xi^* = (\xi - x)1\!\!1_{A\cap C_x} \equiv ((\xi_n - x)1\!\!1_{A\cap C_x}, n \ge 1)\,.$$

За лемою *про перетворення строго стаціонарної послідовності* з функцією $g(y) = y - x$ та інваріантною подією $A \cap C_x$ послідовність ξ^* також є *строго стаціонарною.*

Визначимо $\mu^*_n = \max_{0\le k\le n} S^*_k$, $S^*_n = \sum_{k=1}^n \xi^*_k$, $C^*_n = \{\mu^*_n > 0\}$. Тоді

$$C^*_n = A \cap C_x \cap \{\max_{1\le k\le n}(S_k - kx) > 0\} = A \cap C_x \cap \cup_{k=1}^n \{S_k > kx\} =$$
$$A \cap C_x \cap \{\max_{1\le k\le n}(S_k/k) > x\} \uparrow A \cap C_x \cap \{\sup_{k\ge 1}(S_k/k) > x\} = A \cap C_x,$$

при $n \to \infty$, де остання рівність випливає з означення C_x.

З *максимальної ергодичної теореми Гарсіа,* що застосована до послідовності ξ^* та події $C^*_n = \{\mu^*_n > 0\}$, виводимо нерівність $\mathbf{E}\xi^*_1 1\!\!1_{C^*_n} \ge 0$. Оскільки $C^*_n \uparrow A \cap C_x$, то $\xi^*_1 1\!\!1_{C^*_n} \to \xi^*_1 1\!\!1_{A\cap C_x}, n \to \infty$, причому збіжність є мажоровано інтегровною: $|\xi_1^* \mathbb{I}_{A\cap C_x}| \le |\xi_1 - x|$. Тому за *теоремою Лебега про мажоровану збіжність* внаслідок наведеної вище нерівності отримуємо нерівність $\mathbf{E}\xi_1^* \mathbb{I}_{A\cap C_x} = \lim_{n\to\infty} \mathbf{E}\xi_1^* \mathbb{I}_{C_n^*} \ge 0$

Отже, за означенням $\xi_1^* \equiv \xi_1 - x$ має місце нерівність

$$\mathbf{E}\xi_1 \mathbb{I}_{A\cap C_x} \ge x\mathbf{P}(A\cap C_x),\ \forall A \in \mathfrak{I}[\xi].$$

Розглядом послідовність випадкових величин $\xi^* = (y - \xi)\mathbb{I}_{A\cap D_y}$ з інваріантною подією $D_y \equiv \{\underline{\lim}_{n\to\infty} S_n/n < y\}$, аналогічно виводимо, що

$$\mathbf{E}\xi_1 \mathbb{I}_{A\cap D_y} \le y\mathbf{P}(A\cap D_y),\ \forall A \in \mathfrak{I}[\xi].$$

Оберемо $x > y$ та підставимо в першу з наведених нерівностей інваріантну подію $A = D_y$, а у другу – $A = C_x$. У результаті отримаємо такі нерівності:

$$x\mathbf{P}(C_x \cap D_y) \le \mathbf{E}\xi_1 \mathbb{I}_{C_x\cap D_y} \le y\mathbf{P}(C_x\cap D_y).$$

Оскільки $x > y$, з цієї нерівності виводимо, що $\mathbf{P}(C_x \cap D_y) = 0$.

Розглянемо подію $E = \{\exists \lim_{n\to\infty} S_n/n\}$. Її доповнення має вигляд

$$\overline{E} = \{\underline{\lim}_{n\to\infty} S_n/n < \overline{\lim}_{n\to\infty} S_n/n\} = \cup_{y<x,\ x,y\in\mathbb{Q}} D_y \cap C_x,$$

де $\mathbb{Q}$ – множина раціональних чисел. Тому за *напівадитивністю* ймовірності $\mathbf{P}(\overline{E}) = 0$ та $\mathbf{P}(E) = 1$. Оскільки клас *інваріантних подій* є сигма-алгеброю, і $C_x \cap D_y \in \mathfrak{I}[\xi]$, то $E \in \mathfrak{I}[\xi]$.

Крім того, випадкова величина $\underline{\lim}_{n\to\infty} S_n/n$ також є *інваріантною величиною*, як доведено вище. Тому добуток

$$\zeta \equiv \mathbb{I}_E \underline{\lim}_{n\to\infty} S_n/n = \mathbb{I}_E \lim_{n\to\infty} S_n/n$$

є інваріантною випадковою величиною та до того ж визначений коректно для всіх $\omega \in \Omega$, оскільки границя у правій частині існує для всіх $\omega \in E$.

Залишається довести, що $\zeta = \mathbf{E}(\xi_1 \mid \mathfrak{I}[\xi])$ м.н.

Для цього зафіксуємо інваріантну подію: $B \in \mathfrak{I}[\xi]$ та підставимо у отримані вище нерівності $A = B \cap \{\zeta > x\}$ та $A = B\cap\{\zeta < y\}$ відповідно. За означенням ζ події $\{\zeta > x\}$ та C_x відрізняються лише на підмножину $\overline{E}$, що має нульову ймовірність. Тому з першої нерівності виводимо при всіх $B \in \mathfrak{I}[\xi]$

$$\mathbf{E}\xi_1 \mathbb{I}_{B\cap\{\zeta>x\}} = \mathbf{E}\xi_1 \mathbb{I}_{B\cap\{\zeta>x\}\cap C_x} \ge$$
$$x\mathbf{P}(B\cap\{\zeta > x\}\cap C_x) = x\mathbf{P}(B\cap\{\zeta > x\}).$$

Аналогічно з другої нерівності отримуємо

$$\mathbf{E}\xi_1 \mathbb{I}_{B\cap\{\zeta<y\}} = \mathbf{E}\xi_1 \mathbb{I}_{B\cap\{\zeta<y\}\cap D_y} \le$$
$$y\mathbf{P}(B\cap\{\zeta < y\}\cap D_y) = y\mathbf{P}(B\cap\{\zeta < y\}).$$

Нехай $\varepsilon > 0$. Позначимо $\Delta_k = (k\varepsilon, k\varepsilon+\varepsilon]$ при $k \in \mathbb{Z}$. Оскільки величина ζ – інваріантна, з першої вищенаведеної нерівності для $A \in \mathfrak{I}[\xi]$ маємо

$$\mathbf{E}\zeta 1\!\!\mathrm{I}_A = \sum\nolimits_{k\in\mathbb{Z}} \mathbf{E}\zeta 1\!\!\mathrm{I}_{A\cap\{\zeta\in\Delta_k\}} \le \sum\nolimits_{k\in\mathbb{Z}} \mathbf{E}(k\varepsilon+\varepsilon) 1\!\!\mathrm{I}_{A\cap\{\zeta\in\Delta_k\}} =$$

$$\sum\nolimits_{k\in\mathbb{Z}} k\varepsilon \mathbf{P}(A\cap\{\zeta\in\Delta_k\}\cap\{\zeta > k\varepsilon\}) + \varepsilon\mathbf{P}(A) \le$$

$$\sum\nolimits_{k\in\mathbb{Z}} \mathbf{E}\xi_1 1\!\!\mathrm{I}_{A\cap\{\zeta\in\Delta_k\}\cap\{\zeta>k\varepsilon\}} + \varepsilon = \sum\nolimits_{k\in\mathbb{Z}} \mathbf{E}\xi_1 1\!\!\mathrm{I}_{A\cap\{\zeta\in\Delta_k\}} = \mathbf{E}\xi_1 1\!\!\mathrm{I}_A + \varepsilon,$$

де враховано також нерівність $\zeta 1\!\!\mathrm{I}_{\{\zeta\in\Delta_k\}} \le (k\varepsilon+\varepsilon) 1\!\!\mathrm{I}_{\{\zeta\in\Delta_k\}}$ та включення подій $\{\zeta\in\Delta_k\} \subset \{\zeta > k\varepsilon\}$.

Аналогічно, внаслідок вибору $\Delta_k = [k\varepsilon - \varepsilon, k\varepsilon)$ та використання другої з попередніх нерівностей для всіх $A \in \mathfrak{I}[\xi]$ отримуємо

$$\mathbf{E}\zeta 1\!\!\mathrm{I}_A \ge \mathbf{E}\xi_1 1\!\!\mathrm{I}_A - \varepsilon.$$

Враховуючи довільність $\varepsilon > 0$, виводимо з обох останніх нерівностей тотожність $\mathbf{E}\zeta 1\!\!\mathrm{I}_A = \mathbf{E}\xi_1 1\!\!\mathrm{I}_A,\ \forall A \in \mathfrak{I}[\xi]$. Вимірність ζ відносно сигма-алгебри $\mathfrak{I}[\xi]$ встановлена вище. Отже, випадкова величина ζ задовольняє умови означення *умовного математичного сподівання* відносно сигма-алгебри. Тому рівність $\zeta = \mathbf{E}(\xi_1 \mid \mathfrak{I}[\xi])$ м.н. випливає з теореми *про існування та єдиність умовного математичного сподівання* ⊡

Ергодичність

Границя у *теоремі Біркгофа-Хінчина* є випадковою величиною. Випадок, коли ця границя майже напевне є сталою (що збігається з певним математичним сподіванням), можна трактувати так: асимптотичне середнє за часом дорівнює середньому за фазовим простором. Останню властивість називають ергодичною властивістю.

Означення. *Строго стаціонарна послідовність $\xi = (\xi_n, n \ge 1)$ є* ергодичною*, якщо відповідна сигма-алгебра інваріантних подій є виродженою:* $\forall A \in \mathfrak{I}[\xi]\ \ \mathbf{P}(A) \in \{0,1\}$.

Теорема (ергодична теорема для строго стаціонарної послідовності). *Нехай $\xi = (\xi_n, n \ge 1)$ – строго стаціонарна ергодична послідовність, а борелєва функція $g : \mathbb{R} \to \mathbb{R}$ така, що $\mathbf{E}\,|g(\xi_1)| < \infty$. Тоді*

$$n^{-1}\sum\nolimits_{k=1}^{n} g(\xi_k) \to \mathbf{E}g(\xi_1), n \to \infty, \text{ м.н.}$$

Доведення. Розглянемо послідовність $\eta = g(\xi) \equiv (g(\xi_n), n \ge 1)$. За теоремою *про перетворення строго стаціонарної послідовності* η – також строго стаціонарна. Крім того, за означенням *інваріантних подій*

$$\mathfrak{I}[\eta] \equiv \{\{\eta \in B\}, B \in \mathfrak{I}[\mathbb{R}^\infty]\} = \ \{\{\xi \in g^{(-1)}(B)\}, B \in \mathfrak{I}[\mathbb{R}^\infty]\} \subset \mathfrak{I}[\xi],$$

оскільки $g^{(-1)}(B) \in \mathfrak{I}[\mathbb{R}^\infty]$. Тому послідовність η одночасно є ергодичною та інтегровною за умовою теореми. Отже, за *теоремою Біркгофа-Хінчина*

$$n^{-1} \sum\nolimits_{k=1}^{n} \eta_k \to^{P1} \zeta \equiv \mathbf{E}(\eta_1 \mid \mathfrak{I}[\eta]).$$

Останнє умовне математичне сподівання майже напевне збігається зі сталою $\mathbf{E}\eta_1$, оскільки вона вимірна відносно сигма-алгебри $\mathfrak{I}[\eta]$ та внаслідок включення $\mathbf{P}(C) \in \{0,1\}$ задовольняє умову балансу (б) з означення *умовного математичного сподівання*: $\mathbf{E}\eta_1 1\!\!\mathrm{I}_C = \mathbf{E}((\mathbf{E}\eta_1) 1\!\!\mathrm{I}_C), \forall C \in \mathfrak{I}[\eta]$ $\boxdot$

Вправи

(1) Довести, що інваріантні множини належать залишковій сигма-алгебрі: $\mathfrak{I}(\mathbb{R}^\infty) \subset \cap_{n\geq 1}(\mathbb{R}^n \times \mathfrak{B}(\mathbb{R}^\infty))$.

(2) Перевірити, що твердження теореми Біркгофа-Хінчина виконується при розширенні сигма-алгебри $\mathfrak{I}[\xi]$ на клас всіх випадкових подій нульової ймовірності: $\mathfrak{I}[\xi, \mathbf{P}] = \{A \in \mathfrak{F} : \exists B \in \mathfrak{I}[\xi],\ \mathbf{P}(A\Delta B) = 0\}$.

(3) Вивести з теореми про закон нуля та одиниці Колмогорова ергодичність послідовності незалежних випадкових величин.

(4) Довести твердження ергодичної теореми для послідовності випадкових величин $\eta_n = g(\xi_n, \xi_{n+1}, ..., \xi_{n+m-1})$, що утворена строго стаціонарною ергодичною послідовністю ξ та борелєвою функцією $g : \mathbb{R}^m \to \mathbb{R}$.

(5) Для строго стаціонарної послідовності $(\xi_n, n \geq 1)$ виконується при всіх $k, l \geq 1$, $B \in \mathfrak{B}(\mathbb{R}^k)$, $C \in \mathfrak{B}(\mathbb{R}^l)$ така умова перемішування при $n \to \infty$:

$\mathbf{P}((\xi_1, ..\xi_k) \in B, (\xi_{n+1}, ..\xi_{n+l}) \in C) \to \mathbf{P}((\xi_1, ..\xi_k) \in B)\mathbf{P}((\xi_1, ..\xi_l) \in C)$,
Довести, що дана послідовність ергодична. Вивести звідси ергодичність послідовності незалежних однаково розподілених величин.

(6) Для строго стаціонарної послідовності $(\xi_n, n \geq 1)$ залишкова сигма-алгебра $\overline{\lim}_{n\to\infty}\sigma[\xi_n] \equiv \cap_{n\geq 1}\sigma[\xi_k, k \geq n]$ містить лише події нульової чи одиничної ймовірності. Довести, що дана послідовність ергодична.

(7) Строго стаціонарна послідовність $(\xi_n, n \geq 1)$ є гауссовою (тобто будь-яка скінченна її підпослідовність є нормальним вектором), причому $\mathbf{E}\xi_n = 0$ та $\mathbf{E}\xi_n\xi_{n+k} = r(k) \to 0, k \to \infty$. Довести, що ця послідовність є ергодичною.

(8) Навести приклад строго стаціонарної не ергодичної послідовності.

(9) Нехай імовірнісний простір $(\Omega, \mathfrak{F}, \mathbf{P}) = ((0,1], \mathfrak{B}(0,1], \mu)$, де міра μ має вигляд $\mu(B) = (\ln 2)^{-1} \int_B (1+x)^{-1} dx$, функція $a(x) = x^{-1} - [x^{-1}]$ при $x \in (0,1]$. Тоді послідовність n-кратних суперпозицій $\xi_n(x) = [a(a(...a(x)...))]$ є строго стаціонарною ($\xi_n(x)$ є n-м елементом у розкладі x у неперервний дріб).

2.8. Уточнення центральної граничної теореми

Центральну граничну теорему можна уточнити у напрямку дослідження швидкості збіжності функції розподілу центрованих та нормованих сум незалежних величин при необмеженому зростанні числа доданків.

2.8.1. Розклад Еджуорта в центральній граничній теоремі

Розглянемо послідовність $(\xi_n, n \geq 1)$ незалежних однаково розподілених величин, що є центрованими та нормованими: $\mathbf{E}\xi_n = 0, \mathbf{D}\xi_n = 1$. Тоді за *класичною центральною граничною теоремою* суми $S_n = \xi_1 + ... + \xi_n$ після нормування слабко збігаються до стандартної нормальної величини: $S_n/\sqrt{n} \to^W \zeta \simeq N(0,1), n \to \infty$. Це твердження еквівалентне поточковій збіжності відповідних функцій розподілу $F_n(x) \to \Phi(x), \forall x \in \mathbb{R}$. Дана збіжність широко використовується, зокрема, у статистиці. Тому важливими є питання про швидкість збіжність у наведеному співвідношенні, а також про можливість прискорення збіжності. Цим питанням присвячений даний розділ.

Для спрощення буде розглянуто локальну центральну граничну теорему, в якій йдеться про збіжність щільностей. Доведення спирається на метод *характеристичних функцій*, та такі леми.

Лема (про розклад логарифму характеристичної функції). *Нехай величина ξ має характеристичну функцію $\varphi(s) = \mathbf{E}\exp(is\xi)$ та скінченні моменти $\mu_k = \mathbf{E}\xi^k, k = \overline{1,3}$, $\mathbf{E}|\xi|^3 < \infty$. Тоді*

$$\ln\varphi(s) = is\sigma_1 + (is)^2\sigma_2/2 + (is)^3\sigma_3/6 + o(s^3), \ s \to 0,$$

*де коефіцієнти σ_k називаються **семіінваріантами (кумулянтами)** величини ξ і мають вигляд:*

$$\sigma_1 = \mu_1 = \mathbf{E}\xi, \ \sigma_2 = \mu_2 - \mu_1^2 = \mathbf{D}\xi, \ \sigma_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 = \mathbf{E}(\xi - \mu_1)^3.$$

Доведення. Так само, як і в теоремі *про властивості характеристичної функції*, з абсолютної інтегровності $\mathbf{E}|\xi|^3 < \infty$ на підставі відповідного розкладу Тейлора функції $\exp(isx)$ виводимо зображення

$$\varphi(s) = 1 + is\mu_1 + (is)^2\mu_2/2 + (is)^3\mu_3/6 + o(s^3), \ s \to 0.$$

Звідси, зокрема, $\varphi(s) - 1 \to 0$. Після застосування формули розкладу $\ln(1+z) = z - z^2/2 + z^3/3 + o(z^3), z \to 0$, у точці $z = \varphi(s) - 1$ та підстановки у результат наведеного вище зображення отримуємо шуканий розклад $\boxdot$

Лема (про характеристичну функцію абсолютно неперервного розподілу). *Нехай ξ_1 – абсолютно неперервна величина з щільністю f_1 і характеристичною функцією φ. Тоді*

$$\theta(\varepsilon) \equiv \sup_{|s| \geq \varepsilon} |\varphi(s)| < 1, \ \forall \varepsilon > 0.$$

Доведення. За теоремою *про основні властивості характеристичної функції* $|\varphi(s)| \leq 1$. Тому $\theta(\varepsilon) \leq 1$. Припустимо, що $\theta(\varepsilon) = 1$. За лемою Рімана – Лебега про перетворення Фур'є від інтегровної функції $f_1(x)$ має місце збіжність $\varphi(s) \to 0, |s| \to \infty$. До того ж $\varphi \in C_b(\mathbb{R})$. Тому верхня межа в означенні $\theta(\varepsilon) = 1$ має досягатись для деякого s з $|s| \geq \varepsilon$.

З рівняння $|\varphi(s)| = 1$ виводимо, що $\varphi(s)\exp(i\alpha) = 1$ для $\alpha \in \mathbb{R}$. Тому

$$0 = 1 - \mathrm{Re}(\varphi(s)\exp(i\alpha)) = \int_{-\infty}^{\infty} (1 - \cos(sx + \alpha)) f_1(x) dx \geq 0.$$

Оскільки $s \neq 0$, то функція $1 - \cos(sx + \alpha) > 0$ для майже всіх x за мірою Лебега. Тому з останньої рівності випливає, що $f_1(x) = 0$ майже всюди, що суперечить визначенню *щільності* розподілу $\boxdot$

Теорема (про розклад Еджуорта). *Нехай $(\xi_n, n \geq 1)$ – незалежні однаково розподілені випадкові величини, $\mathbf{E}\xi_1 = 0, \mathbf{D}\xi_1 = 1, \mathbf{E}\,|\xi_1|^3 < \infty$ і $\mathbf{E}\xi_1^3 = \mu_3$. Припустимо також, що характеристична функція одного доданка $\varphi(t) \equiv \mathbf{E}\exp(it\xi_1)$ абсолютно інтегровна: $\int_{\mathbb{R}} |\varphi(t)|\, dt < \infty$.*

Тоді щільність $f_n(x)$ нормованих сум $S_n/\sqrt{n} = (\xi_1 + ... + \xi_n)/\sqrt{n}$ задовольняє асимптотичне зображення

$$f_n(x) = u(x) + u(x)(x^3 - 3x)\mu_3/6\sqrt{n} + o(1/\sqrt{n}), \ n \to \infty,$$

рівномірно за $x \in \mathbb{R}$, де $u(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ – щільність стандартного нормального розподілу.

Зауваження. Існування щільності f_n випливає за теоремою *про формулу обертання для характеристичної функції* з умови інтегровності характеристичної функції. Останню умову можна послабити.

Зауваження. З теореми Еджуорта можна зробити висновок: швидкість збіжності у центральній граничній теоремі за скінченності третього моменту $\mathbf{E}\,|\xi_1|^3$ дорівнює $O(1/\sqrt{n})$, і врахування поправки Еджуорта (другий доданок у правій частині) дозволяє підвищити вказану швидкість. Можна довести, що за скінченності четвертого моменту зазначений підвищений порядок (третій доданок) дорівнює $O(1/n)$.

Доведення. Позначимо через $\varphi_n(t) = \mathbf{E}\exp(itS_n/\sqrt{n})$ – характеристичну функцію нормованої суми. За теоремою *про властивості характеристичної функції*

$$\int_{-\infty}^{\infty} \exp(itx) f_n(x)dx \equiv \varphi_n(t) = \varphi^n(t/\sqrt{n}).$$

Розглянемо характеристичну функцію $v(t) = \exp(-t^2/2)$ стандартного нормального розподілу, що визначається з рівності

$$v(t) = \int_{-\infty}^{\infty} \exp(itx) u(x)dx.$$

Диференціюванням за t з цієї рівності отримуємо тотожність

$$v'''(t) + 3v'(t) = \int_{-\infty}^{\infty} \exp(itx)(-ix^3 + 3ix)u(x)dx.$$

Звідси виводимо, що для всіх $t \in \mathbb{R}$

$$\int_{-\infty}^{\infty} \exp(itx)(x^3 - 3x)u(x)dx = i(-t^3 + 3t - 3t)v(t) = (it)^3 v(t).$$

Позначимо через

$$\Delta_n(x) = f_n(x) - u(x) - (x^3 - 3x)u(x)\mu_3/6\sqrt{n}$$

різницю між доданками лівої та правої частин зображення теореми. Твердження теореми еквівалентне зображенню $\Delta_n(x) = o(1/\sqrt{n})$, при $n \to \infty$, рівномірно за x.

Визначимо функцію

$$\delta_n(t) \equiv \varphi^n(t/\sqrt{n}) - v(t) - (it)^3 v(t)\mu_3/6\sqrt{n}.$$

Унаслідок лінійності інтеграла в означенні характеристичної функції, з попередніх тотожностей робимо висновок, що

$$\delta_n(t) = \int_{-\infty}^{\infty} \exp(itx)\Delta_n(x)dx,$$

де функцію $\Delta_n(x)$ визначено вище.

За очевидним узагальненням теореми *про формулу обертання для характеристичної функції,* (б), враховуючи умову інтегровності характеристичної функції, з останньої тотожності виводимо співвідношення

$$\Delta_n(x) = (2\pi)^{-1} \int_{-\infty}^{\infty} \exp(-itx)\delta_n(t)dt.$$

Тим самим доведення теореми зводиться до задачі оцінювання інтеграла в нерівності

$$\sup_{x\in\mathbb{R}} |\Delta_n(x)| \le (2\pi)^{-1} \int_{-\infty}^{\infty} |\delta_n(t)|\, dt.$$

Для оцінки інтегралу у правій частині зафіксуємо довільне $\varepsilon > 0$. Позначимо інтервал $T_n = \{t : |t| < \varepsilon\sqrt{n}\}$ та інтеграли

$$I_n \equiv \int_{-\infty}^{\infty} |\delta_n(t)|\, dt = \int_{\overline{T_n}} |\delta_n(t)|\, dt + \int_{T_n} |\delta_n(t)|\, dt \equiv I_{1n} + I_{2n}.$$

З означення I_{1n} та $\delta_n(t)$ виводимо для деякої сталої c співвідношення

$$I_{1n} \le \textstyle\int_{\overline{T_n}} |\varphi^n(t/\sqrt{n})|\, dt + \int_{\overline{T_n}} |v(t)|\, (1 + c\,|t|^3) dt \le$$

$$\sup_{t\in\overline{T_n}} |\varphi^{n-1}(t/\sqrt{n})| \textstyle\int_{\overline{T_n}} |\varphi(t/\sqrt{n})|\, dt + \int_{\overline{T_n}} |v(t)|\, (1 + c\,|t|^3) dt \le$$

$$\theta^{n-1}(\varepsilon)\sqrt{n} \textstyle\int_{\mathbb{R}} |\varphi(s)|\, ds + O(\exp(-\varepsilon^2 n/3)) = O(\rho^n(\varepsilon)), n \to \infty,$$

для довільного $\rho(\varepsilon) \in (\max(\theta(\varepsilon), \exp(-\varepsilon^2/3),\ 1)$. Тут величина $\theta(\varepsilon) < 1$ за лемою *про характеристичну функцію абсолютно неперервного розподілу.*

З метою оцінки I_{2n} позначимо

$$\beta(s) \equiv \ln\varphi(s) + s^2/2, \quad \gamma(s) \equiv \ln\varphi(s) + s^2/2 - (is)^3\mu_3/6.$$

За лемою *про розклад логарифму характеристичної функції* справедливі такі зображення:

$$\beta(s) = O(s^3),\ \gamma(s) = o(s^3),\ s \to 0,$$

оскільки внаслідок умов $\mu_1 = 0, \mu_2 = 1$ відповідні *семіінваріанти* у вказаній лемі дорівнюють $\sigma_1 = 0$, $\sigma_2 = 1$, $\sigma_3 = \mu_3$.

За означенням функцій u, v та β, γ

$$I_{2n} = \textstyle\int_{T_n} v(t)\, |\exp(n\beta(t/\sqrt{n})) - 1 - (it)^3\mu_3/6\sqrt{n}|\, dt =$$

$$\textstyle\int_{T_n} v(t)\, |\exp(n\beta(t/\sqrt{n})) - 1 - n\beta(t/\sqrt{n}) + n\gamma(t/\sqrt{n})|\, dt \le$$

$$\textstyle\int_{T_n} v(t)\, |\exp(n\beta(t/\sqrt{n})) - 1 - n\beta(t/\sqrt{n})|\, dt + \int_{T_n} v(t)\, |n\gamma(t/\sqrt{n})|\, dt \le$$

$$\textstyle\int_{T_n} v(t)\, |n\beta(t/\sqrt{n})|^2 \exp(n\,|\beta(t/\sqrt{n})|) dt + \int_{T_n} v(t) n\, |\gamma(t/\sqrt{n})|\, dt \equiv I_{21n} + I_{22n},$$

де використана елементарна нерівність для експоненти на комплексній площині $|\exp(z) - 1 - z| \le |z|^2 \exp(|z|)$.

Оскільки в області $t \in T_n$ виконується нерівність $|t|/\sqrt{n} \le \varepsilon$, та $\beta(s) = O(s^3), s \to 0$, то $n\,|\beta(t/\sqrt{n})| \le nC_1(|t|/\sqrt{n})^3 \le C_1 t^2\varepsilon$ для деякої сталої C_1. Отже, для всіх досить малих $\varepsilon > 0$ показник під знаком експоненти в означенні I_{21n} задовольняє нерівність $n\,|\beta(t/\sqrt{n})| \le t^2/3$. З урахуванням вище наведеної нерівності для цього ж показника отримуємо

$$I_{21n} \le \textstyle\int_{T_n} v(t) n^2 C_1^2 (|t|/\sqrt{n})^6 \exp(t^2/3) dt \le$$

$$n^{-1} C_1^2 \textstyle\int_{\mathbb{R}} t^6 \exp(-t^2/6) = O(1/n),\ n \to \infty.$$

Нарешті, для оцінки I_{22n} розглянемо функцію $\tau(\varepsilon) = \sup_{|s|\le\varepsilon} |\gamma(s)/s^3|$. Оскільки $\gamma(s) = o(s^3), s \to 0$, то $\tau(\varepsilon) \to 0, \varepsilon \to 0$. З урахуванням означення τ та оцінки $|t|/\sqrt{n} \le \varepsilon$ при $t \in T_n$ виводимо нерівність

$$I_{22n} \le \int_{T_n} v(t)n \left|t/\sqrt{n}\right|^3 \tau(\varepsilon)dt \le C_2\tau(\varepsilon)/\sqrt{n}.$$

Остаточно з нерівностей для $I_n, I_{1n}, I_{2n}, I_{21n}, I_{22n}$ отримуємо при кожному $\varepsilon > 0$ нерівність

$$\overline{\lim}_{n\to\infty}\sqrt{n}I_n \le \overline{\lim}_{n\to\infty}\sqrt{n}(I_{n1} + I_{21n} + I_{22n}) \le$$
$$\overline{\lim}_{n\to\infty}\sqrt{n}\left(O(\rho^n(\varepsilon)) + O(1/n) + C_2\tau(\varepsilon)\right) = C_2\tau(\varepsilon),$$

де враховано, що $\rho(\varepsilon) < 1$. Оскільки ліва частина нерівності не залежить від ε, а $\tau(\varepsilon) = o(1), \varepsilon \to 0$, звідси виводимо, що $\overline{\lim}_{n\to\infty}\sqrt{n}I_n = 0$. Отже, $I_n = o(1/\sqrt{n}), n \to \infty$ $\boxdot$

Вправи

(1) Вивести з нерівностей при $t \in \mathbb{R}$
$\left|\exp(it) - \sum_{k=0}^n (it)^k/k!\right| \le \min(2\,|t|^n/n!,\ |t|^{n+1}/(n+1)!)$
відповідні нерівності для характеристичної функції $\varphi_\xi(t) = \mathbf{E}\exp(it\xi)$.

(2) Нехай $(\xi_n, n \ge 1)$ – незалежні однаково розподілені випадкові величини з $\mathbf{E}\xi_1 = 0, \mathbf{D}\xi_1 = 1$ та абсолютно інтегровною на $\mathbb{R}$ характеристичною функцією φ. Тоді щільність $f_n(x)$ нормованих сум $S_n/\sqrt{n}$ задовольняє локальну центральну граничну теорему: $f_n(x) \to u(x), n \to \infty$, рівномірно за x.

2.8.2. Нерівність Беррі-Ессеєна

Асимптотичний розклад Еджуорта для розподілу сум незалежних величин має якісний характер: він встановлює порядок швидкості збіжності $O(1/\sqrt{n})$ при $n \to \infty$, однак не містить чисельних границь для відповідного множника при $1/\sqrt{n}$. Нерівність Беррі-Ессеєна закриває цю прогалину.

Доведення вказаної нерівності, як і теореми Еджуорта, спирається на метод характеристичних функцій – оскільки характеристична функція суми незалежних величин є значно простішим об'єктом, ніж функція розподілу. Між простором функцій розподілу та простором характеристичних функцій за теоремою *про однозначну відповідність між функціями розподілу та характеристичними функціями* існує взаємно однозначна відповідність. За *теоремою Леві про критерій слабкої збіжності* ця відповідність є неперервною, якщо збіжність функцій розподілу є *слабкою збіжністю*, а збіжність характеристичних функцій визначити як поточкову збіжність. Тому доведення потрібної нерівності проводиться у два етапи: (а) оцінюється модуль неперервності вказаного відображення, що зводиться до певного співвідношення між відповідними метриками у просторах функцій розподілу та характеристичних функцій, (б) оцінюється близькість у обраній метриці характеристичних функцій суми незалежних величин та стандартної нормальної величини.

Перший із перерахованих етапів становить предмет такої нерівності.

Теорема (про співвідношення метрик у просторах функцій розподілу та характеристичних функцій). *Нехай F, G – функції розподілу з характеристичними функціями φ, γ відповідно, причому G має обмежену щільність $g(x) \le m, \forall x \in \mathbb{R}$. Тоді для кожного $T > 0$:*

$$\sup_{x\in\mathbb{R}} |F(x) - G(x)| \le \frac{1}{\pi} \int_{-T}^{T} \left| \frac{\varphi(t) - \gamma(t)}{t} \right| dt + \frac{24m}{\pi T}.$$

Зауваження. З даної нерівності випливає, зокрема, достатність у *теоремі Леві про критерій слабкої збіжності.* Дійсно, якщо $\varphi_n(t) \to \gamma(t)$ поточково, а відповідна функція розподілу G задовольняє наведені умови, то вибором досить великого T з подальшим вибором n праву частину нерівності можна зробити як завгодно малою. Однак цінність даної теореми значно більша, оскільки вона дає можливість оцінити швидкість збіжності кількісно.

Доведення теореми спирається на таку лему.

Лема (про нерівність згладжування). *Нехай функції розподілу F, G такі ж, як і у теоремі про співвідношення метрик у просторах функцій розподілу та характеристичних функцій, $\Delta(x) \equiv F(x) - G(x)$, W_T – функція розподілу з параметром $T > 0$ та щільністю*

$$w_T(x) = (1 - \cos(Tx))/\pi T x^2, x \in \mathbb{R},$$

а функція $\Delta_T(x)$ дорівнює різниці згорток функцій розподілу:

$$\begin{aligned}\Delta_T(x) &= \Delta * W_T(x) \equiv \int_{\mathbb{R}} \Delta(x-y) w_T(y) dy = \\ &\int_{\mathbb{R}} F(x-y) dW_T(y) - \int_{\mathbb{R}} G(x-y) dW_T(y).\end{aligned}$$

Тоді справедлива нерівність

$$\sup_{x\in\mathbb{R}} |\Delta(x)| \le 2 \sup_{x\in\mathbb{R}} |\Delta_T(x)| + 24m \ / \ \pi T.$$

Доведення. Позначимо $\delta \equiv \sup_{x\in\mathbb{R}} |\Delta(x)|$. Можна вважати, що $\delta > 0$. Оскільки $\Delta(\pm\infty) = 0$, а функція Δ має границі справа та зліва $\Delta(x \pm 0)$, то знайдеться $x_0 \in \mathbb{R}$ таке, що $|\Delta(x_0 - 0)| = \delta$ або ж $|\Delta(x_0 + 0)| = \delta$.

Припустимо, що $\Delta(x_0 - 0) = \delta$. Інші випадки вигляду $\Delta(x_0 \pm 0) = \pm\delta$ розглядаються аналогічно.

Нехай $s \geq 0$. Тоді

$$\Delta(x_0 + s) = F(x_0 + s) - G(x_0 + s) =$$
$$F(x_0) - G(x_0) + F(x_0 + s) - F(x_0) - G(x_0 + s) + G(x_0) \geq$$
$$\delta + 0 - \int_{x_0}^{x_0+s} g(y)dy \geq \delta - \int_{x_0}^{x_0+s} m\ dy = \delta - ms.$$

Визначимо $h \equiv \delta/2m$, $x \equiv x_0+h$, $y \equiv h-s$. Тоді з попередньої нерівності виводимо, що при $s \in [0, 2h]$

$$\Delta(x - y) = \Delta(x_0 + s) \geq \delta - ms = \delta/2 + my, \quad \forall y : |y| \leq h.$$

За означенням δ

$$\Delta(x - y) \geq -\delta,\ \forall y : |y| > h.$$

З останніх двох нерівностей отримуємо

$$\Delta_T(x) = \int_{|y|\leq h} \Delta(x - y)w_T(y)dy + \int_{|y|>h} \Delta(x - y)w_T(y)dy \geq$$
$$\int_{|y|\leq h}(\delta/2 + my)w_T(y)dy - \delta \int_{|y|>h} w_T(y)dy =$$
$$\delta/2 - 3\delta \int_{y>h} w_T(y)dy \geq \delta/2 - 3\delta \int_{y>h} 2(\pi T y^2)^{-1}dy =$$
$$\delta/2 - 6\delta/\pi T h = \delta/2 - 12m/\pi T.$$

Звідси за означенням величини δ та верхньої межі

$$\sup_{x\in\mathbb{R}} |\Delta_T(x)| \geq \sup_{x\in\mathbb{R}} |\Delta(x)| /2 - 12m/\pi T \quad \boxdot$$

Доведення теореми. Будемо використовувати позначення, що містяться у доведенні леми.

Розглянемо такі перетворення Фур'є функцій обмеженої варіації:

$$\delta_T(t) = \int_{-\infty}^{\infty} \exp(itx)d\Delta_T(x),\ \ \varpi_T(t) = \int_{-\infty}^{\infty} \exp(itx)dW_T(x).$$

Оскільки Δ_T за означенням є різницею згорток: $\Delta_T = F * W_T - G * W_T$, то за теоремою *про властивості характеристичної функції*

$$\delta_T(t) = \varphi(t)\varpi_T(t) - \gamma(t)\varpi_T(t) = (\varphi(t) - \gamma(t))\varpi_T(t),$$

де враховане означення Δ_T та лінійність інтеграла.

Нескладні обчислення з використанням теореми *про формулу обертання для характеристичної функції* призводять до такого виразу для характеристичної функції $\varpi_T(t)$:

$$\varpi_T(t) = (1 - |t| /T)\mathbb{1}_{|t|\leq T}.$$

Звідси, зокрема, робимо висновок, що функція $\delta_T(t)$ дорівнює нулю при $|t| > T$, а отже, інтегровна на $\mathbb{R}$. Тому за теоремою *про формулу обертання для характеристичної функції*, (б), функція $\Delta_T(x)$ має щільність $\Delta'_T(x)$, яка дорівнює

$$\Delta'_T(x) = (2\pi)^{-1}\int_{-\infty}^{\infty}\exp(-itx)\delta_T(t)dt = (2\pi)^{-1}\int_{-T}^{T}\exp(-itx)\delta_T(t)dt.$$

Можна вважати, що функція $(\varphi(t) - \gamma(t))/t$ інтегровна в околі нуля, інакше права частина нерівності згладжування є нескінченною. За цього припущення розглянемо функцію

$$\widetilde{\Delta}_T(x) = (2\pi)^{-1}\int_{-T}^{T}\exp(-itx)(-it)^{-1}\delta_T(t)dt.$$

Оскільки її похідна за x дорівнює інтегралу від похідної, що абсолютно збігається, причому ця похідна збігається з щільністю $\Delta'_T(x)$, то справедлива тотожність $\widetilde{\Delta}_T(x) = \Delta_T(x)$ для всіх x, враховуючи рівність початкових значень $\widetilde{\Delta}_T(-\infty) = \Delta_T(-\infty) = 0$.

Тому з останнього зображення виводимо, що

$$\sup\nolimits_{x\in\mathbb{R}}|\Delta_T(x)| \le (2\pi)^{-1}\int_{-T}^{T}\left|\delta_T(t)t^{-1}\right|dt \le (2\pi)^{-1}\int_{-T}^{T}|(\varphi(t) - \gamma(t))/t|\,dt,$$

де використано зображення для $\delta_T(t)$ та нерівність $|\varpi_T(t)| \le 1$.

Після підстановки останньої нерівності у твердження леми про нерівність згладжування отримуємо нерівність теореми $\boxdot$

Теорема (про нерівність Беррі – Ессеєна). *Нехай (ξ_n) – послідовність незалежних однаково розподілених випадкових величин, що центровані і нормовані: $\mathbf{E}\xi_1 = 0, \mathbf{D}\xi_1 = 1$ та $\rho \equiv \mathbf{E}\left|\xi_1^3\right| < \infty$. Тоді для функції розподілу $F_n(x) = \mathbf{P}((\xi_1 + ... + \xi_n)/\sqrt{n} < x)$ має місце нерівність*

$$\sup\nolimits_{x\in\mathbb{R}}|F_n(x) - \Phi(x)| < \frac{33}{4}\frac{\rho}{\sqrt{n}},$$

де Φ є функцією стандартного нормального розподілу.

Зауваження. Абсолютну сталу $33/4$ у правій частині можна зменшити за рахунок більш точних оцінок. Зокрема, відомий російський математик В.М. Золотарьов довів таку нерівність з константою 0.91. Однак з розкладу Еджуорта випливає, що швидкість збіжності $O(n^{-1/2})$ без додаткових умов посилити не можна.

Доведення. Позначимо через φ та φ_n характеристичні функції величин ξ_1 та $(\xi_1 + ... + \xi_n)/\sqrt{n}$ відповідно, через $\gamma(t) = \exp(-t^2/2)$ – характеристичну функцію стандартного нормального розподілу. За теоремою *про властивості характеристичної функції* $\varphi_n(t) = \varphi^n(t/\sqrt{n})$.

Для доведення теореми скористаємося теоремою *про співвідношення метрик у просторах функцій розподілу та характеристичних функцій,* у якій оберемо $F \equiv F_n, G \equiv \Phi, \varphi \equiv \varphi_n, \gamma \equiv \gamma$, $T = \sqrt{n}/\rho$, та $m = G'(0) = 1/\sqrt{2\pi}$. З цієї теореми отримуємо нерівність

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq \pi^{-1} I_n + \frac{24}{\pi\sqrt{2\pi}} \frac{\rho}{\sqrt{n}},$$

де $I_n = \displaystyle\int_{-T}^{T} \left| \left(\varphi^n(t/\sqrt{n}) - \gamma(t) \right) t^{-1} \right| dt = \int_{-T}^{T} \gamma(t) \, |t|^{-1} \left| \exp(n\beta(t/\sqrt{n})) - 1 \right| dt$, а функція $\beta(s) \equiv \ln \varphi(s) + s^2/2$.

Зазначимо, що $1 = \mathbf{D}\xi_1 = \mathbf{E}\xi_1^2 \leq (\mathbf{E}\, |\xi_1|^3)^{2/3} = \rho^{2/3}$ внаслідок *нерівності Ляпунова*, звідки $\rho \geq 1$.

Далі, з елементарних нерівностей

$$|\exp(iy) - 1 - iy| \leq y^2/2, \quad \left|\exp(iy) - 1 - iy + y^2/2\right| \leq |y|^3/6$$

при $y = s\xi_1$ обчисленням математичного сподівання та врахуванням центрованості і нормованості величини ξ_1 виводимо, що

$$|\varphi(s) - 1| \leq s^2/2, \ \left|\varphi(s) - 1 + s^2/2\right| \leq \rho\, |s|^3/6.$$

Звідси при $|s| \leq 1/\rho \leq 1$ з розкладу логарифмічної функції у ряд Тейлора отримуємо

$$|\beta(s)| = |\ln \varphi(s) + s^2/2| = \left|\varphi(s) - 1 + s^2/2 - \textstyle\sum_{n\geq 2} (1 - \varphi(s))^n/n\right| \leq$$
$$\rho\, |s|^3/6 + \textstyle\sum_{n\geq 2} s^{2n}/2^{n+1} = \rho\, |s|^3/6 + s^4/(8(1 - s^2/2)) \leq$$
$$\rho\, |s|^3/6 + s^4/4 \leq \rho\, |s|^3/6 + \rho\, |s|^3/4 = \tfrac{5}{12}\rho\, |s|^3 \leq \tfrac{5}{12} s^2.$$

З використанням останньої та передостанньої оцінки для $|\beta(s)|$ при $s = t/\sqrt{n}$ та нерівності $|\exp(z) - 1| \leq |z| \exp(|z|)$ при $z = n\beta(t/\sqrt{n})$ оцінимо доданок I_n, що визначений вище:

$$I_n \leq \textstyle\int_{-T}^{T} \gamma(t)\, |t|^{-1} |n\beta(t/\sqrt{n})| \exp(|n\beta(t/\sqrt{n})|) dt \leq$$
$$\textstyle\int_{-T}^{T} \gamma(t)\, |t|^{-1} n \frac{5}{12} \rho\, |t|^3 n^{-3/2} \exp(n \frac{5}{12} t^2 n^{-1}) dt =$$
$$\tfrac{\rho}{\sqrt{n}} \tfrac{5}{12} \textstyle\int_{-T}^{T} t^2 \exp(-\frac{1}{12} t^2) dt \leq \tfrac{\rho}{\sqrt{n}} \tfrac{5}{12} \int_{\mathbb{R}} t^2 \exp(-\frac{1}{12} t^2) dt = \tfrac{\rho}{\sqrt{n}} 5\sqrt{3\pi}.$$

Остаточно з нерівності на початку доведення виводимо, що

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{\rho}{\sqrt{n}} \left(\frac{5\sqrt{3\pi}}{\pi} + \frac{24}{\pi\sqrt{2\pi}} \right) < \frac{33}{4} \frac{\rho}{\sqrt{n}} \ \boxdot$$

Вправи

(1) Довести, що абсолютна стала у правій частині нерівності Беррі-Ессеєна не менша за $(2\pi)^{-1/2} \approx 0.3989$. *Вказівка*: обрати випадкові величини з розподілом $\mathbf{P}(\xi_n = \pm 1) = 1/2$ та довести, що при $n = 2k$ для таких величин має місце зображення: $F_n(0) - \Phi(0) = (2\pi n)^{-1/2} + o(n^{-1/2}), n \to \infty$.

(2) Випадкові величини $(\xi_n, n \ge 1)$ незалежні , $\mathbf{E}\xi_n = 0$, $\sigma_n^2 \equiv \sum_{k=1}^n \mathbf{E}\xi_k^2$, та $\mathbf{E}\left|\xi_n\right|^{2+\delta} < \infty$ при $\delta \in (0,1]$. Довести для деякої сталої A нерівність

$\sup_{x\in\mathbb{R}} \left|\mathbf{P}((\xi_1 + ... + \xi_n)/\sigma_n < x) - \Phi(x)\right| \le A\sigma_n^{-2-\delta} \sum_{k=1}^n \mathbf{E}\left|\xi_k\right|^{2+\delta}$.

2.9. Нескінченно подільні розподіли

2.9.1. Означення нескінченної подільності

У *граничній теоремі Пуассона для стандартних серій* доведено, що граничним розподілом для сум незалежних величин може виступати не тільки нормальний розподіл. У зв'язку з цим виникає питання про те, які ж саме розподіли можуть бути граничними. Наступне означення визначає такі граничні розподіли для сум серій незалежних та однаково розподілених величин.

Означення. *Випадкова величина ζ та її характеристична функція φ називаються* нескінченно подільними, *якщо для кожного $n \ge 1$ знайдеться послідовність $(\xi_{nk}, k = \overline{1,n})$ незалежних однаково розподілених величин зі спільною характеристичною функцією φ_n, причому*

$$\xi_{n1} + ... + \xi_{nn} \to^W \zeta,\ n \to \infty,$$

або ж, що еквівалентно, $\varphi_n^n(t) \to \varphi(t), n \to \infty, \forall t \in \mathbb{R}$.

Згадана еквівалентність є очевидним наслідком *теореми Леві про критерій слабкої збіжності*.

Зауваження. Можна довести, що заміна умови однакової розподіленості в означенні нескінченно подільного розподілу на умову рівномірної малості: $\max_{1\le k\le n} \xi_{nk} \to^P 0, n \to \infty$, веде до еквівалентного означення.

2.9.2. Канонічні міри

Означення. *Сигма-скінченна міра μ на $\mathfrak{B}(\mathbb{R})$ є* канонічною, *якщо:*
(а) $\mu(I) < \infty$ для кожного обмеженого інтервалу $I \subset \mathbb{R}$,
(б) $\int_{|x|\ge 1} x^{-2}\mu(dx) < \infty$.

Наприклад, міра Лебега є канонічною. Наведені дві умови еквівалентні одній умові $\int_{\mathbb{R}}(1+x^2)^{-1}\mu(dx) < \infty$.

Визначимо такий клас неперервних функцій:

$$C_{b2}(\mathbb{R}) = \left\{g \in C(\mathbb{R}) : \sup_{x\in\mathbb{R}}(1+x^2)\,|g(x)| < \infty\right\}.$$

Зауважимо, що за означенням інтеграл $\int_{\mathbb{R}} g(x)\mu(dx)$ коректно визначений для кожної *канонічної* міри μ та кожної функції $g \in C_{b2}(\mathbb{R})$.

Природним є узагальнення *слабкої збіжності* ймовірнісних мір.

Означення. *Послідовність скінчених мір* (μ_n) *слабко збігається до скінченої міри* μ ($\mu_n \to^W \mu$), *якщо* $\int_{\mathbb{R}} g\, d\mu_n \to \int_{\mathbb{R}} g\, d\mu, n \to \infty, \forall g \in C_b(\mathbb{R})$.

Аналогічно визначається збіжність канонічних мір.

Означення. *Послідовність канонічних мір* (μ_n) *узагальнено слабко збігається до канонічної міри* μ *(позначення* $\mu_n \to^{W2} \mu$*), якщо* $\int_{\mathbb{R}} g\, d\mu_n \to \int_{\mathbb{R}} g\, d\mu, n \to \infty$, *для всіх* $g \in C_{b2}(\mathbb{R})$.

Кожна з наведених збіжностей породжує своє поняття компактності.

Означення. Клас $\mathfrak{M}$ скінченних мір (канонічних мір) на $\mathfrak{B}(\mathbb{R})$ є слабко компактним (узагальнено слабко компактним), якщо для довільної послідовності $(\mu_n) \subset \mathfrak{M}$ знайдеться підпослідовність $(\mu_{n_k}) \subset (\mu_n)$ та скінченна (канонічна) міра μ такі, що $\mu_{n_k} \to^W \mu$ (відповідно $\mu_{n_k} \to^{W2} \mu$).

Доведемо критерії, що є аналогами *теореми Прохорова про критерій слабкої компактності* для узагальненої збіжності.

Теорема (про умови слабкої компактності класу обмежених мір). *Нехай* $\mathfrak{M}$ *– клас скінченних мір на* $\mathfrak{B}(\mathbb{R})$ *такий, що:*

(а) $\sup_{\mu\in\mathfrak{M}}\mu(\mathbb{R}) < \infty$,

(б) $\lim_{c\to\infty}\sup_{\mu\in\mathfrak{M}}\mu(\overline{[-c,c)}) = 0$.

Тоді цей клас є слабко компактним.

Доведення. Позначимо через $m \equiv \overline{\lim}_{n\to\infty}\mu_n(\mathbb{R})$.

Припустимо, що $m = 0$. Тоді твердження теореми виконується для $n_k = k$ та нульової міри $\mu(B) \equiv 0$.

Нехай $m > 0$ та для деякої підпослідовності $m_k \equiv \mu_{n_k}(\mathbb{R}) \to m, k \to \infty$. З (б) виводимо, що функції розподілу $F_k(x) \equiv m_k^{-1}\mu_{n_k}((-\infty, x))$ задовольняють умову *теореми Прохорова про критерій слабкої компактності.* За цією теоремою існує слабко збіжна до деякої функції розподілу F підпослідовність $(F_{n_j}) \subset (F_{n_k})$. Визначимо скінченну міру $\mu(B) = mF(B)$. Тоді для $g \in C_b(\mathbb{R})$ при $j \to \infty$ отримуємо

$$\int_{\mathbb{R}} g(x)\mu_{n_j}(dx) = m_j \int_{\mathbb{R}} g(x)F_{n_j}(dx) \to m \int_{\mathbb{R}} g(x)F(dx) = \int_{\mathbb{R}} g(x)\mu(dx) \boxdot$$

Теорема (про умови узагальненої слабкої компактності класу канонічних мір). *Нехай* $\mathfrak{M}$ *– клас канонічних мір на* $\mathfrak{B}(\mathbb{R})$ *такий, що:*

(а) $\sup_{\mu\in\mathfrak{M}}\mu(I)<\infty$ *для кожного обмеженого інтервалу* $I\subset\mathbb{R}$,

(б) $\lim_{c\to\infty}\sup_{\mu\in\mathfrak{M}}\int_{|x|\geq c}x^{-2}\mu(dx)=0.$

Тоді цей клас є узагальнено слабко компактним.

Доведення. Розглянемо при кожному $m\geq 1$ сім'ю звужень мір: $(\mu(\cdot\cap[-m,m]),\mu\in\mathfrak{M})$. Ці звуження утворюють обмежену множину скінченних мір внаслідок умови (а). Тому за теоремою *про умови слабкої компактності класу обмежених мір* при кожному m знайдеться послідовність $(\mu_{mn},n\in K_m)\subset\mathfrak{M}$ та скінченна міра ν_m на $\mathfrak{B}([-m,m])$ такі, що має місце збіжність звужень $\mu_{mn}\mid_{[-m,m]}\to^W\nu_m,n\to\infty,n\in K_m$.

Якщо будувати такі послідовності послідовно за зростанням m, обираючи кожну наступну як підпослідовність попередньої, то можна вважати, що $(\mu_{m+1,n},n\in K_{m+1})\subset(\mu_{mn},n\in K_m)$. В цьому разі звуження міри ν_{m+1} на інтервал $[-m,m]$ збігатиметься з ν_m. Оберемо діагональну послідовність Кантора $\mu_n\equiv\mu_{n,n}$ для n з діагональної множини Кантора $K\equiv(\min K_m,m\geq 1)$, та визначимо $\mu=\nu_1+\sum_{m\geq 1}(\nu_{m+1}-\nu_m)=\lim_{m\to\infty}\nu_m$.

Нехай неперервна функція g фінітна, тобто $g(x)=0$ при $|x|>m$. Тоді з включення $(\mu_{nn},n\geq m,n\in K)\subset(\mu_{mn},n\in K_m)$ для $n\geq m$, отримуємо при $n\to\infty,n\in K$, збіжність

$$\int_{\mathbb{R}}g(x)\mu_n(dx)=\int_{-m}^{m}g(x)\mu_{nn}(dx)=\int_{-m}^{m}g(x)\mu_{mn}(dx)\to\int_{-m}^{m}g(x)\nu_m(dx)=$$
$$=\int_{\mathbb{R}}g(x)\nu_m(dx)=\int_{\mathbb{R}}g(x)\mu(dx).$$

Нехай $\rho_c(x)$ – неперервна функція зі значеннями у $[0,1]$ така, що $\rho_c(x)=0$ при $|x|>c+1$ та $\rho_c(x)=1$ при $|x|\leq c$.

За означенням $\mu([-m,m])=\nu_m(\mathbb{R})<\infty$. З останньої збіжності отримуємо також нерівність при $c\geq 1$

$$\int_{\mathbb{R}}\rho_c(x)(1+x^2)^{-1}\mu(dx)=\lim_{n\to\infty}\int_{\mathbb{R}}\rho_c(x)(1+x^2)^{-1}\mu_n(dx)\leq$$
$$\sup_{\mu\in\mathfrak{M}}\int_{\mathbb{R}}(1+x^2)^{-1}\mu(dx)\leq\sup_{\mu\in\mathfrak{M}}\Big(\mu([-1,1]+2\int_{|x|\geq 1}x^{-2}\mu(dx)\Big)<\infty.$$

Оскільки права частина не залежить від c, то з *теореми Лебега про монотонну збіжність* при $c\to\infty$ виводимо збіжність інтегралу $\int_{|x|\geq 1}x^{-2}\mu(dx)$. Отже, μ є *канонічною* мірою.

Нехай $g\in C_{b2}(\mathbb{R})$ та $L=\sup_{x\in\mathbb{R}}(1+x^2)\left|g(x)\right|$. Тоді

$$\overline{\lim}_{n\to\infty}\left|\int_{\mathbb{R}}g(x)\mu_n(dx)-\int_{\mathbb{R}}g(x)\mu(dx)\right|\leq$$
$$\overline{\lim}_{n\to\infty}\left|\int_{\mathbb{R}}g(x)\rho_c(x)\mu_n(dx)-\int_{\mathbb{R}}g(x)\rho_c(x)\mu(dx)\right|+$$

$$\overline{\lim}_{n\to\infty} \int_{|x|>c} |g(x)|\,\mu_n(dx) + \int_{|x|>c} |g(x)|\,\mu(dx) \le$$

$$0 + L\sup_{\mu\in\mathfrak{M}} \int_{|x|\ge c} x^{-2}\mu(dx) + L\int_{|x|\ge c} x^{-2}\mu(dx) \to 0$$

при $c \to \infty$, де врахована фінітність неперервної функції $g(x)\rho_c(x)$, умова (б) теореми та *канонічність* міри μ. Оскільки ліва частина нерівності не залежить від c, то вона дорівнює нулю ⊡

2.9.3. Зображення Леві-Хінчина

Теорема (про зображення Леві – Хінчина для нескінченно подільної характеристичної функції). *Функція φ є характеристичною функцією нескінченно подільного розподілу тоді й тільки тоді, коли знайдеться канонічна міра μ та стала $a \in \mathbb{R}$ такі, що*

$$\varphi(t) = \exp\left(ita + \int_{\mathbb{R}} g_t(x)\mu(dx)\right),\ t \in \mathbb{R},$$

де $g_t(x) \equiv (\exp(itx) - 1 - it\sin x))/x^2$.

Зауваження. Функцію $\sin x$ у означенні g_t можна замінити на довільну обмежену неперервну функцію, що еквівалентна $x + o(x), x \to 0$.

Зауваження. Можна довести, що функція φ є характеристичною функцією *нескінченно подільного* розподілу тоді й тільки тоді, коли знайдеться скінченна міра Π з $\Pi(\{0\}) = 0$, та сталі a, b такі, що має місце зображення Колмогорова-Леві-Хінчина:

$$\varphi(t) = \exp\left(ita - b^2t^2/2 + \int_{\mathbb{R}} \left(\exp(itx) - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\Pi(dx)\right).$$

Еквівалентний варіант зображення Леві-Хінчина виведений у теорії випадкових процесів з незалежними приростами:

$$\varphi(t) = \exp\left(ita - b^2t^2/2 + \int_{\mathbb{R}} \left(\exp(itx) - 1 - itx\mathbb{1}_{|x|\le 1}\right)K(dx)\right),$$

де σ-скінченна міра K така, що $K(\{0\}) = 0$, і $\int_{\mathbb{R}} \min(1, x^2)K(dx) < \infty$.

Доведення *необхідності* в теоремі.

Припустимо, що φ – *нескінченно подільна* характеристична функція, тобто $\varphi(t) = \lim_{n\to\infty} \varphi_n^n(t), \forall t \in \mathbb{R}$, для характеристичних функцій $\varphi_n(t)$.

Оскільки φ – характеристична функція, то знайдеться $\theta > 0$ таке, що $|\varphi(t)| \ge 1/2$ при $|t| \le \theta$. Зафіксуємо таке θ.

Для $|t| \le \theta$ та $t \in \mathbb{R}$ коректно визначені неперервні обмежені функції

$$\psi(t) \equiv \ln\varphi(t), \qquad \psi_n(t) \equiv n(\varphi_n(t) - 1).$$

Лема 1. $\psi_n(t) \to \psi(t)$ *при* $n \to \infty$ *рівномірно за* $|t| \le \theta$.

Доведення. Унаслідок теореми *про рівномірну збіжність характеристичних функцій* $\varphi_n^n(t) \to \varphi(t), n \to \infty$, рівномірно за t на кожному обмеженому інтервалі. Зокрема, $|\varphi_n^n(t)| \ge 1/3$ при $|t| \le \theta$, починаючи з деякого номера. Тому коректно визначені функції $\ln \varphi_n(t)$.

Зі збіжності $n \ln \varphi_n(t) \to \psi(t), n \to \infty$, рівномірно за $|t| \le \theta$ виводимо таку саму рівномірну збіжність $\varphi_n(t) \to 1$. Отже, твердження леми є наслідком еквівалентності $\ln z \sim z - 1$ при $z \to 1$ $\boxdot$

Розглянемо сигма-скінченні міри

$$\mu_n(B) = \int_B nx^2 F_n(dx),\ \ B \in \mathfrak{B}(\mathbb{R}),$$

де F_n – міра Лебега – Стілтьєса для функції розподілу, що відповідає характеристичній функції φ_n.

Лема 2. *Для всіх* $t \in \mathbb{R}$ *має місце зображення*

$$\psi_n(t) = ita_n + \int_{\mathbb{R}} g_t(x) \mu_n(dx),$$

де $a_n \equiv \int_{\mathbb{R}} (\sin x)\ nF_n(dx)$.

Доведення є очевидним наслідком рівності $\exp(itx) - 1 - it\sin x = x^2 g_t(x)$ та тотожності

$$\int_{\mathbb{R}} f(x) \mu_n(dx) = \int_{\mathbb{R}} f(x) nx^2 F_n(dx),$$

для борелєвих обмежених функцій $f = g_t$, що виконується для простих f за означенням, та подовжується на функції $f \in C_{b2}(\mathbb{R})$ за *теоремою Лебега про мажоровану збіжність* $\boxdot$

Лема 3. При кожному $h > 0$ справедлива тотожність

$$-\frac{1}{2h} \int_{-h}^{h} \psi_n(t) dt = \int_{\mathbb{R}} \left(1 - \frac{\sin xh}{xh}\right) nF_n(dx) = \int_{\mathbb{R}} \left(1 - \frac{\sin xh}{xh}\right) x^{-2} \mu_n(dx).$$

Доведення. За означенням ψ_n

$$-(2h)^{-1} \int_{-h}^{h} \psi_n(t) dt = (2h)^{-1} n \int_{-h}^{h} (1 - \varphi_n(t)) dt =$$

$$(2h)^{-1} \int_{-h}^{h} \int_{\mathbb{R}} (1 - \exp(itx)) nF_n(dx) dt =$$

$$(2h)^{-1} \int_{\mathbb{R}} nF_n(dx) \int_{-h}^{h} (1 - \exp(itx))\, dt = \int_{\mathbb{R}} (1 - (xh)^{-1} \sin xh)\, nF_n(dx).$$

Остання рівність у лемі доводиться так само, як і в лемі 2 $\boxdot$

Лема 4. *Клас мір* $(\mu_n, n \ge 1)$ *задовольняє умови (а),(б) теореми про умови узагальненої слабкої компактності класу канонічних мір.*

Доведення. (а) Для кожного інтервалу $I = [-c, c]$ визначимо

$$\delta(c,h) = \inf_{x\in I}(1-(\sin xh)/xh)/h^2x^2 > 0,$$

для досить малих $h > 0$, оскільки неперервна функція $1-(\sin y)/y$ строго додатна при $y \neq 0$, а в околі нуля дорівнює $y^2/6 + o(y^2)$.

З означення $\delta(c,h)$ та леми 3 виводимо, що для $0 \le h \le \theta$

$$\delta(c,h)h^2 \sup_{n\ge1} \mu_n(I) \le \sup_{n\ge1} \int_I (1-\tfrac{\sin xh}{xh})x^{-2}\mu_n(dx) \le$$

$$\sup_{n\ge1} \int_{\mathbb{R}} (1-\tfrac{\sin xh}{xh})x^{-2}\mu_n(dx) = \sup_{n\ge1}\left(-(2h)^{-1}\int_{-h}^{h}\psi_n(t)dt\right) < \infty,$$

оскільки функції ψ_n обмежені при $|t| \le \theta$ при кожному n та $\psi_n \to \psi$ при $n \to \infty$ рівномірно за t, за лемою 1, де функція $\psi \equiv \ln\varphi$ обмежена на інтервалі $[-h, h]$ при $h \le \theta$, як відзначено вище.

(б) Розглянемо величини $\varepsilon_n(c) \equiv \int_{|x|\ge c} x^{-2}\mu_n(dx)$. Послідовність $\varepsilon_n(c)$ не зростає за c та $\varepsilon_n(c) \to 0, c \to \infty$, оскільки міра μ_n – канонічна.

Визначимо $h = 2/c$ для $c > 0$. Тоді $(\sin xh)/xh \le 1/2$ для всіх $|x| \ge c$. Звідси за лемою 3 при $h \le \theta$

$$\varepsilon_n(c) = \int_{|x|\ge c} x^{-2}\mu_n(dx) \le 2\int_{\mathbb{R}}(1-\tfrac{\sin xh}{xh})x^{-2}\mu_n(dx) = -\tfrac{2}{2h}\int_{-h}^{h}\psi_n(t)dt,$$

$$\overline{\lim}_{n\to\infty}\varepsilon_n(c) \le \overline{\lim}_{n\to\infty}\tfrac{2}{2h}\int_{-h}^{h}(-\psi_n(t))dt = -\tfrac{2}{2h}\int_{-h}^{h}\psi(t)dt \to 0, h \to 0,$$

оскільки $\psi_n(t) \to \psi(t)$, і $\psi(t) \equiv \ln\varphi(t) \to 0$ при $t \to 0$.

Отже, визначена послідовність задовольняє умови теореми *про верхні межу та границю монотонної послідовності.* Тому $\sup_{n\ge1}\varepsilon_n(c) \to 0$ при $c \to \infty$, що доводить твердження леми $\boxdot$

За лемою 4 внаслідок теореми *про умови узагальненої слабкої компактності класу канонічних мір* знайдеться підпослідовність $K \subset \mathbb{N}$ та канонічна міра μ такі, що $\int_{\mathbb{R}} g_t(x)\mu_n(dx) \to \int_{\mathbb{R}} g_t(x)\mu(dx)$ при $n \to \infty$ так, що $n \in K$, для всіх $t \in \mathbb{R}$, оскільки функції $g_t \in C_{b2}(\mathbb{R})$.

За лемою 1 $\psi_n(t) \to \psi(t)$ при $n \to \infty$ рівномірно за $|t| \le \theta$, тому для таких t з попереднього твердження та з зображення леми 2 робимо висновок, що послідовність ita_n збігається при $n \to \infty, n \in K$. Отже, існує скінченна границя $a \equiv \lim_{n\to\infty, n\in K} a_n$.

Ще одним застосуванням зображення леми 2 вже для всіх $t \in \mathbb{R}$ при $n \to \infty, n \in K$, виводимо існування скінченної границі

$$\lim_{n\to\infty, n\in K}\psi_n(t) = \omega(t) \equiv ita + \int_{\mathbb{R}} g_t(x)\mu(dx), \forall t \in \mathbb{R},$$

де μ – канонічна міра, що побудована вище. Оскільки $\psi_n \equiv n(\varphi_n - 1)$, звідси виводимо, що $\varphi_n(t) \to 1, n \to \infty$. Унаслідок еквівалентності $\ln z \sim z - 1$

при $z \to 1$, підстановкою $z = \varphi_n(t)$ з останнього співвідношення отримуємо: $\ln \varphi_n^n(t) = n \ln \varphi_n(t) \sim n(\varphi_n(t) - 1) \to \omega(t),\ n \to \infty$. Тому за неперервністю експоненти $\varphi_n^n(t) \to \exp(\omega(t))$ при $n \to \infty$, та $\varphi(t) = \exp(\omega(t))$, звідки випливає наведене зображення для φ. *Необхідність* доведена.

Доведення *достатності* в теоремі.

Доведемо, що права частина наведеного зображення є характеристичною функцією для довільної канонічної міри μ. Для цього позначимо $\sigma^2 = \mu(\{0\})$ та розглянемо звуження $\mu_n(B) = \mu(B \cap \{x : n^{-1} \le |x| \le n\})$, що є скінченною мірою та має нульове звуження на деякий окіл нуля. Зауважимо, що функція

$$\varphi_n(t) \equiv \exp\Big(ita - \sigma^2 t^2/2 + \int_{\mathbb{R}} g_t(x)\mu_n(dx)\Big) =$$

$$\exp\Big(it\Big(a - \int_{\mathbb{R}} x^{-2} \sin x\ \mu_n(dx)\Big) - \sigma^2 t^2/2 + \int_{\mathbb{R}} (\exp(itx) - 1)x^{-2}\mu_n(dx)\Big)$$

є характеристичною функцією суми нормальної величини та незалежного від неї зсунутого на сталу значення *складного процесу Пуассона* з функцією розподілу стрибків, що пропорційна $x^{-2}\mu_n(dx)$.

Оскільки за означенням *канонічної* міри μ та функції g_t має місце збіжність $\int_{\mathbb{R}} g_t d\mu_n \to \int_{\mathbb{R}\setminus\{0\}} g_t d\mu,\ n \to \infty$, то поточково $\varphi_n(t) \to \varphi(t)$, де функція φ неперервна в нулі: $\lim_{t\to 0} \varphi(t) = \varphi(0) = 1$ за *теоремою Лебега про мажоровану збіжність*. Тому функція φ є характеристичною за *теоремою Леві про критерій слабкої збіжності*.

Якщо μ – *канонічна* міра, то міра $n^{-1}\mu$ є канонічною. Тому функція

$$\varphi_n(t) \equiv \exp\big(itan^{-1} + \textstyle\int_{\mathbb{R}} g_t(x) n^{-1}\mu(dx)\big)$$

є характеристичною, як доведено вище. Оскільки $\varphi(t) = \varphi_n^n(t)$, то функція φ є *нескінченно подільною* $\boxdot$

Вправи

(1) Нескінченно подільна характеристична функція не має коренів.

(2) Якщо φ_1, φ_2 – нескінченно подільні характеристичні функції, то їх добуток $\varphi_1(t)\varphi_2(t)$ – також нескінченно подільна характеристична функція.

(3) Поточкова границя нескінченно подільних характеристичних функцій, що неперервна в нулі, є нескінченно подільною характеристичною функцією.

(4) Нескінченно подільна величина є слабкою границею послідовностей сум незалежних величин, що мають розподіли Пуассона на решітках $a + b\mathbb{Z}_+$.

(5) Випадкова величина ξ має розподіл (а) нормальний, (б) Пуассона, причому $\xi = \xi_1 + \xi_2$, де доданки незалежні та мають нескінченно подільні розподіли. Довести, що ξ_k мають розподіли (а) та (б).

(6) Довести, що функція $\varphi(t) = \exp(-c\,|t|^{\alpha})$ при $c > 0, \alpha \in (0,2]$ є нескінченно подільною характеристичною функцією. Знайти канонічне зображення для цієї функції. *Вказівка*: розглянути незалежні симетричні випадкові величини такі, що $\mathbf{P}(|\xi_1| > x) = x^{-\alpha}, x \geq 1$ та довести, що характеристична функція нормованих сум $(\xi_1 + ...\xi_n)n^{-1/\alpha}$ збігається до $\varphi(t)$.

(7) Гама-розподіл, розподіли Пуассона та Коші є нескінченно подільними.

(8) Біноміальний та рівномірний розподіли не є нескінченно подільними.

(9) Випадкові величини $(\xi_n, n \geq 1)$ незалежні однаково розподілені, цілочисельна величина ν не залежить від них і є нескінченно подільною, а суми $S_n = \xi_1 + ... + \xi_n$. Довести, що величина S_ν є нескінченно подільною.

(10) Нехай φ – нескінченно подільна характеристична функція, $\psi = \ln\varphi$, а S – функція розподілу симетричної квадратично інтегровної випадкової величини. Довести, що функція $\omega = \psi - \psi * S$ після ділення на сталу $\omega(0)$ є характеристичною.

(11) Випадкові величини $(\xi_n, n \geq 1)$ незалежні однаково розподілені, мають щільності $\exp(-|x|)/2$ та характеристичну функцію $1/(1+t^2)$. Довести, що ряд $\xi = \sum_{n\geq 1} \xi_n/n$ збігається м.н., а характеристична функція $\varphi_\xi(t) = \pi t/\sinh(\pi t)$ є нескінченно подільною. Знайти її канонічне зображення.

(12) Нехай φ – нескінченно подільна характеристична функція. Довести, що $(1 - \ln\varphi(t))^{-1}$ є нескінченно подільною характеристичною функцією.

(13) Функція $p(s) = \sum_{n\geq 0} p_n s^n$ така, що $p_n \geq 0$, $p_0 > 0$, $p(1) = 1$, причому $\ln p(s)/p_0$ розкладається у ряд Тейлора з невід'ємними коефіцієнтами. Довести, що для довільної характеристичної функції φ суперпозиція $p(\varphi(t))$ є нескінченно подільною характеристичною функцією. Зокрема, дане твердження виконується для $p(s) = (1 - as)(1 - bs)^{-1}$ при $0 \leq a < b < 1$.

(14) Характеристична функція φ називається стійкою, якщо для довільних додатних a, b знайдуться $c, d > 0$ такі, що $\varphi(at)\varphi(bt) = \exp(-itc)\varphi(dt)$. (а) Довести, що стійка характеристична функція є нескінченно подільною. (б) Характеристична функція φ є стійкою тоді й тільки тоді, коли знайдеться послідовність незалежних однаково розподілених величин $(\xi_n, n \geq 0)$ з характеристичною функцією φ такі, що при кожному $n \geq 1$ для деяких сталих $a_n, b_n > 0$ має місце збіжність розподілів $(\xi_1 + ... + \xi_n - a_n)/b_n \simeq \xi_0$. (в) Характеристична функція φ є стійкою тоді й тільки тоді, коли знайдеться послідовність незалежних однаково розподілених величин $(\xi_n, n \geq 0)$ такі, що при кожному n для деяких сталих $a_n, b_n > 0$ має місце слабка збіжність $(\xi_1 + ... + \xi_n - a_n)/b_n \to^W \xi_0$. (г) Стійким характеристичним функціям відповідають нескінченно подільні функції, для яких канонічна міра у зображенні Леві-Хінчина має щільність $C_\pm\,|x|^{1-\alpha}\,1\!\!1_{\,x \lessgtr 0}$ при деякому $\alpha \in (0,2]$.

2.10. Випадкові блукання

Випадкове блукання є найпростішим випадковим процесом.

Означення. Випадковим блуканням $(S_n, n \geq 0)$ *на* $\mathbb{Z}$ *називається послідовність сум незалежних у сукупності однаково розподілених цілозначних випадкових величин* $(\xi_n, n \geq 1)$:

$$S_n \equiv \sum\nolimits_{k=1}^{n} \xi_k, \quad \textit{де за означенням} \quad S_0 = 0.$$

Величину S_n можна трактувати як положення частинки в момент часу n, якщо вона в кожний момент k стрибком зміщується з поточного положення S_{k-1} на величину ξ_k, тобто $S_k = S_{k-1} + \xi_k,\ k \geq 1$.

Послідовність точок (n, S_n) розглядається як траєкторія частинки в координатах час – простір.

Означення. *В момент* $n \geq 1$ *відбувається* повернення *в* 0, *якщо реалізується подія* $U_n \equiv \{S_n = 0\}$.

У момент $n \geq 1$ *відбувається* перше повернення *в* 0, *якщо справджується подія* $F_n \equiv \{S_1 \neq 0, ..., S_{n-1} \neq 0,\ S_n = 0\}$.

Означення. *Повернення в* 0 *відбудеться коли-небудь, якщо реалізується подія* $U \equiv \cup_{n \geq 1} U_n$.

Зауваження. *Справедлива рівність* $U \equiv \cup_{n\geq 1} U_n \ = \cup_{n \geq 1} F_n$, *оскільки повернення коли-небудь є те саме, що повернення коли-небудь вперше.*

На події U визначений час до першого повернення

$$\tau \equiv \inf(n \geq 1 : S_n = 0),$$

а саме: $\tau(\omega) = n$ на елементарних подіях $\omega \in F_n$. Будемо вважати, що $\tau = \infty$ на доповненні U. За означенням $U = \{\tau < \infty\}$.

2.10.1. Рекурентність блукання та її критерій

Означення. *Випадкове блукання* $(S_n, n \geq 0)$ *називається* рекурентним, *якщо* $\mathbf{P}(U) = 1$, *тобто* $\mathbf{P}(\tau < \infty) = 1$. *У протилежному випадку блукання називається* транзієнтним.

Для аналізу рекурентності блукання введемо такі позначення

$$u_n = \mathbf{P}(U_n),\ f_n = \mathbf{P}(F_n),\ u_0 = 1,\ f_0 = 0.$$

Теорема (про рівняння відновлення). *Імовірності повернення* u_n *та ймовірності першого повернення* f_n *задовольняють при* $n \geq 1$ *рівняння відновлення*

$$u_n = \sum\nolimits_{k=1}^{n}\ f_k\ u_{n-k}.$$

Означення. *Права частина рівняння відновлення називається* згорткою послідовностей u_n *та* f_n *і позначається як* $(f * u)_n$.

Доведення. Оскільки $\{S_n = 0\} \subset \{\tau \le n\}$, то справедлива тотожність: $\{S_n = 0\} = \{S_n = 0,\ \tau \le n\}$. Унаслідок *попарної несумісності* подій $\{\tau = k\}$ звідси отримуємо

$$u_n = \mathbf{P}(S_n = 0) = \mathbf{P}(S_n = 0,\ \tau \le n) =$$

$$\textstyle\sum_{k=1}^n \mathbf{P}(\tau = k,\ S_n = 0) = \sum_{k=1}^n \mathbf{P}(\tau = k,\ S_n - S_k = 0) =$$

$$\textstyle\sum_{k=1}^n \mathbf{P}(\tau = k)\, \mathbf{P}(\xi_{k+1} + ... + \xi_n = 0) =$$

$$\textstyle\sum_{k=1}^n \mathbf{P}(\tau = k)\, \mathbf{P}(\xi_1 + ... + \xi_{n-k} = 0) =$$

$$\textstyle\sum_{k=1}^n \mathbf{P}(F_k)\, \mathbf{P}(S_{n-k} = 0) = \sum_{k=1}^n\ f_k\, u_{n-k},$$

де враховано незалежність подій $\{\tau = k\} = \{S_1 \ne 0..., S_{k-1} \ne 0, S_k = 0\}$ та $\{S_n - S_k = 0\} = \{\xi_{k+1} + ... + \xi_n = 0\}$, яка випливає з теореми *про векторні перетворення незалежних величин*, а також *однакова розподіленість* випадкових векторів $(\xi_1, ..., \xi_{n-k})$ та $(\xi_{k+1}, ..., \xi_n)$ $\boxdot$

Означення. Генератрисою послідовності $(a_n, n \ge 0)$ *називається сума степеневого ряду*

$$A(z) = \sum\nolimits_{n \ge 0} a_n z^n.$$

Зауваження. Якщо послідовність $(a_n,\ n \ge 0)$ обмежена, то генератриса визначена та аналітична всередині одиничного кола $\{z : |z| < 1\}$. Для генератрис виконується більшість властивостей *генератрис випадкових величин*.

Теорема (про генератрису ймовірностей повернення). *Якщо послідовності* $(u_n, n \ge 0), (f_n, n \ge 0)$ *задовольняють рівняння відновлення, та* $u_0 = 1, f_0 = 0$, *то для кожного* $z \in \mathbb{Z}$ *з* $|z| < 1$ *генератриси*

$$u(z) \equiv \sum\nolimits_{n \ge 0} u_n z^n,\ f(z) \equiv \sum\nolimits_{n \ge 0} f_n z^n$$

задовольняють рівняння

$$u(z) = 1 + u(z) f(z).$$

Доведення. Помножимо n-те рівняння на z^n та додамо результати

$$\textstyle\sum_{n \ge 1} u_n z^n = u(z) - 1 = \sum_{n \ge 1} z^n \sum_{k=1}^n\ f_k\, u_{n-k} =$$

$$\textstyle\sum_{k=1}^\infty\ f_k z^k \sum_{n \ge k} z^{n-k} u_{n-k} = \sum_{k=1}^\infty\ f_k z^k \sum_{n \ge 0} z^n u_n = f(z) u(z).$$

Збіжність рядів обумовлена обмеженістю послідовностей u_n та f_n $\boxdot$

Теорема (про критерій рекурентності випадкового блукання). *Випадкове блукання* $(S_n,\ n \ge 0)$ *на* $\mathbb{Z}$ *рекурентне тоді й тільки тоді, коли*

$$U \equiv \sum\nolimits_{n\ge 0} \mathbf{P}(S_n = 0) = \infty.$$

У випадку збіжності цього ряду $\mathbf{P}(\tau < \infty) = 1 - 1/U$.

Доведення. Оскільки $f(1) = \sum_{n\ge 0} f_n = \mathbf{P}(\tau < \infty)$, то $f(1) = 1$ тоді й тільки тоді, коли блукання рекурентне.

З іншого боку, за теоремою Абеля про суму ряду з невід'ємними коефіцієнтами (чи за *теоремою Лебега про монотонну збіжність*) має місце тотожність $u(1-) = u(1) \le \infty$. Тому з теореми *про генератрису ймовірностей повернення* при $0 \le z \uparrow 1$ дістанемо

$$u(1) = 1 + u(1)f(1), \quad 1 = 1/u(1) + f(1).$$

Отже, твердження $f(1) < 1$ та $u(1) = U \equiv \sum_{n\,\ge\,0} u_n < \infty$ еквівалентні $\boxdot$

2.10.2. Блукання Бернуллі

Означення. *Випадковим* блуканням Бернуллі $(S_n, n \ge 0)$ *на* $\mathbb{Z}$ *називається блукання зі стрибками* $(\xi_n, n \ge 1)$ *такими, що*

$$\mathbf{P}(\xi_n = 1) = p = 1 - \mathbf{P}(\xi_n = -1) = 1 - q.$$

Теорема (про рекурентність блукання Бернуллі). *Блукання Бернуллі рекурентне тоді й тільки тоді, коли* $p = q = 1/2$.

Доведення. За формулою Стірлінга маємо

$$u_{2n} = C_{2n}^n\, p^n (1-p)^n \sim (4p(1-p))^n / \sqrt{\pi n},\ n \to \infty,$$

де $4p(1-p) < 1$ при $p \ne 1/2$ та $4p(1-p) = 1$ при $p = 1/2$ $\boxdot$

Теорема (про ймовірності повернення). *Для блукання Бернуллі ймовірності першого повернення будь-коли та в момент* $2n$ *дорівнюють*

$$\mathbf{P}(\tau < \infty) = 1 - |p - q|\,, \quad f_{2n} = \frac{2}{2n-1} C_{2n-1}^n (pq)^n.$$

Доведення. Обчислимо при $|z| < 1$ за формулою розкладу в ряд Тейлора *генератрису послідовності ймовірностей повернення*

$$u(z) = \sum\nolimits_{n\ge 0} z^{2n} C_{2n}^n (pq)^n = (1 - 4pqz^2)^{-1/2},$$

звідки за теоремою *про генератрису ймовірностей повернення*

$$f(z) = 1 - 1 \ / \ u(z) = 1 - (1 - 4pqz^2)^{+1/2}.$$

Звідси при $z \uparrow 1$ дістанемо першу тотожність

$$\mathbf{P}(\tau < \infty) = f(1-) = 1 - (1-4pq)^{+1/2} = 1 - |p-q|\,.$$

Другу тотожність обчислюємо за означенням *генератриси послідовності* за допомогою розкладу в ряд Тейлора функції $(1-4pqz^2)^{+1/2}$ $\boxdot$

Вправи

(1, теорема Пойа) Симетричне блукання Бернуллі (з одиничними стрибками) у просторі $\mathbb{Z}^d$ рекурентне тоді й тільки тоді, коли $d \le 2$.

(2) Імовірність того, що блукання Бернуллі $(S_n, n \ge 0)$ коли-небудь досягне точку $x > 0$, дорівнює 1 при $p \ge q$, та $(p/q)^x$ при $p < q$. За останнім припущенням величина $\sup_{n\ge 0} S_n$ має геометричний розподіл з параметром p/q.

(3) Випадкове блукання має такий розподіл стрибків: $\mathbf{P}(\xi_1 = -2) = 1/3$, $\mathbf{P}(\xi_1 = 1) = 2/3$. Довести, що воно є рекурентним.

(4) Блукання Бернуллі $(S_n, n \ge 0)$ таке, що $S_0 = x \in (a,b)$. Позначимо момент виходу $\tau_x = \inf(n \ge 1 : S_n \notin (a,b))$. (а) Скласти систему рівнянь для $p_x = \mathbf{P}(\tau_x < \infty, S_{\tau_x} = b)$ та знайти $p_x = (\theta^x - \theta^a)/(\theta^b - \theta^a)$, якщо $\theta \equiv q/p \ne 1$. (б) Аналогічно знайти $m_x = \mathbf{E}\tau_x = (bp_x + a(1-p_x) - x)/(p-q)$.

2.10.3. Критерій рекурентності через характеристичні функції

Застосуємо до аналізу *рекурентності* поняття *характеристичної функції*. Позначимо через $p_n = P(\xi_1 = n), n \in \mathbb{Z}$, розподіл стрибка і розглянемо його характеристичну функцію:

$$\varphi(t) = \varphi_{\xi_1}(t) = \mathbf{E}\exp(it\xi_1) = \sum\nolimits_{n\in\mathbb{Z}} \exp(itn)p_n,$$

Ця функція має період 2π, оскільки $\xi_1 \in \mathbb{Z}$, та $\exp(2\pi i\xi_1) = 1$.

Лема (про формулу обертання характеристичної функції цілочисельної величини). *Для всіх $a \in \mathbb{Z}$ справедлива формула:*

$$\mathbf{P}(\xi_1 = a) = p_a = (2\pi)^{-1}\int_{-\pi}^{\pi} \exp(-iat)\varphi_{\xi_1}(t)dt.$$

Доведення зводиться до обчислення інтегралу з урахуванням того, що інтеграл від гармонічної функції $\exp(itn)$ на відрізку $[-\pi,\pi]$ дорівнює $2\pi\delta_{n0}$, де δ_{ij} – символ Кронекера $\boxdot$

Означення. Періодом блукання (S_n) *називається число*

$$d = \max\Big(h \ge 1 : \sum\nolimits_{n\in\mathbb{Z}} \mathbf{P}(\xi_1 = nh) = 1\Big) = \text{ н.с.д.}(n \in \mathbb{Z} : p_n > 0).$$

Блукання називається негратчатим, *якщо* $d = 1$.

За означенням періоду $\xi_1 \in d\mathbb{Z} \equiv \{dn, n \in \mathbb{Z}\}$ *м.н.*

Лема (про період випадкового блукання).

(а) Період блукання дорівнює

$$d = 2\pi/r, \text{ де } r = \min(s > 0 : \varphi(s) = 1).$$

(б) Для неґратчатого блукання $\varphi(s) \neq 1$ *при* $0 < |s| < 2\pi$.

Доведення. (а) Доведемо, що $r = 2\pi/d$.

Для $s = 2\pi/d$ маємо $\varphi(s) = \sum_{n\in d\mathbb{Z}} \exp(i(n/d)(sd))p_n = \sum_{n\in d\mathbb{Z}} p_n = 1$.

Нехай $s \in (0, 2\pi/d)$. Досить довести, що $\varphi(s) \neq 1$. Для цього знайдемо

$$\mathrm{Re}(1 - \varphi(s)) = \sum\nolimits_{n\in d\mathbb{Z}}(1 - \cos(ns))p_n = \sum\nolimits_{n\in d\mathbb{Z}} 2\sin^2(ns/2))p_n.$$

Якщо $\varphi(s) = 1$ і $p_n > 0$, то $ns \in 2\pi\mathbb{Z}$ і $n \in d\mathbb{Z}, n \neq 0$, звідки $|sd| \geq 2\pi$, що суперечить зробленому припущенню. При $d = 1$ звідси виводимо (б) $\boxdot$

Лема (про характеристичну функцію стрибка неґратчатого блукання). *Якщо блукання* (S_n) *є неґратчатим, то для деякого* $\alpha > 0$

$$\mathrm{Re}(1 - \varphi(s)) \geq \alpha s^2, \quad \forall\, |s| \leq \pi.$$

Доведення. Позначимо

$$\alpha \equiv \inf(\mathrm{Re}(1 - \varphi(s))/s^2 : |s| \leq \pi).$$

Припустимо, що $\alpha = 0$. Тоді відповідна мінімізуюча послідовність s_n не має 0 як граничну точку, адже за нерівністю з леми *про період випадкового блукання* відповідне граничне значення для деякого $n \in \mathbb{Z}$.додатне:

$$\underline{\lim}_{s\to 0} \mathrm{Re}(1 - \varphi(s))/s^2 \geq \underline{\lim}_{s\to 0} 2\sin^2(ns/2))p_n/s^2 = n^2 p_n/2 > 0.$$

Якщо така гранична точка належить області $\{s : \delta \leq |s| \leq \pi\}$, то внаслідок неперервності функція $\mathrm{Re}(1 - \varphi(s))$ дорівнює нулю в цій точці. Однак за лемою *про період випадкового блукання* та за умовою неґратчатості функція $1 - \varphi(s)$ не має ненульових коренів на $[-\pi, \pi]$. Тому від супротивного $\alpha > 0$ $\boxdot$

Означення. *Часом до першого досягнення точки* $a \in \mathbb{Z}$ *випадковим блуканням* (S_n) *називається узагальнена випадкова величина*

$$\tau_a = \inf(n \geq 1 : S_n = a) \in \mathbb{N} \cup \{\infty\}, \quad \inf(\varnothing) \equiv \infty.$$

Означення. *Позначимо для* $a \in \mathbb{Z}$ *генератрису послідовності*

$$r_a(z) \equiv \sum\nolimits_{n\geq 0} z^n \mathbf{P}(S_n = a).$$

Лема (про генератрису часу до першого досягнення). *Для всіх* $a \in \mathbb{Z}\backslash\{0\}$ *та* $z \in (0,1)$ *мають місце тотожності*

$$\mathbf{E}z^{\tau_a} = r_a(z) \ / \ r_0(z),$$

$$\sum\nolimits_{n\geq 0} z^n \mathbf{P}(S_n = 0, \tau_a \leq n) = r_0(z)\mathbf{E}z^{\tau_a}\mathbf{E}z^{\tau_{-a}}.$$

Доведення. Оскільки події $\{\tau_a = k\}$ та $\{S_n - S_k = 0\}$ незалежні, аналогічно теоремі *про рівняння відновлення* виводимо рівність

$$\mathbf{P}(S_n = a) = \sum\nolimits_{k=1}^{n} \mathbf{P}(\tau_a = k,\ S_n - S_k = 0) = \sum\nolimits_{k=1}^{n} \mathbf{P}(\tau_a = k)\mathbf{P}(S_{n-k} = 0).$$

Після множення обох частин на z^n та підсумовування за $n \geq 0$, аналогічно доведенню теореми *про генератрису часу до першого досягнення* отримуємо першу рівність леми.

Аналогічно з урахуванням першої виводиться друга рівність:

$$\sum\nolimits_{n\geq 0} z^n \mathbf{P}(S_n = 0, \tau_a \leq n) = \sum\nolimits_{n\geq 0} z^n \sum\nolimits_{k=1}^{n} \mathbf{P}(S_n = 0, \tau_a = k) =$$

$$\sum\nolimits_{n\geq 0} z^n \sum\nolimits_{k=1}^{n} \mathbf{P}(S_n - S_k = -a, \tau_a = k) =$$

$$\sum\nolimits_{n\geq 0} z^n \sum\nolimits_{k=1}^{n} \mathbf{P}(S_{n-k} = -a)\mathbf{P}(\tau_a = k) =$$

$$\sum\nolimits_{k\geq 0} z^k \mathbf{P}(\tau_a = k) \sum\nolimits_{n\geq k} z^{n-k}\mathbf{P}(S_{n-k} = -a) = \mathbf{E}z^{\tau_a} r_{-a}(z) = \mathbf{E}z^{\tau_a}\mathbf{E}z^{\tau_a} r_0(z) \boxdot$$

Теорема (про критерій рекурентності блукання через характеристичну функцію). *Нехай випадкове блукання* (S_n) *на* $\mathbb{Z}$ *є неґратчатим, а* $\varphi(t)$ *– характеристична функція стрибка. Для його рекурентності необхідне і достатнє виконання однієї з еквівалентних умов:*

$$\text{(а)} \quad \lim\nolimits_{z\uparrow 1} \int_{-\pi}^{\pi} \mathrm{Re}(1 - z\ \varphi(t))^{-1} dt = \infty,$$

$$\text{(б)} \quad \int_{-\pi}^{\pi} \mathrm{Re}(1 - \varphi(t))^{-1} dt = \infty,$$

причому кожний з цих інтегралів можна обчислювати у будь-якому околі нуля.

Доведення. Введемо при $0 < z \leq 1, t \in [-\pi, \pi]$ позначення

$$R_z(t) = \mathrm{Re}(1 - z\varphi(t))^{-1}.$$

Зауважимо, що функції $R_z(t) = (1 - z\,\mathrm{Re}\,\varphi(t))\,|1 - z\varphi(t)|^{-2}$ невід'ємні при $0 < z \leq 1$, оскільки $|\varphi(t)| \leq 1$. Крім того, за лемою *про період випадкового блукання* та з неперервності характеристичної функції робимо висновок, що функція $R_1(t)$ неперервна та обмежена на $[-\pi, \pi]$ поза будь яким околом нуля. Звідси, зокрема, випливає обґрунтованість останнього твердження теореми.

(а) За теоремою *про властивості характеристичної функції* обчислимо

$$\varphi_{S_n}(t) = \mathbf{E}\exp(itS_n) = \mathbf{E}\exp(it(\xi_1 + ... + \xi_n)) = \varphi^n(t).$$

Звідси за лемою *про формулу обертання характеристичної функції цілочисельної величини*

$$\mathbf{P}(S_n = 0) = (2\pi)^{-1} \int_{-\pi}^{\pi} \exp(-i0t)\varphi^n(t)dt.$$

Множенням на z^n та обчисленням суми при $n \geq 0$ звідси виводимо

$$r_0(z) = (2\pi)^{-1} \sum\nolimits_{n\geq 0} z^n \int_{-\pi}^{\pi} \varphi^n(t)dt = (2\pi)^{-1} \int_{-\pi}^{\pi} (1 - z\varphi(t))^{-1}dt,$$

де підінтегральна функція неперервна та обмежена при $0 < z < 1$, внаслідок неперервності характеристичної функції та оцінок $|1 - z\varphi(t)| \geq 1 - z\,|\varphi(t)| \geq 1 - z > 0$. Крім того, оскільки функція $r_0(z)$ – дійсна, то інтеграл від уявної частини справа дорівнює нулю. Тому підінтегральну функцію у наведеній тотожності можна замінити на її дійсну частину:

$$r_0(z) = (2\pi)^{-1} \int_{-\pi}^{\pi} R_z(t)dt.$$

Тому за *теоремою Лебега про монотонну збіжність* при $z \uparrow 1$ отримуємо тотожність

$$\lim\nolimits_{z\uparrow 1} r_0(z) = \lim\nolimits_{z\uparrow 1} \sum\nolimits_{n\geq 0} z^n\mathbf{P}(S_n = 0) = \sum\nolimits_{n\geq 0} \mathbf{P}(S_n = 0) = r_0(1) \leq \infty.$$

Отже, ліва та права її частини є нескінченними одночасно. За теоремою *про критерій рекурентності випадкового блукання* рекурентність еквівалентна розбіжності ряду $\sum_{n\geq 0} \mathbf{P}(S_n = 0)$, тому твердження (а) доведене.

(б) *Достатність*. Скористаємося зауваженням до *теореми Лебега про мажоровану збіжність* та відзначеною вище невід'ємністю $R_z(t)$:

$$\underline{\lim}_{z\uparrow 1} \int_{-\pi}^{\pi} R_z(t)dt \geq \int_{-\pi}^{\pi} \underline{\lim}_{z\uparrow 1} R_z(t)dt = \int_{-\pi}^{\pi} R_1(t)dt = \infty.$$

за умови (б). Отже, за твердженням достатності (а) блукання рекурентне.

Необхідність. Для доведення від супротивного припустимо, що функція $R_1(t)$ інтегровна:

$$\int_{-\pi}^{\pi} R_1(t)dt < \infty.$$

Нехай $a \in \mathbb{Z}\backslash\{0\}, z \in (0,1)$. Обчислимо з урахуванням леми *про генератрису часу до першого досягнення* та леми *про формулу обертання характеристичної функції цілочисельної величини* S_n :

$$\sum\nolimits_{n\geq 0} z^n\mathbf{P}(S_n = 0, \tau_a > n) = r_a(z) - \sum\nolimits_{n\geq 0} z^n\mathbf{P}(S_n = 0, \tau_a \leq n) =$$

$$r_0(z)\,(1-\mathbf{E}z^{\tau_a}\mathbf{E}z^{\tau_{-a}}) \le r_0(z)\,(2-\mathbf{E}z^{\tau_a}-\mathbf{E}z^{\tau_{-a}}) =$$
$$2r_0(z)-r_a(z)-r_{-a}(z) =$$
$$\textstyle\sum_{n\ge 0} z^n\,(2\mathbf{P}(S_n=0)-\mathbf{P}(S_n=a)-\mathbf{P}(S_n=-a)) =$$
$$(2\pi)^{-1}\textstyle\int_{-\pi}^{\pi}\sum_{n\ge 0} z^n\,(2-\exp(-iat)-\exp(iat))\,\varphi^n(t)dt =$$
$$(2\pi)^{-1}\textstyle\int_{-\pi}^{\pi} 2(1-\cos at)(1-z\varphi(t))^{-1}dt = \pi^{-1}\int_{-\pi}^{\pi}\frac{1-\ \cos at}{1-\varphi(t)}\frac{1-\varphi(t)}{1-z\varphi(t)}dt.$$

За лемою *про характеристичну функцію стрибка негратчатого блукання* перший дріб під знаком останнього інтеграла обмежений на $[-\pi,\pi]$. Крім того, модуль функції $(1-u)(1-zu)^{-1}$ не перевищує 2 на одиничному колі $\{u\in\mathbb{C}:|u|\le 1\}$ при $0<z\le 1$. Тому, переходячи до границі в попередній нерівності при $z\uparrow 1$, з *теореми Лебега про мажоровану збіжність* отримуємо

$$\sum\nolimits_{n\ge 0}\mathbf{P}(S_n=0,\tau_a>n)\le\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{1-\cos at}{1-\varphi(t)}dt=\frac{1}{\pi}\int_{-\pi}^{\pi}(1-\cos at)R_1(t)dt,$$

де остання рівність виконується, оскільки її ліва частина та перший множник під знаком інтегралу є дійсними.

Оскільки при кожному фіксованому $n>0$

$$\mathbf{P}(S_n=0,\tau_a\le n)\le\sum\nolimits_{k\le n}\mathbf{P}(S_k=a)\to 0,\ a\to\infty,$$

то $\mathbf{P}(S_n=0,\tau_a>n)\to\mathbf{P}(S_n=0),\ a\to\infty$. Тому при кожному $N>0$

$$\textstyle\sum_{n\le N}\mathbf{P}(S_n=0)=\lim_{a\to\infty}\sum_{n\le N}\mathbf{P}(S_n=0,\tau_a>n)\le$$
$$\textstyle\pi^{-1}\lim_{a\to\infty}\int_{-\pi}^{\pi}(1-\cos at)R_1(t)dt=\pi^{-1}\int_{-\pi}^{\pi}R_1(t)dt,$$

де остання рівність є наслідком інтегровності функції $R_1(t)$ та леми Рімана – Лебега з курсу математичного аналізу. Спрямовуючи тут $N\to\infty$, переконуємося у збіжності ряду $\sum_{n\ge 0}\mathbf{P}(S_n=0)\le\pi^{-1}\int_{-\pi}^{\pi}R_1(t)dt<\infty$. Отримали суперечність з рекурентністю за теоремою *про критерій рекурентності випадкового блукання*, що і доводить (б) $\boxdot$

Теорема (про рекурентність блукання з інтегровними стрибками). *Нехай* $(S_n, n\ge 1)$ *– негратчате випадкове блукання на* $\mathbb{Z}$, *що має інтегровні стрибки:* $\mathbf{E}\,|\xi_1|<\infty$. *Це блукання є рекурентним тоді й тільки тоді, коли* $\mathbf{E}\xi_1=0$.

Доведення. Позначимо $\mu=\mathbf{E}\xi_1$, $p_n=P(\xi_1=n)$, $\varphi(t)=\mathbf{E}\exp(it\xi_1)$, $u(t)=\mathrm{Re}\,\varphi(t)$, $v(t)=\mathrm{Im}\ \varphi(t)$. За теоремою *про властивості характеристичної функції* $\varphi(s)=1+i\mu s+o(s),\ s\to 0$, звідки $u(s)=\mathbf{E}\cos(s\xi_1)=1+o(s)$,

$v(s) = \mu s + o(s),\ s \to 0$. Далі, $1 - u(s) \geq \alpha s^2$ при деякому $\alpha > 0$ для $|s| \leq \pi$ за лемою *про характеристичну функцію стрибка неґратчатого блукання.*

За означенням має місце тотожність

$$R_z(t) = \mathrm{Re}(1 - z\varphi(t))^{-1} = (1 - zu(t))\left\{(1 - zu(t))^2 + z^2 v^2(t)\right\}^{-1}.$$

Необхідність. Нехай $\mu \neq 0$. Обчислимо з урахуванням останнього зображення при $z = 1$:

$$R_1(t) = (1 - \mathbf{E}\cos(t\xi_1))/(o(t^2) + \mu^2 t^2) =$$
$$\mu^{-2} t^{-2} \textstyle\sum_{n\in\mathbb{Z}} (1 - \cos(tn)) p_n + o(1),\ t \to 0.$$

Як зазначено у теоремі *про критерій рекурентності блукання через характеристичну функцію,* з умовою (б), для доведення відсутності рекурентності досить перевірити інтегровність $R_1(t)$ у деякому околі нуля. Оскільки функція $s^{-2}(1 - \cos(s))$ абсолютно інтегровна на $\mathbb{R}$, то для досить малого $\delta > 0$

$$\textstyle\int_{|t|\leq\delta} R_1(t) dt \leq \mu^{-2} \sum_{n\in\mathbb{Z}} \int_{|t|\leq\delta} t^{-2}(1 - \cos(tn)) p_n\ dt + 1 =$$
$$\mu^{-2} \textstyle\sum_{n\ \in\ \mathbb{Z}} |n|\, p_n \int_{|s|\leq|n|\delta} s^{-2}(1 - \cos(s))\ dt + 1 < \infty,$$

де враховано абсолютну збіжність ряду $\sum_{\ n\in\mathbb{Z}} np_n$, згідно з інтегровністю ξ_1, при внесенні інтеграла під знак суми. Отже, блукання не є рекурентним, що і доводить необхідність від супротивного.

Достатність. Припустимо, що $\mu = 0$.

Тоді $(1 - u(t))^2 + v^2(t) = o(t^2),\ t \to 0$. Тому

$$\varepsilon^2(\delta) \equiv \inf_{\ |t|\leq\delta}\ t^{-2}((1 - u(t))^2 + v^2(t)) = o(1),\ \delta \to 0.$$

Крім того, для кожного $z \in (0, 1)$ з *нерівності Коші* виводимо, що

$$(1 - zu(t))^2 + z^2 v^2(t) = (1 - z + z(1 - u(t)))^2 + z^2 v^2(t) =$$
$$z^2\left((1 - u(t))^2 + v^2(t)\right) + (1 - z)^2 + 2z(1 - z)(1 - u(t)) \leq$$
$$2z^2\left((1 - u(t))^2 + v^2(t)\right) + 2(1 - z)^2.$$

Звідси та з урахуванням означення $\varepsilon(\delta)$

$$\textstyle\int_{|t|\leq\delta} R_z(t) dt = \int_{|t|\leq\delta} (1 - zu(t))\left\{(1 - zu(t))^2 + z^2 v^2(t)\right\}^{-1} dt \geq$$
$$\tfrac{1}{2} \textstyle\int_{|t|\leq\delta} (1 - z)\left\{z^2\left((1 - u(t))^2 + v^2(t)\right) + (1 - z)^2\right\}^{-1} dt \geq$$
$$\tfrac{1}{2} \textstyle\int_{|t|\leq\delta} (1 - z)\left\{z^2 t^2 \varepsilon^2(\delta) + (1 - z)^2\right\}^{-1} dt =$$
$$(z\varepsilon(\delta))^{-1} \arctan(\delta z \varepsilon(\delta)/(1 - z)).$$

Тому з урахуванням невід'ємності $R_z(t)$

$$\lim\nolimits_{z\uparrow 1}\int_{|t|\le\pi} R_z(t)dt \ge \lim\nolimits_{z\uparrow 1}\int_{|t|\le\delta} R_z(t)dt \ge \varepsilon^{-1}(\delta)\arctan(\infty) = \varepsilon^{-1}(\delta)\pi/2.$$

Оскільки ліва частина не залежить від δ, то

$$\lim\nolimits_{\ z\uparrow 1}\int_{|t|\le\pi} R_z(t)dt \ge \lim\nolimits_{\delta\to 0}\varepsilon^{-1}(\delta)\pi/2 = \infty.$$

Отже, блукання є рекурентним за умовою (а) теореми *про критерій рекурентності блукання через характеристичну функцію* $\boxdot$

Вправи

(1) Стрибки випадкового блукання мають симетричний розподіл, причому $\mathbf{P}(\xi_1 = n) \sim cn^{-\alpha}, n \to \infty$. Довести, що при $1 < \alpha < 2$ блукання транзієнтне, а при $\alpha > 2$ – рекурентне.

(2) Зі збіжності $\int_{-\pi}^{\pi}(1-\mathrm{Re}\,\varphi(t))^{-1}dt < \infty$ випливає транзієнтність блукання.

(3) Нехай $(S_n, n \ge 0)$ – випадкове блукання зі стійким розподілом стрибків, що мають характеристичну функцію $\exp(-|t|^{\alpha}), \alpha \in (0, 2]$. Довести, що це блукання є рекурентним тоді й тільки тоді, коли $\alpha \ge 1$.

(4) Довести, що клас симетричних розподілів $q = (\mathbf{P}(\xi_1 = n), n \in \mathbb{Z})$, для яких відповідне випадкове блукання є рекурентним, є опуклою множиною (для кожних двох своїх точок q_k містить відрізок $\alpha q_1 + (1-\alpha)q_2, \alpha \in (0, 1)$).

(5) Узагальнити теорему про критерій рекурентності блукання на випадок блукань у просторі $\mathbb{R}^d$.

2.11. Граничні задачі для випадкових блукань

У теорії граничних задач вивчаються так звані граничні функціонали від випадкових процесів, а саме: їх послідовні максимуми, мінімуми, та відповідні моменти досягнення. Відомий з комплексного аналізу метод факторизації Вінера-Хопфа дозволяє знайти низку аналітичних співвідношень для характеристичних функцій граничних функціоналів від випадкових блукань.

У даному розділі будемо розглядати послідовність $(\xi_n, n \ge 1)$ незалежних однаково розподілених випадкових величин. Їх суми позначатимемо через $S_n = \xi_1 + ... + \xi_n$, $S_0 = 0$, а максимуми та мінімуми сум – через

$$\mu_n = \max\nolimits_{1\le k\le n} S_k,\ \ \mu_n^* = \min\nolimits_{1\le k\le n} S_k.$$

Нехай $\varphi(t) = \mathbf{E}\exp(it\xi_1)$ – характеристична функція одного доданка.

2.11.1. Граничні функціонали та факторизаційні тотожності

Розглянемо такі граничні функціонали:

$$\tau_+ = \inf(n \ge 1 : S_n \ge 0), \quad \tau_- = \inf(n \ge 1 : S_n \le 0),$$

$$\overline{\tau}_+ = \inf(n \ge 1 : S_n > 0), \quad \overline{\tau}_- = \inf(n \ge 1 : S_n < 0),$$

що називаються відповідно верхніми і нижніми **рекордними моментами**, та верхніми і нижніми **строгими рекордними моментами**. Тут за означенням $\inf \varnothing = \infty$.

На подіях, де скінченні відповідні моменти: $\{\tau_\pm < \infty\}$ або $\{\overline{\tau}_\pm < \infty\}$, коректно визначені також **рекордні висоти** та **строгі рекордні висоти**:

$$\sigma_+ = S_{\tau_+} 1\!\!1_{\{\tau_+ < \infty\}}, \quad \sigma_- = S_{\tau_-} 1\!\!1_{\{\tau_- < \infty\}},$$

$$\overline{\sigma}_+ = S_{\overline{\tau}_+} 1\!\!1_{\{\overline{\tau}_+ < \infty\}}, \quad \overline{\sigma}_- = S_{\overline{\tau}_-} 1\!\!1_{\{\overline{\tau}_- < \infty\}}.$$

Надалі у випадку, коли формулювання теорем та означень міститимуть символи ($\pm$, $\mp$, $\lesseqgtr$, $\gtreqless$, $\substack{\le \\ >}$, $\substack{\ge \\ <}$), їх слід розуміти як пару тверджень: для ($+$, $-$, $<$, $>$, $\le$, $\ge$) та ($-$, $+$, $>$, $<$, $\ge$, $\le$).

Лема (про властивості рекордних висот та моментів).

(а) На подіях $\{\tau_\pm < \infty\}$ та $\{\overline{\tau}_\pm < \infty\}$ виконуються відповідно нерівності $\{\sigma_\pm \substack{\ge \\ <} 0\}$ та $\{\overline{\sigma}_\pm \gtrless 0\}$.

(б) Для кожного $n \ge 1$ справедливі тотожності

$$\{\tau_+ > n\} = \{\mu_n < 0\}, \quad \{\tau_- > n\} = \{\mu_n^* > 0\},$$

$$\{\overline{\tau}_+ > n\} = \{\mu_n \le 0\}, \quad \{\overline{\tau}_- > n\} = \{\mu_n^* \ge 0\}.$$

Доведення очевидне $\boxdot$

Визначимо при $t \in \mathbb{R}, z \in \mathbb{C}$ з $|z| < 1$ такі сумісні перетворення:

$$\varphi_\pm(t,z) = \mathbf{E}\left(\exp(it\sigma_\pm) z^{\tau_\pm} 1\!\!1_{\{\tau_\pm < \infty\}}\right),$$

$$\psi_\pm(t,z) = \textstyle\sum_{n \ge 0} z^n \mathbf{E}\left(\exp(itS_n) 1\!\!1_{\{\tau_\mp > n\}}\right),$$

$$\varkappa_\pm(z) \equiv 1 - \mathbf{E}(z^{\tau_\pm} 1\!\!1_{\{\sigma_\pm = 0, \tau_\pm < \infty\}}),$$

$$\overline{\varphi}_\pm(t,z) = \mathbf{E}\left(\exp(it\overline{\sigma}_\pm) z^{\overline{\tau}_\pm} 1\!\!1_{\{\overline{\tau}_\pm < \infty\}}\right),$$

$$\overline{\psi}_\pm(t,z) = \textstyle\sum_{n \ge 0} z^n \mathbf{E}\left(\exp(itS_n) 1\!\!1_{\{\overline{\tau}_\mp > n\}}\right),$$

$$\overline{\varkappa}_\pm(z) \equiv \textstyle\sum_{n \ge 0} z^n \mathbf{P}(S_n = 0, \ \overline{\tau}_\mp > n),$$

які є сумами абсолютно збіжних при $|z| < 1$ рядів внаслідок обмеженості гармонічної функції від чисто уявного аргументу it.

Лема (про аналітичне продовження перетворень від граничних функціоналів). *При кожному* $z \in \mathbb{C}$ *з* $|z| < 1$ *функції* $\varphi_{\pm}(t,z)$, $\psi_{\pm}(t,z)$, $\overline{\varphi}_{\pm}(t,z)$, $\overline{\psi}_{\pm}(t,z)$ *мають аналітичне та обмежене продовження на напівплощину* $\{t \in \mathbb{Z} : \mathrm{Im}\ t \gtrless 0\}$, *та є неперервними на її границі.*

Крім того, $\psi_{\pm}(\pm i\infty, z) = 1$ *та* $\overline{\varphi}_{\pm}(\pm i\infty, z) = 0$.

Доведення. Зауважимо, що для кожного $x \gtrless 0$ функція $\exp(itx)$ не перевищує 1 за абсолютною величиною та є аналітичною всередині області $D_{\pm} \equiv \{t \in \mathbb{Z} : \mathrm{Im}\ t \ \gtrless 0\}$.

За лемою *про властивості рекордних висот та моментів* $\sigma_+ \geq 0$ на множині $\{\tau_+ < \infty\}$. Тому функція $\mathbf{E}\left(\exp(it\sigma_+)1\!\!1_{\{\tau_+<\infty\}}\right)$ є аналітичною і обмеженою при $\mathrm{Im}\ t \geq 0$ та неперервною на границі $\mathrm{Im}\ t = 0$. Унаслідок обмеженості $|z^{\tau_+}| \leq 1$ таку ж властивість має функція $\varphi_+(t,z)$.

Аналогічно перевіряємо аналітичність і обмеженість у D_+ функції $\overline{\varphi}_+(t,z)$. Крім того, за лемою *про властивості рекордних висот та моментів* $\overline{\sigma}_+ > 0$ на події $\{\overline{\tau}_+ < \infty\}$. Отже, за *теоремою Лебега про мажоровану збіжність* $\overline{\varphi}_+(+is,z) = \mathbf{E}\left(\exp(-s\overline{\sigma}_+)z^{\overline{\tau}_+}1\!\!1_{\{\overline{\tau}_+<\infty\}}\right) \to 0, s \to \infty$. Тому $\overline{\varphi}_+(+i\infty, z) = 0$.

Далі, за означенням моменту τ_- на множині $\{\tau_- > n\}$ при $n \geq 1$ виконується нерівність $S_n > 0$. Тому доданок $\mathbf{E}\left(\exp(itS_n)1\!\!1_{\{\tau_->n\}}\right)$ в означенні функції $\psi_+(t,z)$ аналітичний і обмежений в області $t \in D_+$ та неперервний на його границі. Отже, внаслідок рівномірної збіжності відповідного ряду при $|z| < 1$ функція $\psi_+(t,z)$ також має такі самі властивості. Одночасно при $t = is$, $s \to +\infty$ вказаний доданок прямує до нуля: $\mathbf{E}\left(\exp(-sS_n)1\!\!1_{\{\tau_->n\}}\right) \to 0$, для всіх $n \geq 1$, оскільки величина S_n строго додатна на множині інтегрування. Звідси виводимо рівність $\psi_+(i\infty, z) = 1$.

Аналогічно перевіряються всі інші твердження $\boxdot$

Теорема (про факторизаційні тотожності через перетворення рекордних функціоналів). *Для всіх* $t \in \mathbb{R}$, *та* $z \in \mathbb{C}$ *з* $|z| < 1$ *справедливі такі тотожності:*

(а) $(1 - \varphi_+(t,z)) \ / \ \psi_-(t,z) = 1 - z\varphi(t) = (1 - \varphi_-(t,z)) \ / \ \psi_+(t,z)$,

(б) $(1 - \overline{\varphi}_+(t,z)) \ / \ \overline{\psi}_-(t,z) = 1 - z\varphi(t) = (1 - \overline{\varphi}_-(t,z)) \ / \ \overline{\psi}_+(t,z)$.

Доведення

(а) Оскільки доданки (ξ_n) незалежні та однаково розподілені, то за теоремою *про властивості характеристичної функції*

$$\varphi^n(t) = \mathbf{E}\exp(itS_n) =$$

$$\mathbf{E}(\exp(itS_n)1\!\!1_{\{\tau_+>n\}}) + \textstyle\sum_{k=1}^n \mathbf{E}(\exp(itS_n)1\!\!1_{\{\tau_+=k\}}) =$$

$$\mathbf{E}(\exp(itS_n)1\!\!1_{\{\tau_+>n\}}) + \textstyle\sum_{k=1}^n \mathbf{E}(\exp(it(S_n - S_k))\exp(itS_k)1\!\!1_{\{\tau_+=k\}}) =$$
$$\mathbf{E}(\exp(itS_n)1\!\!1_{\{\tau_+>n\}}) + \textstyle\sum_{k=1}^n \mathbf{E}(\exp(it(S_n - S_k))\mathbf{E}(\exp(itS_k)1\!\!1_{\{\tau_+=k\}}) =$$
$$\mathbf{E}(\exp(itS_n)1\!\!1_{\{\tau_+>n\}}) + \textstyle\sum_{k=1}^n \varphi^{n-k}(t)\exp(itS_{\tau_+}1\!\!1_{\{\tau_+=k\}}).$$

Тут використано незалежність випадкових величин $\exp(itS_k)1\!\!1_{\{\tau_+=k\}}$ та $S_n - S_k$, що випливає з теореми *про векторні перетворення незалежних величин*, оскільки вказані величини є функціями відповідно векторів $(\xi_1, ..., \xi_k)$ та $(\xi_{k+1} + ... + \xi_n)$. Одночасно врахована однакова розподіленість $S_n - S_k \simeq S_{n-k}$, що випливає з незалежності та однакової розподіленості доданків (ξ_n).

Множенням останньої тотожності на z^n та підсумовуванням по $n \geq 0$ отримуємо

$$(1 - z\varphi(t))^{-1} = \textstyle\sum_{n\geq 0} z^n \mathbf{E}(\exp(itS_n)1\!\!1_{\{\tau_+>n\}}) +$$
$$\textstyle\sum_{n\geq 0} z^n \sum_{k=1}^n \varphi^{n-k}(t)\ \mathbf{E}\exp(itS_{\tau_+}1\!\!1_{\{\tau_+=k\}}) =$$
$$\psi_-(t,z) + \textstyle\sum_{k\geq 0} z^k \mathbf{E}\exp(itS_{\tau_+}1\!\!1_{\{\tau_+=k\}})\ \sum_{n\geq k} z^{n-k}\varphi^{n-k}(t) =$$
$$\psi_-(t,z) + \varphi_+(t,z)(1 - z\varphi(t))^{-1},$$

де враховано абсолютна збіжність рядів внаслідок $|z| < 1$ та обмеженості характеристичних функцій. Цим доведено першу тотожність (а) леми. Друга тотожність доводиться аналогічно.

(б) Застосування замість τ_+ моменту τ_- у попередніх рівностях призводить до тотожності, що еквівалентна рівностям (б):

$$(1 - z\varphi(t))^{-1} = \psi_+(t,z) + \varphi_-(t,z)(1 - z\varphi(t))^{-1} \boxdot$$

Теорема (про співвідношення для перетворень від граничних функціоналів). *Для всіх* $t \in \mathbb{Z}$ *з* $\mathrm{Im}\ t \geqslant 0$, *та* $z \in \mathbb{C}$ *з* $|z| < 1$ *виконуються тотожності*

(а) $$(1 - \varphi_\pm(t,z))\ \psi_\pm(t,z) = \varkappa_+(z) = \varkappa_-(z),$$

(б) $$(1 - \overline{\varphi}_\pm(t,z))\ \overline{\psi}_\pm(t,z) = \overline{\varkappa}_+(z) = \overline{\varkappa}_-(z),$$

(в) $$(1 - \varphi_\pm(t,z))\ \overline{\psi}_\pm(t,z) = 1,$$

(г) $$(1 - \overline{\varphi}_\mp(t,z))\ \psi_\mp(t,z) = 1,$$

(д) $$\overline{\varkappa}_\pm(z) = 1/\varkappa_\pm(z),$$

(е) $$1 - \overline{\varphi}_\pm(t,z) = (1 - \varphi_\pm(t,z))\ /\ \varkappa_\pm(z).$$

Доведення

(а) З тверджень (а) теореми *про факторизаційні тотожності через перетворення рекордних функціоналів* виводимо, що для всіх $t \in \mathbb{R}$, $|z| < 1$ має місце тотожність

$$(1 - \varphi_+(t,z))\ \psi_+(t,z) = (1 - \varphi_-(t,z))\ \psi_-(t,z).$$

За лемою *про аналітичне продовження перетворень від граничних функціоналів* ліва частина цієї тотожності аналітична при $\{t \in \mathbb{Z} : \mathrm{Im}\ t > 0\}$ та неперервна на границі $\{t \in \mathbb{Z} : \mathrm{Im}\ t = 0\}$, а права частина – аналітична в області $\{t \in \mathbb{Z} : \mathrm{Im}\ t < 0\}$ та неперервна на границі $\{t \in \mathbb{Z} : \mathrm{Im}\ t = 0\}$. Оскільки при $t \in \mathbb{R}$ вказані функції збігаються, то за теоремою про єдиність аналітичного продовження вони є відповідними звуженнями на $\mathbb{R}$ цілої функції. Унаслідок леми *про аналітичне продовження перетворень від граничних функціоналів* ця функція – обмежена, тому вона є сталою і не залежить від t. Позначимо цю сталу через $\varkappa(z)$. Тоді для всіх $t \in \mathbb{C}, |z| < 1$ виводимо, що

$$(1 - \varphi_+(t,z))\ \psi_+(t,z) = \varkappa(z) = (1 - \varphi_-(t,z))\ \psi_-(t,z).$$

Підставимо у ліву рівність $t = is$ і спрямуємо $s \to \infty$. З леми *про аналітичне продовження перетворень від граничних функціоналів* маємо

$$\varkappa(z) = \psi_+(+i\infty, z)\ \lim_{s\to\infty}(1 - \varphi_+(is,z)) =$$
$$1 - \lim_{s\to\infty} \mathbf{E}\left(\exp(-s\sigma_+) z^{\tau_+} 1\!\mathrm{I}_{\{\tau_+<\infty\}}\right) =$$
$$1 - \mathbf{E}(z^{\tau_+} 1\!\mathrm{I}_{\{\sigma_+=0,\tau_+<\infty\}}) = \varkappa_+(z),$$

де передостання рівність випливає з *теореми Лебега про мажоровану збіжність*, оскільки $\lim_{s\to\infty} \mathbf{E}\left(\exp(-s\sigma_+) z^{\tau_+} 1\!\mathrm{I}_{\{\tau_+<\infty\}} 1\!\mathrm{I}_{\{\sigma_+>0\}}\right) = 0$.

Так само граничним переходом $t = is, s \to -\infty$ отримуємо з правої тотожності рівність $\varkappa(z) = \varkappa_-(z)$.

(б) З рівностей (б) теореми *про факторизаційні тотожності через перетворення рекордних функціоналів* виводимо для всіх $t \in \mathbb{R}, |z| < 1$ тотожності

$$(1 - \overline{\varphi}_+(t,z))\ \overline{\psi}_+(t,z) = 1 - z\varphi(t) = (1 - \overline{\varphi}_-(t,z))\ \overline{\psi}_-(t,z).$$

Як і в (а), звідси виводимо існування сталої $\overline{\varkappa}(z)$ такої, що для всіх $t \in \mathbb{C}, |z| < 1$ справедливі тотожності:

$$(1 - \overline{\varphi}_+(t,z))\ \overline{\psi}_+(t,z) = \overline{\varkappa}(z) = (1 - \overline{\varphi}_-(t,z))\ \overline{\psi}_-(t,z).$$

Підставимо у ліву рівність $t = is$ і спрямуємо $s \to \infty$. З леми *про аналітичне продовження перетворень від граничних функціоналів* маємо

$$\overline{\varkappa}(z) = \big(1 - \overline{\varphi}_+(+i\infty, z)\big) \lim_{s\to\infty} \overline{\psi}_+(t, z) =$$
$$\sum\nolimits_{n\ge 0} z^n \lim_{s\to\infty} \mathbf{E}\left(\exp(-sS_n)1\!\!1_{\{\overline{\tau}_->n\}}\right) =$$
$$\sum\nolimits_{n\ge 0} z^n \mathbf{P}(S_n = 0,\ \overline{\tau}_- > n) = \overline{\varkappa}_+(z),$$

де використано також *теорему Лебега про мажоровану збіжність*.

Так само граничним переходом $t = is, s \to -\infty$ отримуємо з правої тотожності рівність $\overline{\varkappa}(z) = \overline{\varkappa}_-(z)$.

(в,г,д) З порівняння тотожностей (а) та (б) теореми *про факторизаційні тотожності через перетворення рекордних функціоналів* виводимо, що

$$(1 - \varphi_+(t, z))\ /\ \psi_-(t, z) = 1 - z\varphi(t) = (1 - \overline{\varphi}_-(t, z))\ /\ \overline{\psi}_+(t, z),$$
$$(1 - \varphi_+(t, z))\ \overline{\psi}_+(t, z) = \widetilde{\varkappa}(z) = (1 - \overline{\varphi}_-(t, z))\ \psi_-(t, z)$$

при всіх $t \in \mathbb{C}$ для деякої сталої $\widetilde{\varkappa}(z)$. Підставимо сюди вже доведені тотожності (а),(б):

$$(1 - \varphi_+(t, z))\ \overline{\psi}_+(t, z) = \widetilde{\varkappa}(z) = (\overline{\varkappa}_-(z)/\overline{\psi}_-(t, z))(\varkappa_-(z)/(1 - \varphi_-(t, z))).$$

Спрямувавши тут $t \to \pm i\infty$, за лемою *про аналітичне продовження перетворень від граничних функціоналів* отримуємо

$$1 = \widetilde{\varkappa}(z) = (\overline{\varkappa}_-(z))(\varkappa_-(z)).$$

З останньої тотожності та попередніх виводимо твердження (в,г,д).

(е) З правої рівності (в) та з рівності (а) виводимо, що

$$1/(1 - \overline{\varphi}_+(t, z)) = \psi_+(t, z) = \varkappa_+(z)/(1 - \varphi_+(t, z)) \quad \boxdot$$

Теорема (про факторизаційні тотожності для граничних функціоналів). *Для всіх* $t \in \mathbb{R}$, $z \in \mathbb{C}$ *з* $|z| < 1$ *справедливі такі рівності:*

$$\text{(а)}\quad 1 - z\varphi(t) = (1 - \varphi_+(t, z))(1 - \varphi_-(t, z))\ /\ \varkappa_+(z),$$
$$\text{(б)}\quad 1 - z\varphi(t) = (1 - \overline{\varphi}_+(t, z))(1 - \overline{\varphi}_-(t, z))\ /\ \overline{\varkappa}_+(z),$$
$$\text{(в)}\quad 1 - z\varphi(t) = (1 - \varphi_+(t, z))(1 - \overline{\varphi}_-(t, z)),$$
$$\text{(г)}\quad 1 - z\varphi(t) = (1 - \varphi_-(t, z))(1 - \overline{\varphi}_+(t, z)).$$

Доведення

(а) Підставимо у ліву тотожність (а) теореми *про факторизаційні тотожності через перетворення рекордних функціоналів* рівність (а) теореми *про співвідношення для перетворень від граничних функціоналів*:

$$1 - z\varphi(t) = (1 - \varphi_+(t, z))\ /\ \psi_-(t, z) = (1 - \varphi_+(t, z))\ /(\varkappa_+(z)/(1 - \varphi_-(t, z))),$$

що і доводить співвідношення (а).

(б) Доведення цілком аналогічне (а).

(в) Дана тотожність виводиться з (а) та твердження (е) теореми *про співвідношення для перетворень від граничних функціоналів.* Аналогічне доведення (г) $\boxdot$

Наслідок (про максимум блукання з інтегровними стрибками). *Нехай* $\mathbf{E}\,|\xi_1| < \infty$.

(а) Якщо $\mathbf{E}\xi_1 < 0$, *то* $\mu_\infty \equiv \sup_{k\ge 1} S_k < \infty$ *м.н., та*

$$\mathbf{E}\sigma_- = \mathbf{E}\xi_1/\mathbf{P}(\mu_\infty \le 0).$$

(б) Якщо $\mathbf{E}\xi_1 = 0$, *то* $\mu_\infty > 0$ *м.н., та* $\overline{\tau}_+ < \infty$ *м.н. і* $\tau_- < \infty$ *м.н.*

Доведення

(а) З теореми *про факторизаційні тотожності для граничних функціоналів,* (г), граничним переходом $z \uparrow 1$ за *теоремою Лебега про мажоровану збіжність* отримуємо при $t \in \mathbb{R}$ тотожність

$$1 - \varphi(t) = (1 - \mathbf{E}\left(\exp(it\sigma_-)1\!\!1_{\{\tau_- < \infty\}}\right))\ (1 - \mathbf{E}\left(\exp(it\overline{\sigma}_+)1\!\!1_{\{\overline{\tau}_+ < \infty\}}\right)).$$

Зазначимо, що за *критерієм Колмогорова посиленого закону великих чисел* $S_n/n \to^{P1} \mathbf{E}\xi_1 < 0$. Тому $\mathbf{P}(\underline{\lim}_{n\to\infty}\{S_n \le 0\}) = 1$, звідки $\mathbf{P}(\tau_- < \infty) = 1$ та $\mu_\infty < \infty$ м.н. З урахуванням теореми *про властивості характеристичної функції* виводимо існування границі

$$-i\mathbf{E}\xi_1 = \lim_{t\to 0}(1 - \varphi(t))/t =$$
$$\lim_{t\to 0} t^{-1}(1 - \mathbf{E}\,(\exp(it\sigma_-)))\ (1 - \mathbf{P}(\overline{\tau}_+ < \infty)) =$$
$$-i\mathbf{E}\sigma_-\mathbf{P}(\overline{\tau}_+ = \infty) = -i\mathbf{E}\sigma_-\mathbf{P}(\mu_\infty \le 0),$$

де остання рівність є наслідком недодатності величини σ_-.

(б) Для $\varepsilon > 0$ позначимо $\xi'_n = \xi_n - \varepsilon$, $S'_n, \mu'_\infty, \sigma'_-$ – відповідні функціонали. Тоді $\mathbf{E}\xi'_n = -\varepsilon$, $\mu'_\infty \le \mu_\infty$, $\sigma'_- \le 0$, $\sigma'_- = \xi'_1$ на множині $\{\xi'_1 \le 0\}$. Тому внаслідок твердження (а)

$$\mathbf{P}(\overline{\tau}_+ = \infty) = \mathbf{P}(\mu_\infty \le 0) \le \mathbf{P}(\mu'_\infty \le 0) = (\mathbf{E}\xi'_1)/\mathbf{E}\sigma'_- = -\varepsilon/\mathbf{E}\sigma'_- \le$$

$$-\varepsilon/\mathbf{E}(\xi'_1 1\!\!1_{\{\xi'_1 \le 0\}}) = -\varepsilon/\mathbf{E}(\xi_1 - \varepsilon)1\!\!1_{\{\xi_1 \le \varepsilon\}} \le -\varepsilon/\mathbf{E}\xi_1 1\!\!1_{\{\xi_1 \le \varepsilon\}} \le \varepsilon/m(\delta),$$

для $\varepsilon \in (0, \delta)$, де $m(\delta) = -\mathbf{E}\xi_1 1\!\!1_{\{\xi_1 \le \delta\}} > 0$ при деякому $\delta > 0$. Оскільки ймовірності у лівій частині не залежать від ε, з отриманої нерівності виводимо, що ці ймовірності нульові. Скінченність м.н. τ_- випливає з оцінки $\tau_- \le \overline{\tau}'_+$, де останній момент будується за величинами $(-\xi_n)$ $\boxdot$

2.11.2. Натуральна факторизація

Одночасно з наведеними у теоремі *про факторизаційні тотожності для граничних функціоналів* мають місце такого ж типу факторизації, але вже через розподіли послідовних сум $S_n = \xi_1 + ... + \xi_n$.

Лема (про натуральну факторизацію). *Для всіх* $t \in \mathbb{R}$ *та* $z \in \mathbb{C}$ *з* $|z| < 1$ *справедлива тотожність*

$$1 - z\varphi(t) = w_0(z)w_+(t,z) \ / \ w_-(t,z),$$

де

$$w_+(t,z) = \exp\left(-\sum\nolimits_{n\geq 1} \frac{z^n}{n}\mathbf{E}(\exp(itS_n)1\!\!1_{\{S_n>0\}})\right),$$

$$w_-(t,z) = \exp\left(+\sum\nolimits_{n\geq 1} \frac{z^n}{n}\mathbf{E}(\exp(itS_n)1\!\!1_{\{S_n<0\}})\right),$$

$$w_0(z) = \exp\left(-\sum\nolimits_{n\geq 1} \frac{z^n}{n}\mathbf{P}(S_n = 0)\right),$$

причому всі функції $w_\pm(t,z)$ *– аналітичні та обмежені на напівплощині* $\{t \in \mathbb{Z} : \mathrm{Im}\ t \gtrless 0\}$, *і є неперервними на її границі.*

Крім того, $w_\pm(\pm i\infty, z) = 1$ *та* $w_\pm(t,z) \neq 0$.

Доведення. Аналітичність $w_\pm(t,z)$ є наслідком обмеженості та аналітичності експоненти $\exp(ita)$ при $\mathrm{Im}\ t \gtrless 0$ та $a \gtrless 0$.

Оскільки $|z\varphi(t)| \leq |z| < 1$ для вказаних t, z, то за формулою розкладу логарифмічної функції у ряд Тейлора

$$1 - z\varphi(t) = \exp(\ln(1 - z\varphi(t))) =$$

$$\exp\left(-\sum\nolimits_{n\geq 1} \frac{z^n}{n}\varphi^n(t)\right) = w_0(z)w_+(t,z) \ / \ w_-(t,z),$$

де враховано тотожність

$$\varphi^n(t) = \mathbf{E}\exp(itS_n) =$$

$$\mathbf{E}\exp(itS_n)1\!\!1_{\{S_n>0\}} + \mathbf{E}\exp(itS_n)1\!\!1_{\{S_n=0\}} + \ \mathbf{E}\exp(itS_n)1\!\!1_{\{S_n<0\}}.$$

Нормованість множників: $w_\pm(\pm i\infty, z) = 1$ випливає зі збіжності перетворень $\mathbf{E}(\exp(-sS_n)1\!\!1_{\{S_n \gtrless 0\}}) \to 0, s \to \pm\infty$. Останнє твердження леми є наслідком відсутності коренів у експоненційної функції на $\mathbb{C}$ $\boxdot$

Вправи

(1) Знайти для симетричного блукання Бернуллі $w_0(z) = (1 + \sqrt{1 - z^2})/2$.

(2) Величини $(\xi_n, n \geq 1)$ незалежні однаково розподілені, $\mathbf{E}\xi_1 = 0, \mathbf{D}\xi_1 = 1$, а $S_n = \xi_1 + ... + \xi_n$. Тоді $\max_{1\leq k\leq n} S_k/\sqrt{n} \to^W |\zeta| .n \to \infty$, де $\zeta \simeq N(0,1)$.

Теорема (про натуральні зображення для перетворень від граничних функціоналів). *Для всіх* $t \in \mathbb{R}$ *та* $z \in \mathbb{C}$ *з* $|z| < 1$:

$$\text{(а)} \quad 1 - \varphi_+(t,z) = w_0(z) w_+(t,z),$$

$$\text{(б)} \quad 1 - \varphi_-(t,z) = w_0(z) / w_-(t,z),$$

$$\text{(в)} \quad 1 - \overline{\varphi}_+(t,z) = w_+(t,z),$$

$$\text{(г)} \quad 1 - \overline{\varphi}_-(t,z) = 1/w_-(t,z),$$

$$\text{(д)} \quad \varkappa_\pm(z) = 1/\overline{\varkappa}_\pm(z) = w_0(z).$$

Доведення

(а,г,д) За теоремою *про факторизаційні тотожності для граничних функціоналів,* (в), та лемою *про натуральну факторизацію* для $t \in \mathbb{R}$, $|z| < 1$ отримуємо тотожності

$$(1 - \varphi_+(t,z))(1 - \overline{\varphi}_-(t,z)) = 1 - z\varphi(t) = w_0(z) w_+(t,z) \;/\; w_-(t,z),$$

звідки виводимо, що

$$(1 - \overline{\varphi}_-(t,z)) w_-(t,z) = w_0(z) w_+(t,z) \;/\; (1 - \varphi_+(t,z)).$$

Ліва частина цієї тотожності аналітична в області $\{t \in \mathbb{Z} : \mathrm{Im}\ t < 0\}$ та неперервна на її границі $\{t \in \mathbb{Z} : \mathrm{Im}\ t = 0\}$, а права частина – аналітична в області $\{t \in \mathbb{Z} : \mathrm{Im}\ t > 0\}$ та неперервна на границі $\{t \in \mathbb{Z} : \mathrm{Im}\ t = 0\}$. Оскільки при $t \in \mathbb{R}$ вказані функції збігаються, то за теоремою про єдиність аналітичного продовження вони є відповідними звуженнями на $\mathbb{R}$ цілої функції, що внаслідок обмеженості є сталою. За лемою *про аналітичне продовження перетворень від граничних функціоналів* $1 - \overline{\varphi}_-(-i\infty, z) = 1$, а за лемою *про натуральну факторизацію* $w_-(-i\infty, z) = 1$. Отже, вказана стала дорівнює 1. Тому з наведеної тотожності випливають рівності (а,г). Оскільки права частина цієї тотожності при $t \to +i\infty$ дорівнює $w_0(z) \;/\; \varkappa_+(z)$ за теоремою *про співвідношення для перетворень від граничних функціоналів,* (а), то справедливі тотожності (д).

Твердження (б,в) доводяться аналогічно ⊡

Теорема (про скінченність граничних функціоналів). *Припустимо, що* $\mathbf{P}(\xi_1 = 0) < 1$.

(1) Завжди збігається числовий ряд

$$\sum\nolimits_{n \ge 1} n^{-1} \mathbf{P}(S_n = 0) = -\ln(1 - \mathbf{P}(\sigma_\pm = 0, \tau_\pm < \infty)).$$

(2) Наступні умови еквівалентні:

(а) $\mu_\infty < \infty$ *м.н., (б)* $\mathbf{P}(\mu_\infty \le 0) > 0$, *(в)* $\sum\nolimits_{n \ge 1} n^{-1} \mathbf{P}(S_n > 0) < \infty$.

(3) $\mathbf{P}(-\infty < \mu^*_\infty,\ \mu_\infty < \infty) = 0.$

(4) Якщо величина ξ_1 *інтегровна і* $\mathbf{E}\xi_1 < 0$*, то* $\mathbf{P}(-\infty = \mu^*_\infty,\ \mu_\infty < \infty) = 1.$

(5) Якщо величина ξ_1 *інтегровна і* $\mathbf{E}\xi_1 = 0$*, то* $\mathbf{P}(-\infty = \mu^*_\infty,\ \mu_\infty = \infty) = 1.$

Доведення

(1) Оберемо $t = 0$ у рівності (д) теореми *про натуральні зображення для перетворень від граничних функціоналів* і спрямуємо $z \uparrow 1$:

$$\exp\Big(-\sum\nolimits_{n\geq 1} n^{-1}\mathbf{P}(S_n = 0)\Big) = w_0(1) = \varkappa_\pm(1) = 1 - \mathbf{P}(\sigma_\pm = 0, \tau_\pm < \infty).$$

Припустимо, що ряд у лівій частині розбігається. Тоді права частина останньої тотожності дорівнює нулю. Якщо $\mathbf{P}(\xi_1 > 0) > 0$, то

$$0 = 1 - \mathbf{P}(\sigma_+ = 0, \tau_+ < \infty) \geq \mathbf{P}(\sigma_+ > 0, \tau_+ < \infty) \geq \mathbf{P}(\xi_1 > 0) > 0.$$

Отримана суперечність доводить збіжність вказаного ряду. Альтернативне припущення $\mathbf{P}(\xi_1 < 0) > 0$ розглядається аналогічно.

(2) (а$\Longrightarrow$б) Для деяких $c > 0$ та $n \geq 1$:

$$\mathbf{P}(\mu_\infty \leq 0) \geq \mathbf{P}^n(\xi_1 \leq -c)\mathbf{P}(\sup\nolimits_{k>n}(S_k - S_n) \leq nc) =$$
$$\mathbf{P}^n(\xi_1 \leq -c)\mathbf{P}(\mu_\infty \leq nc) > 0$$

(б$\Longrightarrow$а) Випадкова подія $\{\mu_\infty < \infty\}$ належить *залишковій сигма-алгебрі* $\overline{\lim}_{n\to\infty}\sigma[\xi_n]$. Тому за теоремою *про закон нуля та одиниці Колмогорова* $\mathbf{P}(\mu_\infty < \infty) \in \{0, 1\}$. Отже, з $0 < \mathbf{P}(\mu_\infty \leq 0) \leq \mathbf{P}(\mu_\infty < \infty)$ виводимо, що $\mathbf{P}(\mu_\infty < \infty) = 1$.

(б$\Longleftrightarrow$в). Оберемо $t = 0$ у тотожності (в) теореми *про натуральні зображення для перетворень від граничних функціоналів* та спрямуємо $z \uparrow 1$:

$$\exp\big(-\sum\nolimits_{n\geq 1} n^{-1}\mathbf{P}(S_n > 0)\big) = w_+(0,1) = 1 - \overline{\varphi}_+(0,1) =$$
$$1 - \mathbf{P}(\overline{\tau}_+ < \infty) = \mathbf{P}(\overline{\tau}_+ = \infty) = \mathbf{P}(\mu_\infty \leq 0).$$

(3) З твердження (а) для послідовностей (ξ_n) та $(-\xi_n)$ виводимо, що одночасне виконання подій з (3) еквівалентне одночасній збіжності рядів вигляду: $\sum_{n\geq 1} n^{-1}\mathbf{P}(S_n \gtrless 0)$, що суперечить твердженню (а) теореми.

(4) За наслідком *про максимум блукання з інтегровними стрибками,* (а), $\mu_\infty < \infty$ м.н.,.тому за твердженням (3) $\mu^*_n = -\infty$ м.н.

(5) Подія $\{\mu_\infty < \infty\}$ належить *залишковій сигма-алгебрі* $\overline{\lim}_{n\to\infty}\sigma[\xi_n]$. Припустимо, що $\mathbf{P}(\mu_\infty < \infty) > 0$. Тоді за теоремою *про закон нуля та одиниці Колмогорова* $\mathbf{P}(\mu_\infty < \infty) = 1$. Унаслідок (2) звідси виводимо, що

$\mathbf{P}(\mu_\infty \le 0) > 0$. Однак за наслідком *про максимум блукання з інтегровними стрибками,* (б), остання ймовірність нульова. Отже, $\mathbf{P}(\mu_\infty < \infty) = 0$ від супротивного. Застосування доведеного твердження до послідовності $(-\xi_n)$ призводить до рівності $\mu^*_\infty = -\infty$ м.н. $\boxdot$

2.11.3. Односторонній закон великих чисел

Теорема (про лівосторонній посилений закон великих чисел).
(а) Нерівність $\overline{\lim}_{n\to\infty} S_n/n \le c$ *м.н. для сталої* $c \in \mathbb{R}$ *виконується тоді й тільки тоді, коли для кожного* $b > c$ *збігається ряд:*

$$\sum\nolimits_{n\ge 1} n^{-1}\mathbf{P}(S_n > bn) < \infty.$$

(б) Справедлива тотожність

$$\overline{\lim}_{n\to\infty} S_n/n = \inf\left(a \in \mathbb{R} : \sum\nolimits_{n\ge 1} n^{-1}\mathbf{P}(S_n > an) < \infty\right) \text{ м.н.}$$

Доведення

(а) Розглянемо для $b \in \mathbb{R}$ послідовність $\xi'_n = \xi_n - b$, $S'_n = S_n - nb$, $\mu'_\infty = \sup_{n\ge 1} S'_n$.

Достатність. Зі збіжності ряду, що наведений в умові, випливає збіжність ряду $\sum_{n\ge 1} n^{-1}\mathbf{P}(S'_n > 0)$, звідки за теоремою *про скінченність граничних функціоналів,* (2), виводимо, що $\mu'_\infty < \infty$ м.н. За означенням μ'_∞ при всіх n справедлива нерівність $S_n \le nb + \mu'_\infty$. Тому $\overline{\lim}_{n\to\infty} S_n/n \le b$ м.н. при кожному $b > c$, та $\overline{\lim}_{n\to\infty} S_n/n \le c$ м.н.

Необхідність. Для кожного $b > c$ мають місце включення:

$$\{\overline{\lim}_{n\to\infty} S_n/n \le c\} \subset \{\overline{\lim}_{n\to\infty} (S_n - nb) < 0\} \subset$$

$$\underline{\lim}_{n\to\infty} \{S_n - nb \le 0\} = \underline{\lim}_{n\to\infty} \{S'_n \le 0\} \subset \{\sup\nolimits_{n\ge 1} S'_n < \infty\},$$

де за умовою перша подія має одиничну ймовірність. Тому і остання подія виконується м.н., звідки з теореми *про скінченність граничних функціоналів,* (2), виводимо збіжність ряду з умови.

(б) Величина $\overline{\lim}_{n\to\infty} S_n/n$ вимірна відносно *залишкової сигма-алгебри* $\overline{\lim}_{n\to\infty} \sigma[\xi_n]$, а тому за теоремою *про закон нуля та одиниці Колмогорова* є виродженою: $\overline{\lim}_{n\to\infty} S_n/n = c$ м.н. для деякої сталої $c \in \mathbb{R}$. Позначимо через b вказану в умові нижню межу. Оскільки $\sum_{n\ge 1} n^{-1}\mathbf{P}(S_n > an) < \infty$ для всіх $a > b$, то внаслідок твердження (а) $c = \overline{\lim}_{n\to\infty} S_n/n \le a$, звідки виводимо нерівність $c \le b$. Аналогічно з $\sum_{n\ge 1} n^{-1}\mathbf{P}(S_n > an) = \infty$ при $a < b$ виводимо, що $c \ge b$ $\boxdot$

2.12. Гіллясті процеси

Гіллястий процес є стохастичною моделлю таких явищ, як спонтанне розмноження частинок, зростання популяцій, поширення інфекції тощо.

Припустимо, що в початковий момент наявна одна частинка. Вона живе певний відрізок часу, потім гине та одночасно породжує замість себе випадкову кількість таких самих частинок. Ці частинки утворюють перше покоління нащадків. Кожна з новоутворених частинок еволюціонує за такими ж правилами, утворюючи друге покоління і так далі. Постулюється, що події (щодо часу загибелі та кількості нащадків), які пов'язані з різними частинками з одного чи різних поколінь, *незалежні в сукупності*.

Означення. *Нехай* $(\xi_k^n,\ k \ge 1, n \ge 0)$ *подвійна послідовність незалежних у сукупності однаково розподілених невід'ємних цілозначних випадкових величин.* **Гіллястим процесом** *називається послідовність випадкових величин* $(\zeta_n,\ n \ge 0)$*, що обчислюються рекурентно*

$$\zeta_0 = 1,\ \ \zeta_{n+1} = \sum\nolimits_{k=1}^{\zeta_n} \xi_k^n,\ n \ge 0.$$

Величина ζ_n задає загальну кількість частинок у n-ому поколінні, а ξ_k^n інтерпретується як кількість нащадків k-ої частинки з n-ого покоління.

Зміст останнього рівняння очевидний: загальна кількість частинок наступного покоління дорівнює сумі кількостей всіх нащадків всіх наявних частинок попереднього покоління.

Зауважимо, що за означенням із **виродження** популяції у n-му поколінні: $\zeta_n = 0$ випливає повна виродженість процесу: $\zeta_k = 0,\ \forall k \ge n.$

Для аналізу гіллястих процесів використаємо *метод генератрис.*

2.12.1. Генератриса гіллястого процесу

Теорема (про генератриси гіллястого процесу). *Нехай гіллястий процес* $(\zeta_n, n \ge 0)$ *має генератрису* $\varphi(z) = \mathbf{E} z^{\xi_1^1}$ *числа нащадків однієї частинки, а* $\varphi_n(z) = \mathbf{E} z^{\zeta_n}$ *– генератриса кількості частинок* n*-го покоління. Тоді для всіх* $n \ge 0$:

(а) $\varphi_{n+1}(z) = \varphi_n(\varphi(z)),\ \ \varphi_0(z) = z,$

(б) $\varphi_{n+1}(z) = \varphi(\varphi_n(z)).$

Доведення

(а) є очевидним наслідком означення гіллястого процесу та теореми *про властивості генератрис*, пункт (е). Незалежність кількості доданків ζ_n та самих доданків ξ_k^n випливає з теореми *про векторні перетворення* *незалежних величин*, оскільки величини ζ_n будується за значеннями попередніх кількостей нащадків $(\xi_k^m,\ k \geq 1, m \leq n-1)$, що не залежать від $(\xi_k^n, k \geq 1)$ за умови *незалежності в сукупності* величин $(\xi_k^n,\ k \geq 1, n \geq 0)$.

(б) З попереднього твердження за індукцією отримуємо

$$\varphi_n(z) = \varphi(\varphi(...\varphi(z)...)),$$

де в правій частині міститься n-кратна суперпозиція. Очевидно, що остання послідовність задовольняє тотожність (б) $\boxdot$

2.12.2. Властивості гіллястих процесів

Означення. Критичним показником *гіллястого процесу* $(\zeta_n, n \geq 0)$ *називається середня кількість нащадків однієї частинки:*

$$\mu = \mathbf{E}\xi_1^1.$$

Гіллястий процес називається

(а) субкритичним, *якщо* $\mu < 1$,

(б) критичним, *якщо* $\mu = 1$,

(в) суперкритичним, *якщо* $\mu > 1$.

Означення. *Імовірністю виродження у* n*-му поколінні називається ймовірність відсутності частинок у цьому поколінні:*

$$\pi_n = \mathbf{P}(\zeta_n = 0),$$

а ймовірністю виродження – *границя*

$$\pi = \lim_{n\to\infty} \pi_n = \mathbf{P}(\cup_{n\,\geq\,0}\{\zeta_n = 0\}).$$

Зауваження. За означенням *гіллястого процесу* події $\{\zeta_n = 0\}$ не спадають, тому $\{\zeta_n = 0\} \uparrow \cup_{n\,\geq\,0}\{\zeta_n = 0\}, n \to \infty$, і за неперервністю ймовірності $\pi_n \uparrow \pi, n \to \infty$.

Теорема (про ймовірність виродження гіллястого процесу).

(а) Якщо гіллястий процес є субкритичним або (б) критичним, тоді ймовірність виродження є одиничною: $\pi = 1$,

(в) Якщо гіллястий процес – суперкритичний, то ймовірність виродження $\pi < 1$ *є єдиним на інтервалі* $[0, 1)$ *розв'язком рівняння*

$$\pi = \varphi(\pi).$$

Доведення. Оскільки за означенням генератриси має місце рівність $\varphi_n(0) = \mathbf{P}(\zeta_n = 0) = \pi_n$, то з теореми *про генератриси гіллястого процесу*, пункт (б), отримуємо при $z = 0$ рекурентне рівняння

$$\pi_{n+1} = \varphi(\pi_n),$$

причому $\pi_0 = 0$. Переходячи тут до границі при $n \to \infty$, з неперервності φ робимо висновок, що *ймовірність виродження* π завжди є коренем рівняння

$$p = \varphi(p), \quad p \geq 0.$$

Оскільки $\varphi(1) = 1$, а функція φ як сума ряду з невід'ємними коефіцієнтами опукла донизу на $\mathbb{R}_+$, то дане рівняння завжди має не більше двох коренів, один з яких дорівнює 1.

Доведемо, що ймовірність π є найменшим із невід'ємних коренів. Дійсно, якщо p – якийсь невід'ємний корінь, то $p \geq \pi_0 = 0$. Тоді за індукцією $p \geq \pi_n$, оскільки з $p \geq \pi_{n-1}$ та з монотонності φ випливає нерівність

$$p = \varphi(p) \geq \varphi(\pi_{n-1}) = \pi_n.$$

Отже, $\pi = \lim \pi_n \leq p$, що і доводить мінімальність π.

В умовах твердження (а) маємо $\varphi'(1) = \mu < 1$, тому в лівому околі одиниці має місце зображення

$$\varphi(z) = 1 + (z-1)\mu + o(|z-1|) > z.$$

Оскільки з $\mu < 1$ випливає $\varphi(0) > 0$, то з опуклості φ робимо висновок, що $\varphi(z) > z$ при всіх $z \in [0, 1)$. Отже, найменший корінь $p = 1$, тому виконується рівність $\pi = 1$.

У випадку (б) пряма $y = z$ є дотичною до графіка $\varphi(z)$ у точці $z = 1$. Тому за опуклістю φ знову $\varphi(z) > z$ при $z \in [0, 1)$ та $p = 1$, $\pi = 1$.

Отже, у випадках (а) і (б) мінімальний корінь рівняння $p = \varphi(p)$, $p \geq 0$, дорівнює 1 і за доведеним вище $\pi = 1$.

(в) За умови $\mu > 1$ у лівому околі одиниці

$$\varphi(z) = 1 + (z-1)\mu + o(|z-1|) < z.$$

Одночасно $\varphi(0) \geq 0$. Тому з неперервності φ випливає існування кореня рівняння $p = \varphi(p)$, $p \in [0, 1)$, а з опуклості – його єдиність (адже 1 – інший корінь) $\boxdot$

Теорема (про середню чисельність поколінь). *Нехай* $(\zeta_n, n \geq 0)$ *гіллястий процес. Середня чисельність* n*-го покоління* $\mu_n = \mathbf{E}\zeta_n$ *дорівнює*

$$\mu_n = \mu^n,$$

причому при $n \to \infty$ *справедливі такі співвідношення:*

(а) для субкритичного процесу $\mu_n \to 0$,

(б) для критичного процесу $\mu_n = 1$,

(в) для суперкритичного процесу $\mu_n \to \infty$.

Доведення. З властивості (г) теореми *про властивості генератрис* та з теореми *про генератриси гіллястого процесу* обчислимо

$$\mu_{n+1} = \varphi'_{n+1}(1) = \ (\varphi(\varphi_n(z)))' \,|_{z=1}= \ \varphi'_n(1)\varphi'(\varphi_n(1)) = \mu_n\mu.$$

Звідси за індукцією $\mu_n = \mu^n$ ⊡

Вправи

(1) Якщо (ζ_n) – гіллястий процес, а $d \in \mathbb{N}$, то (ζ_{nd}) – гіллястий процес.

(2) Довести, що побудована за методом Ньютона-Рафсона послідовність $p_0 = 0$, $p_{m+1} = p_m + (\varphi(p_m) - p_m)/(\varphi'(p_m) - 1), m \geq 1$, збігається при $m \to \infty$ до ймовірності виродження π, та обчислити швидкість збіжності.

(3) Довести, що за припущенням геометричного розподілу числа нащадків частинки: $\mathbf{P}(\xi_1^1 = k) = ab^{k-1}, k \geq 1$, з деякими $a, b > 0, a + b < 1$, генератриси φ_n є дробово-раціональними функціями, та обчислити ці функції.

(4) Знайти ймовірність виродження для гіллястого процесу такого, що кількість нащадків $\xi_1^1 \in \{0, 1, 2\}$.

(5) Вивести узагальнення теореми про ймовірність виродження на випадок, коли генератриса кількості нащадків дорівнює φ для парних поколінь та ψ – для непарних. Чи зміняться відповіді, якщо φ та ψ поміняти місцями ?

(6) Нехай (ζ_n) – гіллястий процес. Довести при $k, m > 0$ нерівність

$$\mathbf{P}(\cup_{n=1}^{m-1}\{\zeta_n > k\} \mid \zeta_m = 0) \leq (\mathbf{P}(\zeta_m = 0))^k.$$

(7) Нехай (ζ_n) – гіллястий процес,.для якого $\mathbf{P}(\xi_1^1 = k) = bc^k$ при $k \geq 1$. Знайти граничний розподіл $\lim_{n\to\infty} \mathbf{P}(\zeta_n = k \mid \zeta_n > 0)$.

(8) Нехай (ζ_n) – гіллястий процес.з генератрисою φ числа нащадків, а $S_n = \sum_{k=0}^n \zeta_k$ і $\psi_n(z) = \mathbf{E}z^{S_n}$. Довести, що $\psi_{n+1}(z) = z\varphi(\psi_n(z))$.

(9) Нехай (ζ_n) – гіллястий процес з $\mu > 1$. Довести, що $\zeta_n/\mu^n \to^{P1} \eta$, де $\mathbf{E}\eta = 1$, $\mathbf{D}\eta = \mathbf{D}\zeta_1/(\mu^2 - \mu)$.

(10) Нехай (ζ_n) – гіллястий процес з $\mu < 1$ та $\mathbf{E}\zeta_1^2 < \infty$. Довести, що $\lim_{n\to\infty} \mathbf{P}(\zeta_n = k \mid \zeta_n \neq 0)$ є дискретним розподілом ймовірностей.

2.13. Мартингальні послідовності

Мартингали широко використовуються у різноманітних застосуваннях теорії ймовірностей, зокрема, у фінансовій математиці.

Нехай T – деяка впорядкована множина цілих чисел, наприклад: $T = \overline{0, N}$, $T = \mathbb{Z}_+$ або $T = \mathbb{Z}$.

Надалі для скорочення будемо вживати такі позначення: $a \wedge b \equiv \min(a, b)$, $a \vee b \equiv \max(a, b)$.

Означення. *Сім'я під-сигма-алгебр $\mathfrak{F}_t \subset \mathfrak{F}, t \in T$, на ймовірнісному просторі $(\Omega, \mathfrak{F}, \mathbf{P})$ називається стохастичним* потоком *(або ж* фільтрацією*), якщо $\mathfrak{F}_s \subset \mathfrak{F}_t$, $\forall s < t,\ s, t \in T$.*

Наприклад, послідовність сигма-алгебр $\mathfrak{F}_t = \sigma[\xi_s, 0 \le s \le t]$, що породжені послідовністю $(\xi_t, t \ge 0)$, є потоком.

Означення. *Сім'я випадкових величин $(\xi_t, t \in T)$ називається* узгодженою *зі стохастичним потоком $(\mathfrak{F}_t, t \in T)$, якщо при кожному $t \in T$ величина ξ_t вимірна відносно сигма-алгебри $\mathfrak{F}_t$.*

2.13.1. Означення та приклади мартингалів

Означення. *Сім'я інтегровних випадкових величин $(\xi_t, t \in T)$ називається* мартингалом *відносно стохастичного потоку $(\mathfrak{F}_t, t \in T)$, якщо*

(а) вона узгоджена з цим потоком,

(б) $\xi_s = \mathbf{E}(\xi_t \mid \mathfrak{F}_s)$ м.н. для всіх $s < t,\ s, t \in T$.

Якщо замість рівностей (б) виконуються нерівності:

(б′) $\xi_s \le \mathbf{E}(\xi_t \mid \mathfrak{F}_s)$ м.н. для всіх $s < t,\ s, t \in T$,

то послідовність $(\xi_t, t \in T)$ називається субмартингалом, *а при виконанні протилежних нерівностей $(\ge)$ –* супермартингалом.

За означенням мартингал одночасно є суб- та супермартингалом.

Зауваження. Якщо в означенні мартингальної послідовності (ξ_n) не вказано, який саме стохастичний потік $(\mathfrak{F}_n)$ мається на увазі, то слід вважати, що це – натуральний потік: $\mathfrak{F}_n = \sigma[\xi_k, k \le n]$.

Зауваження. Умову (б) в означенні мартингалу можна замінити на умову $\mathbf{E}(\xi_t - \xi_s \mid \mathfrak{F}_s) = 0$ м.н., оскільки за теоремою *про властивості умовного сподівання* $\mathbf{E}(\xi_s \mid \mathfrak{F}_s) = \xi_s$ м.н. За лемою *про знаковизначеність випадкової величини* вказана умова еквівалентна рівностям

$$\mathbf{E}(\xi_t - \xi_s)1\!\!\mathrm{I}_A = 0,\ \forall A \in \mathfrak{F}_s,$$

оскільки її ліва частина дорівнює $\mathbf{E}(\mathbf{E}(\xi_t \mid \mathfrak{F}_s) - \xi_s)1\!\!\mathrm{I}_A$.

Аналогічно, субмартингальна умова (б′) у означенні еквівалентна нерівностям: $\mathbf{E}(\xi_t - \xi_s \mid \mathfrak{F}_s) \ge 0$ м.н., або ж: $\mathbf{E}(\xi_t - \xi_s)1\!\!\mathrm{I}_A \ge 0,\ \forall A \in \mathfrak{F}_s$.

Лема (про критерій мартингальної властивості). *Інтегровна послідовність $(\xi_n, n \in T)$ є мартингалом (субмартингалом) тоді й тільки тоді, коли вона узгоджена з фільтрацією $(\mathfrak{F}_n)$ та виконуються умови: $\xi_n = (\le)\ \mathbf{E}(\xi_{n+1} \mid \mathfrak{F}_n)$ м.н. для всіх $n \in T, n < \sup T$.*

Для виконання останньої рівності (нерівності) необхідно і достатньо, щоб $\mathbf{E}(\xi_{n+1} - \xi_n)1\!\!\mathrm{I}_A = (\ge)\ 0,\ \forall A \in \mathfrak{F}_n$.

Доведення проводиться за індукцією. Якщо умова (б) виконується для $s < t < \sup T$, то за теоремою *про властивості умовного сподівання*

$$\xi_s = \mathbf{E}(\xi_t \mid \mathfrak{F}_s) = \mathbf{E}(\mathbf{E}(\xi_{t+1} \mid \mathfrak{F}_t) \mid \mathfrak{F}_s) = \mathbf{E}(\xi_{t+1} \mid \mathfrak{F}_s) \text{ м.н.}$$

Отже, ця умова справедлива також для пари $s, t+1$ ⊡

Приклади

1. *Випадкові блукання*: нехай $(\zeta_n, n \geq 1)$ послідовність незалежних випадкових величин з $\mathbf{E}\zeta_n = 0$. Тоді послідовність $\xi_n = \sum_{k=1}^n \zeta_k$ є *мартингалом* відносно *потоку* $\mathfrak{F}_n = \sigma[\zeta_k, k \leq n]$, оскільки

$$\mathbf{E}(\xi_{n+1} - \xi_n \mid \mathfrak{F}_n) = \mathbf{E}(\zeta_{n+1} \mid \mathfrak{F}_n) = \mathbf{E}\zeta_{n+1} = 0 \text{ м.н.}$$

внаслідок незалежності ζ_{n+1} і $\mathfrak{F}_n$, та умови центрованості доданків. Якщо умову центрованості замінити на $\mathbf{E}\zeta_n \geq 0$ (для чого достатньо невід'ємності доданків), то послідовність ξ_n буде *субмартингалом.*

2. *Випадкові еволюції*: якщо $(\zeta_n, n \geq 1)$ послідовність незалежних додатних випадкових величин з $\mathbf{E}\zeta_n = 1$. Тоді послідовність $\xi_n = \prod_{k=1}^n \zeta_k$ є мартингалом відносно потоку $\mathfrak{F}_n = \sigma[\zeta_k, k \leq n]$, оскільки

$$\mathbf{E}(\xi_{n+1} \mid \mathfrak{F}_n) = \mathbf{E}(\xi_n \zeta_{n+1} \mid \mathfrak{F}_n) = \xi_n \mathbf{E}(\zeta_{n+1} \mid \mathfrak{F}_n) = \xi_n \mathbf{E}(\zeta_{n+1}) = \xi_n \quad \text{м.н.}$$

за теоремою *про властивості умовного сподівання.*

3. *Умовні щільності*: нехай Q – довільна міра на $\mathfrak{F}$, що *абсолютно неперервна відносно міри* $P : Q(A) = 0$ для всіх $A \in \mathfrak{F}$ з $P(A) = 0$, а $(\mathfrak{F}_n)$ – стохастичний потік. Тоді послідовність щільностей $\xi_n = \frac{dQ_n(\cdot)}{dP_n(\cdot)}$ (що існують за *теоремою Радона-Нікодима*) звужень Q_n міри Q на $\mathfrak{F}_n$ відносно звуження P_n міри P є мартингалом відносно $\mathfrak{F}_n$. Вимірність ξ_n відносно $\mathfrak{F}_n$ очевидна. З означення щільності отримуємо при $B_n \in \mathfrak{F}_n$ також умову балансу в означенні *умовного математичного сподівання:*

$$\mathbf{E}\xi_{n+1} 1\!\!1_{B_n} = \int_{B_n} \xi_{n+1} dP_{n+1} = Q_{n+1}(B_n) = Q_n(B_n) = \int_{B_n} \xi_n dP_n = \mathbf{E}\xi_n 1\!\!1_{B_n}.$$

4. *Мартингали Леві:* якщо ξ – інтегровна величина, а $(\mathfrak{F}_n)$ – стохастичний потік, то послідовність $\xi_n = \mathbf{E}(\xi \mid \mathfrak{F}_n)$ є мартингалом за теоремою *про властивості умовного сподівання.*

2.13.2. Перетворення мартингалів

Означення. *Послідовність випадкових величин $(V_n, n \in T)$ називається* **передбачуваною** *відносно потоку $(\mathfrak{F}_n, n \in T)$, якщо величина V_n вимірна відносно сигма-алгебри $\mathfrak{F}_{n-1}$ для всіх $n > \min T$, а величина $V_{\min T}$ – невипадкова.*

Приклад. Якщо $(\xi_n, n \geq 1)$ відображає динаміку курсу акцій на кінець n-го біржового дня, а V_n – кількість таких акцій, призначених брокером на n-ий день для купівлі або продажу, то послідовність (V_n) є передбачуваною відносно потоку $(\sigma[\xi_k, k \leq n])$.

Теорема (про перетворення мартингальних послідовностей). *Нехай $(\mathfrak{F}_t, t \in T)$ – фіксований стохастичний потік, відносно якого розглядаються наведені нижче мартингальні послідовності.*

(1) Якщо $(\xi_t, t \in T)$ – мартингал (субмартингал), а $S \subset T$, то послідовність $(\xi_t, t \in S)$ – також мартингал (субмартингал).

(2) Якщо $(\xi_t^{(1)}, t \in T)$ та $(\xi_t^{(2)}, t \in T)$ – мартингали (субмартингали), а $c_k \in \mathbb{R}$ (відповідно $c_k \in \mathbb{R}_+$), то $(c_1\xi_t^{(1)} + c_2\xi_t^{(2)}, t \in T)$ – мартингал (відповідно субмартингал).

(3) Послідовність $(\xi_t, t \in T)$ є супермартингалом тоді й тільки тоді, коли $(-\xi_t, t \in T)$ – субмартингал.

(4) Якщо $(\xi_t^{(1)},\ t \in T)$ та $(\xi_t^{(2)},\ t \in T)$ – два субмартингали, то послідовність $(\xi_t^{(1)} \vee \xi_t^{(2)},\ t \in T)$ – також субмартингал.

(5) Нехай $(\xi_t, t \in T)$ – мартингал, борелєва функція g опукла донизу, і величини $g(\xi_t)$ – інтегровні. Тоді $(g(\xi_t), t \in T)$ є субмартингалом.

(6) Нехай $(\xi_t, t \in T)$ – субмартингал, борелєва функція g опукла донизу, не спадає і така, що величини $g(\xi_t)$ інтегровні. Тоді послідовність $(g(\xi_t), t \in T)$ є субмартингалом.

(7) Нехай $(\xi_n, n \in \overline{0, N})$ – мартингал, а послідовність $(V_n, n \in \overline{0, N})$ є передбачуваною. Якщо величини $\zeta_n \equiv V_0\xi_0 + \sum_{k=1}^{n} V_k(\xi_k - \xi_{k-1})$ є інтегровними, то вони утворюють мартингал $(\zeta_n, n \in \overline{0, N})$.

(8) Якщо $(\xi_n, n \in \overline{0, N})$ – субмартингал (супермартингал), а послідовність $(V_n, n \in \overline{0, N})$ є передбачуваною, невід'ємною та обмеженою, то величини $\zeta_n \equiv V_0\xi_0 + \sum_{k=1}^{n} V_k(\xi_k - \xi_{k-1})$ утворюють субмартингал (відповідно супермартингал).

Наслідок. *Якщо (ξ_t) – мартингал, то субмартингалами є послідовності: (1) $(\xi_t \vee a)$ при $a \in \mathbb{R}$, (2) $(|\xi_t|^\alpha)$ при $\alpha \geq 1$, (3) (ξ_t^+). Дійсно, функції $x \vee a$, $|x|^\alpha, x^+$ є опуклими донизу.*

Доведення

(1) Твердження є очевидним наслідком означення.

(2) Лінійне перетворення зберігає умову (а) вимірності в означенні мартингалу. Умова (б) на умовні ймовірності виконується внаслідок лінійності в теоремі *про загальні властивості умовного сподівання*.

(3) Це твердження випливає з означення.

(4) Величина $\xi_t \equiv \xi_t^{(1)} \vee \xi_t^{(2)}$ є $\mathfrak{F}_t$-вимірною за теоремою *про вимірність функції від випадкового вектора.* Для перевірки умови (б′) використаємо монотонність умовного сподівання: $\xi_s^{(k)} \le \mathbf{E}(\xi_t^{(k)} \mid \mathfrak{F}_s) \le \mathbf{E}(\xi_t \mid \mathfrak{F}_s)$ при $k = 1, 2,\ s < t$. Звідси $\xi_s = \xi_s^{(1)} \vee \xi_s^{(2)} \le \mathbf{E}(\xi_t \mid \mathfrak{F}_s)$, що доводить (4).

(5) За нерівністю Йенсена (7) з теореми *про властивості умовного сподівання* отримуємо $g(\xi_s) = g(\mathbf{E}(\xi_t \mid \mathfrak{F}_s)) \le \mathbf{E}(g(\xi_t) \mid \mathfrak{F}_s)$ м.н.

(6) Аналогічно за нерівністю Йенсена з урахуванням монотонності g виводимо нерівність $g(\xi_s) \le g(\mathbf{E}(\xi_t \mid \mathfrak{F}_s)) \le \mathbf{E}(g(\xi_t) \mid \mathfrak{F}_s)$ м.н.

(7) Вимірність ζ_n відносно $\mathfrak{F}_n$ очевидна. З означення виводимо тотожність $\zeta_n - \zeta_{n-1} = V_n(\xi_n - \xi_{n-1})$. Її ліва частина є інтегровною за умовою,.а величини ξ_n також інтегровні за означенням. Тому на підставі $\mathfrak{F}_{n-1}$-вимірності V_n за теоремою *про властивості умовного сподівання,* отримуємо

$$\mathbf{E}(\zeta_n - \zeta_{n-1} \mid \mathfrak{F}_{n-1}) =$$
$$\mathbf{E}(V_n(\xi_n - \xi_{n-1}) \mid \mathfrak{F}_{n-1}) = V_n\mathbf{E}((\xi_n - \xi_{n-1}) \mid \mathfrak{F}_{n-1}) = 0 \text{ м.н.}$$

(8) Доведення проводиться аналогічно до (7), оскільки інтегровність $V_n\xi_n$ випливає з інтегровності ξ_n та обмеженості V_n, а справедливість необхідних нерівностей випливає з невід'ємності V_n ⊡

Теорема (теорема Дуба про зображення узгодженої послідовності). *Для кожної інтегровної послідовності* $(\xi_n, n \ge 0)$, *що узгоджена з потоком* $(\mathfrak{F}_n, n \ge 0)$, *знайдеться мартингал* $(\zeta_n, n \ge 0)$ *та передбачувана послідовність* $(V_n, n \ge 0)$ *відносно* $(\mathfrak{F}_n)$ *такі, що*

$$\xi_n = \zeta_n + V_n \quad \text{м.н. } \forall n \ge 0.$$

Це зображення є єдиним з точністю до рівності м.н. за умови $V_0 = 0$.

Якщо (ξ_n) *– субмартингал, то послідовність* (V_n) *є м.н. неспадною.*

Означення. *Послідовність* (V_n) *у зображенні Дуба називається* компенсатором *процесу* $(\xi_n, n \ge 0)$.

Доведення. Визначимо $\zeta_0 = \xi_0,\ V_0 = 0$ та рекурентно при $n \ge 1$

$$\zeta_n = \zeta_0 + \sum\nolimits_{k=1}^{n}(\xi_k - \mathbf{E}(\xi_k \mid \mathfrak{F}_{k-1})),$$

$$V_n = \sum\nolimits_{k=1}^{n}(\mathbf{E}(\xi_k \mid \mathfrak{F}_{k-1}) - \xi_{k-1}).$$

Тоді за означенням $\zeta_n + V_n = \xi_n$, величина (ζ_n) є $\mathfrak{F}_n$-вимірною, а послідовність (V_n) – *передбачуваною.* Оскільки за теоремою *про властивості умовного сподівання*

$$\mathbf{E}((\zeta_{n+1} - \zeta_n) \mid \mathfrak{F}_n) = \mathbf{E}((\xi_{n+1} - \mathbf{E}(\xi_{n+1} \mid \mathfrak{F}_n)) \mid \mathfrak{F}_n) = 0 \text{ м.н.},$$

то (ζ_n) – мартингал за лемою *про критерій мартингальної властивості.*

Для доведення єдиності з рівняння $\zeta_n + V_n = \zeta'_n + V'_n$ обчисленням умовного сподівання відносно $\mathfrak{F}_{n-1}$ отримуємо $\zeta_{n-1}+V_n = \zeta'_{n-1}+V'_n$, звідки $\zeta_n - \zeta_{n-1} = \zeta'_n - \zeta'_{n-1}$ і $\zeta_n = \zeta'_n$ внаслідок означення $\zeta_0 = \zeta'_0 = \xi_0$.

Послідовність (V_n) м.н. не спадає за означенням *субмартингалу* $\boxdot$

Вправи

(1) Якщо $(\xi_t, t \in T)$ – мартингал, то послідовність $\mathbf{E}\xi_s = \mathbf{E}\xi_t$ є сталою, а для субмартингалу ця послідовність не спадає.

(2) Випадкові величини $(\zeta_n, n \ge 1)$ інтегровні. Довести, що послідовність $\xi_n = \zeta_1 + ... + \zeta_n$ є мартингалом відносно потоку $\mathfrak{F}_n = \sigma[\zeta_k, k \le n]$ тоді й тільки тоді, коли (ζ_n) є мартингал-приростами, тобто $\mathbf{E}\zeta_{n+1} g_n(\zeta_1, ..., \zeta_n) = 0$ для всіх обмежених борелєвих g_n та всіх $n \ge 1$.

(3) Послідовність величин $(\xi_n, n \ge 1)$ є мартингалом відносно потоку $(\mathfrak{F}_n)$. Довести, що вона є мартингалом відносно натурального потоку $(\sigma[\xi_k, k \le n])$.

(4, М.Й.Ядренко) Випадкові величини (ξ_n) незалежні рівномірно розподілені на $(0,1)$, а $S_n = \xi_1 + ... + \xi_n$, $\tau = \inf(n : S_n > 1)$. Довести, що $\mathbf{P}(\tau > n) = 1/n!$, $\mathbf{E}\tau = e$, $\mathbf{E}S_\tau = e/2$.

(5) Випадкові величини $(\zeta_n, n \ge 1)$ незалежні, $\mathbf{P}(\zeta_n = 0) = \mathbf{P}(\zeta_n = 2) = 1/2$. Довести, що мартингал $\xi_n = \prod_{k=1}^n \zeta_k$ не є мартингалом Леві.

(6) Послідовність $(\xi_n, n \ge 1)$ є мартингалом, $\xi_0 = 0$, а $\zeta_n = \xi_n - \xi_{n-1}, n \ge 1$. Якщо $\sum_{n\ge 1} \mathbf{E}\zeta_n^2/n^2 < \infty$, то $\xi_n/n \to^{P1} 0, n \to \infty$.

(7) Величини $(\zeta_n, n \ge 1)$ незалежні та інтегровні, $\mathbf{E}\zeta_n = 0$. Довести, що: (а) для кожного $k \ge 1$ послідовність $\xi_n = \sum_{1 \le i_1 < i_2 < ... < i_k \le n} \zeta_{i_1} ... \zeta_{i_k}$ є мартингалом, (б) для сталих $(c_{nk}, 0 \le k < n)$ частинні суми ряду $\sum_{n \ge 1} \sum_{k<n} c_{nk}\zeta_k\zeta_n$ утворюють мартингал.

(8) Величини $(\zeta_n, n \ge 1)$ незалежні, інтегровні, $\mathbf{E}\zeta_n = 0$, а $(\eta_n, n \ge 1)$ – довільна незалежна від (ζ_n) послідовність інтегровних випадкових величин. Довести, що величини $\xi_n = \sum_{k=1}^n \zeta_k \eta_k$ утворюють мартингал.

(9) Величини $(\zeta_n, n \ge 1)$ незалежні, $\mathbf{P}(\zeta_n = \pm 1) = 1/2$. Обчислити компенсатор та відповідний мартингал у розкладі Дуба для субмартингала $\xi_n = |S_n|$, де $S_n = \zeta_1 + ... + \zeta_n$. Довести, що $\mathbf{E} \sum_{k=1}^{n-1} 1\!\!\mathrm{I}_{\{S_k = 0\}} = \mathbf{E}\, |S_n| \sim \sqrt{2n/\pi}, n \to \infty$.

(10) Випадкові величини $(\zeta_n, n \ge 1)$ незалежні та квадратично інтегровні, $\mathbf{E}\zeta_n = 0$, а $\eta_n = \sum_{k=1}^n \zeta_k$. Довести, що послідовність $\xi_n = \eta_n^2 - \mathbf{E}\eta_n^2$ є мартингалом, тобто $(\mathbf{E}\eta_n^2, n \ge 1)$ є компенсатором субмартингалу $(\eta_n^2, n \ge 1)$.

(11) Нехай $(\xi_n, n \ge 0)$ – мартингал відносно потоку $(\mathfrak{F}_n)$, що є квадратично інтегровним: $\mathbf{E}\xi_n^2 < \infty$, для всіх n. Довести, що компенсатор субмартингалу $(\xi_n^2, n \ge 0)$ дорівнює $V_n = \sum_{k=1}^n \mathbf{E}((\xi_k - \xi_{k-1})^2 \mid \mathfrak{F}_{n-1})$.

(12) Величини $(\zeta_n, n \ge 1)$ незалежні, $\mathbf{P}(\zeta_n = 1) = p$, $\mathbf{P}(\zeta_n = -1) = q$, $p + q = 1$, а $\eta_n = \sum_{k=1}^n \zeta_k$. Довести, що $\xi_n \equiv (q/p)^{\eta_n}$ – мартингал.

(13) Нехай $(\zeta_n, n \geq 0)$ – гіллястий процес з $\zeta_0 = 1$ та $\mathbf{E}\zeta_1 = \mu > 0$. Довести, що $(\zeta_n \mu^{-n})$ – мартингал.

(14) Пересвідчитись, що в інтерпретації прикладу до означення передбачуваної послідовності, послідовність у теоремі про перетворення мартингальних послідовностей, (7), збігається з сумарним прибутком (втратами) брокера від операцій з акціями на кінець n-го дня.

(15) Випадкові величини $(\zeta_n, n \geq 1)$ незалежні, однаково розподілені та інтегровні. Визначимо $\xi_{-n} = (\zeta_1 + ... + \zeta_n)/n$ при $n \geq 1$. Довести, що послідовність $(\xi_n, n \leq -1)$ є мартингалом.

2.13.3. Моменти зупинки

Означення. *Випадкова величина τ зі значеннями у множині T є* моментом зупинки *відносно потоку $(\mathfrak{F}_t,\ t \in T)$, якщо $\{\tau \leq t\} \in \mathfrak{F}_t$ для всіх $t \in T$. Відповідна величина τ зі значеннями у множині $T \cup \{\infty\}$ називається* марковським моментом.

Приклад. Якщо $(\xi_n, n \geq 1)$ незалежні випадкові величини, що породжують потік $\mathfrak{F}_n = \sigma[\xi_k, k \leq n]$, і $S_n = \xi_1 + ... + \xi_n$, то випадкова величина

$$\tau = \inf(n \leq N : S_n \geq a) \wedge N$$

є *моментом зупинки*. Дійсно, при $n < N$ подія $\{\tau \leq n\} = \cup_{k=1}^n \{S_k \geq a\} \in \mathfrak{F}_n$, і за означенням $\{\tau \leq N\} = \Omega \in \mathfrak{F}_N$.

Зауваження. Якщо τ – момент зупинки, то випадкові події $\{\tau < t\}$, $\{\tau = t\}$, $\{\tau > t\}$ належать $\mathfrak{F}_t$ при $t \in T$.

Доведення випливає зі співвідношень $\{\tau < t\} = \cup_{k \geq 1}\{\tau \leq t - 1/k\}$, $\{\tau = t\} = \{\tau \leq t\} \backslash \{\tau < t\}$, $\{\tau > t\} = \overline{\{\tau \leq t\}}$ $\boxdot$

Теорема (про перетворення моментів зупинки). *Нехай τ_n – невід'ємні моменти зупинки. Тоді: (а) $\tau_1 \vee \tau_2$, (б) $\tau_1 + \tau_2$, (в) $\tau_1 \wedge \tau_2$ також є моментами зупинки. (г) Якщо $\tau_n \uparrow \tau$ або $\tau_n \downarrow \tau$ при $n \to \infty$, то τ – марковський момент.*

Доведення

(а) $\{\tau_1 \vee \tau_2 \leq t\} = \{\tau_1 \leq t\} \cap \{\tau_2 \leq t\} \in \mathfrak{F}_t$.

(б) $\{\tau_1 + \tau_2 \leq t\} = \cup_{s=0}^t (\{\tau_1 \leq t - s\} \cap \{\tau_2 = s\}) \in \mathfrak{F}_t$.

(в) $\{\tau_1 \wedge \tau_2 \leq t\} = \{\tau_1 \leq t\} \cup \{\tau_2 \leq t\} \in \mathfrak{F}_t$.

(г) $\{\tau \leq t\} = \cup_{n \geq 1}\{\tau_n \leq t\}$ при $\tau_n \downarrow \tau$,

$\{\tau > t\} = \cup_{n \geq 1}\{\tau_n > t\}$ при $\tau_n \uparrow \tau$ $\boxdot$

Означення. *Нехай τ – момент зупинки відносно потоку $(\mathfrak{F}_t, t \in T)$. Сигма-алгеброю подій, що передують моменту τ, називається такий* *клас випадкових подій:*

$$\mathfrak{F}_\tau \equiv \{A \in \mathfrak{F} : A \cap \{\tau \le t\} \in \mathfrak{F}_t,\ \forall t \in T\}.$$

Цей клас є сигма-алгеброю, оскільки: (1) $\Omega \in \mathfrak{F}$ за означенням *моменту зупинки.* (2) Якщо $A \in \mathfrak{F}_\tau$, то $\overline{A} \cap \{\tau \le t\} = \{\tau \le t\} \backslash (A \cap \{\tau \le t\}) \in \mathfrak{F}_t$. (3) Для $A_k \in \mathfrak{F}_\tau$ маємо $(\cup_{k\ge1} A_k) \cap \{\tau \le t\} = \cup_{k\ge1}(A_k \cap \{\tau \le t\}) \in \mathfrak{F}_t$ $\boxdot$

Теорема (про події, що передують моментам зупинки).

(а) Якщо $\tau_k, k = 1, 2,$ – моменти зупинки відносно стохастичного потоку $(\mathfrak{F}_t, t \in T)$, і $\tau_1 \le \tau_2$, то $\mathfrak{F}_{\tau_1} \subset \mathfrak{F}_{\tau_2}$.

(б) Нехай послідовність $(\zeta_t, t \in T)$ та потік $(\mathfrak{F}_t, t \in T)$ такі, що ζ_t є $\mathfrak{F}_t$-вимірними, а τ – момент зупинки відносно $(\mathfrak{F}_t)$. Тоді випадкова величина

$$\zeta_\tau \equiv \sum\nolimits_{t \in T} \zeta_t \mathbb{I}_{\{\tau = t\}}$$

є вимірною відносно сигма-алгебри $\mathfrak{F}_\tau$.

Доведення

(а) Нехай $A \in \mathfrak{F}_{\tau_1}$. Оскільки $\{\tau_2 \le t\} \subset \{\tau_1 \le t\}$, то

$$A \cap \{\tau_2 \le t\} = (A \cap \{\tau_1 \le t\}) \cap \{\tau_2 \le t\} \in \mathfrak{F}_t,\ \forall t \in T.$$

(б) $\{\zeta_\tau < a\} \cap \{\tau \le t\} = \cup_{s \in T, s \le t}\{\zeta_s < a\} \cap \{\tau = s\} \in \mathfrak{F}_t,\ \forall t \in T$ $\boxdot$

Вправи

(1) Обчислити $\mathfrak{F}_\tau$ для сталого моменту зупинки $\tau \equiv c \in T$.

(2) Нехай $A \in \mathfrak{F}_\tau$. Довести, що $A \cap \{\tau < t\} \in \mathfrak{F}_t,\ \ A \cap \{\tau = t\} \in \mathfrak{F}_t$.

(3) Довести, що момент зупинки τ вимірний відносно $\mathfrak{F}_\tau$.

(4) Нехай τ, σ – моменти зупинки. Довести, що: (а) $\mathfrak{F}_\tau \subset \mathfrak{F}_\sigma$ при $\tau \le \sigma$, (б) $\{\tau \le \sigma\} \in \mathfrak{F}_\tau \cap \mathfrak{F}_\sigma$, (в) $\mathfrak{F}_{\tau \wedge \sigma} = \mathfrak{F}_\tau \cap \mathfrak{F}_\sigma$.

(5) Якщо (τ_n) – моменти зупинки, то величини $\inf \tau_n, \sup \tau_n, \overline{\lim} \tau_n, \underline{\lim} \tau_n$ є марковськими моментами.

(6) Нехай (ξ_n) – мартингал відносно $(\mathfrak{F}_n)$, а τ – момент зупинки відносно цього потоку. Довести, що $(\xi_{n \wedge \tau})$ – також мартингал.

(7) Нехай $(\tau_k, k = \overline{1, n})$ – неспадна послідовність моментів зупинки відносно потоку $(\mathfrak{F}_n)$, а величини ξ_k обмежені та $\mathfrak{F}_{\tau_k}$-вимірні. Довести, що послідовність $\zeta_t = \sum_{k \le n} \xi_k \mathbb{I}_{\{t > \tau_k\}}$ є передбачуваною.

(8) Визначимо σ-алгебру $\mathfrak{F}_{\tau+0}$ заміною події $\{\tau \le t\}$ у означенні на подію $\{\tau < t\}$. Знайти аналоги тверджень теореми про перетворення моментів зупинки.

(9, теорема Гальмаріно) Нехай $\Omega = C[0, 1]$, $\mathfrak{F} = \mathfrak{B}[\Omega]$, $\mathfrak{F}_t = \sigma[\{\omega : \omega_s < x\}, s \le t, x \in \mathbb{R}]$. Довести, що невід'ємна борелева функція $\tau = \tau(\omega) : \Omega \to \mathbb{R}$ є марковським моментом відносно потоку $(\mathfrak{F}_t)$ тоді й тільки тоді, коли з $\omega_s = \omega'_s$ при всіх $s \le t$ та з $\tau(\omega) \le t$ випливає, що $\tau(\omega) = \tau(\omega')$.

(10) Величина ξ набуває ненульових значень, а $\tau \in \mathbb{Z}_+$. Довести, що послідовність $(\xi 1\!\!1_{\{\tau \le n\}}, n \ge 0)$ є узгодженою з заданим потоком $(\mathfrak{F}_n)$ тоді й тільки тоді, коли τ є моментом зупинки відносно цього потоку, а ξ є $\mathfrak{F}_\tau$-вимірною. Знайти відповідні умови для послідовності $(\xi 1\!\!1_{\{\tau < n\}}, n \ge 0)$.

2.13.4. Випадкова зупинка субмартингалу

Теорема (теорема Дуба про вільний вибір). *Нехай* $(\xi_n, n = \overline{0, N})$ – *мартингал (субмартингал) відносно потоку* $(\mathfrak{F}_n)$. *Якщо* $\tau_1 \le \tau_2 \le \ldots \le \tau_m$ *– моменти зупинки відносно цього потоку зі значеннями у інтервалі* $\overline{0, N}$, *то послідовність* $(\xi_{\tau_k}, 1 \le k \le m)$ *є мартингалом (відповідно субмартингалом) відносно потоку* $(\mathfrak{F}_{\tau_k}, 1 \le k \le m)$.

Зокрема, $\xi_{\tau_k} = (\le)\mathbf{E}(\xi_{\tau_{k+1}} \mid \mathfrak{F}_{\tau_k})$, *м.н.* $\forall k < m$.

Доведення. Послідовність сигма-алгебр $(\mathfrak{F}_{\tau_k})$ є стохастичним *потоком* за теоремою *про події що передують моментам зупинки*, з якої випливає також $\mathfrak{F}_{\tau_k}$-вимірність величин ξ_{τ_k}.

Визначимо рекурентно при $j = \overline{0, N-1}$ випадкові величини, які внаслідок теореми *про перетворення моментів зупинки* є *моментами зупинки:*

$$\tau_{k0} = \tau_k, \ \tau_{k,j+1} = (\tau_k + j) \wedge \tau_{k+1}.$$

Зауважимо, що за означенням $\tau_{k,j+1} - \tau_{k,j} \in \{0, 1\}$. Оскільки $\tau_{k+1} \le N$, то $\tau_{kN} = \tau_{k+1}$. Крім того, при фіксованому k послідовність τ_{kj} не спадає за j. Отже, подвійна послідовність $(\tau_{kj}, \ 0 \le j \le N, \ 0 \le k \le m)$ також є неспадною послідовністю моментів зупинки зі значеннями у $\overline{0, N}$. Ця послідовність містить (τ_k) як підпослідовність та після перенумерації задовольняє умову $\tau_{i+1} - \tau_i \in \{0, 1\}$. Тому за теоремою *про перетворення мартингальних послідовностей* досить довести твердження теореми за умови, що $\tau_{k+1} - \tau_k \in \{0, 1\}$.

З урахуванням наведеної вище властивості вимірності величин ξ_{τ_k} за лемою *про критерій мартингальної властивості* досить перевірити при всіх $0 \le k < m$ рівності (для випадку субмартингалу – нерівності виду ≥ 0):

$$\mathbf{E}(\xi_{\tau_{k+1}} - \xi_{\tau_k}) 1\!\!1_A = 0, \ \forall A \in \mathfrak{F}_{\tau_k}.$$

Для цього розглянемо при $A \in \mathfrak{F}_{\tau_k}$ останнє математичне сподівання з урахуванням умови $\tau_{k+1} - \tau_k \in \{0, 1\}$:

$$\mathbf{E}(\xi_{\tau_{k+1}} - \xi_{\tau_k}) 1\!\!1_A = \mathbf{E}(\xi_{\tau_{k+1}} - \xi_{\tau_k}) 1\!\!1_A 1\!\!1_{\{\tau_{k+1} - \tau_k = 1\}} =$$

$$\mathbf{E} \sum\nolimits_{s \ge 0} 1\!\!1_{\{\tau_k = s\}} (\xi_{\tau_{k+1}} - \xi_s) 1\!\!1_A 1\!\!1_{\{\tau_{k+1} - s = 1\}} =$$

$$\mathbf{E}\sum\nolimits_{s\ge 0} 1\!\!\mathrm{I}_{\{\tau_k=s\}}(\xi_{s+1}-\xi_s)1\!\!\mathrm{I}_A 1\!\!\mathrm{I}_{\{\tau_{k+1}=s+1\}} =$$
$$\mathbf{E}\sum\nolimits_{s\ge 0}\mathbf{E}\left((\xi_{s+1}-\xi_s)1\!\!\mathrm{I}_{A\cap\{\tau_k=s\}}1\!\!\mathrm{I}_{\{\tau_{k+1}>s\}}\mid \mathfrak{F}_s\right) =$$
$$\mathbf{E}\sum\nolimits_{s\ge 0}\mathbf{E}((\xi_{s+1}-\xi_s)\mid \mathfrak{F}_s)1\!\!\mathrm{I}_{A\cap\{\tau_k=s\}}1\!\!\mathrm{I}_{\{\tau_{k+1}>s\}} = 0.$$

Тут враховані включення $A\cap\{\tau_k=s\}\in\mathfrak{F}_s$, $\{\tau_{k+1}>s\}\in\mathfrak{F}_s$, з означення *моменту зупинки*, та теорема *про властивості умовного сподівання* $\boxdot$

З теореми про вільний вибір випливає низка нерівностей для максимуму мартингальних послідовностей.

Теорема (нерівність Колмогорова для субмартингалу).

(а) Нехай $(\xi_n, n\in\overline{0,N})$ *– субмартингал відносно потоку* $(\mathfrak{F}_n)$. *Тоді для кожного* $\varepsilon>0$ *справедлива нерівність*

$$\mathbf{P}\left(\max\nolimits_{0\le n\le N}\xi_n\ge\varepsilon\right)\le \mathbf{E}\xi_N^+\ /\varepsilon.$$

(б) Для субмартингалу $(\xi_n, n\in\mathbb{Z}_+)$ *при кожному* $\varepsilon>0$

$$\mathbf{P}(\sup\nolimits_{n\ge 0}\xi_n\ge\varepsilon)\le\sup\nolimits_{N\ge 0}\mathbf{E}\xi_N^+\ /\varepsilon.$$

Доведення

(а) Розглянемо момент зупинки $\tau = N\wedge\inf(n\le N:\xi_n\ge\varepsilon)$.

За *теоремою Дуба про вільний вибір* послідовність (ξ_τ,ξ_N) є субмартингалом відносно відповідного потоку.

Оскільки $\{\max_{0\le n\le N}\xi_n\ge\varepsilon\}=\{\xi_\tau\ge\varepsilon\}$, то

$$\varepsilon\mathbf{P}(\max\nolimits_{0\le n\le N}\xi_n\ge\varepsilon)=\mathbf{E}\varepsilon 1\!\!\mathrm{I}_{\{\xi_\tau\ge\varepsilon\}}\le\mathbf{E}\xi_\tau 1\!\!\mathrm{I}_{\{\xi_\tau\ge\varepsilon\}}\le\mathbf{E}\xi_N 1\!\!\mathrm{I}_{\{\xi_\tau\ge\varepsilon\}}\le\mathbf{E}\xi_N^+,$$

де середня нерівність випливає з леми *про критерій мартингальної властивості*, оскільки $\{\xi_\tau\ge\varepsilon\}\in\mathfrak{F}_\tau$, а інші є наслідками нерівностей між величинами під знаком математичного сподівання.

(б) Звуження (ξ_n) на $\overline{0,N}$ є субмартингалом. З нерівності (а) для нього заміною правої частини на її верхню межу і граничним переходом $N\to\infty$ отримуємо нерівність (б) $\boxdot$

Теорема (нерівність Колмогорова для мартингалу).

(а) Нехай $(\xi_n, n\in\overline{0,N})$ *– мартингал відносно потоку* $(\mathfrak{F}_n)$. *Тоді для кожного* $p\ge 1$ *та всіх* $\varepsilon>0$ *виконується нерівність*

$$\mathbf{P}(\max\nolimits_{0\le n\le N}|\xi_n|\ge\varepsilon)\le\mathbf{E}\left|\xi_N\right|^p\ /\varepsilon^p.$$

(б) Для мартингалу $(\xi_n, n\in\mathbb{Z}_+)$ *при всіх* $p\ge 1$, $\varepsilon>0$

$$\mathbf{P}(\sup\nolimits_{n\ge 0}|\xi_n|\ge\varepsilon)\le\sup\nolimits_{N\ge 0}\mathbf{E}\left|\xi_N\right|^p\ /\varepsilon^p.$$

Доведення

(а) За наслідком теореми *про перетворення мартингальних послідовностей* послідовність $\zeta_n \equiv |\xi_n|^p$ є субмартингалом. Застосування *нерівності Колмогорова для субмартингалу* до цієї послідовності зі сталою $\varepsilon^p > 0$ призводить до нерівності (а).

(б) Звуження (ξ_n) на $\overline{0, N}$ є мартингалом. Тому нерівність (б) є очевидним наслідком (а) після граничного переходу $N \to \infty$ $\boxdot$

Вправи

(1) На прикладі симетричного блукання Бернуллі з моментом досягнення стану 1 перевірити суттєвість умови обмеженості $\tau \leq N$ в теоремі Дуба про вільний вибір.

(2) Послідовність інтегровних величин $(\xi_n, 0 \leq n \leq N)$ узгоджена з потоком $(\mathfrak{F}_n)$. Довести, що вона утворює мартингал (субмартингал) тоді й тільки тоді, коли $\mathbf{E}\xi_\tau = (\leq)\mathbf{E}\xi_\nu$ для довільних моментів зупинки $\tau \leq \nu \leq N$.

(3) Нехай $(\xi_n, n \geq 0)$ – $(\mathfrak{F}_n)$-субмартингал такий, що $\xi_n \leq \mathbf{E}(\xi \mid \mathfrak{F}_n)$ м.н. для інтегровної ξ, а τ, ν – моменти зупинки. Тоді $\xi_{\tau\wedge\nu} \leq \mathbf{E}(\xi_\nu \mid \mathfrak{F}_\tau)$ м.н.

(4) Якщо (ξ_n) – мартингал, а τ – момент зупинки, то $\mathbf{E}\,|\xi_\tau| \leq \lim_{n\to\infty} \mathbf{E}\,|\xi_n|$.

(5) Нехай $(\xi_n, n \geq 0)$ – субмартингал, $\varepsilon > 0$. Довести такі нерівності:

(а) $\varepsilon\mathbf{P}(\min_{k\leq n} \xi_k \leq -\varepsilon) \leq \mathbf{E}\xi_n^+ - \mathbf{E}\xi_0$,

(б) $\varepsilon\mathbf{P}(\max_{k\leq n} |\xi_k| \geq \varepsilon) \leq 3\max_{k\leq n} \mathbf{E}\,|\xi_k|$.

(6) Для мартингалу $(\xi_n, n \geq 0)$ довести при $\alpha > 1$ нерівності:

$\mathbf{E}\,|\xi_n|^\alpha \leq \mathbf{E}\max_{k\leq n} |\xi_k|^\alpha \leq \mathbf{E}\,|\xi_n|^\alpha\,(\alpha/(\alpha-1))^\alpha$.

(7) Для мартингалу чи невід'ємного супермартингалу (ξ_n) довести, що

$\mathbf{E}\max_{k\leq n} |\xi_k| \leq \frac{e}{e-1}(1 + \mathbf{E}\,|\xi_n| \ln\max(|\xi_n|, 1))$.

(8) Нехай $(\xi_n, n \geq 0)$ – мартингал, а τ – момент зупинки такий, що величина ξ_τ інтегровна і $\mathbf{E}\xi_n 1\!\!\mathrm{I}_{\{\tau>n\}} \to 0, n \to \infty$. Довести, що $\mathbf{E}\xi_\tau = \mathbf{E}\xi_0$.

(9) Випадкові величини $(\zeta_n, n \geq 1)$ незалежні однаково розподілені, $\mathbf{E}\zeta_1 = 0$, та $\xi_n = \zeta_1 + ... + \zeta_n$. Довести, що $\mathbf{E}\max_{1\leq k\leq n} |\xi_k| \leq 8\mathbf{E}\,|\xi_n|$.

(10) Нехай $(\xi_n, n \geq 0)$ – субмартингал, g – додатна неспадна опукла донизу функція, і $\varepsilon > 0$. Довести, що $\varepsilon\mathbf{P}(\max_{k\leq n} \xi_k \geq \varepsilon) \leq \mathbf{E}g(\xi_n)/g(\varepsilon)$.

(11) Випадкові величини $(\zeta_n, n \geq 1)$ задовольняють такі умови: $\mathbf{E}\zeta_k^2 < \infty$ і $\mathbf{E}(\zeta_k \mid \zeta_1, ...\zeta_{k-1}) = 0, k > 1$, а $\xi_n = \zeta_1 + ... + \zeta_n$. Довести при $\varepsilon > 0$ нерівність $\mathbf{P}(\max_{1\leq k\leq n} |\xi_k| \geq \varepsilon) \leq \mathbf{E}\xi_n^2 \,/\, \varepsilon^2$.

(12, Тотожності Вальда) Величини $(\zeta_n, n \geq 1)$ незалежні однаково розподілені, $\xi_n = \zeta_1 + ... + \zeta_n$, а τ – інтегровний момент зупинки відносно $(\sigma[\zeta_k, k \leq n])$. Якщо величини ζ_k є (а) інтегровними, то $\mathbf{E}\xi_\tau = \mathbf{E}\tau\mathbf{E}\zeta_1$, (б) квадратично інтегровними, то $\mathbf{E}(\xi_\tau - \tau\mathbf{E}\zeta_1)^2 = \mathbf{E}\tau\mathbf{D}\zeta_1$. (в) Якщо $\varphi(t) \equiv \mathbf{E}\exp(it\zeta_1) \neq 0$, а момент τ обмежений, то $\mathbf{E}\left(\exp(it\xi_\tau)(\varphi(t))^{-\tau}\right) = 1$.

(13) Довести твердження попередньої вправи для різнорозподілених незалежних величин. *Вказівка*: замінити інтегровність τ на $\mathbf{E}\left(\sum_{k=0}^{\tau}\mathbf{E}\,|\zeta_k|\right) < \infty$.

(14) Нехай $(\xi_n, n = \overline{0,N})$ – мартингал, а τ – момент зупинки. Довести, що послідовність $\eta_n = \xi_n 1\!\!\mathrm{I}_{\{n\le\tau\}} + (2\xi_\tau - \xi_n) 1\!\!\mathrm{I}_{\{n>\tau\}}$ – також мартингал.

(15) Нехай $(\xi_n^{(k)}, n \ge 0)$ – субмартингали, а τ – момент зупинки такий, що $\xi_\tau^{(1)} \le \xi_\tau^{(2)}$. Тоді послідовність $\eta_n = \xi_n^{(1)} 1\!\!\mathrm{I}_{\{\tau\le n\}} + \xi_n^{(2)} 1\!\!\mathrm{I}_{\{\tau>n\}}$ є субмартингалом.

2.13.5. Число перетинів субмартингалом смуги

Для послідовності випадкових величин $(\xi_n, n \le N)$, пари дійсних $a < b$ та $k \ge 1$ розглянемо випадкові події

$$A_k^N(a,b) \equiv \bigcup_{1\le s_1<t_1<s_2<\ldots<s_k<t_k\le N} \left\{\xi_{s_1} \ge b,\ \xi_{t_1} \le a,\ \xi_{s_2} \ge b, ..., \xi_{t_k} \le a\right\}.$$

Зауважимо, що події $A_k^N(a,b)$ не зростають за k, та $A_N^N(a,b) = \varnothing$.

Означення. Число перетинів *$\theta_N(a,b)$ згори донизу смуги (a,b) послідовністю $(\xi_n, n \le N)$ дорівнює $k \ge 1$, якщо справджується випадкова подія*
$A_k^N(a,b)\backslash A_{k+1}^N(a,b)$, та дорівнює нулю на доповненні $A_1^N(a,b)$.

Очевидно, що величина $\theta_N(a,b)$ визначена коректно для всіх ω.

Теорема (про нерівність Дуба для числа перетинів).

(а) Нехай $(\xi_n, n = \overline{0,N})$ – субмартингал відносно потоку $(\mathfrak{F}_n)$. Тоді при $a < b$ виконується нерівність

$$\mathbf{E}\theta_N(a,b) \le \mathbf{E}(\xi_N - b)^+ \ / \ (b-a).$$

(б) Для субмартингалу $(\xi_n, n \in \mathbb{Z}_+)$ при всіх $a < b$

$$\mathbf{E}\theta_\infty(a,b) \le \sup\nolimits_{N\ge0} \mathbf{E}(\xi_N - b)^+ \ / \ (b-a).$$

де величина $\theta_\infty(a,b)$ визначена як число перетинів згори донизу на всій множині $\mathbb{Z}_+$ – заміною N на ∞ в означенні $A_k^N(a,b)$.

Доведення

(а) Визначимо рекурентно при $k \ge 1$ величини

$$\nu_k = \inf(s > \tau_{k-1} : \xi_s \ge b),\ \ \tau_k = \inf(t > \nu_{k-1} : \xi_t \le a),$$

де $\tau_0 \equiv 0, \nu_0 \equiv 0$, нижні межі обчислюються за $s,t \le N$ і вважається, що $\inf\varnothing = N$. За означенням

$$\{\theta_N(a,b) = m\} = \{\nu_m < N\} \cap \{\xi_{\tau_m} \le a\} \cap (\{\xi_{\nu_{m+1}} < b\} \cup \{\xi_{\tau_{m+1}} > a\}) \subset$$
$$(\cap_{k=1}^m \{\xi_{\nu_k} - \xi_{\tau_k} \ge b-a\}) \cap \{\tau_{m+1} = N\} \cap (\cap_{k=m+2}^N \{\nu_k = \tau_k = N\}).$$

За індукцією перевіряємо, що величини ν_k, τ_k є *моментами зупинки:*

$$\{\nu_{k+1} \le m\} = \cup_{n=1}^{m-1}\{\tau_k = n\} \cap (\cup_{s=n+1}^{m}\{\xi_s \ge b\}) =$$
$$\cup_{n=1}^{m-1} \cup_{s=n+1}^{m} \{\tau_k = n,\ \xi_s \ge b\} \in \mathfrak{F}_m.$$

Розглянемо випадкову величину

$$\zeta = \sum\nolimits_{k=1}^{N}(\xi_{\nu_k} - \xi_{\tau_k}).$$

Оскільки $\nu_k \le \tau_k$ – моменти зупинки, то за *теоремою Дуба про вільний вибір* пара $(\xi_{\nu_k}, \xi_{\tau_k})$ є субмартингалом. Отже, з теореми *про властивості умовного сподівання* та означення *субмартингалу* отримуємо

$$\mathbf{E}\zeta = \sum\nolimits_{k=1}^{N} \mathbf{E}(\xi_{\nu_k} - \xi_{\tau_k}) = \sum\nolimits_{k=1}^{N} \mathbf{E}(\xi_{\nu_k} - \mathbf{E}(\xi_{\tau_k} \mid \mathfrak{F}_{\nu_k})) \le 0.$$

З іншого боку, з наведеного вище включення для події $\{\theta_N(a,b) = m\}$ виводимо такі співвідношення

$$\zeta = \sum\nolimits_{m=0}^{N} \zeta 1\!\!1_{\{\theta_N(a,b)=m\}} =$$

$$\sum\nolimits_{m=0}^{N} 1\!\!1_{\{\theta_N(a,b)=m\}} \left(\sum\nolimits_{k=1}^{m}(\xi_{\nu_k} - \xi_{\tau_k}) + \xi_{\nu_{m+1}} - \xi_{\tau_{m+1}} + \sum\nolimits_{k=m+2}^{N}(\xi_N - \xi_N)\right)$$
$$\ge \sum\nolimits_{m=0}^{N} 1\!\!1_{\{\theta_N(a,b)=m\}} \left(m(b-a) + \xi_{\nu_{m+1}} - \xi_{\tau_{m+1}}\right) =$$
$$(b-a)\theta_N(a,b) + \sum\nolimits_{m=0}^{N} 1\!\!1_{\{\theta_N(a,b)=m\}}(\xi_{\nu_{m+1}} - \xi_N) =$$
$$(b-a)\theta_N(a,b) + \sum\nolimits_{m=0}^{N} 1\!\!1_{\{\theta_N(a,b)=m\}} 1\!\!1_{\{\xi_{\nu_{m+1}} \ge b\}}(\xi_{\nu_{m+1}} - \xi_N) \ge$$
$$(b-a)\theta_N(a,b) + \sum\nolimits_{m=0}^{N} 1\!\!1_{\{\theta_N(a,b)=m\}} 1\!\!1_{\{\xi_{\nu_{m+1}} \ge b\}}(b - \xi_N) \ge$$
$$(b-a)\theta_N(a,b) - \sum\nolimits_{m=0}^{N} 1\!\!1_{\{\theta_N(a,b)=m\}}(\xi_N - b)^+ = (b-a)\theta_N(a,b) - (\xi_N - b)^+,$$

де враховано рівність $\tau_{m+1} = N$ на події $\{\theta_N(a,b) = m\}$, та тотожність $\xi_{\nu_{m+1}} - \xi_N = \xi_N - \xi_N = 0$ на події $\{\xi_{\nu_{m+1}} < b\}$.

З наведених нерівностей виводимо оцінку

$$0 \ge \mathbf{E}\zeta \ge \mathbf{E}((b-a)\theta_N(a,b) - (\xi_N - b)^+) = (b-a)\mathbf{E}\theta_N(a,b) - \mathbf{E}(\xi_N - b)^+,$$

що і доводить твердження (а).

(б) Позначимо через $\theta_N(a,b)$ число перетинів смуги згори донизу для звуження (ξ_n) на $\overline{0,N}$. Тоді $\theta_N(a,b) \uparrow \theta_\infty(a,b), N \to \infty$. Тому нерівність (б) випливає з (а) за *теоремою Лебега про монотонну збіжність* $\boxdot$

2.13.6. Збіжність мартингальних послідовностей

Лема (про критерій збіжності числової послідовності). *Для числової послідовності $x = (x_n, n \geq 0)$ позначимо через $\theta^x(a, b)$ число перетинів згори донизу смуги $a < b$ послідовністю x, що визначається так само, як і величина $\theta_\infty(a, b)$ для $(\xi_n, n < \infty)$.*

Для існування скінченної границі $\lim_{n\to\infty} x_n$ необхідно та достатньо, щоб для всіх $a < b$ зі щільної підмножини $\mathbb{R}$

$$\theta^x(a, b) < \infty, \text{ та } \lim_{m\to\infty} \theta^x(-m, b) = 0, \ \lim_{m\to\infty} \theta^x(a, m) = 0.$$

Доведення. Оскільки величина $\theta^x(a, b)$ набуває цілих невід'ємних значень, то з умови $\lim_{m\to\infty} \theta^x(-m, b) = 0$ виводимо, що $\theta^x(-m, b) = 0$, починаючи з деякого m, тобто послідовність x обмежена знизу. Аналогічно переконуємося у обмеженості x згори. Тому границі $\underline{\lim}_{n\to\infty} x_n \leq \overline{\lim}_{n\to\infty} x_n$ є скінченними. Якщо остання нерівність є строгою, то для деяких $\alpha, \beta > 0$ кількість перетинів смуги $\underline{\lim}_{n\to\infty} x_n + \alpha < \overline{\lim}_{n\to\infty} x_n - \beta$ є нескінченною, що суперечить умові. Отже, існування границі $\lim_{n\to\infty} x_n$ доведено від супротивного. Необхідність наведеної умови очевидна $\boxdot$

Теорема (про збіжність субмартингалу). *Нехай $(\xi_n, n \in \mathbb{Z}_+)$ – субмартингал відносно потоку $(\mathfrak{F}_n, n \in \mathbb{Z}_+)$.*

(а) Якщо $\sup_{n\geq 0} \mathbf{E}\xi_n^+ < \infty$, то м.н. існує $\lim_{n\to\infty} \xi_n \equiv \xi_\infty$, і $\mathbf{E}\,|\xi_\infty| < \infty$.

(б) За умови рівномірної інтегровності послідовності (ξ_n) має місце збіжність у середньому: $\mathbf{E}\,|\xi_n - \xi_\infty| \to 0, n \to \infty$, причому для всіх $n \geq 0$ м.н. справедливі нерівності $\xi_n \leq \mathbf{E}(\xi_\infty \mid \mathfrak{F}_n)$. Тому розширена послідовність $(\xi_n, n \in \mathbb{Z}_+ \cup \{\infty\})$ також є субмартингалом, якщо обрати $\mathfrak{F}_\infty \equiv \sigma[\cup_{n\geq 0}\mathfrak{F}_n]$.

Доведення

(а) З теореми *про нерівність Дуба для числа перетинів*, (б), виводимо, що $\theta_\infty(a, b) < \infty$ м.н. для всіх $a < b$. З тієї ж нерівності при $b \geq 0$ маємо $\mathbf{E}\theta_\infty(a, b) \leq (\sup_{N\geq 0} \mathbf{E}\xi_N^+)/(b - a) \to 0$ при $b \to \infty$ або $a \to -\infty$. Оскільки величина $\theta_\infty(-a, b) \in \mathbb{Z}_+$ не зростає за $a.b \geq 0$, з останньої нерівності випливає, що $\theta_\infty(-a, b) \downarrow 0$ м.н. при $a \to \infty$ або $b \to \infty$. Тому

$$\mathbf{P}\left(\bigcap_{a<b\in\mathbb{Q}} \left\{\theta_\infty(a, b) < \infty, \ \lim_{m\to\infty} \theta_\infty(-m, b) = 0, \ \lim_{m\to\infty} \theta_\infty(a, m) = 0\right\}\right) = 1.$$

Отже, за лемою *про критерій збіжності числової послідовності* має місце рівність $\mathbf{P}(\exists \lim_{n\to\infty} \xi_n \in \mathbb{R}) = 1$. Позначимо через $\xi_\infty \equiv \lim_{n\to\infty} \xi_n$.

За означенням *субмартингалу* послідовність $\mathbf{E}\xi_n$ не спадає, тому

$$\sup_{n\geq 0}\mathbf{E}\,|\xi_n| = \sup_{n\geq 0}(2\mathbf{E}\xi_n^+ - \mathbf{E}\xi_n) \leq 2\sup_{n\geq 0}\mathbf{E}\xi_n^+ - \mathbf{E}\xi_0 < \infty.$$

Звідси за *нерівністю Фату* з *теореми Лебега про мажоровану збіжність* випливає інтегровність

$$\mathbf{E}\,|\xi_\infty| = \mathbf{E}\underline{\lim}_{n\to\infty}|\xi_n| \leq \underline{\lim}_{n\to\infty}\mathbf{E}\,|\xi_n| < \infty.$$

(б) З твердження (а) та теореми *про граничний перехід для рівномірно інтегровної послідовності* виводимо, що $\mathbf{E}\,|\xi_n - \xi_\infty| \to 0, n\to\infty$.

Далі, за означенням *субмартингалу* та *умовного математичного сподівання* $\mathbf{E}\xi_s 1\!\!1_A \leq \mathbf{E}\xi_t 1\!\!1_A, \forall A\in\mathfrak{F}_s$ при кожних $s<t$. За теоремою *про умови рівномірної інтегровності* послідовність $(\xi_t 1\!\!1_A, t\geq 0)$ також є рівномірно інтегровною. Тому за теоремою *про граничний перехід для рівномірно інтегровної послідовності* спрямуванням $t\to\infty$ з останньої нерівності отримуємо: $\mathbf{E}\xi_s 1\!\!1_A \leq \mathbf{E}\xi_\infty 1\!\!1_A, \forall A\in\mathfrak{F}_s$, що внаслідок $\mathfrak{F}_s$-вимірності ξ_s за лемою *про критерій мартингальної властивості* спричиняє нерівність $\xi_n \leq \mathbf{E}(\xi_\infty \mid \mathfrak{F}_n)$.

Останнє твердження теореми є очевидним наслідком попередньої нерівності, якщо врахувати $\mathfrak{F}_\infty$-вимірність величин ξ_n та $\overline{\lim}_{n\to\infty}\xi_n$, та обрати дану границю як $\mathfrak{F}_\infty$-вимірний варіант для ξ_∞ $\boxdot$

Теорема (про збіжність мартингалу). *Нехай* $(\xi_n, n\in\mathbb{Z}_+)$ – *мартингал відносно потоку* $(\mathfrak{F}_n, n\in\mathbb{Z}_+)$, *а* $\mathfrak{F}_\infty = \sigma[\cup_{n\geq 0}\mathfrak{F}_n]$.

(а) Якщо $\sup_{n\geq 0}\mathbf{E}\,|\xi_n| < \infty$, *то м.н. існує* $\lim_{n\to\infty}\xi_n \equiv \xi_\infty$, *причому* $\mathbf{E}\,|\xi_\infty| < \infty$.

(б) Для того, щоб існувала інтегровна $\mathfrak{F}_\infty$*-вимірна величина* ζ *така, що* $\xi_n = \mathbf{E}(\zeta \mid \mathfrak{F}_n)$ *м.н. для всіх* $n\geq 0$, *необхідно і достатньо, щоб послідовність* (ξ_n) *була рівномірно інтегровною.*

(в) За умов (б) $\zeta = \xi_\infty$ *м.н.,* $\mathbf{E}\,|\xi_n - \xi_\infty| \to 0, n\to\infty$, *причому послідовність* $(\xi_n, n\in\mathbb{Z}_+\cup\{\infty\})$ *є мартингалом відносно потоку* $(\mathfrak{F}_n, n\in\mathbb{Z}_+\cup\{\infty\})$.

Доведення

(а) Оскільки мартингал є субмартингалом, а $\mathbf{E}\xi_n^+ \leq \mathbf{E}\,|\xi_n|$, то існування та інтегровність величини $\xi_\infty = \lim_{n\to\infty}\xi_n$ випливає з теореми *про збіжність субмартингалу.*

(б) *Достатність* рівномірної інтегровності. З (а) та теореми *про граничний перехід для рівномірно інтегровної послідовності* випливає збіжність ξ_n *у середньому* до величини ξ_∞. Остання інтегровна згідно з (а) та м.н. дорівнює $\mathfrak{F}_\infty$-вимірній величині $\zeta = \overline{\lim}_{n\to\infty}\xi_n$, оскільки для збіжної послідовності верхня границя дорівнює границі.

Далі, як і вище, з рівності $\mathbf{E}\xi_s 1\!\!1_A = \mathbf{E}\xi_t 1\!\!1_A, \forall A \in \mathfrak{F}_s$ при всіх $s < t$, з теореми *про граничний перехід для рівномірно інтегровної послідовності* спрямуванням $t \to \infty$ виводимо $\mathbf{E}\xi_s 1\!\!1_A = \mathbf{E}\zeta 1\!\!1_A$ і $\xi_n = \mathbf{E}(\zeta \mid \mathfrak{F}_n)$ м.н.

Необхідність. За теоремою *про властивості умовного сподівання* внаслідок зображення (б) для всіх $A \in \mathfrak{F}_n$ справедлива нерівність

$$\mathbf{E}\left|\xi_n\right| 1\!\!1_A \leq \mathbf{E}(\mathbf{E}(|\zeta| \mid \mathfrak{F}_n) 1\!\!1_A) = \mathbf{E}(\mathbf{E}(|\zeta| \, 1\!\!1_A \mid \mathfrak{F}_n)) = \mathbf{E}(|\zeta| \, 1\!\!1_A).$$

Звідси $\mathbf{P}(|\xi_n| \geq c) \leq \mathbf{E}\left|\xi_n\right|/c \leq \mathbf{E}\left|\zeta\right|/c$ для $c > 0$ за *загальною нерівністю Чебишева.* Отже, з $\mathfrak{F}_n$-вимірності ξ_n при $b > 0$ випливає нерівність

$$\mathbf{E}\left|\xi_n\right| 1\!\!1_{\{|\xi_n| \geq c\}} \leq \mathbf{E}\left|\zeta\right| 1\!\!1_{\{|\xi_n| \geq c\}} = \mathbf{E}\left|\zeta\right| 1\!\!1_{\{|\zeta| \geq b, |\xi_n| \geq c\}} + \mathbf{E}\left|\zeta\right| 1\!\!1_{\{|\zeta| < b, |\xi_n| \geq c\}} \leq$$

$$\mathbf{E}\left|\zeta\right| 1\!\!1_{\{|\zeta| \geq b\}} + b\mathbf{P}(|\xi_n| \geq c) \leq \mathbf{E}\left|\zeta\right| 1\!\!1_{\{|\zeta| \geq b\}} + b\mathbf{E}\left|\zeta\right|/c,$$

звідки $\overline{\lim}_{c \to \infty} \sup_{n \geq 0} \mathbf{E}\left|\xi_n\right| 1\!\!1_{\{|\xi_n| \geq c\}} \leq \mathbf{E}\left|\zeta\right| 1\!\!1_{\{|\zeta| \geq b\}} + 0$. Оскільки ліва частина не залежить від b, спрямувавши $b \to \infty$, виводимо, що вона дорівнює нулю. Останнє за теоремою *про умови рівномірної інтегровності* доводить рівномірну інтегровність.

(в) Рівність $\zeta = \xi_\infty$ м.н доведено в частині достатності умови рівномірної інтегровності, справедливість якої також встановлено. Збіжність у середньому випливає з рівномірної інтегровності, як вказано вище $\boxdot$

Вправи

(1) Величини $(\varepsilon_k, k \geq 1)$ незалежні, $\mathbf{P}(\varepsilon_k = \pm 1) = 1/2$, $\xi_k = \varepsilon_1 + ... + \varepsilon_k$. Розглянемо розподіл числа перетинів $q_r = \mathbf{P}(\theta_n(-1,1) = r)$. Довести при $r \geq 1$ тотожність $q_r + q_{r+1} = \mathbf{P}(\max_{1 \leq k \leq n} \xi_k \geq 2r + 1)$.

(2) Незалежні величини (ζ_n) такі, що $\mathbf{P}(\zeta_1 = 0) = 1/2 = \mathbf{P}(\zeta_n = 2)$, а потік $\mathfrak{F}_n = \sigma[\zeta_k, k \leq n]$. Довести, що послідовність $\xi_n = \prod_{k=1}^n \zeta_k$ є мартингалом, однак його не можна подати у вигляді $\xi_n = \mathbf{E}(\xi \mid \mathfrak{F}_n)$ з інтегровною ξ.

(3) Мартингал (ξ_n) з $\sup_n \mathbf{E}\left|\xi_n\right| < \infty$ є різницею невід'ємних мартингалів.

(4) Довести, що невід'ємний супермартингал (мартингал) збігається м.н.

(5) Функція $f \in L_1([0,1))$. Довести, що: (а) послідовність функцій вигляду $f_n(\omega) = \sum_{k=-\infty}^{\infty} 1\!\!1_{\omega \in I_{kn}} 2^n \int_{I_{kn}} f(x)dx$, з $I_{kn} = [k2^{-n}, (k+1)2^{-n})$, є мартингалом на просторі $(\Omega, \mathfrak{F}, \mathbf{P}) = ([0,1), \mathfrak{B}[0,1), L)$, (б) $f_n \to f, n \to \infty$, м.н.

(6) Функція $f : [0,1) \to \mathbb{R}$ задовольняє умову Ліпшиця. Довести, що існує борелєва інтегровна за Лебегом функція g така, що $f(x) = f(0) + \int_0^x g(y)dy$. *Вказівка*: послідовність $\xi_n(\omega) = \sum_{k=0}^{2^n - 1} 2^n (g(t_{n,k+1}) - g(t_{nk})) 1\!\!1_{\{t_{nk} \leq \omega < t_{n,k+1}\}}$ з $t_{nk} \equiv k2^{-n}$ є рівномірно інтегровним мартингалом на просторі $([0,1), \mathfrak{B}[0,1), L)$.

(7) Нехай (ζ_n) послідовність незалежних однаково розподілених інтегровних величин, $S_n = \zeta_1 + ... + \zeta_n$. Довести, що $\mathbf{E}(\zeta_1 \mid \sigma[S_k, k \geq n]) = S_n/n$ м.н. та вивести звідси посилений закон великих чисел: $S_n/n \to^{P1} \mathbf{E}\zeta_1, n \to \infty$.

(8) Мартингал $(\xi_n, n \geq 0)$ такий, що $\left|\xi_{n+1} - \xi_n\right| \leq c$ м.н. Вивести з теореми про збіжність мартингалу, що $\mathbf{P}(\{\overline{\lim}_{n\to\infty} |\xi_n| = \infty\} \cup \{\exists \lim_{n\to\infty} \xi_n \in \mathbb{R}\}) = 1$.

(9) Послідовність σ-алгебр $(\mathfrak{F}_n)$ незростає, $\mathfrak{F}_\infty = \cap_{n\geq 1}\mathfrak{F}_n$, а величина ξ інтегровна. Довести, що $\mathbf{E}(\xi \mid \mathfrak{F}_n) \to \mathbf{E}(\xi \mid \mathfrak{F}_\infty), n \to \infty$, м.н. і в середньому.

(10) Нехай $(\mathfrak{F}_n, n \geq 0)$ – стохастичний потік, $\mathfrak{F}_n \uparrow \mathfrak{F}_\infty, n \to \infty$, і $A \in \mathfrak{F}_\infty$. Довести, що $\mathbf{P}(A \mid \mathfrak{F}_n) \to^{P1} 1\!\!\mathrm{I}_A, n \to \infty$.

(11) Нехай $(\mathfrak{F}_n, n \in \mathbb{Z}_+)$ – стохастичний потік з $\mathfrak{F}_0 = \{\varnothing, \Omega\}$, а $A_n \in \mathfrak{F}_n$. Довести (а) рівність $\overline{\lim}_{n\to\infty} A_n = \left\{\sum_{n\geq 1} \mathbf{P}(A_n \mid \mathfrak{F}_{n-1}) = \infty\right\}$, (б) збіжність $\left(\sum_{k=1}^n 1\!\!\mathrm{I}_{A_k}\right) / \left(\sum_{k=1}^n \mathbf{P}(A_k \mid \mathfrak{F}_{k-1})\right) \to^{P1} 1, n \to \infty$, на множині $\overline{\lim}_{n\to\infty} A_n$.

(12) Сигма-алгебри $(\mathfrak{F}_n, n \geq 0)$ незалежні у сукупності. Довести, що з точністю до подій нульової ймовірності $\cap_{n\geq 1}\sigma[\cup_{k\geq n}\mathfrak{F}_k \cup \mathfrak{F}_0] = \mathfrak{F}_0$.

(13, розклад Рісса) Нехай $(\xi_n, n \in \mathbb{Z}_+)$ – рівномірно інтегровний супермартингал відносно потоку $(\mathfrak{F}_n)$. Тоді $\xi_n = M_n + U_n$, де (M_n) – рівномірно інтегровний мартингал, а (U_n) – потенціал, тобто рівномірно інтегровний невід'ємний супермартингал такий, що $U_n \to^{P1} 0, n \to \infty$.

(14) Випадкові величини $(\xi_n, \zeta_n, n \geq 1)$, невід'ємні, інтегровні, узгоджені з потоком $(\mathfrak{F}_n)$ та $\mathbf{E}(\xi_{n+1} \mid \mathfrak{F}_n) \leq (1 + \zeta_n)\xi_n$ м.н. для всіх $n \geq 1$, причому $\sum_{n\geq 1} \zeta_n < \infty$ м.н. Довести, що ξ_n збігається м.н.

(15) Нехай $(\xi_n, n \in \mathbb{Z}_+)$ – мартингал (субмартингал) відносно натурального потоку $\mathfrak{F}_n \equiv \sigma[\xi_k, k \leq n]$ такий, що $|\xi_0| \leq c$, $\mathbf{E}(\left|\xi_{k+1} - \xi_k\right| \mid \mathfrak{F}_k) \leq c$ м.н. для деякої сталої c, а τ – момент зупинки відносно $(\mathfrak{F}_n)$ і $\mathbf{E}\tau < \infty$. Довести, що випадкова величина ξ_τ – інтегровна та $\xi_\tau = (\leq)\mathbf{E}(\xi_0 \mid \mathfrak{F}_\tau)$ м.н.

(16) Нехай $(\xi_n, n \in \mathbb{Z}_+)$ – мартингал такий, що $\sup \mathbf{E}\xi_n^2 < \infty$. Довести, що $\mathbf{E}\,|\xi_n - \xi|^2 \to 0, n \to \infty$, для деякої ξ.

(17) Випадкові величини $(\zeta_n, n \geq 1)$ незалежні однаково розподілені, а функція $\varphi : \mathbb{R}^k \to \mathbb{R}$ – обмежена борелєва. Визначимо послідовність випадкових величин $\xi_n = (A_n^k)^{-1} \sum_{1\leq i_1, i_2, \ldots i_k \leq n} \varphi(\zeta_{i_1} \ldots \zeta_{i_k}), n \geq k$, де сума обчислюється за всіма наборами різних індексів з $\overline{1, n}$, і нормується на їх кількість. Довести, що (а) послідовність $(\xi_{-n}, n \leq -k)$ – мартингал, (б) $\xi_n \to^{P1} \mathbf{E}\varphi(\zeta_1, \ldots, \zeta_k), n \to \infty$.

2.14. Ланцюги Маркова

Спеціальний клас залежних випадкових величин був розглянутий А.А. Марковим на початку 20 ст. при дослідженні необхідних умов у центральній граничній теоремі. Згодом виявилось, що модель марковської залежності є цілком самодостатньою і може бути застосована в різних прикладних галузях теорії ймовірностей.

Розглянемо послідовність випадкових величин $(\zeta_t,\ t = 0, 1, ...)$, в якій індекс t будемо інтерпретувати як час. Тоді події, що відбуваються в моменти $s < t$, можна інтерпретувати як *минуле* відносно *сучасного* моменту t, а події у моменти $s > t$ – як *майбутнє*.

Марковська залежність означає, що *майбутнє умовно не залежить від минулого за умови відомого сучасного* – тобто залежність відбувається виключно через сучасне:

$$\mathbf{P}(\text{Майбутнє} \cap \text{Минуле} \ \mid \ \text{Сучасне}) =$$

$$\mathbf{P}(\text{Майбутнє} \ \mid \ \text{Сучасне}) \cdot \mathbf{P}(\text{Минуле} \ \mid \ \text{Сучасне})$$

З означення *умовної ймовірності* виводимо, що остання рівність еквівалентна рівності

$$\mathbf{P}(\text{Майбутнє} \mid \text{Минуле} \cap \text{Сучасне}) = \mathbf{P}(\text{Майбутнє} \ \mid \text{Сучасне}).$$

Звідси приходимо до формального означення.

Означення. *Послідовність випадкових величин* $(\zeta_t,\ t = 0, 1, ...)$ *із дискретним простором значень* $E = \{i, j, ...\}$ *називається* однорідним **ланцюгом Маркова,** *якщо для всіх* $t \geq 1,\ i_0, i_1, ..., i_{t-1}, i, j \in E$ *виконуються рівності*

$$\mathbf{P}(\zeta_{t+1} = j \mid \zeta_0 = i_0, ..., \zeta_{t-1} = i_{t-1}, \zeta_t = i) =$$

$$\mathbf{P}(\zeta_{t+1} = j \mid \zeta_t = i) = \mathbf{P}(\zeta_1 = j \mid \zeta_0 = i).$$

Перша з цих рівностей визначає власне **марковську властивість**, а друга – **однорідність за часом**.

Приклади

1. *Незалежні однаково розподілені величини* – для них *умовні ймовірності* збігаються з безумовними.

2. *Суми незалежних однаково розподілених величин* (випадкові блукання). Для них справедлива тотожність

$$\zeta_{t+1} = \zeta_t + \xi_{t+1},$$

де стрибок ξ_{t+1} не залежить від наявного положення ζ_t, оскільки останнє однозначно визначається попередніми стрибками $\xi_s, s \leq t$. Тому умовна ймовірність збігається з безумовною:

$$\mathbf{P}(\zeta_{t+1} = j \mid \zeta_0 = i_0, ..., \zeta_{t-1} = i_{t-1}, \zeta_t = i) =$$

$$\mathbf{P}(i + \xi_{t+1} = j \mid \zeta_0 = i_0, ..., \zeta_{t-1} = i_{t-1}, \zeta_t = i) =$$

$$\mathbf{P}(i + \xi_{t+1} = j) = \ P(i + \xi_{t+1} = j \mid \zeta_t = i)$$

3. *Рекурентні послідовності вигляду*

$$\zeta_{t+1} = g(\zeta_t, \xi_{t+1})$$

із незалежними величинами $(\xi_t, t \geq 1)$ та борелєвими функціями g є ланцюгами Маркова за таких самих міркувань, що і в попередньому пункті.

4. *Серії успіхів у схемі Бернуллі.* Якщо $\xi_k \in \{0,1\}$ результат k-го випробування Бернуллі, то довжина серії успіхів у момент n дорівнює

$$\zeta_n = \inf(k \geq 0 : \xi_{n-k} \neq 0)$$

та задовольняє відповідне марковське рекурентне рівняння.

5. *Модель черги.* Нехай ξ_k – випадкова кількість викликів, що надійшли до сервера на інтервалі $[k, k+1)$, кожний виклик обслуговується за одиницю часу, причому у випадку зайнятості сервера виклики стають у чергу. Якщо ζ_n – кількість заявок у системі в момент n, то марковська властивість є наслідком рівняння

$$\zeta_{n+1} = \max(\zeta_n - 1, 0) + \xi_{n+1}.$$

6. *Гіллясті процеси.*

Вправа. Випадкові величини $(\xi_t, t \geq 0)$ незалежні і $\mathbf{P}(\xi_t = \pm 1) = 1/2$. Довести, що послідовність $\zeta_t = (\xi_t + \xi_{t+1})/2, t \geq 0$, не є ланцюгом Маркова.

2.14.1. Перехідна матриця за один крок

Означення. Матрицею перехідних імовірностей *за один крок марковського ланцюга* $(\zeta_t,\ t \geq 0)$ *називається матриця*

$$P = (p_{ij},\ i,j \in E), \quad p_{ij} = \mathbf{P}(\zeta_1 = j \mid \zeta_0 = i).$$

Початковий розподіл *ланцюга дорівнює* $q_i = \mathbf{P}(\zeta_0 = i)$.

Зауваження. З властивостей умовної ймовірності випливає, що перехідна матриця завжди є стохастичною матрицею, тобто задовольняє умови

$$p_{ij} \geq 0, \ \sum\nolimits_{j \in E} p_{ij} = 1.$$

Теорема (про розподіл траєкторій марковського ланцюга). *Для всіх* $t \geq 1,\ i_0, i_1, ..., i_{t-1}, i_t \in E$ *виконуються рівності*

$$\mathbf{P}(\zeta_0 = i_0, ..., \zeta_{t-1} = i_{t-1}, \zeta_t = i_t) = q_{i_0} p_{i_0 i_1} p_{i_1 i_2} ... p_{i_{t-1} i_t},$$

$$\mathbf{P}(\zeta_1 = i_1, ..., \zeta_{t-1} = i_{t-1}, \zeta_t = i_t \mid \zeta_0 = i_0) = p_{i_0 i_1} p_{i_1 i_2} ... p_{i_{t-1} i_t}.$$

Доведення випливає з теореми *про ймовірність перерізу n подій* (вираз через умовні ймовірності) та з *марковської властивості* ланцюга:

$$\mathbf{P}\left(\cap_{s=0}^{t}\{\zeta_s = i_s\}\right) = \mathbf{P}(\zeta_0 = i_0)\prod\nolimits_{s=1}^{t}\mathbf{P}(\zeta_s = i_s \mid \cap_{r=0}^{s-1}\{\zeta_r = i_r\}) =$$

$$\mathbf{P}(\zeta_0 = i_0)\prod\nolimits_{s=1}^{t}\mathbf{P}(\zeta_s = i_s \mid \zeta_{s-1} = i_{s-1}) = q_{i_0}p_{i_0 i_1}p_{i_1 i_2}...p_{i_{t-1}i_t}.$$

Друге рівняння є наслідком означення *марковської властивості* ⊡

Зауваження. Наведені у теоремі рівняння задають узгоджену послідовність сумісних функцій розподілу, що за *теоремою Колмогорова про побудову ймовірності на множинах послідовностей через сумісні функції розподілу* задають імовірнісну міру на $\mathbb{R}^\infty$, за якою можна побудувати канонічний ланцюг Маркова з *матрицею перехідних імовірностей* за один крок (p_{ij}) та *початковим розподілом* (q_i).

2.14.2. Імовірності переходу за t кроків

Означення. Імовірностями переходу за t кроків *називаються умовні ймовірності*

$$p_{ij}^{(t)} = \mathbf{P}(\zeta_t = j \mid \zeta_0 = i).$$

Теорема (про ймовірності переходу за t кроків).

(а) Справедливі ***рівняння Колмогорова – Чепмена:***

$$p_{ik}^{(t+s)} = \sum\nolimits_{j\in E} p_{ij}^{(t)}p_{jk}^{(s)}, \quad \forall i,k \in E,\ \forall t,s \geq 1.$$

(б) Має місце тотожність

$$P^{(t)} \equiv \left(p_{ij}^{(t)},\ i,j \in E\right) = (P)^t,$$

де права частина є ступенем порядку t матриці P.

Доведення

(а) Суму ймовірностей всіх траєкторій ланцюга довжини $t+s$ вигляду $(i, i_1, ..., i_{t-1}, j, j_1, ..., j_{s-1}, k)$ за теоремою *про розподіл траєкторій марковського ланцюга* можна подати як повторну суму – спочатку за змінними $j_1, ..., j_{s-1}$, далі за змінними $i_1, ..., i_{t-1}$, а потім за j :

$$p_{ik}^{(t+s)} = \sum\nolimits_{i_1,...,.i_{t-1},\ j,\ j_1,...,j_{s-1}\in E} p_{ii_1}p_{i_1 i_2}...p_{i_{t-1}j}\ p_{jj_1}...p_{j_{s-1}k} =$$

$$\sum\nolimits_{j\in E}\left(\sum\nolimits_{i_1,...,i_{t-1}\in E} p_{ii_1}p_{i_1 i_2}...p_{i_{t-1}j}\right)\left(\sum\nolimits_{j_1,...,j_{s-1}\in E} p_{jj_1}...p_{j_{s-1}k}\right).$$

За тією ж теоремою останні дві суми дорівнюють $p_{ij}^{(t)}$ і $p_{jk}^{(s)}$ відповідно.

(б) У матричній формі рівняння (а) має вигляд $P^{(t+s)} = P^{(t)}P^{(s)}$. Звідси (б) виводиться за індукцією, оскільки $P^{(1)} = P$ $\boxdot$

Лема (про нерівність міноризації). *Для всіх $t, s \geq 0$ та $i, j, k \in E$ мають місце нерівності*

$$p_{ik}^{(t+s)} \geq p_{ij}^{(t)} p_{jk}^{(s)}.$$

Доведення очевидне.

Означення. *Для будь-якої події A, що пов'язана з ланцюгом Маркова (ζ_s), для скорочення будемо вживати такі позначення*

$$\mathbf{P}_i(A) \equiv \mathbf{P}(A \mid \zeta_0 = i), \ \mathbf{E}_i\xi \equiv \mathbf{E}(\xi \mid \zeta_0 = i).$$

Вправи

(1) Величини $(\xi_t, t \geq 1)$ незалежні однаково розподілені зі значеннями ± 1, а $S_t = \sum_{s=1}^{t} \xi_s$. Довести, що $\zeta_t \equiv \max_{0 \leq k \leq t} S_k$ – не є ланцюгом Маркова.

(2) Обчислити при кожному $n > 1$ матрицю ймовірностей переходу P^n для випадку з двоелементним простором $E = \{0, 1\}$ та для загальної стохастичної матриці P. Дослідити її асимптотику при $n \to \infty$.

(3) Нехай $(q_k, k = \overline{0, m-1})$ – імовірнісний розподіл, а стохастична матриця $P = (q_{k+j}, 0 \leq k, j \leq m-1)$, де сума $k + j$ обчислюється за модулем m. Обчислити P^m та зайти границю матриць P^t при $t \to \infty$.

(4) Ланцюг Маркова $(\zeta_t, t \geq 0)$ має матрицю перехідних імовірностей P, а $d \in \mathbb{N}$. Довести, що $(\zeta_{td}, t \geq 0)$ – ланцюг Маркова з перехідною матрицею P^d.

2.14.3. Розширена та строга марковська властивість

Для кожного $n \geq 0$ розглянемо сигма-алгебру випадкових подій, породжених значеннями ланцюга до моменту n :

$$\mathfrak{F}_n = \sigma[\zeta_k, 0 \leq k \leq n],$$

та сигма-алгебру, що породжується цими значеннями після моменту n :

$$\mathfrak{F}^n = \sigma[\{\zeta_n \in B_0, ..., \zeta_{n+m} \in B_m\}, \ m \geq 0, B_s \subset E].$$

Зауважимо, що між $\mathfrak{F}^0$ та $\mathfrak{F}^n$ існує ізоморфізм, що полягає у заміні ζ_t на $\zeta_{t+n}, t \geq 0$. Позначимо його через θ^n. За означенням

$$\theta^n\{\zeta_0 \in B_0, ..., \zeta_m \in B_m\} = \{\zeta_n \in B_0, ..., \zeta_{n+m} \in B_m\}.$$

Теорема (про розширену марковську властивість). *Нехай $(\zeta_n, n \geq 0)$ – ланцюг Маркова. Тоді для довільних $n \geq 1$, $k, i \in E$, $A \in \mathfrak{F}^n$, $C \in \mathfrak{F}_n$, $C \subset \{\zeta_n = i\}$ мають місце тотожності*

$$\mathbf{P}(A \mid C, \ \zeta_n = i) = \mathbf{P}(A \mid \zeta_n = i) = \mathbf{P}_i(\theta^{-n}A),$$

$$\mathbf{P}_k(A \mid C,\ \zeta_n = i) = \mathbf{P}_k(A \mid \zeta_n = i) = \mathbf{P}_i(\theta^{-n}A).$$

Доведення. Оскільки права та ліва частини першої тотожності – сигма-адитивні функції від $A \in \mathfrak{F}^n$, то досить довести цю тотожність для подій із класу подій вигляду $A = \{\zeta_n \in B_0, ..., \zeta_{n+m} \in B_m\}$, що породжує $\mathfrak{F}^n$.

Для цього припустимо, що $C = \{\zeta_0 \in C_0, ..., \zeta_n \in C_n\}$ та обчислимо для подій A за теоремою *про розподіл траєкторій марковського ланцюга*

$$\mathbf{P}(C,\ \zeta_n = i,\ \ A) = \mathbf{P}(\zeta_0 \in C_0, ..., \zeta_n \in C_n, \zeta_n = i,\ A) =$$
$$\sum\nolimits_{i_s \in C_s, 0 \le s < n} \sum\nolimits_{j_t \in B_t, 1 \le t \le m} q_{i_0} p_{i_0 i_1} ... p_{i_{n-1} i} 1\!\!1_{i \in B_0} p_{i j_1} ... p_{j_{m-1} j_m} =$$
$$\left(\sum\nolimits_{i_s \in C_s, 0 \le s < n} q_{i_0} p_{i_0 i_1} ... p_{i_{n-1} i}\right) 1\!\!1_{i \in B_0} \left(\sum\nolimits_{j_t \in B_t, 1 \le t \le m} p_{i j_1} ... p_{j_{m-1} j_m}\right) =$$
$$\mathbf{P}(\zeta_0 \in C_0, ..., \zeta_n \in C_n, \zeta_n = i)\ \mathbf{P}(\zeta_0 \in B_0, ..., \zeta_m \in B_m \mid \zeta_0 = i) =$$
$$\mathbf{P}(C,\ \zeta_n = i)\ \mathbf{P}(\theta^{-n}A \mid \zeta_0 = i) = \mathbf{P}(C,\ \zeta_n = i)\ \mathbf{P}(A \mid \zeta_n = i),$$

де використано однорідність ланцюга за часом, включення $i \in C_n$, що є наслідком припущення $C \subset \{\zeta_n = i\}$, та включення $i \in B_0$. Очевидно, що при порушенні останнього включення обидві частини отриманої рівності є нульовими. Оскільки клас подій C наведеного вище вигляду породжує сигма-алгебру $\mathfrak{F}_n$, звідси отримуємо рівність

$$\mathbf{P}(C,\ \zeta_n = i,\ A) = \mathbf{P}(C,\ \zeta_n = i)\ \mathbf{P}(A \mid \zeta_n = i),\ \forall C \in \mathfrak{F}_n.$$

Тому з означення *умовної ймовірності* випливає перша тотожність теореми. Друга є наслідком першої та означення умовної ймовірності $\boxdot$

Розширену марковську властивість можна поширити також і на випадкові моменти часу. Нагадаємо, що означення *моменту зупинки* τ та відповідної сигма-алгебри подій $\mathfrak{F}_\tau$ дано у розділі про мартингальні послідовності.

Теорема (про строго марковську властивість). *Нехай ланцюг Маркова* $(\zeta_n, n \ge 0)$ *породжує стохастичний потік* $(\mathfrak{F}_n \equiv \sigma[\zeta_k, k \le n])$*, а* τ *– момент зупинки відносно цього потоку. Тоді для всіх* $i \in E$, $m \ge 1$, $C \in \mathfrak{F}_\tau$, $C \subset \{\zeta_\tau = i\}$, *і* $B_1, ..., B_m \subset E$ *мають місце тотожності*

$$\mathbf{P}(\zeta_{\tau+1} \in B_1, ..., \zeta_{\tau+m} \in B_m \mid C,\ \zeta_\tau = i) = \mathbf{P}(\zeta_1 \in B_1, ..., \zeta_m \in B_m \mid \zeta_0 = i).$$

Доведення. Обчислимо з урахуванням *марковської властивості*

$$\mathbf{P}(\zeta_{\tau+1} \in B_1, ..., \zeta_{\tau+m} \in B_m,\ C,\ \zeta_\tau = i) =$$
$$\sum\nolimits_{n \ge 0} \mathbf{P}(\zeta_{\tau+1} \in B_1, ..., \zeta_{\tau+m} \in B_m,\ C,\ \zeta_\tau = i,\ \tau = n) =$$
$$\sum\nolimits_{n \ge 0} \mathbf{P}(\zeta_{n+1} \in B_1, ..., \zeta_{n+m} \in B_m,\ C \cap \{\tau = n\},\ \zeta_n = i) =$$

$$\sum\nolimits_{n\geq 0} \mathbf{P}(\cap_{k=1}^m \{\zeta_{n+k} \in B_k\} \mid C \cap \{\tau = n\}, \zeta_n = i)\mathbf{P}(C \cap \{\tau = n\}, \zeta_n = i) =$$
$$\sum\nolimits_{n\geq 0} \mathbf{P}(\cap_{k=1}^m \{\zeta_{n+k} \in B_k\} \mid \zeta_n = i)\ \mathbf{P}(C \cap \{\tau = n\}, \zeta_n = i) =$$
$$\sum\nolimits_{n\geq 0} \mathbf{P}(\cap_{k=1}^m \{\zeta_k \in B_k\} \mid \zeta_0 = i)\ \mathbf{P}(C \cap \{\tau = n\}, \zeta_n = i) =$$
$$\mathbf{P}(\cap_{k=1}^m \{\zeta_k \in B_k\} \mid \zeta_0 = i) \left(\sum\nolimits_{n\geq 0} \mathbf{P}(C \cap \{\tau = n\}, \zeta_\tau = i)\right) =$$
$$\mathbf{P}(\cap_{k=1}^m \{\zeta_k \in B_k\} \mid \zeta_0 = i)\ \mathbf{P}(C, \zeta_\tau = i),$$

де враховані включення $C \cap \{\tau = n\} \in \mathfrak{F}_n$, $C \cap \{\tau = n\} \subset \{\zeta_n = i\}$, що є наслідком означення *моменту зупинки*, та теорему *про розширену марковську властивість* $\boxdot$

2.14.4. Класифікація станів ланцюга Маркова

Означення. *Стан $j \in E$* досяжний *з $i \in E$, позначення $i \to j$, якщо існує $t \geq 0$ таке, що $p_{ij}^{(t)} > 0$. Позначимо через $A(i) = \{j \in E : i \to j\}$ відповідний клас досяжних станів.*

Лема (про транзитивність досяжності). *Якщо $i \to j$ та $j \to k$, то $i \to k$.*

Доведення випливає з леми *про нерівність мінорризації* $\boxdot$

Означення. *Стани $i, j \in E$ сполучаються, позначення $i \longleftrightarrow j$, якщо $i \to j$ та одночасно $j \to i$.*

Теорема (про класи еквівалентності). *Нехай за означенням $p_{ii}^{(0)} = 1$, тобто $i \to i$. Тоді*

(а) відношення $i \longleftrightarrow j$ є відношенням еквівалентності на E,

(б) існує розбиття простору E на ***класи еквівалентності*** *:*

$$E = \bigcup\nolimits_{s\geq 1} E_s, \quad E_s \cap E_t = \varnothing,\ s \neq t,$$

(або ***класи станів*** *E_s), причому включення $i, j \in E_t$ відбуваються при деякому t тоді й тільки тоді, коли $i \longleftrightarrow j$.*

Доведення. Транзитивність відношення $\longleftrightarrow$ випливає з леми *про транзитивність досяжності*, симетрія є наслідком означення, рефлективність випливає з припущення $i \to i$ $\boxdot$

Означення. *Позначимо через $C(i) = \{j \in E : j \longleftrightarrow i\}$ клас станів, який містить даний стан $i \in E$.*

Означення. *Ланцюг Маркова (ζ_t) називається* незвідним, *якщо весь простір E утворює один клас, тобто всі його стани сполучаються.*

Означення. *Стан $i \in E$ називається* істотним станом, *якщо всі стани, що з нього досяжні, сполучаються з ним: $\forall j : i \to j \Longrightarrow j \to i$, тобто у випадку рівності $C(i) = A(i)$.*

Теорема (про істотність класу станів). *Істотність є властивістю класу станів, тобто всередині класу всі стани істотні чи не істотні одночасно. У першому випадку клас називається* ***істотним***.

Доведення є наслідком такого твердження: з *істотного стану досяжним може бути лише істотний стан.* Для його перевірки припустимо від супротивного, що стан i – істотний, $i \to j$, а j – не істотний. Тоді за означенням знайдеться $k \in E$ такий, що $j \to k$ та одночасно $k \nrightarrow j$. За лемою *про транзитивність досяжності* з $i \to j, j \to k$ робимо висновок $i \to k$, звідки $k \to i$, оскільки стан i – істотний. За тією ж транзитивністю з $k \to i, i \to j$ випливає, що $k \to j$. Це суперечить вибору стану k.

Оскільки всередині класу всі стани досяжні з кожного, то наявність хоча б одного істотного стану спричиняє істотність всіх $\boxdot$

Означення. **Періодом** *стану $i \in E$ називається число*

$$d(i) = \text{ н.с.д.} D(i), \quad D(i) \equiv \left\{ n \geq 1 : p_{ii}^{(n)} > 0 \right\},$$

де **н.с.д.**D *– найбільший спільний дільник множини D.*

Лема (про ідеал у множині натуральних чисел). *Нехай множина $D \subset \mathbb{N}$ замкнена відносно додавання та має період $d \equiv$ н.с.д.D. Тоді знайдеться номер N такий, що $\{nd, \ n \geq N, \ n \in \mathbb{N}\} \subset D$.*

Доведення. За алгоритмом Евкліда побудови н.с.д. знайдуться числа $m \geq 1, \ z_1, ..., z_m \in \mathbb{Z}$ та $l_1, ..., l_m \in D$ такі, що $\sum_{s=1}^m z_s l_s = d$. Позначимо $l = l_1 + ... + l_m$. Довільне натуральне n запишемо у формі $n = pl + r$, де $0 \leq r < l$, і $p \to \infty$ при $n \to \infty$. Звідси

$$nd = pld + rd = pd \sum\nolimits_{s=1}^m l_s + r \sum\nolimits_{s=1}^m z_s l_s = \sum\nolimits_{s=1}^m (pd + r z_s) l_s.$$

Вибором досить великого n (а саме, при $pd > r \max |z_s|$) всі цілі множники $pd + r z_s$ можна зробити додатними, тобто натуральними числами. Оскільки D замкнена відносно додавань, включаючи й однакові доданки, то відповідна лінійна комбінація $l_s \in D$ знову належить D $\boxdot$

Теорема (про додатність імовірностей повернення). *Нехай стан $i \in E$ має період $d(i) = d$. Тоді:*

(а) $p_{ii}^{(n)} = 0$ для всіх $n \notin \{td, \ t \in \ \mathbb{N}\}$.

(б) $p_{ii}^{(td)} > 0$ для всіх $t \geq N$, для деякого N.

Доведення. Твердження (а) є очевидним наслідком означення періоду: адже всі елементи $D(i)$ подільні на d.

З леми *про нерівність мінoризації* та з означення множини $D(i)$ виводимо, що ця множина замкнена відносно додавань, тобто з $s, t \in D(i)$

випливає включення $s+t \in D(i)$. Крім того, період $d =$ н.с.д.$D(i)$. Тому (б) є наслідком леми *про ідеал у множині натуральних чисел* $\boxdot$

Теорема (про період класу станів). *Період є характеристикою класу станів, тобто всередині класу всі стани мають однакові періоди.*

Означення. *Незвідний ланцюг Маркова назвемо* **аперіодичним,** *якщо клас станів E має одиничний період.*

Доведення теореми. Якщо $i \longleftrightarrow j$, то знайдуться t, s такі, що $p_{ij}^{(t)} > 0$, $p_{ji}^{(s)} > 0$. За лемою *про нерівність міноризації* та за теоремою *про додатність імовірностей повернення* $p_{jj}^{(s+n+t)} \geq p_{ji}^{(s)} p_{ii}^{(n)} p_{ij}^{(t)} > 0$ для всіх досить великих n, що подільні на період $d(i)$. Тому $s+n+t$ ділиться на $d(j)$ для всіх таких n. Обчисливши різницю двох послідовних номерів, виводимо подільність $d(i)$ на $d(j)$, звідки $d(i) \geq d(j)$. Міняючи i, j місцями, отримуємо рівність $\boxdot$

Теорема (про періодичні підкласи). *Нехай клас станів $C \subset E$ ланцюга Маркова $(\zeta_t, t \geq 0)$ має період $d = d(C) > 1$.*

Тоді існує його розбиття на ***періодичні підкласи****:*

$$C = \bigcup\nolimits_{s=0}^{d-1} C_s, \; C_s \cap C_t = \varnothing, s \neq t,$$

з такими властивостями.

(а) З кожного стану $i \in C_s$ за один крок можливий або вихід із C, або перехід у підклас C_{s+1}:

$$\sum\nolimits_{j \in C \setminus C_{s+1}} p_{ij} = 0,$$

де $C_d \equiv C_0$. Зокрема, для істотного класу C

$$\sum\nolimits_{j \in C_{s+1}} p_{ij} = 1.$$

(б) Якщо $i \in C_s$, $j \in C_t$, то з $p_{ij}^{(nd+r)} > 0$ випливає, що $r - t + s$ ділиться на d. Крім того, $p_{ij}^{(nd+t-s)} > 0$, починаючи з деякого n.

Доведення. Зафіксуємо $k \in C$. Позначимо $D_k(i) = \left\{ n \geq 1 : p_{ki}^{(n)} > 0 \right\}$ для $i \in C$. Оскільки $k \longleftrightarrow i$, то $p_{ik}^{(m)} > 0$ для деякого фіксованого m. З леми *про нерівність міноризації* виводимо, що для довільного $n \in D_k(i)$ сума $n+m \in D_k(k)$ ділиться на d за теоремою *про додатність імовірностей повернення.* Тому всі елементи множини $D_k(i)$ мають однаковий залишок при діленні на d. Позначимо цей залишок через $r(i)$.

Покладемо $C_s = \{i \in C : r(i) = s\}, s = \overline{0, d-1}$. За означенням ці множини попарно несумісні та в об'єднанні утворюють весь клас C.

Для доведення (а) припустимо, що $i \in C_s, j \in C_r$ та $p_{ij} > 0$. Оскільки $r(i) = s$, то знайдеться $n \in D_k(i)$, що має залишок s при діленні на d. Тоді за нерівністю міноризації

$$p_{kj}^{(n+1)} \geq p_{ki}^{(n)} p_{ij} > 0,$$

отже, $n+1 \in D_k(j)$ за означенням. Таким чином, залишок $r(j) = r$ має збігатися з залишком $s+1$, що доводить першу рівність. Друга отримується з першої та з властивості замкненості істотного класу: $\sum_{j \in C} p_{ij} = 1,\ \forall i \in C$.

Нехай виконуються умови (б). Оберемо $m \in D_k(i)$. Тоді m має залишок $r(i) = s$ і за нерівністю міноризації $p_{kj}^{(m+nd+r)} \geq p_{ki}^{(m)} p_{ij}^{(nd+r)} > 0$, отже, $m + nd + r \in D_k(j)$ має залишок $r(j) = t$ при діленні на d. Тому $s + r - t$ ділиться на d. Далі, зафіксуємо m так, щоб $p_{ij}^{(md+t-s)} > 0$. Тоді з нерівності міноризації та за теоремою *про додатність імовірностей повернення* виводимо, що

$$p_{ij}^{(nd+md+t-s)} \geq p_{ii}^{(nd)} p_{ij}^{(md+t-s)} > 0,$$

починаючи з деякого n $\boxdot$

Вправи

(1) Довести, що умову $\zeta_n = i$ під знаком умовної ймовірності у теоремі про розширену марковську властивість, не можна замінити на умову $\zeta_n \in C$.

(2) Ланцюги Маркова з перехідними ймовірностями $P = (p_{ij})$ та $Q = (q_{ij})$ такими, що $p_{ij} = 0 \iff q_{ij} = 0$, мають однакові класи станів.

(3) Нехай C – клас станів, $j \notin C$ та $i \to j$ для деякого $i \in C$. Довести, що $\mathbf{P}_j(\cup_{n \geq 1}\{\zeta_n \in C\}) = 0$.

(4) Навести приклад ланцюга Маркова, всі стани якого – неістотні.

(5) Якщо ланцюг не є скінченним, то істотних класів може не існувати.

(6) Нехай ланцюг скінченний: $|E| = n < \infty$. (а) Якщо $i \to j$, то знайдеться $k \leq n$ таке, що $p_{ij}^{(k)} > 0$. (б) Існує принаймні один істотний клас. (в) Множина E однозначно розбивається у об'єднання декількох істотних класів та множину неістотних станів, можливо, порожню.

(7) Клас C істотний тоді й тільки тоді, коли він є **замкненим класом**, тобто $\sum_{j \in C} p_{ij} = 1$ для всіх $i \in C$.

(8) Звуження ланцюга (ζ_s) на замкнений клас є ланцюгом Маркова.

(9) Якщо стан j є неістотним, то $\mathbf{P}_i(\zeta_n = j) \to 0, n \to \infty$.

(10) Для ланцюга Маркова $(\zeta_n, n \geq 0)$ та обмеженої функції $(g_i, i \in E)$ послідовність $\xi_n = g(\zeta_n) - g(\zeta_0) - \sum_{k=0}^{n-1} h(\zeta_k)$ є мартингалом, де функція $h(i) = \mathbf{E}_i g(\zeta_1) - g(i)$.

2.14.5. Рекурентність

Розглянемо для кожного $j \in E$ момент першого досягнення

$$\tau_j = \inf(n \geq 1 : \zeta_n = j).$$

Якщо відповідна множина n є порожньою, покладемо $\tau_j = \infty$.

Визначимо ймовірності першого досягнення

$$f_{ij}^{(n)} \equiv \mathbf{P}_i(\tau_j = n),\ n \geq 1,\ f_{ij}^{(0)} \equiv 0,$$

та ймовірності досягнення коли-небудь стану j зі стану i :

$$F_{ij} \equiv \sum\nolimits_{n\geq 0} f_{ij}^{(n)} = \mathbf{P}_i(\tau_j < \infty).$$

Остання рівність випливає з зображення $\{\tau_j < \infty\} = \cup_{n\geq 1}\{\tau_j = n\}$.

При $i = j$ визначені ймовірності називають імовірностями першого повернення та ймовірностями повернення відповідно.

Означення. *Стан $i \in E$ є* рекурентним, *якщо $\mathbf{P}_i(\tau_i < \infty) = 1$, або ж $F_{ii} = 1$. У протилежному випадку цей стан називається* транзієнтним.

Теорема (про рівняння першого досягнення). *Для довільних $i, j \in E$ та $n \geq 1$ має місце рівняння першого досягнення*

$$p_{ij}^{(n)} = \sum\nolimits_{s=1}^{n} f_{ij}^{(s)} p_{jj}^{(n-s)},$$

де за означенням $p_{jj}^{(0)} = 1$.

Доведення. Оскільки $\{\zeta_n = j\} \subset \{\tau_j \leq n\}$, то

$$p_{ij}^{(n)} = \mathbf{P}_i(\zeta_n = j) = \mathbf{P}_i(\zeta_n = j, \tau_j \leq n) =$$
$$\sum\nolimits_{s=1}^{n} \mathbf{P}_i(\zeta_n = j, \tau_j = s) = \sum\nolimits_{s=1}^{n} \mathbf{P}_i(\tau_j = s)\mathbf{P}_i(\zeta_n = j \mid \tau_j = s) =$$
$$\sum\nolimits_{s=1}^{n} f_{ij}^{(s)} \mathbf{P}(\zeta_n = j \mid \zeta_0 = i, \zeta_1 \neq j, ..., \zeta_{s-1} \neq j, \zeta_s = j) =$$
$$\sum\nolimits_{s=1}^{n} f_{ij}^{(s)} \mathbf{P}(\zeta_n = j \mid \zeta_s = j) = \sum\nolimits_{s=1}^{n} f_{ij}^{(s)} \mathbf{P}(\zeta_{n-s} = j \mid \zeta_0 = j),$$

де використана теорема *про розширену марковську властивість* та однорідність ланцюга за часом ⊡

Теорема (про критерій рекурентності стану). *Стан $i \in E$ є рекурентним тоді й тільки тоді, коли розбігається ряд з імовірностей повернення:*

$$G_{ii} \equiv \sum\nolimits_{n\geq 0} p_{ii}^{(n)} = \infty.$$

У випадку збіжності цього ряду

$$F_{ii} \equiv \mathbf{P}_i(\tau_i < \infty) = 1 - 1/G_{ii}.$$

Доведення. Визначимо послідовності $u_n = p_{ii}^{(n)},\ f_n = f_{ii}^{(n)}, n \geq 0$, де за означенням $u_0 = 1, f_0 = 0$. За теоремою *про рівняння першого досягнення* ці послідовності задовольняють *рівняння відновлення*

$$u_n = \sum\nolimits_{k=1}^{n} f_k\, u_{n-k},\ n \geq 1.$$

Тому аналогічно доведенню теореми *про генератриси ймовірностей повернення* з розділу про випадкові блукання та її наслідку – теореми *про критерій рекурентності випадкового блукання* – отримуємо шукане твердження $\boxdot$

Теорема (про рекурентні класи станів). *Рекурентність є властивістю класу станів, тобто всередині класу всі стани є рекурентними або транзієнтими одночасно.*

Означення. *Клас станів C називається рекурентним, якщо всі його стани рекурентні.*

Доведення теореми. Якщо $i \longleftrightarrow j$, то знайдуться $t, s \geq 0$ такі, що $p_{ij}^{(t)} > 0, p_{ji}^{(s)} > 0$. За лемою *про нерівність міноризації* $p_{ii}^{(t+n+s)} \geq p_{ij}^{(t)} p_{jj}^{(n)} p_{ji}^{(s)}$. Звідси робимо висновок, що з розбіжності ряду G_{jj} випливає розбіжність ряду G_{ii}. З симетрії i, j виводимо, що рекурентність стану i еквівалентна рекурентності j $\boxdot$

Вправи

(1, теорема Пойа) Блукання Бернуллі на $\mathbb{Z}$ є рекурентним тоді й тільки тоді, коли воно симетричне. Це твердження справедливе для $\mathbb{Z}^2$ та не для $\mathbb{Z}^3$.

(2) Рекурентний клас є істотним. *Вказівка*: $F_{ii} \leq \mathbf{P}_i(\zeta_1 \notin C) < 1$ для деякого $i \in C$, якщо цей клас не є істотним.

(3) Для скінченого незвідного аперіодичного ланцюга Маркова всі елементи матриці перехідних імовірностей за деяку кількість кроків строго додатні.

2.14.6. Імовірності досягнення та відвідувань

Одночасно з *імовірностями досягнення* F_{ij} розглянемо **ймовірності нескінченної кількості відвідувань** стану j зі стану i :

$$Q_{ij} = \mathbf{P}_i \left(\overline{\lim}_{t\to\infty} \{\zeta_t = j\}\right) = \mathbf{P}_i \left(\zeta_t = j \text{ нескінченно часто}\right).$$

Теорема (про ймовірності нескінченного числа відвідувань).
(а) Якщо стан j рекурентний, то $Q_{jj} = 1$ та $Q_{ij} = F_{ij}, \forall\ i \in E$.
(б) Якщо j – транзієнтний стан, то $Q_{ij} = 0, \forall\ i \in E$.

Доведення. Визначимо узагальнені випадкові величини

$$\nu_j^s = \sum\nolimits_{t>s} 1\!\!\mathrm{I}_{\{\zeta_t\ =\ j\}},$$

що дорівнюють кількості відвідань стану j після моменту s. За означенням сигма-алгебри $\mathfrak{F}^{s+1}$ випадкових подій, що породжується значеннями ланцюга не раніше моменту $s+1$, має місце включення $\left\{\nu_j^s \geq n\right\} \in \mathfrak{F}^{s+1}$. Визначимо також імовірності $Q_{ij}^{(n)} = \mathbf{P}_i\left(\nu_j^0 \geq n\right)$. Оскільки

$$\left\{\nu_j^0 \geq n\right\} \downarrow \left\{\nu_j^0 = \infty\right\} \equiv \{\zeta_t = j \text{ нескінченно часто}\} \text{ при } n \to \infty,$$

то за неперервністю ймовірності $Q_{ij} = \lim_{n\to\infty} Q_{ij}^{(n)}$.

При $n \geq 1$ виконується включення $\{\nu_j^0 \geq n\} \subset \{\tau_j < \infty\}$, тому

$$Q_{ij}^{(n)} = \sum\nolimits_{s\geq 1} \mathbf{P}_i\left(\nu_j^0 \geq n,\ \tau_j = s\right) = \sum\nolimits_{s\geq 1} \mathbf{P}_i\left(\nu_j^s \geq n-1,\ \tau_j = s\right) =$$

$$\sum\nolimits_{s\geq 1} \mathbf{P}_i\left(\tau_j = s\right) \mathbf{P}_i\left(\nu_j^s \geq n-1 \mid \tau_j = s\right) =$$

$$\sum\nolimits_{s\geq 1} \mathbf{P}_i\left(\tau_j = s\right) \mathbf{P}\left(\nu_j^s \geq n-1 \mid \zeta_0 = i, \zeta_1 \neq j, ..., \zeta_s = j\right) =$$

$$\sum\nolimits_{s\geq 1} \mathbf{P}_i\left(\tau_j = s\right) \mathbf{P}_j\left(\nu_j^0 \geq n-1 \mid \zeta_0 = j\right) = \sum\nolimits_{s\geq 1} f_{ij}^{(s)} Q_j^{(n-1)} = F_{ij} Q_{jj}^{(n-1)},$$

де враховані означення кількості відвідань ν_j^s та теорема *про розширену марковську властивість.*

За означенням $Q_{jj}^{(1)} = F_{jj}$, тому за індукцією з останньої рівності при $i = j$ отримуємо $Q_{jj}^{(n)} = (F_{jj})^n$, звідки

$$Q_{ij}^{(n)} = F_{ij}\left(F_{jj}\right)^{n-1}, n \geq 1.$$

За означенням *рекурентності* з умови (а) випливає, що $F_{jj} = 1$, а з умови (б), що $F_{jj} < 1$. Тому з $Q_{ij} = \lim_{n\to\infty} Q_{ij}^{(n)}$ у кожному випадку з урахуванням останньої тотожності отримуємо твердження теореми ⊡

Лема (про досяжність рекурентних станів). *Якщо стан j рекурентний та $j \to i$, то $Q_{ij} = F_{ij} = 1$.*

Доведення. Зауважимо, що для кожного $s \geq 0$

$$\{\zeta_t = j \text{ нескінченно часто}\} = \left\{\nu_j^0 = \infty\right\} = \left\{\nu_j^s = \infty\right\}.$$

Тому за теоремою *про ймовірності нескінченного числа відвідувань* та з рекурентності j виводимо такі тотожності:

$$1 = Q_{jj} = \mathbf{P}_j(\nu_j^0 = \infty) = \sum\nolimits_{s\geq 1} \mathbf{P}_j(\nu_j^0 = \infty,\ \tau_i = s) + \mathbf{P}_j(\nu_j^0 = \infty, \tau_i = \infty) \leq$$

$$\sum\nolimits_{s\geq 1} \mathbf{P}_j(\tau_i = s) \mathbf{P}_j(\nu_j^0 = \infty \mid \tau_i = s) + \mathbf{P}_j(\tau_i = \infty) =$$

$$\sum\nolimits_{s\geq 1} f_{ji}^{(s)} \mathbf{P}_j(\nu_j^s = \infty \mid \tau_i = s) + \mathbf{P}_j(\tau_i = \infty) =$$

$$\sum\nolimits_{s\geq 1} f_{ji}^{(s)} \mathbf{P}(\nu_j^s = \infty \mid \zeta_s = i) + 1 - F_{ji} =$$

$$\sum\nolimits_{s\geq 1} f_{ji}^{(s)} Q_{ij} + 1 - F_{ji} = F_{ji} Q_{ij} + 1 - F_{ji},$$

де враховано теорему *про розширену марковську властивість* та означення *ймовірностей досягнення* F_{ji}.

Оскільки $F_{ji} \geq \mathbf{P}_j(\tau_i \leq n) \geq p_{ji}^{(n)} > 0$ для деякого n за умовою досяжності $j \to i$, то з доведеної вище нерівності $1 \leq F_{ji}Q_{ij} + 1 - F_{ji}$ робимо висновок, що $Q_{ij} = 1$, а з нерівності $Q_{ij} \leq F_{ij}$ отримуємо $F_{ij} = 1$ ⊡

Лема (про рівняння першого стрибка). *Для всіх $i, j \in E$ мають місце співвідношення*

$$F_{ij} = p_{ij} + \sum\nolimits_{k \neq j} p_{ik} F_{kj}.$$

Доведення. За означенням та за теоремою *про розширену марковську властивість*

$$F_{ij} = \mathbf{P}_i(\tau_j < \infty) = \mathbf{P}_i(\tau_j < \infty, \zeta_1 = j) + \mathbf{P}_i(\tau_j < \infty, \zeta_1 \neq j) =$$
$$\mathbf{P}_i(\zeta_1 = j) + \sum\nolimits_{k \neq j} \mathbf{P}_i(\tau_j < \infty, \zeta_1 = k) =$$
$$\mathbf{P}_i(\zeta_1 = j) + \sum\nolimits_{k \neq j} \mathbf{P}_i(\zeta_1 = k) \mathbf{P}_i(\tau_j < \infty \mid \zeta_1 = k) = p_{ij} + \sum\nolimits_{k \neq j} p_{ik} F_{kj} \boxdot$$

Теорема (про критерій рекурентності через імовірності досягнення). *Клас C є рекурентним тоді й тільки тоді, коли $F_{ij} = 1$, для всіх $i \neq j,\ i, j \in C$.*

Доведення. *Необхідність* є наслідком леми *про досяжність рекурентних станів*, згідно з якою $F_{ij} = 1 \ \forall i, j \in C$.

Достатність. Клас C є *істотним*, інакше $F_{ij} \leq \mathbf{P}_i(\zeta_1 \in C) < 1$ для деякого $i \in C$. Рекурентність виводиться з леми *про рівняння першого стрибка* та рівності $p_{ik} = 0$ при $k \notin C$, що випливає з істотності C :

$$F_{ij} = p_{ij} + \sum\nolimits_{k \neq j} p_{ik} F_{kj} = p_{ij} + \sum\nolimits_{k \neq j, k \in C} p_{ik} F_{kj} =$$
$$p_{ij} + \sum\nolimits_{k \neq j, k \in C} p_{ik} 1 = p_{ij} + \sum\nolimits_{k \neq j} p_{ik} 1 = 1 \boxdot$$

Означення. *Функція $(v_i, i \in E)$ називається* супергармонічною *для матриці перехідних імовірностей (p_{ij}), якщо при всіх $i \in E$*

$$v_i \geq \sum\nolimits_{j \in E} p_{ij} v_j.$$

Зауважимо, що дану умову завжди задовольняє стала функція $v_i \equiv v$.

Теорема (про алгебраїчний критерій рекурентності). *Незвідний ланцюг Маркова з матрицею перехідних імовірностей (p_{ij}) є рекурентним тоді й тільки тоді, коли кожна невід'ємна супергармонічна функція є сталою.*

Доведення. *Необхідність.* Якщо ланцюг є *транзієнтним*, то за теоремою *про критерій рекурентності через імовірності досягнення* знайдеться пара станів $k \neq j$ з $F_{kj} < 1$. Зафіксуємо стан j і визначимо $v_i = 1\!\!1_{i=j} + 1\!\!1_{i\neq j} F_{ij}$. Тоді ця функція невід'ємна, не стала та є супергармонічною за лемою *про рівняння першого стрибка,* оскільки при $i \neq j$ нерівність у означенні супергармонічної функції є рівністю з леми, а при $i = j$ нерівність випливає зі стохастичності: $1 = \sum_{j\in E} p_{ij} \geq \sum_{j\in E} p_{ij} v_j$.

Достатність. Нехай (v_i) – не стала неві'ємна супергармонічна функція, а клас C рекурентний. Тоді $v_j > 0$ для деякого $j \in C$. Зафіксуємо цей стан та розглянемо функцію $u_i = v_i / v_j, i \in C$. Тоді за означенням

$$u_j = 1, u_i \geq 0, \quad u_i \geq p_{ij} + \sum\nolimits_{k\neq j} p_{ik} u_k, \ \ \forall i \in C.$$

Оскільки матриця (p_{ik}) невід'ємна, то після підстановки замість u_k у праву частину відповідної нерівності (тобто після ітерації) з останньої системи отримаємо правильну нерівність:

$$u_i \geq p_{ij} + \sum\nolimits_{k\neq j} p_{ik} p_{kj} + \sum\nolimits_{k\neq j} \sum\nolimits_{l\neq j} p_{ik} p_{kl} u_l.$$

У результаті $(n+1)$-кратної ітерації з урахуванням невід'ємності $u_k \geq 0$ отримуємо з теореми *про розподіл траєкторій марковського ланцюга* та з означення *ймовірностей першого досягнення*

$$u_i \geq p_{ij} + \sum\nolimits_{k\neq j} p_{ik} p_{kj} + ... + \sum\nolimits_{k_1\neq j} \sum\nolimits_{k_2\neq j} \cdots \sum\nolimits_{k_n\neq j} p_{ik_1} p_{k_1k_2} ... p_{k_{n-1}k_n} p_{k_nj} =$$
$$f_{ij}^{(1)} + f_{ij}^{(2)} + ... + f_{ij}^{(n+1)} = \mathbf{P}_i(\tau_j \leq n+1).$$

Переходячи до границі при $n \to \infty$, звідси маємо

$$u_i \geq \mathbf{P}_i(\tau_j < \infty) = F_{ij} = 1, \ \forall i \in C,$$

за теоремою *про критерій рекурентності через імовірності досягнення,* оскільки клас C – рекурентний за припущенням.

Отже, за означенням функції u_i для всіх станів $i \in C$ виконуються нерівності $v_i \geq v_k > 0$. Переставляючи i, k місцями, виводимо, що $v_i \equiv v_k$, тобто функція v – стала. Отримана суперечність доводить достатність $\boxdot$

Вправи

(1) Всередині не істотного або не рекурентного класу $Q_{ij} = 0$.

(2) Всередині рекурентного класу $Q_{ij} = F_{ij} = 1$.

(3) Довести, що $Q_{ij} = F_{ij}$ для істотного стану j.

(4) Довести, що з другої умови теореми про критерій рекурентності через імовірності досягнення випливає суттєвість класу C.

(5) Для незвідного транзієнтного ланцюга Маркова з перехідною матрицею P довільний стовпчик матриці $\sum_{n\geq 0} P^n$ є супергармонічною функцією.

(6) Задано нескінченну послідовність поліноміальних випробувань Бернуллі з результатами $i = \overline{1,k}$. (а) Знайти ймовірність того, що комбінація i, j сусідніх результатів вперше зустрінеться після n випробувань. (б) Знайти генератрису ймовірностей того, що r-кратне повторення i-го результату зустрінеться першим серед інших. (в) Знайти розподіл найдовшої серії сусідніх результатів вигляду i, що спостерігається у n випробуваннях.

2.14.7. Ергодичність у середньому

Важливою властивістю марковських ланцюгів є ергодична властивість. Вона полягає в тому, що процес "забуває" свій початковий стан і з плином часу переходить у певний стаціонарний режим, тобто його ймовірності переходів за t кроків $p_{ij}^{(t)}$ при великих t перестають залежати як від t, так і від i.

З іншого боку, *ергодичність* зводять до того, що вказані "фінальні" ймовірності збігаються з певними середніми (як у випадку *строго стаціонарних* послідовностей). Тому властивість ергодичності пов'язана з інтегровністю *моментів першого досягнення* $\tau_j \equiv \inf(t \geq 1 : \zeta_t = j)$, тобто зі скінченністю математичних сподівань:

$$m_{ij} \equiv \mathbf{E}_i \tau_j,$$

та з середніми кількостями відвідувань станів:

$$r_{ij}^{(k)} \equiv \sum\nolimits_{t\geq 0} \mathbf{P}_i(\zeta_t = j, \tau_k > t) = \mathbf{E}_i \sum\nolimits_{t=0}^{\tau_k - 1} 1\!\!1_{\{\zeta_t = j\}}.$$

Зауважимо, що при $j = k$ остання сума містить лише перший доданок, оскільки $\tau_k \geq 1$, тому $r_{ik}^{(k)} = \delta_{ik}$.

Лема (про середні моментів досягнення). *Для всіх $i, k \in E$ справедливі тотожності, включаючи випадок розбіжності ряду:*

$$m_{ik} = 1 + \sum\nolimits_{j\neq k} p_{ij} m_{jk}.$$

Доведення. За *формулою повної ймовірності*

$$\mathbf{E}_i \tau_k = \sum\nolimits_j \mathbf{E}_i \tau_k 1\!\!1_{\{\zeta_1 = j\}} = p_{ik} + \sum\nolimits_{j\neq k} \sum\nolimits_{t\geq 1} t\mathbf{P}_i(\tau_k = t, \zeta_1 = j) =$$
$$p_{ik} + \sum\nolimits_{j\neq k} \sum\nolimits_{t\geq 2} t\mathbf{P}_i(\zeta_1 = j, \zeta_2 \neq k, ..., \zeta_t = k) =$$
$$p_{ik} + \sum\nolimits_{j\neq k} \sum\nolimits_{t\geq 2} t\mathbf{P}_i(\zeta_1 = j)\mathbf{P}(\zeta_2 \neq k, ..., \zeta_t = k \mid \zeta_0 = i, \zeta_1 = j) =$$
$$p_{ik} + \sum\nolimits_{j\neq k} \sum\nolimits_{t\geq 2} t p_{ij} \mathbf{P}_j(\tau_k = t - 1) = p_{ik} + \sum\nolimits_{j\neq k} p_{ij} (\mathbf{E}_j \tau_k + 1) =$$

$$p_{ik} + \sum_{j \neq k} p_{ij} + \sum_{j \neq k} p_{ij} m_{jk} = 1 + \sum_{j \neq k} p_{ij} m_{jk},$$

де врахована теорема *про розширену марковську властивість*. Оскільки дані ряди є невід'ємними, то наведені рівності виконуються і у випадку розбіжності за *теоремою Лебега про монотонну збіжність* $\boxdot$

Лема (про середні кількості відвідувань). *Нехай ланцюг (ζ_t) – незвідний та рекурентний, а стан $k \in E$ такий, що $m_{kk} < \infty$. Тоді*

(а) $m_{ik} < \infty$ для всіх i,

(б) $\sum_{j \in E} r_{ij}^{(k)} = m_{ik} < \infty$,

(в) матриці $R^{(k)} \equiv (r_{ij}^{(k)}, i, j \in E)$, $I \equiv (\delta_{ij}, i, j \in E)$, $I_k \equiv (\delta_{kj}, i, j \in E)$ задовольняють тотожність

$$R^{(k)} = I - I_k + R^{(k)} P.$$

Доведення. (а) З леми *про середні моментів досягнення* виводимо для всіх $i \in E$ нерівність

$$m_{ik} = 1 + \sum_{j \neq k} p_{ij} m_{jk} \geq m_{kk}^{-1} \sum_{j \in E} p_{ij} m_{jk},$$

що при $j = k$ випливає з $1 \geq p_{ik}$, а при $j \neq k$ – з нерівності $m_{kk} \geq 1$. Ітеруючи цю нерівність (підстановкою її у праву частину) t разів, виводимо, що при $i = k$

$$m_{kk} = m_{ik} \geq m_{kk}^{-t} \sum_{j \in E} p_{ij}^{(t)} m_{jk}.$$

Оскільки ланцюг незвідний, то $p_{ij}^{(t)} > 0$ для кожного j при деякому t. Тому $m_{jk} < \infty$ для всіх j.

(б) Шукана рівність випливає з тотожностей

$$\sum_{j \in E} r_{ij}^{(k)} = \sum_{t \geq 0} \mathbf{P}_i(\tau_k > t) = \mathbf{E}_i \tau_k = m_{ik}.$$

(в) З (а) виводимо, що ряди у означенні добутку $R^{(k)} P$ збігаються, оскільки рядки матриці $R^{(k)}$ є сумованими послідовностями, а $p_{ij} \leq 1$.

З означення та з теореми *про розширену марковську властивість* виводимо тотожності

$$(R^{(k)} P)_{ij} = \sum_{l \in E} r_{il}^{(k)} p_{lj} = \sum_{l \in E} \Big(\delta_{il} + \sum_{t \geq 1} \mathbf{P}_i(\zeta_t = l, \tau_k > t) \Big) p_{lj} =$$

$$p_{ij} + \sum_{l \neq k} \sum_{t \geq 1} \mathbf{P}_i(\zeta_1 \neq k, \ldots, \zeta_{t-1} \neq k, \zeta_t = l) \mathbf{P}(\zeta_{t+1} = j \mid \zeta_t = l) =$$

$$p_{ij} + \sum_{l \neq k} \sum_{t \geq 1} \mathbf{P}_i(\zeta_1 \neq k, \ldots, \zeta_{t-1} \neq k, \zeta_t = l, \zeta_{t+1} = j) =$$

$$p_{ij} + \sum_{t \geq 1} \mathbf{P}_i(\zeta_1 \neq k, \ldots, \zeta_{t-1} \neq k, \zeta_t \neq k, \zeta_{t+1} = j).$$

Нехай $j \neq k$. Тоді за останньою рівністю

$$(R^{(k)}P)_{ij} = p_{ij} + \sum_{t\geq 1} \mathbf{P}_i(\tau_k \geq t, \zeta_t \neq k, \zeta_{t+1} = j) =$$

$$p_{ij} + \sum_{t\geq 1} \mathbf{P}_i(\tau_k > t+1, \zeta_{t+1} = j) = \sum_{t\geq 1} \mathbf{P}_i(\tau_k > t, \zeta_t = j) =$$

$$(R^{(k)})_{ij} - \delta_{ij} = (R^{(k)} - I + I_k)_{ij}.$$

Нехай $j = k$. Тоді за тією ж рівністю внаслідок рівності $r_{ik}^{(k)} = \delta_{ik}$

$$(R^{(k)}P)_{ik} = p_{ik} + \sum_{t\geq 1} \mathbf{P}_i(\tau_k = t+1) = p_{ik} + \mathbf{P}_i(2 \leq \tau_k < \infty) =$$

$$\mathbf{P}_i(1 \leq \tau_k < \infty) = F_{ik} = 1 = (R^{(k)} - I + I_k)_{ik} \boxdot$$

Для формулювання ергодичної теореми визначимо середні за Чезаро, що дорівнюють середнім частотам відвідувань стану j за час t :

$$\overline{p}_{ij}^{(t)} \equiv \frac{1}{t} \sum_{s=0}^{t-1} p_{ij}^{(s)} = \mathbf{E}_i \left(\frac{1}{t} \sum_{s=0}^{t-1} 1\!\!\mathrm{I}_{\{\zeta_s = j\}} \right), \ \overline{P}^{(t)} = (\overline{p}_{ij}^{(t)}, i, j \in E).$$

Визначимо також Банахів простір сумованих послідовностей

$$l_1(E) = \left\{ \mu = (\mu_j, j \in E) : \|\mu\| \equiv \sum_{j\in E} |\mu_j| < \infty \right\}.$$

Зауважимо, що кожна стохастична матриця Q є стискаючим оператором на $l_1(E)$, оскільки $\|\mu Q\| = \sum_j |\sum_i \mu_i Q_{ij}| \leq \sum_i |\mu_i| \sum_j Q_{ij} = \|\mu\|$. Тут і надалі добуток вектора на матрицю μQ визначимо через $(\mu Q)_j = \sum_i \mu_i Q_{ij}$.

Теорема (ергодична теорема для ланцюга Маркова). *Нехай* (ζ_t) – *незвідний та рекурентний ланцюг Маркова.*

(а) Якщо $m_{kk} < \infty$ *для деякого* $k \in E$, *то існує ймовірнісний розподіл* $\pi = (\pi_j, j \in E)$ *такий, що* $\left\| \alpha \overline{P}^{(t)} - \pi \right\| \to 0, t \to \infty$, *для кожного розподілу* α, *причому* $\pi_k = 1/m_{kk}$ *і розподіл* π *задовольняє лінійну систему рівнянь*

$$\pi = \pi P.$$

(б) Якщо $m_{kk} = \infty$, *то* $\overline{p}_{ik}^{(t)} \to 0, t \to \infty$ *для всіх* i.

Доведення. (а) Визначимо $\pi_j \equiv r_{kj}^{(k)}/m_{kk}$. За лемою *про середні кількості відвідувань,* (б), вектор $\pi \equiv (\pi_j, j \in E)$ є ймовірнісним розподілом. Обчисливши з тотожності (в) цієї леми

$$r_{kj}^{(k)} = (R^{(k)})_{kj} = (I - I_k + R^{(k)}P)_{kj} = \delta_{kj} - \delta_{kj} + \sum_i r_{ki}^{(k)} p_{ij},$$

отримуємо рівняння $\pi = \pi P$. Звідси виводимо також за індукцією $\pi = \pi P^t$ та $\pi = \pi \overline{P}^{(t)}$. Рівність $\pi_k = 1/m_{kk}$ випливає з $r_{ik}^{(k)} = \delta_{ik}$ при $i = k$.

Нехай $\mu = (\mu_j, j \in E)$ – такий дискретний розподіл, що $\mu_j = \pi_j$, починаючи з деякого j. Тоді сума у визначенні добутку $(\mu - \pi)R^{(k)}$ визначена коректно, тому що є скінченною, і є сумованою послідовністю: $\left\|(\mu - \pi)R^{(k)}\right\| < \infty$, оскільки сумованими є всі рядки матриці $R^{(k)}$.

З тотожності (в) леми *про середні кількості відвідувань* виводимо, що

$$\mu\overline{P}^{(t)} - \pi = (\mu - \pi)\overline{P}^{(t)} = (\mu - \pi)(I_k + R^{(k)} - R^{(k)}P)\overline{P}^{(t)} =$$
$$(\mu - \pi)R^{(k)}(I - P)\overline{P}^{(t)} = (\mu - \pi)R^{(k)}(I - P^t)t^{-1},$$

де враховані тотожності $((\mu - \pi)I_k)_j = \sum_i(\mu_i - \pi_i)\delta_{kj} = 0$ та $(I - P)\overline{P}^{(t)} = t^{-1}(I - P)\sum_{s=0}^{t-1} P^s = t^{-1}(I - P^t)$. Оскільки матриця P^t є стискаючим оператором, з доведеної нерівності виводимо, що

$$\left\|\mu\overline{P}^{(t)} - \pi\right\| \le \left\|(\mu - \pi)R^{(k)}\right\| 2t^{-1} \to 0, t \to \infty.$$

Нехай тепер $\alpha = (\alpha_j, j \in E)$ – довільний розподіл, а E_n – множина перших n елементів простору E. Позначимо $\varepsilon_n = \sum_{j \notin E_n} \alpha_j$, $\delta_n = \sum_{j \notin E_n} \pi_j$ та розглянемо розподіли $\mu^n = (\mu_j^n, j \in E)$ з координатами $\mu_j^n = \pi_j$ при $j \notin E_n$ і $\mu_j^n = \alpha_j(1 - \varepsilon_n)/(1 - \delta_n)$ при $j \in E_n$. Оскільки

$$\|\alpha - \mu^n\| \le |1 - (1 - \varepsilon_n)/(1 - \delta_n)| + \varepsilon_n + \delta_n \to 0, \ n \to \infty,$$

то за доведеним вище

$$\overline{\lim}_{t\to\infty} \left\|\alpha\overline{P}^{(t)} - \pi\right\| \le \overline{\lim}_{t\to\infty} \left\|\mu^n\overline{P}^{(t)} - \pi\right\| + \overline{\lim}_{t\to\infty} \left\|(\mu^n - \alpha)\overline{P}^{(t)}\right\| \le$$
$$0 + \|\mu^n - \alpha\| \to 0, n \to \infty.$$

Отже, ліва частина цих нерівностей нульова, що доводить (а).

(б) Визначимо послідовності ймовірностей повернення $p = (p_{kk}^{(t)})$ при $t \ge 0$ і першого повернення $f = (f_{kk}^{(t)})$, $f_{kk}^{(0)} = 0$. Рівняння теореми *про рівняння першого досягнення* при $i = j = k$ має вигляд $p = \delta + f * p$, де $(f * p)_t \equiv \sum_{s=0}^t f_s p_{t-s}$ – *згортка послідовностей* f, p, а $\delta \equiv (\delta_{0t}, t \ge 0)$. Позначимо через 1 послідовність $(1, t \ge 0)$. Тоді за асоціативністю згортки

$$1 = 1 * \delta = 1 * (p - f * p) = 1 * p - 1 * f * p = (1 - 1 * f) * p = g * p.$$

Тут послідовність $g_t = (1 - 1 * f)_t = \sum_{s>t} f_s = \sum_{s>t} f_{kk}^{(s)}$.

Отриману тотожність знову згорнемо з 1 :

$$1 * 1 = 1 * p * g.$$

Зауважимо, що $(1*1)_t = t+1$, а $(1*p)_t = \sum_{s=0}^t p_s = (t+1)\overline{p}_{kk}^{(t+1)}$. Тому з останньої тотожності при $t > n$ отримуємо нерівність

$$t+1 = \sum_{s=0}^t (t-s+1)\overline{p}_{kk}^{(t-s+1)} g_s \ge \sum_{s=0}^n (t-s+1)\overline{p}_{kk}^{(t-s+1)} g_s.$$

Оберемо послідовність $K = \{t_k\} \subset \mathbb{N}$ так, щоб

$$\overline{\lim}_{t\to\infty}\overline{p}_{kk}^{(t+1)} = \lim_{t\to\infty, t\in K}\overline{p}_{kk}^{(t+1)}.$$

Зауважимо, що границя $\lim_{t\to\infty,t\in K}\overline{p}_{kk}^{(t+1-s)}$ не залежить від s при кожному $s \ge 0$, оскільки $t^{-1}\sum_{r=t-s}^t p_{kk}^{(r)} \le t^{-1}(s+1) \to 0, t \to \infty$. Тому після ділення на $t+1$ попередньої нерівності та спрямування $t \to \infty, t \in K$ отримуємо

$$1 \ge \lim_{t\to\infty,t\in K}\sum_{s=0}^n (t+1)^{-1}(t-s+1)\overline{p}_{kk}^{(t-s+1)} g_s =$$

$$\sum_{s=0}^n \lim_{t\to\infty,t\in K}\overline{p}_{kk}^{(t-s+1)} g_s = \sum_{s=0}^n \overline{\lim}_{t\to\infty}\overline{p}_{kk}^{(t+1)} g_s = \overline{\lim}_{t\to\infty}\overline{p}_{kk}^{(t+1)}\sum_{s=0}^n g_s.$$

Оскільки $\sum_{s=0}^n g_s \to \sum_{t=0}^\infty g_t = \sum_{t=0}^\infty \sum_{s>t} f_{kk}^{(s)} = \sum_{s\ge 0} s f_{kk}^{(s)} = \mathbf{E}_k\tau_k = \infty$ за умовою, то з останньої нерівності виводимо, що $\overline{\lim}_{t\to\infty}\overline{p}_{kk}^{(t+1)} = 0$.

Нарешті, з теореми *про рівняння першого досягнення* при $j = k$ згорткою рівняння з послідовністю 1 отримуємо рівність

$$(t+1)\overline{p}_{ik}^{(t+1)} = \sum_{s=0}^t f_{ik}^{(s)}(t-s+1)\overline{p}_{kk}^{(t-s+1)},$$

звідки діленням на $t+1$ та граничним переходом при $t \to \infty$ отримуємо

$$\overline{\lim}_{t\to\infty}\overline{p}_{ik}^{(t+1)} \le \sum_{s=0}^\infty f_{ik}^{(s)}\overline{\lim}_{t\to\infty}\overline{p}_{kk}^{(t-s+1)} = 0 \ \boxdot$$

Означення. *Рекурентний стан $i \in E$ називається* позитивним станом, *якщо $\pi_i \equiv \lim_{t\to\infty}\overline{p}_{ii}^{(t)} > 0$, і* нульовим станом *у протилежному випадку. Граничні значення π_i називаються* ергодичними ймовірностями *ланцюга.*

Теорема (про позитивність класу станів). *Позитивність є характеристикою рекурентного класу станів, тобто всередині класу всі стани одночасно є позитивними, або одночасно нульовими. Рекурентний клас станів C є позитивним тоді й тільки тоді, коли $m_{kk} < \infty$ для деякого $k \in C$, або, що еквівалентно, для всіх $k \in C$.*

Означення. *Незвідний ланцюг називається* позитивним ланцюгом, *якщо його клас станів E є позитивним.*

Доведення теореми. Якщо $i \longleftrightarrow j$, то знайдуться $m, n \ge 0$ такі, що $p_{ij}^{(m)} > 0, p_{ji}^{(n)} > 0$. З леми *про нерівність міноризації* виводимо нерівності

$p_{ii}^{(m+s+n)} \geq p_{ij}^{(m)} p_{jj}^{(s)} p_{ji}^{(n)}$. Звідси підсумовуванням за $s = \overline{0, t-1}$ отримуємо нерівність $(m+t+n)\overline{p}_{ii}^{(m+t+n)} - (m+n)\overline{p}_{ii}^{(m+n)} \geq t p_{ij}^{(m)} \overline{p}_{jj}^{(t)} p_{ji}^{(n)}$. Якщо $\pi_j > 0$, то $\lim_{t\to\infty} \overline{p}_{ii}^{(t)} = \pi_i \geq p_{ij}^{(t)} \pi_j p_{ji}^{(s)} > 0$.

Друге твердження теореми є очевидним наслідком *ергодичної теореми для ланцюга Маркова* $\boxdot$

Лема (про рівняння для ергодичних імовірностей). *Нехай (ζ_t) – незвідний рекурентний ланцюг Маркова. Тоді система рівнянь*

$$u_j = \sum\nolimits_{k\in E} u_k p_{kj}$$

має єдиний з точністю до множника розв'язок (u_k) у класі абсолютно сумованих послідовностей. Він є пропорційним вектору ергодичних імовірностей *(π_j) для позитивного ланцюга:*

$$u_j = \left(\sum\nolimits_{i\in E} u_i\right) \pi_j.$$

та нульовим у протилежному випадку.

Доведення. Підставимо у праву частину вказаної системи рівнянь вираз для u_k, що визначається цією системою:

$$u_j = \sum\nolimits_{k\in E} \left(\sum\nolimits_{l\in E} u_l p_{lk}\right) p_{kj} = \sum\nolimits_{l\in E} u_l \left(\sum\nolimits_{k\in E} p_{lk} p_{kj}\right) = \sum\nolimits_{l\in E} u_l p_{lj}^{(2)},$$

де використано сумованість (u_i) та *рівняння Колмогорова – Чепмена*.

Виконавши попередні перетворення n разів, прийдемо до тотожності

$$u_j = \sum\nolimits_{i\in E} u_i p_{ij}^{(n)}.$$

Після знаходження суми останніх тотожностей за $n = \overline{0, t-1}$ маємо

$$u_j = \sum\nolimits_{i\in E} u_i \overline{p}_{ij}^{(t)}.$$

Звідси за *ергодичною теоремою для ланцюгів Маркова* та за *теоремою Лебега про мажоровану збіжність*

$$u_j = \lim\nolimits_{t\to\infty} \sum\nolimits_{i\in E} u_i \overline{p}_{ij}^{(t)} = \sum\nolimits_{i\in E} u_i \lim\nolimits_{t\to\infty} \overline{p}_{ij}^{(t)} = \left(\sum\nolimits_{i\in E} u_i\right) \pi_j \ \boxdot$$

Теорема (про властивості ергодичних імовірностей). *Нехай (ζ_t) – рекурентний незвідний позитивний ланцюг Маркова.*

(a) Ергодичні ймовірності $(\pi_i, i \in E)$ є дискретним розподілом імовірностей, та єдиним з точністю до множника розв'язком у класі сумованих послідовностей лінійної системи рівнянь

$$\pi_j = \sum\nolimits_{i\in E} \pi_i p_{ij}.$$

(б) Якщо обрати початковий розподіл $\mathbf{P}(\zeta_0 = i) = \pi_i, i \in E$, то послідовність $(\zeta_t, t \geq 0)$ є строго стаціонарною, тобто сумісні розподіли:

$$\mathbf{P}(\zeta_t = i_0, ..., \zeta_{t+n} = i_n) = \pi_{i_0} \prod\nolimits_{s=0}^{n-1} p_{i_s i_{s+1}}, \ \forall t, n \geq 0, \ \forall i_r \in E,$$

не залежать від часового зсуву t.

Доведення

(а) Справедливість системи $\pi = \pi P$ є елементом доведення *ергодичної теореми для ланцюгів Маркова.* Цікаво, що дане твердження є прямим наслідком збіжності перехідних імовірностей.

Позначимо через E_n множину перших n елементів простора E. Тоді за *ергодичною теоремою для ланцюгів Маркова*

$$\sum\nolimits_{j \in E} \pi_j = \lim\nolimits_{n \to \infty} \sum\nolimits_{j \in E_n} \pi_j = \lim\nolimits_{n \to \infty} \lim\nolimits_{t \to \infty} \sum\nolimits_{j \in E_n} \overline{p}_{ij}^{(t)} \leq 1,$$

і послідовність (π_i) є сумованою.

Підсумувавши *рівняння Колмогорова – Чепмена* $p_{ij}^{(s+1)} = \sum_{k \in E} p_{ik}^{(s)} p_{kj}$ по $s = \overline{0, t-1}$, отримуємо нерівність

$$(t+1)\overline{p}_{ij}^{(t+1)} - p_{ij}^{(0)} = t \sum\nolimits_{k \in E} \overline{p}_{ik}^{(t)} p_{kj} \geq t \sum\nolimits_{k \in E_n} \overline{p}_{ik}^{(t)} p_{kj}.$$

Звідси діленням на t за ергодичною теоремою виводимо нерівності

$$\pi_j = \lim\nolimits_{t \to \infty} \overline{p}_{ij}^{(t+1)} \geq \lim\nolimits_{t \to \infty} \sum\nolimits_{k \in E_n} \overline{p}_{ik}^{(t)} p_{kj} = \sum\nolimits_{k \in E_n} \pi_k p_{kj}.$$

Переходячи до границі $n \to \infty$, приходимо до системи нерівностей

$$\pi_j \geq \sum\nolimits_{k \in E} \pi_k p_{kj}.$$

Якщо хоча б одна нерівність тут є строгою, то їх сума теж є строгою нерівністю: $\sum_{j \in E} \pi_j > \sum_{j \in E} \sum_{k \in E} \pi_k p_{kj} = \sum_{k \in E} \pi_k \sum_{j \in E} p_{kj} = \sum_{k \in E} \pi_k$. Оскільки послідовність (π_i) є сумованою, остання нерівність є суперечливою. Отже, всі нерівності у попередній системі є рівностями:

$$\pi_j = \sum\nolimits_{k \in E} \pi_k p_{kj}.$$

Звідси за лемою *про рівняння для ергодичних імовірностей* виводимо рівність $\pi_j = \left(\sum_{i \in E} \pi_i\right) \pi_j$. Тому $\sum_{i \in E} \pi_i = 1$ унаслідок позитивності π_j.

Отже, послідовність (π_i) є сумованою, задовольняює вказану систему та є дискретним розподілом імовірностей. Єдиність розв'язку є наслідком леми *про рівняння для ергодичних імовірностей.*

(б) Як це зроблено у доведенні згаданої леми, з системи рівнянь (а) методом ітерацій виводимо, що $\pi_j = \sum_{i \in E} \pi_i p_{ij}^{(t)}$ для всіх $t \geq 1$. Тому за *формулою повної ймовірності* та означенням *імовірностей переходу за t кроків* $\mathbf{P}(\zeta_t = j) = \sum_{i\in E} \mathbf{P}(\zeta_0 = i)p_{ij}^{(t)} = \pi_j$ для всіх $t \geq 1$. Застосування властивості *однорідності за часом* та теореми *про розподіл траєкторій марковського ланцюга* завершує доведення (б):

$$\mathbf{P}(\zeta_t = i_0, ..., \zeta_{t+n} = i_n) = \mathbf{P}(\zeta_t = i_0)\mathbf{P}(\zeta_{t+1} = i_1, ..., \zeta_{t+n} = i_n \mid \zeta_t = i_0) =$$
$$\pi_{i_0}\mathbf{P}(\zeta_1 = i_1, ..., \zeta_n = i_n \mid \zeta_0 = i_0) \boxdot$$

Теорема (про алгебраїчний критерій позитивності). *Нехай* (ζ_t) – *незвідний ланцюг Маркова. Він є рекурентним і позитивним ланцюгом тоді й тільки тоді, коли система рівнянь*

$$u_j = \sum\nolimits_{k\in E} u_k p_{kj}$$

має ненульовий абсолютно сумований розв'язок.

Доведення. *Необхідність.* Якщо ланцюг є рекурентним та позитивним, то вказана система має ненульовий розв'язок (π_j).

Достатність. Нехай (u_i) – ненульовий сумований розв'язок. Як і в лемі *про рівняння для ергодичних імовірностей,* з даної системи виводимо тотожність $u_j = \sum_{i\in E} u_i p_{ij}^{(n)}$. Тому за теоремою Фубіні

$$\sum\nolimits_i u_i \sum\nolimits_{n\geq 1} p_{ij}^{(n)} = \sum\nolimits_{n\geq 1}\sum\nolimits_i u_i p_{ij}^{(n)} = \sum\nolimits_{n\geq 1} u_j = \pm\infty$$

для деяких j. Оскільки за теоремою *про рівняння першого досягнення*

$$\sum\nolimits_i u_i \sum\nolimits_{n\geq 1} p_{ij}^{(n)} = \sum\nolimits_i u_i \sum\nolimits_{n\geq 1}\sum\nolimits_{s=1}^n f_{ij}^{(s)} p_{jj}^{(n-s)} = \left(\sum\nolimits_i u_i F_{ij}\right)\sum\nolimits_{n\geq 1} p_{jj}^{(n)},$$

де $F_{ij} = \sum_{s\geq 1} f_{ij}^{(s)} \leq 1$, то ряд $\sum_{n\geq 0} p_{jj}^{(n)}$ розбігається при деякому j, і з теореми *про критерій рекурентності стану* випливає рекурентність ланцюга. Якщо цей ланцюг – нульовий, то за лемою *про рівняння для ергодичних імовірностей* розв'язок системи дорівнює $\left(\sum_{i\in E} u_i\right) 0 = 0$ внаслідок теореми *про позитивність класу станів* $\boxdot$

2.14.8. Ланцюг народження та загибелі

Розглянемо ланцюг народження та загибелі на $E = \mathbb{Z}_+$ з перехідними ймовірностями $p_{i,i+1} = p_i, i \geq 0$ та $p_{i,i-1} = q_i, i \geq 1, p_{00} = q_0$, де $p_i + q_i = 1$. Припустимо, що $p_i > 0$ і $q_i > 0$ для всіх $i \geq 0$. Тоді ланцюг є *незвідним.*

Позначимо

$$\delta_k = \prod\nolimits_{i=1}^k \frac{q_i}{p_i}, \quad \theta_k = \prod\nolimits_{i=1}^k \frac{p_{i-1}}{q_i} = \frac{p_0}{p_k \delta_k}.$$

Теорема (критерії рекурентності та позитивності ланцюга народження та загибелі). *Ланцюг народження та загибелі* (ζ_t) *є*

(а) рекурентним тоді й тільки тоді, коли

$$\sum\nolimits_{k\geq 0}\delta_k = \infty,$$

(б) рекурентним і позитивним тоді й тільки тоді, коли

$$\sum\nolimits_{k\geq 0}\theta_k < \infty.$$

У останньому випадку ергодичні ймовірності дорівнюють

$$\pi_i = \theta_i \;/\; \sum\nolimits_{k\geq 0}\theta_k.$$

Доведення. (а) Умова теореми *про алгебраїчний критерій рекурентності* зводиться до сталості супергармонічної невід'ємної функції $(v_i, i \geq 0)$, що є розв'язком системи нерівностей

$$v_i \geq p_i v_{i+1} + q_i v_{i-1},\; i \geq 1, \quad v_0 \geq p_0 v_1 + q_0 v_0.$$

Позначимо через $\Delta_i \equiv v_{i+1} - v_i,\; i \geq 0$, та $\delta_0 \equiv 1$.

З першої нерівності множенням на $p_i + q_i = 1$ виводимо еквівалентну нерівність $\Delta_i \leq (q_i/p_i)\Delta_{i-1},\; i \geq 1$, а з другої – нерівність $\Delta_0 \leq 0$.

Звідси отримуємо $\Delta_i \leq 0, i \geq 0$, і нерівності $\Delta_i \leq \Delta_k\delta_i/\delta_k, i \geq k$.

Якщо функція $(v_i, i \geq 0)$ не є сталою, то $\varepsilon \equiv \Delta_N/\delta_N < 0$ для деякого N. Тоді $\Delta_i \leq \Delta_N\delta_i/\delta_N = \varepsilon\delta_i$ для $i \geq N$. Тому $v_i = v_0 + \sum_{k=0}^{i-1}\Delta_k \to -\infty$ при $i \to \infty$ за умови розбіжності ряду з δ_i. Це суперечить невід'ємності, отже, ланцюг є *рекурентним.*

Припустимо, що ряд з δ_i збігається. Обчислимо рекурентно за заданим $\Delta_0 < 0$ величини $\Delta_i = \delta_i\Delta_0$ з отриманих вище нерівностей, вважаючи їх рівностями. Тоді величини $v_i = v_0 + \Delta_0\sum_{k=0}^{i-1}\delta_k$ утворюватимуть супергармонічну функцію і вибором досить великого $v_0 > 0$ їх можна зробити додатними та різними. Тому ланцюг є *транзієнтним.*

(б) За теоремою *про алгебраїчний критерій позитивності* ця властивість випливає з існування ненульового сумованого розв'язку системи

$$u_i = p_{i-1}u_{i-1} + q_{i+1}u_{i+1},\; i \geq 1, \quad u_0 = q_0 u_0 + q_1 u_{1.}.$$

З першої рівності множенням на $p_i + q_i = 1$ отримуємо рівності вигляду $\Lambda_{i+1} \equiv q_{i+1}u_{i+1} - p_i u_i = \Lambda_i$, а з другої – $\Lambda_0 = 0$. Тому $\Lambda_i = 0$ і $q_{i+1}u_{i+1} = p_i u_i$ для всіх $i \geq 0$. Звідси за індукцією $u_i = \theta_i u_0$. Отже, наявність ненульового сумованого розв'язка еквівалентна збіжності ряду з θ_i $\boxdot$

2.14.9. Ергодична теорема Дебліна

У наступній теоремі, на відміну від теореми про ергодичність у середньому, йдеться про збіжність неусереднених перехідних імовірностей.

Теорема (ергодична теорема Дебліна для аперіодичних ланцюгів). *Нехай (ζ_t) – незвідний аперіодичний рекурентний позитивний ланцюг Маркова. Тоді для всіх $i \in E$ має місце збіжність*

$$\sum\nolimits_{j\in E} \left| p_{ij}^{(t)} - \pi_j \right| \to 0, t \to \infty,$$

де (π_j) – ергодичні ймовірності для ланцюга.

Доведення. Використаємо **метод зклеювання**, що був запропонований В.Дебліном.

Нехай $(\zeta_t', t \geq 0)$ – незалежна копія послідовності (ζ_t) така, що має початкові *ергодичні ймовірності*: $\mathbf{P}(\zeta_0' = k) = \pi_k, k \geq 0$. Тоді за теоремою *про властивості ергодичних імовірностей*, (б), $\mathbf{P}(\zeta_t' = k) = \pi_k, \forall t \geq 0$.

Розглянемо послідовність $\overline{\zeta}_t = (\zeta_t, \zeta_t') \in E^2$. З означення ζ_t' виводимо, що $\overline{\zeta}_t$ є ланцюгом Маркова. Цей ланцюг є незвідним та аперіодичним, оскільки за умовою аперіодичності ланцюгів $(\zeta_t), (\zeta_t')$ та за теоремою *про додатність імовірностей повернення*, (б),

$$\mathbf{P}(\overline{\zeta}_t = (j,l) \mid \overline{\zeta}_0 = (i,k)) = p_{ij}^{(t)} p_{kl}^{(t)} > 0,$$

починаючи з деякого t. Далі, даний ланцюг є *рекурентним* і *позитивним* за теоремою *про алгебраїчний критерій позитивності*, оскільки існує позитивний розв'язок $\pi_j \pi_l$ системи

$$\pi_j \pi_l = \sum\nolimits_{i,k\in E} \pi_i \pi_k p_{ij}^{(1)} p_{kl}^{(1)}.$$

Позначимо $D = \{(i,i), i \in E\} \subset E^2$. Розглянемо *марковський момент*

$$\tau_D = \inf(t \geq 1 : \overline{\zeta}_t \in D).$$

З теореми *про критерій рекурентності через імовірності досягнення* виводимо, що $\tau_D < \infty$ м.н. для довільного початкового розподілу $\overline{\zeta}_0$, оскільки $\{\tau_{(i,i)} < \infty\} \subset \{\tau_D < \infty\}$, а момент досягнення стану (i,i) є скінченним м.н. за теоремою *про критерій рекурентності через імовірності досягнення*.

Розглянемо ймовірність

$$\mathbf{P}(\zeta_t = j, \tau_D \leq t) = \sum\nolimits_{s=1}^{t} \sum\nolimits_{l\in E} \mathbf{P}(\zeta_t = j, \zeta_s = l, \tau_D = s) =$$

$$\sum\nolimits_{s=1}^{t} \sum\nolimits_{l\in E} \mathbf{P}(\zeta_t = j \mid \zeta_s = l) \mathbf{P}(\zeta_s = l, \tau_D = s) =$$

$$\sum_{s=1}^{t}\sum_{l\in E}\mathbf{P}(\zeta'_t=j\mid\zeta'_s=l)\mathbf{P}(\zeta'_s=l,\tau_D=s)=\mathbf{P}(\zeta'_t=j,\tau_D\le t),$$

де враховані теорема *про строго марковську властивість* ланцюга, означення процесу (ζ'_t) та моменту зупинки τ_D, за якими $\zeta_s=\zeta'_s$ при виконанні події $\tau_D=s$.

Звідси виводимо нерівності

$$|\mathbf{P}(\zeta_t=j)-\mathbf{P}(\zeta'_t=j)|=$$

$$\begin{aligned}|\mathbf{P}(\zeta_t &= j,\tau_D\le t)+\mathbf{P}(\zeta_t=j,\tau_D>t)-\\ \mathbf{P}(\zeta'_t &= j,\tau_D\le t)-\mathbf{P}(\zeta'_t=j,\tau_D>t)|=\end{aligned}$$

$$|\mathbf{P}(\zeta_t=j,\tau_D>t)-\mathbf{P}(\zeta'_t=j,\tau_D>t)|\le$$
$$\mathbf{P}(\zeta_t=j,\tau_D>t)+\mathbf{P}(\zeta'_t=j,\tau_D>t).$$

Отже,

$$\sum_{j\in E}|\mathbf{P}(\zeta_t=j)-\mathbf{P}(\zeta'_t=j|\le 2\mathbf{P}(\tau_D>t)\to 0,t\to\infty\ \boxdot$$

Теорема (ергодична теорема для періодичних ланцюгів). *Нехай (ζ_n) – незвідний рекурентний позитивний ланцюг Маркова, що має період $d>1$ та періодичні підкласи $C_s,s=\overline{0,d-1}$ класу станів E.Тоді для всіх $t,r=\overline{0,d-1}$ та всіх $i\in C_t,j\in C_r$ існують граничні ймовірності, що залежать лише від j та r :*

$$\lim_{n\to\infty}p_{ij}^{(nd+r-t)}=\pi_j(r).$$

Доведення. Зауважимо, що за теоремою *про періодичні підкласи* послідовність у лівій частині дорівнює нулю при $j\notin C_r$, а за теоремою *про період класу станів* всі стани мають період d.

Розглянемо ланцюг Маркова $(\zeta_{nd},n\ge 0)$ з *матрицею перехідних імовірностей* P^d. З теореми *про періодичні підкласи* неважко вивести, що цей ланцюг має d *істотних класів станів* $C_s,s=\overline{0,d-1}$, оскільки C_s – *періодичні підкласи* класу E. Звуження цього ланцюга на C_s є незвідним рекурентним аперіодичним ланцюгом Маркова. Це звуження є *позитивним ланцюгом* за теоремою *про позитивність класу станів*, оскільки відповідні моменти повернення m_{ii} для $(\zeta_n,n\ge 0)$ та $(\zeta_{nd},n\ge 0)$ відрізняються лише множником d і є скінченними одночасно.

Тому за *ергодичною теоремою Дебліна для аперіодичних ланцюгів* для всіх s та $i,j\in C_s$ існує $\lim_{n\to\infty}p_{ij}^{(nd)}=\pi_j(s)$, де $(\pi_j(s),j\in C_s)$ є ймовірнісним розподілом. За теоремою *про періодичні підкласи* для всіх $i,j\in E$

$$\lim_{n\to\infty}p_{ij}^{(nd)}=\sum_{s=0}^{d-1}\pi_j(s)1\!\!1_{\ i\in C_s,j\in C_s}.$$

Звідси за *рівняннями Колмогорова – Чепмена* та за *теоремою Лебега про мажоровану збіжність* при $t, r = \overline{0, d-1}$ та $i \in C_t, j \in C_{t+r}$

$$\lim_{n\to\infty} p_{ij}^{(nd+r-t)} = \lim_{n\to\infty} \sum_{k\in E} p_{ik}^{(r-t)} p_{kj}^{(nd)} = \sum_{k\in E} p_{ik}^{(r-t)} \lim_{n\to\infty} p_{kj}^{(nd)} =$$

$$\sum_{k\in E} p_{ik}^{(r-t)} \sum_{s=0}^{d-1} \pi_j(s) 1\!\!1_{k\in C_s, j\in C_s} = \sum_{s=0}^{d-1} \pi_j(s) 1\!\!1_{j\in C_s} \sum_{k\in E} p_{ik}^{(r-t)} 1\!\!1_{k\in C_s} =$$

$$\sum_{s=0}^{d-1} \pi_j(s) 1\!\!1_{j\in C_s} 1\!\!1_{s=r} = \pi_j(r),$$

де за теоремою *про періодичні підкласи* передостання рівність випливає з включення $i \in C_t$, а остання – з $j \in C_r$ $\boxdot$

Вправи

(1) Нехай $\nu_j(n) = \sum_{s=1}^n 1\!\!1_{\{\zeta_s = j\}}$ кількість відвідувань стану j ланцюгом (ζ_n). Довести посилений закон великих чисел: $\nu_j(n)/n \to^{P1} 1/\mathbf{E}_j\tau_j, n \to \infty$.

(2) Довести, що при існуванні стаціонарних імовірностей π_j для незвідного ланцюга Маркова з $\pi_j > 0$ випливає рекурентність стану j.

(3) Матриця перехідних імовірностей ланцюга Маркова двічі стохастична: $\sum_i p_{ij} = 1, \forall j \in E$. Тоді ергодичний розподіл $\pi_j = 1/|E|,\ j \in E$.

(4) Матриця перехідних імовірностей ланцюга Маркова на $\mathbb{Z}_+$ має вигляд: $p_{i0} = q_i, i \geq 0$, та $p_{ij} = (1 - q_i) 1\!\!1_{j=i+1}, i \geq 0, j \geq 1$. Довести, що ланцюг рекурентний тоді й тільки тоді, коли $\sum_{i\geq 0} q_i = \infty$. Знайти умови ергодичності ланцюга, а також відповідні ергодичні ймовірності.

(5) Матриця перехідних імовірностей ланцюга Маркова на $\mathbb{Z}_+$ має вигляд: (а) $p_{0j} = p_j$, $p_{ij} = p_{j-i+1}, i \geq 1, j \geq 0$, (б) $p_{0j} = p_j, j \geq 0$, та $p_{ij} = 1\!\!1_{j=i-1}$, при $i \geq 1, j \geq 0$. Тут $(p_j, j \geq 0)$ – деякий дискретний розподіл. У припущенні додатності $p_j > 0$ встановити необхідні та достатні умови рекурентності, позитивності. Методом генератрис обчислити ергодичний розподіл.

(6) Величини $(\xi_k, k \geq 1)$ незалежні, однаково розподілені, $\mathbf{P}(\xi_1 = j) = f_j$, $j \in \mathbb{N}$, а $S_k = \sum_{i=1}^k \xi_i$. Довести, що послідовність $\zeta_t = \inf(S_k - t,\ S_k \geq t)$ є ланцюгом Маркова на $\mathbb{Z}_+$ (який називається процесом відновлення) та знайти його матрицю перехідних імовірностей. Довести, що цей ланцюг є (а) аперіодичним тоді й тільки тоді, коли *н.с.д.*$(j \geq 1 : f_j > 0) = 1$, (б) ергодичним тоді й тільки тоді, коли $m = \sum j f_j < \infty$, а ергодичний розподіл дорівнює $\pi_j = m^{-1} \sum_{k=j}^\infty f_k$.

(7) Нехай $(\zeta_t, t \geq 0)$ – рекурентний ланцюг Маркова з множиною значень E, а стан $k \in E$ – аперіодичний. Довести, що $\lim_{t\to\infty} p_{ik}^{(t)} = \mathbf{P}_i(\tau_k < \infty)/\mathbf{E}_k\tau_k$.

(8) Нехай ланцюг народження і загибелі $(\zeta_t, t \geq 0)$ є ергодичним, причому початковий розподіл $\mathbf{P}(\zeta_0 = i) = \pi_i$. Довести, що послідовність $\xi_s = \zeta_{t-s}$, $0 \leq s \leq t$, також є ланцюгом Маркова з початковим розподілом π та матрицею перехідних імовірностей за один крок $p_{ij} = \pi_j \mathbf{P}(\zeta_1 = i \mid \zeta_0 = j)/\pi_i$.

(9) Нехай $(\zeta_t, t \geq 0)$ – ергодичний аперіодичний ланцюг Маркова з ергодичним розподілом (π_j), τ – момент зупинки відносно $\sigma[\zeta_s, s \leq t]$, причому $\zeta_\tau = i$ м.н. Довести, що $\mathbf{E}_i \sum_{t=0}^{\tau-1} 1\!\!\mathrm{I}_{\{\zeta_t = j\}} = \pi_j \mathbf{E}_i \tau$.

(10) Знайти необхідні та достатні умови рекурентності та ергодичності ланцюга народження та загибелі, для якого $q_k - 1/2 \sim ck^{-\alpha}$, $k \to \infty$,з $\alpha > 0$.

(11) Ланцюг Маркова $(\zeta_t, t \geq 0)$ з множиною значень E має перехідні імовірності p_{ik}. Тоді послідовність $\xi_t = (\zeta_t, \zeta_{t+1}), t \geq 0$, також є ланцюгом Маркова зі значеннями у E^2 та перехідними ймовірностями $p_{(ij)(kl)} = p_{ik} p_{jl} 1\!\!\mathrm{I}_{j=k}$, $i, j, k, l \in E$. За умови ергодичності (ζ_t) знайти ергодичні ймовірності для (ξ_t).

(12) Нехай $(\zeta_t, t \geq 0)$ – незвідний аперіодичний ланцюг Маркова зі скінченною множиною значень E. Довести, що для моменту зклеювання τ_D з ергодичної теореми для аперіодичних ланцюгів справедлива нерівність $\mathbf{P}(\tau_D > t) \leq a\rho^t$ для деякого $\rho < 1$, і швидкість збіжності у цій теоремі є геометричною.

2.14.10. Ергодична теорема Колмогорова

А.М.Колмогорову належить конструктивний варіант ергодичної теореми, у якому явно оцінюється швидкість збіжності.

Теорема (ергодична теорема Колмогорова для ланцюгів Маркова). *Нехай ланцюг Маркова $(\zeta_t,\ t \geq 0)$ задовольняє умову Колмогорова*

$$\exists \varepsilon > 0,\ \exists r \geq 1,\ \exists l \in E : p_{il}^{(r)} \geq \varepsilon,\ \forall i \in E.$$

Тоді для всіх $i, j \in E$ існує та не залежить від i границя

$$\lim_{t\to\infty}\ p_{ij}^{(t)} = \pi_j,$$

причому $(\pi_j,\ j \in E)$ є дискретним розподілом імовірностей на E.

Цей розподіл називається ***ергодичним розподілом*** *марковського ланцюга і є розв'язком лінійної системи рівнянь*

$$\pi_j = \sum_{i\in E} \pi_i p_{ij},\ \forall j \in E.$$

Доведення. Позначимо

$$m_j^{(s)} \equiv \min_{i\in E} p_{ij}^{(s)} \leq \max_{i\in E} p_{ij}^{(s)} \equiv M_j^{(s)}.$$

З *рівнянь Колмогорова – Чепмена* при $t = 1$ виводимо, що

$$m_j^{(s)} = \min_{i\in E} p_{ij}^{(s)} = \min_{i\in E} \sum_k p_{ik} p_{kj}^{(s-1)} \geq \min_{i\in E} \sum_k p_{ik} m_j^{(s-1)} = m_j^{(s-1)}.$$

Аналогічно отримуємо монотонність $M_j^{(s)} \downarrow$ за аргументом s.

Нехай $q_{ij} = p_{ij}^{(r)} - \varepsilon\delta_{jl}$. Тоді $q_{ij} \geq 0$, $\sum_{j\in E} q_{ij} = 1-\varepsilon$ за умовою. Запишемо різницю *рівнянь Колмогорова – Чепмена* в точках i, k та скористаємося тим, що *матриця перехідних імовірностей* є *стохастичною матрицею*:

$$p_{ij}^{(t)} - p_{kj}^{(t)} = \sum_{h\in E} \left(p_{ih}^{(r)} - p_{kh}^{(r)}\right)\, p_{hj}^{(t-r)} =$$

$$\sum_{h\in E} \left(p_{ih}^{(r)} - p_{kh}^{(r)}\right)\, \left(p_{hj}^{(t-r)} - m_j^{(t-r)}\right) =$$

$$\sum_{h\in E} \left(p_{ih}^{(r)} - \varepsilon\delta_{hl} - p_{kh}^{(r)} + \varepsilon\delta_{hl}\right)\, \left(p_{hj}^{(t-r)} - m_j^{(t-r)}\right) =$$

$$\sum_{h\in E} (q_{ih} - q_{kh})\, \left(p_{hj}^{(t-r)} - m_j^{(t-r)}\right) \leq \sum_{h\in E} q_{ih}\, \left(p_{hj}^{(t-r)} - m_j^{(t-r)}\right) \leq$$

$$\sum_{h\in E} q_{ih}\, \left(M_j^{(t-r)} - m_j^{(t-r)}\right) = (1-\varepsilon)\left(M_j^{(t-r)} - m_j^{(t-r)}\right).$$

Перейдемо в цій нерівності до верхньої межі за i, k, та знайдемо верхню оцінку

$$M_j^{(t)} - m_j^{(t)} \leq (1-\varepsilon)\left(M_j^{(t-r)} - m_j^{(t-r)}\right).$$

Аналогічно виводиться оцінка знизу

$$M_j^{(t)} - m_j^{(t)} \geq -(1-\varepsilon)\left(M_j^{(t-r)} - m_j^{(t-r)}\right).$$

З цих нерівностей за індукцією дістанемо

$$\left|M_j^{(t)} - m_j^{(t)}\right| \leq (1-\varepsilon)^{[t\,/\,r]} \to 0,\ t \to \infty.$$

Позначимо

$$\pi_j = \lim_{t\to\infty} M_j^{(t)} = \lim_{t\to\infty} m_j^{(t)}.$$

Границі тут існують через монотонність послідовностей $M_j^{(t)}$, $m_j^{(t)}$, а рівність є наслідком попередньої нерівності. З цієї ж монотонності отримуємо $\pi_j \in [m_j^{(t)}, M_j^{(t)}]$, тому з означення $m_j^{(t)} \leq p_{ij}^{(t)} \leq M_j^{(t)}$ виводимо, що

$$\left|p_{ij}^{(t)} - \pi_j\right| \leq \left|M_j^{(t)} - m_j^{(t)}\right| \leq (1-\varepsilon)^{[t\,/\,r]},$$

звідки $\lim_{t\to\infty} p_{ij}^{(t)} = \pi_j$.

Оскільки $\sum_{j\leq N} \pi_j = \lim_{t\to\infty} \sum_{j\leq N} p_{ij}^{(t)} \leq 1$ для всіх N, то $\sum_{j\in E} \pi_j \leq 1$.

Перейдемо до границі при $t \to \infty$ в нерівності

$$p_{ik}^{(t+1)} = \sum_{j\in E} p_{ij}^{(t)}\, p_{jk} \geq \sum_{j\leq N} p_{ij}^{(t)}\, p_{jk},$$

та оберемо $N \to \infty$. Отримаємо нерівності $\pi_k \geq \sum_{j\in E} \pi_j\ p_{jk}$. Якби хоча б одна з цих нерівностей є строгою, то їх сума – теж строга нерівність:

$$\sum\nolimits_{k\in E} \pi_k > \sum\nolimits_{k\in E} \sum\nolimits_{j\in E} \pi_j\ p_{jk} = \sum\nolimits_{j\in E} \pi_j \sum\nolimits_{k\in E}\ p_{jk} = \sum\nolimits_{j\in E} \pi_j.$$

Звідси від супротивного дістанемо наведену систему рівнянь для π_j $\boxdot$

Вправи

(1) З умов ергодичної теореми Колмогорова випливає аперіодичність ланцюга.

(2) Довести, що (а) для скінченного простору значень E з властивості ергодичності: $\exists \lim_{t\to\infty}\ p_{ij}^{(t)} = \pi_j > 0, \forall i, j \in E$, випливає умова ергодичної теореми Колмогорова, (б) при $|E| = \infty$ навести приклад ергодичного ланцюга, для якого вказана умова не виконується.

(3) Ланцюги Маркова X та Y мають скінченний простір значень E, перехідні матриці P і Q, та вектори ергодичних імовірностей π, ν. (а) Довести стійкість стаціонарних імовірностей: $\|\nu - \pi\| = O(\|Q - P\|),\ \|Q - P\| \to 0$. (б) Знайти матрицю R таку, що справедливе скориговане граничне зображення: $\|\nu - \pi - \pi R(Q - P)\| = o(\|Q - P\|),\ \|Q - P\| \to 0$. Тут норми векторів та матриць визначаються як $\|\mu\| = \sum_{k\in E} |\mu_k|\,,\ \|M\| = \sup_{i\in E} \sum_{j\in E} |M_{ij}|\,.$

2.15. Процес Пуассона

Вище ми розглядали лише послідовності випадкових величин, що є функціями дискретних моментів часу. Процес Пуассона є моделлю для випадкових потоків подій, які відбуваються в неперервному часі. Наприклад, для потоку викликів, що надходять до телефонної станції чи мережевого сервера.

Означення. Випадковим процесом *на* $\mathbb{R}$ *називається сім'я випадкових величин* $(\xi(t) = \xi(t, \omega),\ t \in \mathbb{R})$, *що залежать від дійсного параметра* t. *Для кожної елементарної події* ω *дійсна функція* $(\xi(t, \omega),\ t \in \mathbb{R})$ *називається* траєкторією *процесу.*

Як і завжди, аргумент ω у позначенні процесу часто не вказується.

Звичайно опис випадкового процесу починається з задання класу його можливих траєкторій.

Означення. Лічильним процесом, *або ж* точковим процесом *на* $\mathbb{R}_+$ *називається випадковий процес* $(\nu(t),\ t \in \mathbb{R}_+)$, *такий, що:*

(а) $\nu(0) = 0,\ \nu(t) \in \mathbb{Z}_+$,

(б) має кусково-сталі і неперервні зліва за t *траєкторії при всіх* ω,

(в) має одиничні прирости: $\nu(t + 0) - \nu(t) \in \{0, 1\},\ \forall t, \omega$.

Означення. Стохастичним потоком *подій на* $\mathbb{R}_+$ *називається довільна зростаюча необмежена послідовність додатних випадкових величин:* $0 < \tau_1 < \tau_2 < ... < \tau_n < ...,\ \lim \tau_n = \infty$ *м.н.*

Теорема (про відповідність між потоком та лічильним процесом). *Існує взаємно-однозначна відповідність між стохастичними потоками* $(\tau_n,\ n \ge 1)$ *та лічильними процесами* $\nu(t)$ *на* $\mathbb{R}_+$. *Ця відповідність задається формулами*

$$\nu(t) = \sum\nolimits_{n\ge 1} 1\!\!\mathrm{I}_{\{\tau_n < t\}} = |\{n \ge 1 : \tau_n < t\}|\,,$$

$$\tau_n = \inf(t > 0 : \nu(t) = n).$$

Доведення для скінчених послідовностей очевидне, для зліченних отримується граничним переходом зі скінчених $\boxdot$

Означення. *Нехай* $\xi(t)$ *– випадковий процес. Його* приростом на інтервалі $[s, t)$ *називається випадкова величина*

$$\xi[s, t) \equiv \xi(t) - \xi(s).$$

Означення. *Випадковий процес* $\xi(t)$ *має* незалежні прирости, *якщо для довільних* $0 \le t_1 < t_2 < ... < t_n$ *випадкові величини*

$$\xi[t_1, t_2),\ \xi[t_2, t_3), ...,\ \xi[t_{n-1}, t_n)$$

незалежні в сукупності.

Означення. *Випадковий процес* $\xi(t)$ *має* однорідні прирости, *якщо функція розподілу приросту* $\xi[t, t+h)$ *залежить лише від* h *і не залежить від аргумента* $t \in \mathbb{R}_+$.

Зауваження. Сам приріст $\xi[t, t+h)$, звичайно, може залежати від t. Так, наприклад, прирости на одиничних інтервалах послідовності сум незалежних *однаково розподілених* величин є незалежними та однорідними, однак різними – адже вони збігаються з відповідним доданками.

2.15.1. Процес Пуассона та його розподіл

Означення. *Випадковий процес* $\nu(t),\ t \in \mathbb{R}_+$, *називається* процесом Пуассона, *якщо цей процес:*

(а) є лічильним процесом,

(б) має незалежні прирости та однорідні прирости.

Теорема (про розподіл процесу Пуассона). *Нехай* $\nu(t)$ *– процес Пуассона, який не є тотожним нулем. Існує стала* $0 < \lambda < \infty$, *така, що* *для всіх* $t \ge 0$, $n \ge 0$ *виконуються рівності*

$$\mathbf{P}(\nu(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t},$$

$$\mathbf{E}\nu(t) = \lambda t,\ \mathbf{D}\nu(t) = \lambda t,\ \lambda = \lim_{h \to 0} h^{-1} \mathbf{P}(\nu(h) = 1).$$

Означення. *Параметр* λ *називається* інтенсивністю *процесу.*

Зауваження. Твердження теореми лишаються справедливими, якщо замість умов (б),(в) на траєкторію в означенні *лічильного процесу* виконується така аналітична умова ординарності:

$$\text{(о)} \quad \mathbf{P}(\nu(h) > 1) = o(h),\ h \to 0.$$

Доведення. Позначимо

$$p_n(t) = \mathbf{P}(\nu(t) = n).$$

Нехай $t, s > 0$. Оскільки величини $\nu(t) = \nu[0, t)$ та $\nu[t, t+s)$ незалежні, то виконуються рівності

$$p_0(t+s) = \mathbf{P}(\nu(t+s) = 0) =$$

$$\mathbf{P}(\nu(t) = 0,\ \nu[t, t+s) = 0) = \mathbf{P}(\nu(t) = 0)\ \mathbf{P}(\nu[t, t+s) = 0) =$$

$$\mathbf{P}(\nu(t) = 0)\mathbf{P}(\nu[0, 0+s) = 0) = \mathbf{P}(\nu(t) = 0)\mathbf{P}(\nu(s) = 0) = p_0(t)p_0(s),$$

де використані також *однорідність приростів*, та рівність $\nu(0) = 0$ з властивості (а) означення *лічильного процесу*.

Отже, функція $p_0(t)$ мультиплікативна. Звідси виводимо, що

$$p_0(1/n) = (p_0(1))^{1/n},\quad p_0(m/n) = (p_0(1))^{m/n}.$$

Оскільки процес не вироджений, то $p_0(m) < 1$ для деякого m, а тому $p_0(1) < 1$. Якщо $p_0(1) = 0$, то $p_0(t) = 0$ при всіх t, і $\mathbf{P}(\nu(h) \ge 1) = 1$ для $h > 0$. У цьому випадку з незалежності та однорідності приростів виводимо, що $\mathbf{P}(\nu(1) \ge n) = 1$ для всіх n. Це суперечить скінченності процесу.

Отже, стала $\lambda \equiv -\ln p_0(1)$ скінченна і додатна. Оскільки $p_0(r) = \exp(-\lambda r)$ для всіх раціональних r, а функція $p_0(t)$ не зростає і $p_0(0) = 1$, то ця функція неперервна в нулі. Тоді з мультиплікативності $p_0(t+h) = p_0(t)p_0(h)$ випливає неперервність справа $p_0(t)$. Переходячи до границі $r \downarrow t$, з $p_0(r) = \exp(-\lambda r)$ дістанемо $p_0(t) = \exp(-\lambda t)$ для всіх $t \ge 0$.

Доведемо, що з умов (б),(в) в означенні *лічильного процесу* випливає умова *ординарності* (о) зауваження. За доведеним вище $\mathbf{P}(\nu(h) > 0) = 1 - p_0(h) = 1 - \exp(-\lambda h) \le \lambda h$. Оскільки $\nu(h) = \sum_{k=0}^{2^n-1} \nu_k(h)$, де доданки

$\nu_k(h) \equiv \nu[kh2^{-n}, (k+1)h2^{-n})$ *незалежні в сукупності, однаково розподілені* і за умовою (в) не перевищують 1, починаючи з деякого номера n, то при виконанні умови $\nu(h) > 1$ принаймні два з них будуть додатними:

$$\mathbf{P}(\nu(h) > 1) \leq \lim_{n\to\infty} \mathbf{P}(\cup_{i\,<\,j}\{\nu_i(h) > 0,\ \nu_j(h) > 0\}) \leq$$
$$\lim_{n\to\infty} 2^{n-1}(2^n - 1)\left(\mathbf{P}(\nu_0(h2^{-n}) > 0)\right)^2 \leq \lambda^2 h^2 = o(h),\ h \to 0.$$

Отже, умова (о) виконана.

Враховуючи умову (а) в означенні лічильного процесу $\nu(t)$, розглянемо при $n \geq 1$ таку тотожність:

$$\{\nu(t+h) = n\} = \{\nu(t) = n,\ \nu[t, t+h) = 0\} \bigcup$$
$$\{\nu(t) = n-1,\ \nu[t, t+h) = 1\} \bigcup \{\nu(t+h) = n,\ \nu[t, t+h) > 1\}.$$

Оскільки три події в правій частині *попарно несумісні*, з незалежності та *однорідності приростів* отримуємо

$$p_n(t+h) = p_n(t)p_0(h) + p_{n-1}(t)p_1(h) + O(p(\nu[t, t+h) > 1)) =$$
$$p_n(t)p_0(h) + p_{n-1}(t)(1 - p_0(h) - o(h)) + o(h),\ h \to 0,$$

де двічі використана умова (о) *ординарності*. Після підстановки в останнє рівняння $p_0(h) = \exp(-\lambda h) = 1 - \lambda h + o(h)$ маємо

$$(p_n(t+h) - p_n(t))/h = -\lambda\, p_n(t) + \lambda\, p_{n-1}(t) + o(1),\ h \to 0,$$

причому $o(1)$ у правій частині є рівномірним за t. Отже, функції p_n диференційовні справа, неперервні та диференційовні внаслідок відзначеної рівномірності, і задовольняють такі диференціальні рівняння:

$$p_n'(t) = -\lambda\, p_n(t) + \lambda\, p_{n-1}(t),\ p_n(0) = 0,\ \forall n \geq 1.$$

З використанням інтегруючого множника $\exp(\lambda t)$ звідси виводимо для функцій $q_n(t) \equiv \exp(\lambda t) p_n(t)$ систему рівнянь

$$q_n'(t) = \lambda\, q_{n-1}(t),\ q_n(0) = 0,\ n \geq 1,$$

звідки з урахуванням доведеної вище рівності $q_0(t) = 1$ рекурентно обчислюємо $q_n(t) = (\lambda t)^n/n!$, $\ p_n(t) = (\lambda t)^n\, e^{-\lambda t}\, /n!$.

Останні рівності теореми є наслідками першої $\boxdot$

Наслідок (про інтенсивності переходу процесу Пуассона). *Перехідні ймовірності процесу Пуассона з інтенсивністю λ задовольняють такі інфінітезимальні зображення при $n \geq 0$:*

$$\mathbf{P}(\nu(t+h) = n+1 \mid \nu(t) = n) = \lambda h + o(h),\ h \to 0,$$
$$\mathbf{P}(\nu(t+h) > n+1 \mid \nu(t) = n) = o(h),\ h \to 0,$$
$$\mathbf{P}(\nu(t+h) = n \mid \nu(t) = n) = 1 - \lambda h + o(h),\ h \to 0.$$

Зауваження. З даних умов також випливає твердження теореми *про розподіл процесу Пуассона* (**вправа**), тому іноді саме їх використовують як означення цього процесу.

Доведення. Оскільки події $\{\nu(t+h) - n \in B\}$ за умови $\{\nu(t) = n\}$ збігаються з $\{\nu[t, t+h) \in B\}$ та за властивістю *незалежності приростів* не залежать від цієї умови, то з урахуванням *однорідності приростів* робимо висновок, що ймовірності у лівих частинах дорівнюють відповідно $\mathbf{P}(\nu(h) = 1)$, $\mathbf{P}(\nu(h) = 0)$, $\mathbf{P}(\nu(h) > 1)$. Тому наведені зображення отримуємо з формули для розподілу процесу, що доведена в теоремі.

Твердження зауваження фактично обґрунтовані у останній частині доведення теореми *про розподіл процесу Пуассона,* оскільки з припущень зауваження за формулою повної ймовірності отримуємо

$$p_0(t+h) = p_0(t)(1 - \lambda h + o(h)),\ h \to 0,$$
$$p_n(t+h) = p_n(t)(1 - \lambda h + o(h)) + p_{n-1}(t)(\lambda h + o(h)) + o(h),\ h \to 0,$$

звідки обчислюються $p_n(t)$ $\boxdot$

2.15.2. Траєкторії процесу Пуассона

Теорема (про властивості траєкторій процесу Пуассона). *Нехай* $(\tau_n,\ n \ge 1)$ *– стохастичний потік, що пов'язаний з процесом Пуассона* $\nu(t)$, *і* $\theta_n \equiv \tau_n - \tau_{n-1}$, $n \ge 1$ *– інтервали між стрибками процесу, де* $\tau_0 = 0$. *Тоді випадкові величини* $(\theta_n,\ n \ge 1)$ *незалежні в сукупності і мають однаковий показниковий розподіл із параметром* λ, *який дорівнює інтенсивності процесу.*

Зауваження. *Лічильний процес*, що утворений *стохастичним потоком* сум *незалежних* однаково *показниково розподілених* випадкових величин, має *незалежні прирости* та *однорідні прирости*, тобто є *процесом Пуассона.* Це твердження (**Вправа**) випливає з властивості *відсутності післядії* для показникового розподілу.

Доведення. Для випадкової величини $\theta_1 = \tau_1$ за теоремою *про розподіл процесу Пуассона* маємо

$$\mathbf{P}(\theta_1 \ge t) = \mathbf{P}(\tau_1 \ge t) = \mathbf{P}(\nu(t) = 0) = \exp(-\lambda t),$$

отже, *розподіл* θ_1 – показниковий із параметром λ.

Для доведення твердження при $n > 1$ визначимо сумісні функції розподілу

$$H_n(x_1, ..., x_n) = \mathbf{P}(\tau_1 < x_1, ..., \tau_n < x_n)$$

та множину $D_n = \{x = (x_1, ..., x_n) : x_1 \le x_2 \le ... \le x_n\}$.

Позначимо через $\Gamma_k(D)$ – клас функцій обмеженої варіації на $D \subset \mathbb{R}^n$ таких, що вони не залежать від аргументу $x_k, k = \overline{1, n}$.

Припустимо, що $x \notin D_n$, тобто $x_k > x_{k+1}$ для деякого $k = \overline{1, n}$. Тоді $\{\tau_k < x_k, \tau_{k+1} < x_{k+1}\} = \{\tau_{k+1} < x_{k+1}\}$, тому $H_n \in \Gamma_k(\mathbb{R}^n \backslash D_n)$.

Нехай тепер $x \in D_n$. За означенням

$$H_n(x_1, ..., x_n) = \mathbf{P}(\nu(x_1) \ge 1, ..., \nu(x_n) \ge n) =$$
$$\textstyle\sum_{k=1}^{n-1} \mathbf{P}(\nu(x_1) = k)\ \mathbf{P}(\nu[x_1, x_{k+1}) \ge 1, ..., \nu[x_1, x_n) \ge n - k) +$$
$$\mathbf{P}(\nu(x_1) \ge n) = \mathbf{P}(\nu(x_1) = 1) H_{n-1}(x_2 - x_1, ..., x_n - x_1) +$$
$$\textstyle\sum_{k=2}^{n-1} \mathbf{P}(\nu(x_1) = k) H_{n-k}(x_{k+1} - x_1, ..., x_n - x_1) + \mathbf{P}(\nu(x_1) \ge n) =$$
$$\lambda x_1 \exp(-\lambda x_1) H_{n-1}(x_2 - x_1, ..., x_n - x_1) + R_n(x),$$

де враховано незалежність та однорідність приростів процесу, а функція $R_n \in \Gamma_2(D_n)$.

З останньої тотожності за індукцією виводимо, що

$$H_n(x_1, ..., x_n) = -\lambda^{n-1} x_1 (x_2 - x_1) ... (x_{n-1} - x_{n-2}) \exp(-\lambda x_n) + S_n(x),$$

де функція S_n є сумою функцій із класів $\Gamma_k(D_n)$. Звідси послідовним диференціюванням за аргументами $x_n, x_{n-1}, ... x_1$ знаходимо *сумсну щільність* вектора $(\tau_1, ..., \tau_n)$:

$$f_{\tau_1, ..., \tau_n}(x_1, ..., x_n) = \lambda^n \exp(-\lambda x_n) 1\!\!\mathrm{I}_{0 \le x_1 \le x_2 \le ... \le x_n}.$$

Тому *сумісна характеристична функція* інтервалів $\theta_k = \tau_k - \tau_{k-1}$ між стрибками дорівнює

$$\mathbf{E} \exp\left(i \textstyle\sum_{k=1}^n t_k \theta_k\right) = \mathbf{E} \exp\left(i \textstyle\sum_{k=1}^n (t_k - t_{k-1}) \tau_k\right) =$$
$$\textstyle\int_{0 \le x_1 \le x_2 \le ... \le x_n} \lambda^n \exp\left(i \sum_{k=1}^n (t_k - t_{k-1}) x_k - \lambda x_n\right) dx_1 ... dx_n =$$
$$\textstyle\int_{0 \le y_1, ... 0 \le y_n} \lambda^n \exp\left(i \sum_{k=1}^n t_k y_k - \lambda \sum_{k=1}^n y_k\right) dy_1 ... dy_n = \prod_{k=1}^n \lambda / (\lambda - i t_k),$$

де використана заміна змінних $x_k = \sum_{i=1}^k y_i$.

Вираз у правій частині є сумісною характеристичною функцією незалежних *однаково розподілених* показникових величин із параметром λ. Тому шукане твердження випливає з теореми *про однозначну відповідність між функціями розподілу та характеристичними функціями* $\boxdot$

Вправи

(1) Довести посилений закон великих чисел: $\nu(t)/t \to^{P1} \lambda, t \to \infty$.

(2) Довести, що процес Пуассона є стохастично неперервним, тобто має місце збіжність $\nu(t+h) \to^P \nu(t)$, $h \to 0$.

(3) Нехай $(\tau_n, n \geq 1)$ – стохастичний потік для процесу Пуассона з параметром λ, а $(\delta_n, n \geq 1)$ – незалежна від нього послідовність незалежних величин таких, що $\mathbf{P}(\delta_n = 1) = p = 1 - \mathbf{P}(\delta_n = 0)$. Довести, що послідовність $(\delta_n \tau_n, n \geq 1)$ після усунення нульових значень є стохастичним потоком для процесу Пуассона з параметром λp. *Вказівка*: скористатись зауваженням до теореми про властивості траєкторій процесу Пуассона.

(4) Нехай $(\tau_n^{(i)}, n \geq 1), i = 1, 2,$ – стохастичні потоки для незалежних процесів Пуассона з параметрами λ_i. Тоді послідовність $\left(\{\tau_n^{(1)}, n \geq 1\} \cup \{\tau_n^{(2)}, n \geq 1\}\right)$ після впорядкування за зростанням є стохастичним потоком для процесу Пуассона з параметром $\lambda_1 + \lambda_2$. *Вказівка*: скористатись означенням процесу Пуассона.

(5) Для послідовності $(g(n))$ визначимо $f(n,t) = \mathbf{E} g(n + \nu_t)$. Довести, що $(\partial/\partial t) f(n,t) = \lambda g(n+1) - \lambda g(n), n \geq 0$.

(6) Нехай $\nu_k(t), t \geq 0, k = 1, 2,$ пара незалежних процесів Пуассона з параметрами λ_k. Довести, що ймовірність перетину двовимірним випадковим процесом $(\nu_1(t), \nu_2(t))$ прямої $\{(i,j) : i + j = n\}$ саме у точці (i,j) дорівнює $C_n^i p^i (1-p)^{n-i}$, де $p = \lambda_1/(\lambda_1 + \lambda_2)$.

2.16. Процес народження та загибелі

Процес Пуассона є моделлю процесу чистого народження, якщо миттєвий приріст процесу інтерпретувати як народження нової частинки в популяції. Процес народження та загибелі дозволяє моделювати ситуації, коли одночасно можливою є також і загибель частинок. Цей процес є узагальненням поняття *ланцюга Маркова* на неперервний час.

Означення. *Випадковий процес* $(\zeta_t, t \in \mathbb{R}_+)$ *з дискретним простором значень* $E = \{i, j, ...\}$ *називається* **однорідним марковським процесом,** *якщо для всіх* $n \geq 1,\ i_0, i_1, ..., i_{n-1}, i, j \in E,\ 0 \leq t_0 < ... < t_{n-1} < t < t+s$ *виконуються рівності*

$$\mathbf{P}(\zeta_{t+s} = j \mid \zeta_{t_0} = i_0, ..., \zeta_{t_{n-1}} = i_{n-1},\ \zeta_t = i) =$$

$$\mathbf{P}(\zeta_{t+s} = j \mid \zeta_t = i) = \mathbf{P}(\zeta_s = j \mid \zeta_0 = i) \equiv p_{ij}(s).$$

Як і для *ланцюгів Маркова*, перша рівність відображає марковську властивість процесу, друга – однорідність за часом, а третя визначає **ймовірності переходу** процесу за проміжок часу s із стану i до стану j. Імовірнісні властивості процесу визначаються поведінкою функцій $p_{ij}(s)$.

Означення. *Випадковий процес* $(\zeta_t,\ t\in\mathbb{R}_+)$ *називається* процесом народження та загибелі, *якщо:*

(1) він є однорідним марковським процесом зі значеннями у $\mathbb{Z}_+$,

(2) для деяких невід'ємних чисел λ_i (інтенсивностей народжень) *і* μ_i (інтенсивностей загибелі), $i\ge 0$, *справедливі такі інфінітезимальні зображення для ймовірностей одиничних стрибків:*

$$\mathbf{P}(\zeta_h=i+1\mid \zeta_0=i)=\lambda_i h+o(h),\ h\to 0,\ \forall i\ge 0,$$

$$\mathbf{P}(\zeta_h=i-1\mid \zeta_0=i)=\mu_i h+o(h),\ h\to 0,\ \forall i> 0,$$

(3) імовірності інших стрибків є нескінченно малими:

$$\mathbf{P}(|\zeta_h-i|>1\mid \zeta_0=i)=o(h),\ h\to 0.$$

Наслідок (про ймовірність невиходу процесу народження та загибелі). *З умов (1)-(3) попереднього означення випливає зображення*

$$\mathbf{P}(\zeta_h=i\mid \zeta_0=i)=1-a_i h+o(h),\ h\to 0,\ \forall i\ge 0,$$

де сумарна інтенсивність

$$a_i\equiv\lambda_i+\mu_i,\ i\ge 0,\quad \mu_0\equiv 0.$$

Доведення. Обчислимо

$$\mathbf{P}(\zeta_h=i\mid \zeta_0=i)=1-\mathbf{P}(|\zeta_h-i|>1\mid \zeta_0=i)-$$
$$\mathbf{P}(\zeta_h=i+1\mid \zeta_0=i)-\mathbf{P}(\zeta_h=i-1\mid \zeta_0=i)=$$
$$1-(\lambda_i+\mu_i)h+o(h),\ h\to 0.$$

Очевидно, що при $i=0$ останній доданок у середній частині даного рівняння відсутній, тому для справедливості зображення в даному випадку слід покласти $\mu_0=0$ ⊡

Означення. Часом перебування *процесу в стані* $\zeta_0=i$ *називається випадкова величина*

$$\tau=\inf(t>0:\zeta_t\ne\zeta_0).$$

Теорема (про траєкторії процесу народження та загибелі). *Нехай* $(\zeta_t,\ t\in\mathbb{R}_+)$ *– процес народження та загибелі, що має неперервні справа траєкторії, а* τ *– час перебування в стані* $\zeta_0=i$. *Тоді для всіх* $i\ge 0$, $j\ne i$ *виконуються рівності*

$$\mathbf{P}(\tau<t\mid \zeta_0=i)=1-\exp(-a_i t),$$

$$\mathbf{P}(\tau<t,\ \zeta_\tau=j\mid \zeta_0=i)=(1-\exp(-a_i t))\ p_{ij},$$

де стохастична матриця (p_{ij}) *має вигляд*

$$p_{ij} = \begin{cases} \lambda_i/a_i, & j = i+1 \\ \mu_i/a_i, & j = i-1 \\ 0, & |j-i| \neq 1, i \neq 0 \end{cases}$$

Зауваження. Твердження теореми можна переформулювати так: час перебування в початковому стані $\zeta_0 = i$ та положення процесу після першого стрибка незалежні, час перебування має *показниковий розподіл* із параметром a_i, а положення після стрибка збігається зі станом після переходу на один крок вкладеного *ланцюга Маркова*, що має *матрицю перехідних імовірностей* за один крок $(p_{ij},\ i,j \geq 0)$.

Наслідок. *Нехай* $(\xi_n,\ n \geq 0)$ *– ланцюг Маркова з матрицею перехідних імовірностей за один крок* $(p_{ij},\ i,j \geq 0)$*, випадкові величини* $(\theta_n, n \geq 1)$ *при фіксованих* $(\xi_n,\ n \geq 0)$ *незалежні в сукупності і мають показниковий розподіл із параметром* $a_{\zeta_{n-1}}$ *відповідно, причому* $\tau_n = \theta_1 + ... +\ \theta_n$*. Тоді процес*

$$\zeta_t = \sum\nolimits_{n \geq 0} \xi_n\ \mathbb{I}_{\{\tau_n \leq\ t\ < \tau_{n+1}\}},$$

є процесом народження та загибелі, що має вказані в означенні інфінітезимальні зображення.

Доведення теореми. Позначимо

$$t_{nk} = k2^{-n},\ \tau_n = \inf\left(t_{nk} > 0 : \zeta_{t_{nk}} \neq \zeta_0\right).$$

Оскільки $\{t_{nk}\} \subset \{t_{n+1,k}\}$, то послідовність τ_n не зростає. Траєкторії процесу ζ_t набувають цілих значень та неперервні справа. Тому знайдеться *м.н.* додатне $\varepsilon(\omega) > 0$ таке, що $\zeta_{\tau+s} = \zeta_\tau$ для всіх $s \in [0, \varepsilon)$. З неперервності траєкторій процесу справа виводимо, що з імовірністю 1 $\zeta_{\tau_n} = \zeta_\tau$ починаючи з деякого номера, отже $\tau_n \downarrow \tau,\ \zeta_{\tau_n} \to \zeta_\tau,\ n \to \infty$. Тому за *неперервністю ймовірності*

$$\mathbf{P}(\tau < t,\ \zeta_\tau = j \mid \zeta_0 = i) = \lim\nolimits_{n\to\infty} \mathbf{P}(\tau_n < t,\ \zeta_{\tau_n} = j \mid \zeta_0 = i).$$

Як і вище, позначимо

$$p_{ij}(t) = \mathbf{P}(\zeta_t = j \mid \zeta_0 = i).$$

Визначимо номери $k_n(t) = [t\ 2^n] = \sup(k : t_{nk} < t)$ та події

$$A^i_{nk} = \{\zeta_0 = i,\ \zeta_{t_{n1}} = i, ...,\ \zeta_{t_{nk}} = i\}.$$

Для часу перебування обчислимо за формулою *про ймовірність перерізу n подій* та за марковською властивістю процесу

$$\mathbf{P}(\tau \geq t \mid \zeta_0 = i) = \lim_{n\to\infty} \mathbf{P}(\tau_n \geq t \mid \zeta_0 = i) =$$

$$\lim_{n\to\infty} \mathbf{P}(A^i_{n,\ k_n(t)} \mid \zeta_0 = i) = \lim_{n\to\infty} \prod_{l=1}^{k_n(t)} \mathbf{P}(\zeta_{t_{nl}} = i \mid A^i_{n,\ l-1}) =$$

$$\lim_{n\to\infty} \prod_{l=1}^{k_n(t)} \mathbf{P}(\zeta_{t_{nl}} = i \mid \zeta_{t_{n,l-1}} = i) = \lim_{n\to\infty} (p_{ii}(2^{-n}))^{k_n(t)} =$$

$$\lim_{n\to\infty} (1 - a_i 2^{-n} + o(2^{-n}))^{t2^n} = \exp(-ta_i),$$

де використаний наслідок *про ймовірність невиходу процесу народження та загибелі*. Перше твердження доведене.

Якщо ж $j \neq i$, то ймовірність виходу дорівнює

$$\mathbf{P}(\tau_n < t,\ \zeta_{\tau_n} = j \mid \zeta_0 = i) = \sum_{k\ =1}^{k_n(t)} \mathbf{P}(\tau_n = t_{nk},\ \zeta_{\tau_n} = j \mid \zeta_0 = i) =$$

$$\sum_{k\ =1}^{k_n(t)} \mathbf{P}(A^i_{n,k-1} \cap \{\zeta_{t_{nk}} = j\} \mid \zeta_0 = i) =$$

$$\sum_{k\ =1}^{k_n(t)} \mathbf{P}(\zeta_{t_{nk}} = j \mid A^i_{n,k-1}) \prod_{l=1}^{k-1} \mathbf{P}(\zeta_{t_{nl}} = i \mid A^i_{n,l-1}) =$$

$$\sum_{k\ =1}^{k_n(t)} \mathbf{P}(\zeta_{t_{nk}} = j \mid \zeta_{t_{n,k-1}} = i) \prod_{l=1}^{k-1} \mathbf{P}(\zeta_{t_{nl}} = i \mid \zeta_{t_{n,l-1}} = i) =$$

$$\sum_{k\ =1}^{k_n(t)} p_{ij}(2^{-n})(p_{ii}(2^{-n}))^{k-1} = p_{ij}(2^{-n})\left(1 - (p_{ii}(2^{-n}))^{k_n(t)}\right)\ /\ (1 - p_{ii}(2^{-n})).$$

Границю другого множника в правій частині обчислено вище:

$$\lim_{n\to\infty}\left(1 - (p_{ii}(2^{-n}))^{k_n(t)}\right) = 1 - \exp(-ta_i).$$

З означення p_{ij}, умов (2),(3) та наведеного вище наслідку *про ймовірність невиходу процесу народження та загибелі* $p_{ii}(h)$ отримуємо також

$$\lim_{n\to\infty} p_{ij}(2^{-n})\ /\ (1 - p_{ii}(2^{-n})) = p_{ij}.$$

Це доводить друге твердження теореми ⊡

Наступна теорема дає можливість обчислити ймовірності переходу через інфінітезимальні характеристики λ_i, μ_i. Зауважимо, що для будь-якого початкового розподілу $q_k = \mathbf{P}(\zeta_0 = k)$ безумовний розподіл процесу визначається за *формулою повної ймовірності*

$$\mathbf{P}(\zeta_t = j) = \sum\nolimits_{k\geq 0} \mathbf{P}(\zeta_0 = k)\mathbf{P}(\zeta_t = j \mid \zeta_0 = k) = \sum\nolimits_{k\geq 0} q_k\ p_{kj}(t).$$

Теорема (про систему диференціальних рівнянь Колмогорова). *Нехай* $(\zeta_t, t \in \mathbb{R}_+)$ *є процесом народження та загибелі з інфінітезимальними характеристиками* (λ_i, μ_i)*, причому асимптотичне зображення в умові (3) означення є рівномірним за* i.

Тоді розподіл процесу $p_k(t) = \mathbf{P}(\zeta_t = k)$ *при довільних початкових умовах є диференційовною функцією часу* t *та задовольняє при* $k \ge 0$ *зворотну* ***систему диференціальних рівнянь Колмогорова***

$$p_k'(t) = \lambda_{k-1}\ p_{k-1}(t) + \mu_{k+1}\ p_{k+1}(t) - (\lambda_k + \mu_k)\ p_k(t),$$

де за означенням $\lambda_{-1} = 0, \mu_0 = 0.$

Доведення. При $h > 0$ за означенням *умовної ймовірності*

$$\mathbf{P}(\zeta_t = j,\ \zeta_{t+h} = k) = p_j(t)\mathbf{P}(\zeta_{t+h} = k \mid \zeta_t = j) = p_j(t)p_{jk}(h).$$

Тому за *формулою повної ймовірності*

$$p_k(t+h) = \mathbf{P}(\zeta_{t+h} = k) =$$
$$\mathbf{P}(\zeta_t = k-1,\ \zeta_{t+h} = k) + \mathbf{P}(\zeta_t = k+1,\ \zeta_{t+h} = k) +$$
$$\mathbf{P}(\zeta_t = k,\ \zeta_{t+h} = k) + \mathbf{P}(|\zeta_t - k| > 1,\ \zeta_{t+h} = k) =$$
$$p_{k-1}(t)\ p_{k-1,k}(h) + p_{k+1}(t)\ p_{k+1,k}(h) + p_k(t)\ p_{kk}(h) + \textstyle\sum_{j:\ |k-j|>1} p_j(t)p_{jk}(h).$$

За умовою

$$\sup_{j,k:\ |k-j|>1} p_{jk}(h) \le \sup_j \mathbf{P}(|\zeta_h - j| > 1 \mid \zeta_0 = j) = o(h), h \to 0.$$

Тому останній доданок у попередній сумі дорівнює $o(h)$.

Підставимо в перші два доданки зображення з умови (2), а в третій – рівність наслідку *про ймовірність невиходу процесу народження та загибелі*:

$$p_k(t+h) = p_{k-1}(t)(\lambda_{k-1}h + o(h)) + p_{k+1}(t)\ (\mu_{k+1}h + o(h)) +$$
$$p_k(t)(1 - (\lambda_k + \mu_k)h + o(h)) + o(h),\ h \to 0.$$

Звідси

$$\frac{p_k(t+h) - p_k(t)}{h} = p_{k-1}(t)\lambda_{k-1} + p_{k+1}(t)\ \mu_{k+1} - p_k(t)(\lambda_k + \mu_k) + \frac{o(h)}{h},\ h \to 0.$$

Оскільки права частина має границю при $h \to 0$, а величина $o(h)/h$ є рівномірно за t малою, то функції $p_k(t)$ диференційовні та задовольняють *систему диференціальних рівнянь Колмогорова.*

При $k = 0$ за умовою в правій частині відсутні доданки з $p_{k-1}(t)$ та μ_0. Тому слід вважати, що $\lambda_{-1} = 0, \mu_0 = 0$ $\boxdot$

Означення. *Початковий розподіл* $q_k = \mathbf{P}(\zeta_0 = k)$ *називається* **стаціонарним розподілом** *процесу, якщо* $\mathbf{P}(\zeta_t = k) = q_k$ *для всіх* $t \ge 0,\ k \ge 0.$

Теорема (про рівняння для стаціонарних імовірностей). *Нехай* $(\zeta_t,\ t\in\mathbb{R}_+)$ *є процесом народження та загибелі з* ***інфінітезимальними*** *характеристиками* (λ_i,μ_i), *і зображення у (3) є рівномірним за* i.

Процес $(\zeta_t,\ t\in\mathbb{R}_+)$ *має стаціонарний розподіл* $(q_k, k\geq 0)$ *тоді й тільки тоді, коли система рівнянь*

$$\pi_{k-1}\lambda_{k-1}+\pi_{k+1}\ \mu_{k+1}-\pi_k(\lambda_k+\mu_k)=0,\ k\geq 0,$$

має розв'язок $(\pi_k, k\geq 0)$ *у класі дискретних розподілів імовірностей. Цей розв'язок збігається зі стаціонарним розподілом:* $q_k=\pi_k$.

Доведення. Необхідність є наслідком теореми *про систему диференціальних рівнянь Колмогорова*, оскільки при виборі стаціонарного розподілу як початкового функції $p_k(t)=\pi_k$ сталі і мають нульові похідні. Достатність не доводиться $\boxdot$

Теорема (про існування стаціонарних імовірностей). *Система рівнянь із теореми про рівняння для стаціонарних імовірностей має ненульовий розв'язок* $(\pi_k, k\geq 0)$ *у класі дискретних імовірнісних розподілів тоді й тільки тоді, коли збігається ряд*

$$\theta=\sum\nolimits_{k\geq 0}\theta_k<\infty,\quad \theta_k=\prod\nolimits_{i=0}^{k-1}\frac{\lambda_i}{\mu_{i+1}},$$

і у випадку збіжності стаціонарний розподіл має вигляд

$$\pi_k=\theta_k\ /\ \theta,\ k\geq 0.$$

Доведення. Позначимо $\delta_k=\pi_{k+1}\ \mu_{k+1}-\pi_k\lambda_k,\ k\geq 0.$

Система рівнянь із теореми *про рівняння для стаціонарних імовірностей* еквівалентна системі $\delta_0=0,\ \delta_{k+1}=\delta_k,\ k\geq 0.$

Отже , з $\pi_k=\pi_{k-1}\lambda_{k-1}/\mu_k$ рекурентно отримуємо $\pi_k=\theta_k\pi_0$.

Якщо $\theta=\infty$, з цих рівностей дістанемо $\pi_k=\pi_0=0$. У протилежному випадку знаходимо шуканий розподіл за означенням θ $\boxdot$

Теорема (про ергодичність процесу народження та загибелі). *За умови існування стаціонарного розподілу* $(\pi_k, k\geq 0)$ *процесу народження та загибелі існує границя, що не залежить від початкового стану:*

$$\lim\nolimits_{t\to\infty}\ p_{ik}(t)=\pi_k,\quad \forall i,k\geq 0.$$

Доведення не наводиться.

Вправи

(1) Довести, що процес народження та загибелі $(\zeta_t, t\geq 0)$ з інтенсивностями $\lambda_n=\lambda$, $\mu_n=0$ є процесом Пуассона.

(2) Довести, що процес народження та загибелі $(\zeta_t, t \geq 0)$ з інтенсивностями $\mu_n = 0$ можна зобразити у вигляді $\zeta_t = \sum_{n\geq 1} n 1\!\!\mathrm{I}_{\{S_{n-1}<t\leq S_n\}}$, де $S_n = \sum_{k=1}^{n} \xi_k$, а ξ_k має показниковий розподіл $Exp(\lambda_k)$. Якщо $\lambda_n = n\lambda$ (процес Юла) та $\zeta_0 = k$, то $\mathbf{E} z^{\zeta_t} = (z \exp(-t)) \, (1 - z + z \exp(-t))^k$.

(3) Знайти стаціонарний розподіл у системі $M/M/n$, що відповідає процесу народження та загибелі з інтенсивностями $\lambda_n = \lambda$, $\mu_n = \mu \min(n, m)$.

(4) Нехай (ν_t) – процес Пуассона. Знайти середнє та коваріаційну функцію процесу $((-1)^{\nu_t}, t \geq 0)$. Чи є цей процес марковським ?

(5) Нехай $(\zeta_t, t \geq 0)$ – процес народження та загибелі з інтенсивностями: $\lambda_n = \lambda$, $\mu_n = n\mu$ (система обслуговування $M/M/\infty$), і $\zeta_0 = k$. Довести тотожність $\mathbf{E} z^{\zeta_t} = \exp\left(\lambda(z-1)(1-\exp(-\mu t))/\mu\right) (1 + (z-1)\exp(-\mu t))^k$.

(6) Нехай $(\zeta_t, t \geq 0)$ – процес народження та загибелі з інтенсивностями $\lambda_n = n\lambda$, $\mu_n = n\mu$. Довести, що розподіл числа частинок у момент першої загибелі є геометричним, та знайти його параметр.

(7) Нехай $(\zeta_t, t \geq 0)$ – процес народження та загибелі з інтенсивностями $\lambda_n = n\lambda + a$, $\mu_n = n\lambda$. Позначимо $u(t) = \mathbf{P}(\zeta_t = 0 \mid \zeta_0 = 1)$ (а) Методом першого стрибка довести, що ця функція задовольняє диференціальне рівняння Рікатті: $u'(t) + 2\lambda u(t) = \lambda + \lambda u^2(t)$. (б) Вивести, що $u(t) = \lambda t/(1 + \lambda t)$.

2.17. Складний процес Пуассона

На відміну від звичайного, прирости складного процесу Пуассона не обов'язково є одиничними.

Означення. *Нехай* $(\xi_n, n \geq 1)$ *– незалежні у сукупності однаково розподілені випадкові величини, а* $\nu(t)$ *– незалежний від них процес Пуассона з параметром* λ. **Складним процесом Пуассона,** *що породжується процесом* ν *та стрибками* $(\xi_n, n \geq 1)$, *називається випадковий процес*

$$\xi(t) \equiv \sum\nolimits_{n=1}^{\nu(t)} \xi_n = \sum\nolimits_{n\geq 1} \xi_n \; 1\!\!\mathrm{I}_{\{n \,\leq\, \nu(t)\}}, \quad t \geq 0.$$

За означенням, *траєкторії* процесу $\xi(t)$ кусково-сталі, неперервні зліва, їх стрибки відбуваються в моменти стрибків процесу $\nu(t)$, а величини стрибків задаються послідовністю ξ_n.

Приклад. Розглянемо модель страхової фірми, яка регулярно отримує від застрахованих осіб **страхові премії** з інтенсивністю c за одиницю часу, та проводить **страхові виплати** випадкових обсягів $(\xi_n, n \geq 1)$ у моменти страхових випадків, що утворюють *стохастичний потік*, модельований *процесом Пуассона.* Якщо початковий страховий фонд фірми дорівнює x, то на момент часу t страховий фонд становитиме

$$\eta(t) = x + ct - \xi(t),$$

де $\xi(t)$ – *складний процес Пуассона.* Випадковий процес $\eta(t)$ називається класичним процесом ризику.

Теорема (про властивості складного пуассонівського процесу). *Складний процес Пуассона має незалежні та однорідні прирости, причому для всіх $u \in \mathbb{R}$ виконується рівність*

$$\mathbf{E}\exp(iu\xi(t)) = \exp(\lambda t(\varphi(u) - 1)),$$

де $\varphi(u) = \mathbf{E}\exp(iu\xi_1)$ – характеристична функція стрибка, та

$$\mathbf{E}\xi(t) = \lambda t\mathbf{E}\xi_1, \quad \mathbf{D}\xi(t) = \lambda t\mathbf{E}\xi_1^2.$$

Доведення. Незалежність та однорідність приростів є наслідком цих властивостей *процесу Пуассона*, оскільки приріст процесу має вигляд

$$\xi[s,t) = \sum\nolimits_{\nu(s)<n\le\nu(t)} \xi_n = \sum\nolimits_{n=1}^{\nu[s,t)} \xi_{\nu(s)\,+\,n},\ \ t \ge s \ge 0,$$

де величини $\nu(s), \nu[s,t), \xi_n$ *незалежні в сукупності.*

Вираз для характеристичної функції виводимо з теореми *про властивості генератрис*, пункт (е), при $z = \exp(iu)$:

$$\mathbf{E}\exp(iu\xi(t)) = \varphi_{\nu(t)}(\mathbf{E}\exp(iu\xi_1)) = \varphi_{\nu(t)}(\varphi(u)) =$$

$$\sum\nolimits_{n\ge0}(\varphi(u))^n\frac{(\lambda t)^n}{n!}e^{-\lambda t} = \exp(\lambda t\varphi(u) - \lambda t),$$

де використана також теорема *про розподіл процесу Пуассона.*

Далі, за теоремою *про властивості характеристичних функцій*

$$i\mathbf{E}\xi(t) = \varphi'_{\xi(t)}(0) = \lambda t\varphi'(u)\exp(\lambda t\varphi(u) - \lambda t)\,|_{u=0} = \lambda t\varphi'(0) = i\lambda t\mathbf{E}\xi_1,$$

$$-\mathbf{E}\xi^2(t) = \varphi''_{\xi(t)}(0) = (\lambda t\varphi''(u) + (\lambda t\varphi'(u))^2)\exp(\lambda t\varphi(u) - \lambda t)\,|_{u=0} =$$
$$\lambda t\varphi''(0) + (\lambda t\varphi'(0))^2 = -\lambda t\mathbf{E}\xi_1^2 - (\lambda t\mathbf{E}\xi_1)^2,$$

звідки отримуємо останні рівності теореми $\boxdot$

За допомогою граничного переходу зі *складного процесу Пуассона* можна отримати неперервний процес із незалежними та *однорідними приростами.* Для цього розглянемо складний процес Пуассона з малими симетричними приростами: $\mathbf{P}(\xi_n = \pm\varepsilon) = 1/2,\ \varepsilon \to 0$, та одночасно з великою кількістю стрибків за одиницю часу: $\lambda \to \infty$. За теоремою *про властивості складного пуассонівського процесу* характеристична функція такого процесу має вигляд

$$\mathbf{E}\exp(iu\xi(t)) = \exp(\lambda t(\cos(\varepsilon u) - 1)).$$

Для існування її границі при $\varepsilon \to 0$, $\lambda \to \infty$ оберемо $\lambda \sim \varepsilon^{-2}$. Тоді

$$\lim_{\varepsilon\to 0,\ \lambda\to\infty} \mathbf{E}\exp(iu\xi(t)) = \exp(-u^2 t/2),$$

тобто слабка границя $\xi(t)$ повинна мати *нормальний розподіл* $N(0,t)$.

Вправи

(1) У припущенні інтегровності стрибка ξ_1 довести посилений закон великих чисел: $\xi(t)/t \to^{P1} \lambda\mathbf{E}\xi_1, t \to \infty$.

(2) У припущеннях $\mathbf{E}\xi_1 = m, \mathbf{D}\xi_1 = \sigma^2$ довести центральну граничну теорему: $(\xi(t) - mt)/\sqrt{\sigma^2 t} \to^W \zeta \simeq N(0,1)$.

(3) Нехай $\sum_{n=1}^{\nu(t)} \xi_n$ – складний процес Пуассона, а множини $B_i \in \mathfrak{B}[\mathbb{R}]$, $i = \overline{1,k}$, попарно несумісні. Визначимо $\xi(t,B) \equiv \sum_{n=1}^{\nu(t)} \xi_n \mathbb{1}_{\xi_n \in B}$. Довести, що (а) випадковий процес $\xi(t,B_i)$ є узагальненим процесом Пуассона та знайти його характеристичну функцію, (б) випадкові величини $(\xi(t,B_i), i = \overline{1,k})$ незалежні.

(4) Нехай $\xi(t)$ – складний процес Пуассона з інтегровними стрибками, а процес $\eta(t) = x + ct - \xi(t)$. Знайти умову, за якої $\sup_{t>0} \eta(t) < \infty$ м.н.

(5) Нехай $\xi(t)$ – складний процес Пуассона з інтенсивністю λ, що задовольняє умову Крамера: $\varphi(u) \equiv \mathbf{E}\exp(u\xi_1) < \infty$ для деяких $u > 0$. Визначимо процес $\eta(t) = x + ct - \xi(t)$. (а) Довести, що при $0 \le s < t$ виконується тотожність $\mathbf{E}\exp(-u(\eta(t) - \eta(s))) = \exp((t-s)g(u))$, де функція $g(u) = \lambda(\varphi(u) - 1) - uc$. (б) Визначимо момент банкрутства $\tau_x = \inf(t > 0 : \eta(t) < 0)$, де $x = \eta(0) > 0$. Довести при $u \ge 0$ нерівність $\mathbf{P}(\tau_x \le t) \le \sup_{0\le s\le t} \exp(-ux + sg(u))$. (б) За умови $c > \lambda\mathbf{E}\xi_1$ існує єдине $\alpha > 0$ таке, що $g(\alpha) = 0$ і $\mathbf{P}(\tau_x \le t) \le \exp(-\alpha x)$.

2.18. Вінерівський процес

На початку 20 ст. Норберт Вінер побудував математичну модель для процесу хаотичного теплового (Броунівського) руху мікрочастинок у рідині. Цей процес характеризується неперервністю траєкторій, незалежністю та однорідністю приростів, нульовим середнім зміщенням за будь-який час та дифузійним характером руху, що проявляється виключно у змінах середнього квадратичного відхилення частинок від початкового стану.

Означення. Вінерівським, (Вінерівим)*, або* Броунівським *процесом називається випадковий процес* $(w(t),\ t \in \mathbb{R}_+)$*, який:*

(а) має неперервні траєкторії,

(б) має незалежні прирости та однорідні прирости,

(в) $w(0) = 0$, $\mathbf{E}w(t) = 0$, $\mathbf{D}w(t) < \infty$.

2.18.1. Розподіл вінерівського процесу

Теорема (про розподіл вінерівського процесу). *Нехай $w(t)$ – вінерів процес. Тоді існує стала $\sigma \geq 0$ така, що для всіх $t > 0$*

$$w(t) \simeq N(0, \sigma^2 t),\ \mathbf{E}w(t) = 0,\ \mathbf{D}w(t) = \sigma^2 t.$$

Зауваження. При виконанні решти припущень умову (а) неперервності траєкторій в означенні вінерівського процесу можна замінити на більш просту аналітичну умову

$$\mathbf{P}(|w(h)| \geq \varepsilon) = o(h),\ h \to 0,\ \text{для кожного } \varepsilon > 0.$$

Доведення теореми. Нехай $t > 0$ фіксоване. Розглянемо *прирости*

$$w_{nk} \equiv w[t(k-1)/n,\ tk/n),$$

де $w[s,t) \equiv w(t) - w(s)$. За умови (б) випадкові величини $(w_{nk,}\ k = \overline{1,n})$ *незалежні в сукупності* та *однаково розподілені*, $w_{n1} = w(t/n), \mathbf{E}w_{nk} = 0$.

Лема 1. *За припущень (б), (в) неперервність траєкторій процесу $w(t)$ еквівалентна умові, що сформульована у зауваженні.*

Доведення необхідності. З неперервності траєкторій процесу на компактному інтервалі $[0, t]$ випливає рівномірна неперервність, з якої, у свою чергу, слідує збіжність до нуля модуля неперервності на цьому інтервалі: $\max_{1\leq k\leq n} |w_{nk}| \to^P 0$ при $n \to \infty$, тобто умова

$$\mathbf{P}\left(\max{}_{1\leq k\leq n} |w_{nk}| \geq \varepsilon\right) \to 0,\ n \to \infty,\ \forall \varepsilon > 0.$$

Згідно з незалежністю та *однорідністю приростів*

$$\mathbf{P}\left(\max{}_{1\leq k\leq n} |w_{nk}| < \varepsilon\right) = \exp(-n\ \ln \mathbf{P}(|w_{n1}| < \varepsilon)).$$

Тому збіжність до нуля модуля неперервності еквівалентна співвідношенню $\ln \mathbf{P}(|w_{n1}| < \varepsilon) = o(1/n),\ n \to \infty$, що внаслідок зображення $\ln(1+\varepsilon) = \varepsilon + o(\varepsilon), \varepsilon \to 0$, еквівалентне співвідношенню:

$$\mathbf{P}(|w_{n1}| \geq \varepsilon) \equiv \mathbf{P}(|w(t/n)| \geq \varepsilon) = o(1/n),\ n \to \infty\ \boxdot$$

Лема 2. *Існують $\varepsilon_n \to 0$ такі, що $n\mathbf{P}(|w_{n1}| \geq \varepsilon_n) \to 0,\ n \to \infty$.*

Доведення. За лемою 1 для кожного $\varepsilon > 0$ справедливе зображення

$$\mathbf{P}(|w_{n1}| \geq \varepsilon) = o(1/n),\ n \to \infty.$$

Виходячи з нього, побудуємо неспадну послідовність $n_k \geq k$ таку, що

$$n\mathbf{P}(|w_{n1}| \geq 1/k) \leq 1/k$$

для всіх $n \geq n_k$ і покладемо $\varepsilon_n = 1/k$ для $n_k \leq n < n_{k+1}$. Послідовність ε_n є шуканою, оскільки при $n_k \leq n < n_{k+1}$

$$n\mathbf{P}(|w_{n1}| \geq \varepsilon_n) = n\mathbf{P}(|w_{n1}| \geq 1/k) \leq 1/k \to 0, \ n \to \infty \ \boxdot$$

Лема 3. *За умов теореми*
(а) функція $\sigma^2(t) \equiv Dw(t)$ *адитивна і не спадає,*
(б) $\sigma^2(t)$ *неперервна,*
(в) $\sigma^2(t) = \sigma^2 t$ *для всіх* $t \geq 0$ *та деякої сталої* $\sigma \geq 0$.

Доведення. (а) За означенням приросту

$$w(t+s) = w(t) + w[t, t+s),$$

де величини в правій частині незалежні, та $w[t, t+s) \simeq w(s)$ за *однорідністю приростів*. Тому адитивність випливає з теореми *про дисперсію суми незалежних величин*. Монотонність є наслідком невід'ємності дисперсії

$$\sigma^2(t+s) = \sigma^2(t) + \sigma^2(s) \geq \sigma^2(t).$$

(б) Розглянемо монотонну границю $\delta \equiv \lim_{h \downarrow 0} \sigma^2(h)$. Якщо $\delta > 0$, то з монотонності випливало б, що $\sigma^2(h) \geq \delta$ при всіх $h > 0$ та внаслідок адитивності $\sigma^2(1) = n\sigma^2(1/n) \geq n\delta \to \infty$, що суперечить скінченності дисперсії в означенні процесу (в). Тому $\delta = 0$ і $\sigma^2(h) \to 0, \ h \to 0$, звідки

$$\sigma^2(t \pm h) - \sigma^2(t) \ = \pm\, \sigma^2(h) \to 0, \ h \to 0.$$

(в) Позначимо $\sigma^2 \equiv \sigma^2(1)$. З адитивності та нормованості виводимо, що $\sigma^2(m/n) = m/n \ \mathbf{D}w(1) = (m/n)\sigma^2$ для натуральних m, n. Переходячи тут до границі $m/n \uparrow t$, з неперервності дістанемо рівність (в) $\boxdot$

Можна вважати, що $\sigma^2 > 0$ – інакше процес є тотожнім нулем майже напевне. Позначимо

$$\xi_{nk} = w_{nk} 1\!\!1_{\{|w_{nk}| \,\leq\, \varepsilon_n\}}.$$

Випадкові величини $(\xi_{nk}, \ k = \overline{1, n})$ незалежні за теоремою *про перетворення незалежних величин*, *однаково розподілені* та $|\xi_{nk}| \leq \varepsilon_n$.

Розглянемо суми

$$\xi_n = \sum\nolimits_{k=1}^{n} \xi_{nk}.$$

З нерівності

$$\mathbf{P}\left(\xi_n \neq w(t)\right) \leq \mathbf{P}\left(\cup_{k=1}^{n} \{\xi_{nk} \neq w_{nk}\}\right) \leq$$
$$\sum\nolimits_{k=1}^{n} \mathbf{P}(|w_{nk}| > \varepsilon_n) = n\mathbf{P}(|w_{n1}| > \varepsilon_n) \to 0, \ n \to \infty,$$

та леми 2 виводимо, що $\xi_n \to^P w(t), \ n \to \infty$.

Позначимо

$$\sigma_n^2 \equiv \mathbf{D}\xi_n, \quad \mu_n \equiv \mathbf{E}\xi_n.$$

Дисперсія тут існує та обмежена, оскільки

$$\mathbf{D}\xi_n = \sum\nolimits_{k=1}^{n} \mathbf{D}\xi_{nk} \le \sum\nolimits_{k=1}^{n} \mathbf{E}\xi_{nk}^2 \le \sum\nolimits_{k=1}^{n} \mathbf{E}w_{nk}^2 = n\mathbf{E}w_{n1}^2 = n\sigma^2 t/n,$$

де використана лема 3(в) та *однакова розподіленість* величин w_{nk} і w_{n1}.

(а) Припустимо, що $\underline{\lim}_{n\to\infty}\, \sigma_n^2 = 0$. Тоді $\xi_n - \mu_n \to^P 0$ по деякій підпослідовності, отже $w(t) - \mu_n = (\xi_n - \mu_n) - (\xi_n - w(t)) \to^P 0$. Оскільки величина $w(t)$ не залежить від n, а за теоремою *про співвідношення слабкої збіжності та збіжності за ймовірністю* має місце слабка збіжність $w(t) - \mu_n \to^W 0$, що спричиняє збіжність характеристичних функцій:

$$\mathbf{E}\exp(isw(t))\exp(-is\mu_n) \to 1, n \to \infty, \forall s \in \mathbb{R},$$

то числова послідовність μ_n повинна збігатися до деякої сталої. Тому $w(t)$ є сталою *майже напевне*. Це суперечить невиродженості $w(t)$ – як показано вище, $\mathbf{D}w(t) = \sigma^2 t > 0$.

(б) Отже, $\underline{\lim}_{n\to\infty}\, \sigma_n^2 > 0$, звідки $\sigma_n^2 \ge \delta^2 > 0$, починаючи з деякого номера. Враховуючи *однакову розподіленість* величин ξ_{nk}, та нерівність $|\xi_{nk}| \le \varepsilon_n$, підрахуємо показник з *умови Ліндеберга* для *загальної послідовності серій* $(\xi_{nk},\ k = \overline{1,n},\ n \ge 1)$:

$$L_n(\varepsilon) = \sigma_n^{-2} n\mathbf{E}(\xi_{n1} - \mathbf{E}\xi_{n1})^2 \mathbb{1}_{\{|\xi_{n1} - \mathbf{E}\xi_{n1}| \ge \varepsilon\sigma_n\}} \le \delta^{-2} n 4\varepsilon_n^2 \mathbb{1}_{2\varepsilon_n \ge \varepsilon\delta} = 0,$$

починаючи з деякого номера, оскільки $\varepsilon_n \to 0$.

Отже, внаслідок *центральної граничної теореми Ліндеберга для загальних серій* має місце слабка збіжність

$$(\xi_n - \mu_n)/\sigma_n \to^W \zeta \simeq N(0,1),$$

Враховуючи обмеженість σ_n^2, оберемо підпослідовність номерів n таку, що $\sigma_n \to b > 0$. Звідси за теоремою *про добуток слабко збіжної послідовності зі збіжною* дістанемо $\xi_n - \mu_n \to^W b\zeta$. За *теоремою Леві про критерій слабкої збіжності* виводимо існування границі *характеристичних функцій*

$$\mathbf{E}\exp(isb\zeta) = \lim\nolimits_{n\to\infty} \mathbf{E}\exp(is(\xi_n - \mu_n)) = \lim\nolimits_{n\to\infty} \exp(-is\mu_n)\mathbf{E}\exp(is\xi_n).$$

З іншого боку, границя

$$\lim\nolimits_{n\to\infty} \mathbf{E}\exp(is\xi_n) = \mathbf{E}\exp(isw(t))$$

існує та не дорівнює нулю при досить малих s, оскільки $\xi_n \to^P w(t)$, як показано вище, а характеристична функція неперервна в нулі. Отже, існує границя $\lim_{n\to\infty} \mu_n = \mu$. Тому

$$\mathbf{E}\exp(isw(t)) = \lim\nolimits_{n\to\infty} \mathbf{E}\exp(is\mu_n)\ \lim\nolimits_{n\to\infty} \mathbf{E}\exp(is(\xi_n - \mu_n)) =$$
$$\exp(is\mu)\varphi_{b\,\zeta}(s) = \exp(is\mu - s^2 b^2/2),$$

тобто $w(t) \simeq N(\mu, b^2)$. Оскільки $\mathbf{E}w(t) = 0$ та $\mathbf{D}w(t) = \sigma^2 t$ за лемою 3, то $w(t) \simeq N(0, \sigma^2 t)$ за теоремою *про інтерпретацію параметрів нормального розподілу* $\boxdot$

2.18.2. Властивості траєкторій вінерівського процесу

Теорема (про властивості траєкторій вінерівського процесу). *Нехай $w(t)$ – вінерів процес, а $t_{nk} = k2^{-n}$. Тоді:*

(а) $\mathbf{P}(\cup_{t\in[0,1]}\{\exists w'(t)\}) = 0$,

(б) $Var_{[0,1]} w \equiv \lim_{n\to\infty} \sum_{k=1}^{2^n} |w[t_{n,k-1}, t_{nk})| = \infty$ *майже напевне,*

(в) $\sum_{k=1}^{2^n} (w[t_{n,k-1}, t_{nk}))^2 \to^P \sigma^2, n \to \infty$.

Доведення. Вважатимемо, що $\sigma^2 = 1$.

(а) Розглянемо події $A = \cup_{t\in[0,1]}\{\exists w'(t)\}$,

$$B_{mn} = \cup_{t\in[0,1]} \cap_{s:\ |t-s|\le 2^{-n+1}} \{|w(t) - w(s)| \le m\,|t-s|\},$$

$$C_{mn} = \cup_{k=1}^{2^n-2}\{\delta_{nk} \le m2^{-n+2}\},$$

де $\delta_{nk} = \max_{i=0,1,2} |w[t_{n,k+i-1}, t_{n,k+i})|$.

Оскільки $A \subset \cup_{n\ge1} \cup_{m\ge1} B_{mn}$, то досить довести рівності $\mathbf{P}(B_{mn}) = 0$.

Має місце включення $B_{mn} \subset C_{mn}$. Дійсно, нехай $\omega \in B_{mn}$, тобто для деякого $t \in [0,1]$ та всіх s з $|t-s| \le 2^{-n+1}$ виконується нерівність $|w(t) - w(s)| \le m\,|t-s|$. Нехай $t \in [t_{nk}, t_{n,k+1})$ для деякого k. Тоді для кожного $i = 0,1,2$

$$|w[t_{n,k+i-1}, t_{n,k+i})| \le |w(t) - w(t_{n,k+i-1})| + |w(t) - w(t_{n,k+i})| \le$$

$$m\,|t - t_{n,k+i-1}| + m\,|t - t_{n,k+i}| \le 2m2^{-n+1} = m2^{-n+2},$$

за означенням B_{mn}, оскільки $|t_{n,k+i} - t| \le 2^{-n+1}$. Тому $\omega \in C_{mn}$.

Далі, оцінимо через *напівадитивність* та *монотонність імовірності*, та за властивостями *незалежності приростів* та *однорідності приростів*

$$\mathbf{P}(B_{mn}) \le \mathbf{P}(C_{mn}) \le \sum_{k=1}^{2^n-2} \mathbf{P}(\delta_{nk} \le m2^{-n+2}) =$$

$$(2^n - 2)\mathbf{P}(\delta_{n1} \le m2^{-n+2}) \le 2^n \left(\mathbf{P}(|w[0, t_{n1})| \le m2^{-n+2})\right)^3 \le$$

$$2^n \left((2\pi 2^{-n})^{-1/2} 2 \cdot m2^{-n+2}\right)^3 \le c2^{-n/2} \to 0, n \to \infty,$$

де враховано верхню межу $(2\pi\sigma^2)^{-1/2}$ для нормальної щільності $N(\mu, \sigma^2)$.

Зауважимо, що події B_{mn} не спадають за n. Тому з отриманої збіжності $\mathbf{P}(B_{mn}) \to 0, n \to \infty$, виводимо, що $\mathbf{P}(B_{mn}) = 0$ для всіх n. Оскільки параметр m тут є довільним, то твердження (а) доведене.

(б) Позначимо

$$w_{nk} = w[t_{n,k-1}, t_{nk}), \quad \zeta_n = \sum\nolimits_{k=1}^{2^n} |w_{nk}| \,.$$

Тоді $\zeta_n \uparrow \zeta \equiv \mathrm{Var}_{[0,1]} w, \ n \to \infty$. Крім того, $w_{nk} \simeq N(0, 2^{-n})$, тому випадкова величина $\eta \equiv 2^{n/2} w_{n1}$ має *стандартний нормальний розподіл.* Оскільки за властивостями *незалежності приростів* та *однорідності приростів* величини $(w_{nk}, k = \overline{1, 2^n})$ *незалежні в сукупності* та *однаково розподілені*, то за теоремами *про перетворення незалежних величин* та *про математичне сподівання добутку незалежних величин* для $s > 0$:

$$\mathbf{E} \exp(-s\zeta_n) = (\mathbf{E} \exp(-s\,|w_{n1}|))^{2^n} =$$

$$\left(\mathbf{E} \exp(-s2^{-n/2} \left|2^{n/2} w_{n1}\right|)\right)^{2^n} = \left(\mathbf{E} \exp(-s2^{-n/2}\,|\eta|)\right)^{2^n} =$$

$$(1 - s2^{-n/2} \mathbf{E}\,|\eta| + o(2^{-n/2}))^{2^n} \to 0, n \to \infty.$$

Оскільки $\zeta_n \uparrow \zeta, n \to \infty$, звідси отримуємо

$$\mathbf{P}(\zeta < \infty) = \lim\nolimits_{s\downarrow 0} \mathbf{E} \exp(-s\zeta) = \lim\nolimits_{s\downarrow 0} \lim\nolimits_{n\to\infty} \mathbf{E} \exp(-s\zeta_n) = 0.$$

(в) Нехай $(\eta_k, \ k \geq 1)$ послідовність незалежних *стандартних нормальних* величин. Оскільки $(w_{nk}, 1 \leq k \leq 2^n)$ незалежні і *однаково розподілені*, причому величини $2^{n/2} w_{n1} \simeq \eta_k \simeq N(0,1)$, то

$$\mathbf{P}\left(\left|\sum\nolimits_{k=1}^{2^n} (w_{nk})^2 - 1\right| \geq \varepsilon\right) = \mathbf{P}\left(\left|2^{-n} \sum\nolimits_{k=1}^{2^n} (2^{n/2} w_{nk})^2 - 1\right| \geq \varepsilon\right)$$

$$= \mathbf{P}\left(\left|2^{-n} \sum\nolimits_{k=1}^{2^n} \eta_k^2 - 1\right| \geq \varepsilon\right) \to 0, n \to \infty,$$

за *теоремою Чебишева про закон великих чисел,* оскільки η_k^2 незалежні однаково розподілені, $\mathbf{E}\eta_k^2 = 1$ $\boxdot$

Вправи. Нехай $w(t)$ – вінерівський процес.

(1) Для довільних $0 < t_1 < ... < t_n$ знайти сумісну щільність випадкових величин $(w(t_1), ..., w(t_n))$.

(2) Довести, що вінерівськими є такі перетворення $w(t)$: (а) $-w(t), t \geq 0$, (б) $tw(1/t), t > 0, \ w(0) \equiv 0$, (в) $w(t + c) - w(c), t \geq 0$, (г) $w(c) - w(c - t)$, $0 \leq t \leq c$, (д) $c^{-1} w(c^2 t)$.

(3)Визначимо $t_{nk} = k2^{-n}$. Довести, що прирости та подвійні прирости $\zeta_{0k} \equiv w(t_{0k}) - w(t_{0,k-1})$, $\zeta_{nk} \equiv w(t_{n,2k-1}) - (w(t_{n-1,k-1}) + w(t_{n-1,k}))/2, n, k \geq 1$, є незалежними з нормальним розподілом. Обчислити $w(t_{nk})$ через (ζ_{nk}).

(4) Знайти розподіл величин: (а) $w(t) + w(s), t < s$, (б) $\int_0^1 w(s) ds$.

(5) Довести, що для кожного $\alpha < 1/2$ траєкторії $w(t)$ м.н. задовольняють умову Гельдера порядку α.

(6) Довести, що процес $(|w(t)|, t \geq 0)$ (а) є марковським, (б) має такий самий розподіл, що і процес $\sup_{s \leq t} w(s) - w(t)$.

(7) Нехай τ – момент зупинки відносно потоку $(\mathfrak{F}_t = \sigma[w(s), s \leq t])$. Довести, що процес $w(t+\tau) - w(\tau)$ є вінерівським та не залежить від $.\mathfrak{F}_\tau$.

(8) Для $t \in (0,1)$.знайти умовну щільність $w(t)$ за умови $w(1) = 0$.

(9) Визначимо $\tau_b = \inf(t > 0 : w(t) \geq b)$ при $b > 0$. (а) Довести, що $\tau_b < \infty$ м.н. (б) Довести, що процес $\varpi(t) = w(t)1\!\!\mathrm{I}_{\{t<\tau_b\}} + (2b - w(t))1\!\!\mathrm{I}_{\{t \geq \tau_b\}}$ також є вінерівським.(в) Вивести принцип відбиття: $\mathbf{P}(\tau_b < t) = 2\mathbf{P}(w(t) \geq b)$. (г) Обчислити $\mathbf{E}\exp(-s\tau_b) = \exp(-b\sqrt{2s})$. (д) Довести, що $(\tau_b, b > 0)$ є випадковим процесом з незалежними приростами.

(10) (а) Довести, що $\inf(t > 0 : w(t) = 0) = 0$ м.н. (б) Нехай τ_1 – максимальний нуль $w(t)$ на відрізку $[0,1]$. Довести, що $\mathbf{P}(\tau_1 < t) = 2\pi^{-1}\arcsin\sqrt{t}$. (в) Знайти щільність мінімального нуля $w(t)$ на відрізку $[1,\infty)$.

(11) Для $x \in (a,b)$ позначимо через $u(x)$ імовірність того, що процес $x + w(t)$ досягне рівня b раніше, ніж рівня a. (а) Довести, що

$u(x) = (2\pi)^{-1/2}\int_a^b \exp(-(x-y)^2/2t)u(y)dy + o(t^2),\ t \to 0.$

(б) Вивести звідси, що функція u задовольняє при $x \in (a,b)$ диференціальне рівняння $u''(x) = 0$ та крайові умови $u(a) = 0, u(b) = 1$. (в) Знайти $u(x)$.

(12) Позначимо через $f(x,t,y)$ щільність випадкової величини $x + at + bw(t)$ для $x, a \in \mathbb{R}, b > 0$. (а) Довести, що ця функція задовольняє диференціальне рівняння $(\partial/\partial t)f(x,t,y) = a(\partial/\partial x)f(x,t,y) + (b/2)(\partial^2/\partial x^2)f(x,t,y)$ для всіх x, y та $t > 0$. (б) Довести, що для довільної $g \in C_b$ наведене рівняння задовольняє також функція $f(t,x,y) = \mathbf{E}g(x + at + bw(t))$.

(13) Випадкові процеси (а) $(w^2(t) - t, t \geq 0)$, (б) $(w_t^3 - 3tw_t, t \geq 0)$, (в) $(w_t^4 - 6tw_t^2 + 3t^2, t \geq 0)$, (г) $(\exp(w(t) - t/2), t \geq 0)$ є мартингалами.

(14) Довести, що $\overline{\lim}\,(\underline{\lim})_{t\to\infty} t^{-1/2}w(t) = +(-)\infty$ м.н.

Розділ 3

Математична статистика

Вступ

Первісна задача математичної статистики є в певному сенсі оберненою до основної задачі теорії ймовірностей. У теорії ймовірностей ми постулювали існування теоретично *повністю визначеного ймовірнісного простору* та виводили ті чи інші властивості подій і *випадкових величин*. Однак намагання застосувати певну ймовірнісну модель на практиці часто стикається з тією обставиною, що теоретичний (гіпотетичний) розподіл подій та випадкових величин є невідомим. Якнайбільше можна припустити, що цей розподіл міститься в деякому відомому класі (наприклад, є нормальним). Тому необхідну інформацію про теоретичні ймовірності доводиться отримувати з того ж самого *стохастичного експерименту*, проводячи *спостереження* над його проміжними результатами. Наприклад, так, як це робилося у *формулі Байєса* в курсі теорії ймовірностей.

Математична статистика – це розділ математики, який базується на теорії ймовірностей та призначений для формулювання і доведення **статистичних висновків** про властивості *ймовірнісного простору* за результатами спостережень над відповідним стохастичним експериментом.

Висновок про ймовірнісний простір можна віднести до однієї з груп:

(1) висновок про *кількісне* значення деякої величини (параметра),

(2) *якісний* висновок про значення параметру чи іншу властивість імовірнісного простору.

Тому умовно задачі математичної статистики можна розбити на групи:

(1) задачі **статистичного оцінювання,** що спрямовані на побудову кількісних **оцінок** невідомих параметрів,

(2) задачі **перевірки статистичних гіпотез,** в яких встановлюються якісні властивості ймовірнісного простору.

Приклади

1. Спостерігається послідовність *випробувань Бернуллі*. Треба кількісно оцінити ймовірність успіху в одному випробуванні.

2. Проводяться підкидання монети. За результатами спостережень необхідно зробити якісний висновок про відносну симетрію монети.

Математична статистика містить велику кількість спеціальних розділів. Серед них:

(а) *непараметрична статистика* – в якій невідомими "параметрами" виступають загальні функції розподілу;

(б) *теорія оптимальних незміщених оцінок*, де в класі незміщених оцінок дисперсія є мірою якості і знаходяться оптимальні оцінки;

(в) *теорія оцінок максимальної вірогідності*, де пропонується універсальний метод для побудови якісних статистичних оцінок параметрів;

(г) *статистичні висновки для нормальних спостережень*, що ґрунтуються на спеціальних властивостях вибіркових статистик;

(д) *регресійний аналіз*, в якому вивчаються задачі встановлення функціональної залежності між числовими змінними при наявності похибок;

(е) *дисперсійний аналіз*, де аналізується залежність між стохастичними якісними факторами;

(ж) *непараметрична перевірка гіпотез*, яка містить класичну задачу про відповідність спостережень наперед заданій функції розподілу;

(з) *параметрична перевірка гіпотез*, де перевіряються припущення щодо значень параметрів;

(і) *теорія найбільш потужних критеріїв*, в якій вивчаються оптимальні процедури перевірки статистичних критеріїв;

(к) *послідовний статистичний аналіз*, в якому статистичні висновки робляться безпосередньо в процесі надходження спостережень;

(л) *теорія планування статистичного експерименту*, де пропонуються раціональні схеми збору статистичних даних для їх подальшої обробки, з урахуванням лімітів чи затрат на кожне спостереження;

тощо.

У свою чергу, математична статистика є теоретичною основою для таких дисциплін, як *прикладна статистика, біометрика, соціологія, технометрика, аналіз даних, економетрика, фінансовий аналіз* та інших. Її результати використовують при розробці комп'ютерних статистичних пакетів, таких як SAS, SPSS, MS Statistics тощо.

Зважаючи на вищу технічну складність теорем математичної статистики в порівнянні з курсом теорії ймовірностей, деякі з наведених нижче доведень спираються на гранично спрощені чи навіть не точно сформульовані припущення. Однак треба мати на увазі, що в кожній із розглянутих схем існують точно сформульовані та доведені математичні твердження. Отже, основна мета даного розділу полягає в тому, щоб дати первісне уявлення про методи та результати математичної статистики.

При викладі наведено прямі посилання на відповідні теореми курсу теорії ймовірностей за їх назвами, що містяться у предметному покажчику.

Даний розділ містить матеріал семестрового курсу математичної статистики для студентів III курсу спеціальностей математика, статистика, що розрахований на 51 годину лекцій та 17 годин практичних занять. За браком часу такі теми, як теорема Глівенка-Кантеллі, асимптотична нормальність оцінок методу моментів, властивості інформації та нерівність Крамера-Рао для векторного параметра, критерій хі-квадрат для складних гіпотез, кореляційний аналіз, критерії з монотонним відношенням вірогідностей, послідовний аналіз, поліноміальна регресія, теорема Хінчина про зображення коваріаційної функції, лінійний прогноз у гільбертовому просторі, регулярні стаціонарні послідовності – викладаються без доведень з розрахунком на самостійну роботу студентів.

3.1. Статистичний простір, вибірка. Статистики та оцінки

Означення. Статистичним простором *називається трійка*

$$(\Omega,\ \mathfrak{F},\ (\mathbf{P}_\theta,\ \theta\in\Theta)),$$

що складається з таких елементів:

Ω – *простір елементарних подій деякого стохастичного експерименту,*

$\mathfrak{F}\subset 2^\Omega$ – *сигма-алгебра випадкових подій – підмножин* Ω,

$(\mathbf{P}_\theta,\ \theta\in\Theta)$ – *деяка параметрична сім'я ймовірностей на* $\mathfrak{F}$,

Θ – параметричний простір – *множина довільної природи.*

Вважається, що вигляд залежності *ймовірностей* $\mathbf{P}_\theta(A)$ від випадкових подій $A\in\mathfrak{F}$ при заданому значенні параметра θ повністю відомий, у той час як сам параметр θ – невідомий статистику.

Найчастіше $\Theta\subset\mathbb{R}^d$, тобто ймовірнісний розподіл на $\mathfrak{F}$ вважається відомим повністю за винятком d числових параметрів – координат θ. У цьому випадку говорять про параметричну статистику. Якщо ж множина

Θ є підмножиною функціонального простору (наприклад, простору всіх *функцій розподілу*), то можна говорити про непараметричну статистику.

Означення. *Якщо ξ – випадкова величина на ймовірнісному просторі $(\Omega,\ \mathfrak{F},\ \mathbf{P}_\theta)$, математичне сподівання ξ (абстрактний інтеграл Лебега) за ймовірністю $\mathbf{P}_\theta$ будемо позначати через*

$$\mathbf{E}_\theta\xi = \int_\Omega \xi(\omega)\mathbf{P}_\theta(d\omega).$$

Аналогічний зміст має позначення $\mathbf{D}_\theta\xi$ для дисперсії.

3.1.1. Статистична вибірка

Статистичні висновки про ймовірність $\mathbf{P}_\theta$ будемо робити на підставі спостережень, тобто значень певних випадкових величин, векторів та інших функцій від елементарних подій.

Означення. *Статистичною* вибіркою *називається довільна вимірна функція $X : \Omega \to S$ зі значеннями у вимірному* вибірковому просторі *$(S,\ \Sigma,\ \lambda)$, де:*

S – деяка множина (вибірковий простір),

Σ – сигма-алгебра підмножин S,

λ – деяка сигма-скінченна міра на Σ.

Вважається, що значення $X(\omega) = x$ є відомим для статистика (тобто *спостерігається*) і може використовуватись для отримання статистичних висновків.

Надалі вибірковим простором буде обиратися переважно евклідів простір $S = \mathbb{R}^n$ із *борелєвою сигма-алгеброю* $\Sigma = \mathfrak{B}(\mathbb{R})$, тому під вибіркою слід розуміти звичайний *випадковий вектор*, що спостерігається в *стохастичному експерименті*. У більшості випадків мірою λ слугує або точкова міра – відносно неї кожна одноточкова множина з певного класу має одиничне значення міри (у випадку дискретної вибірки X), або ж *міра Лебега*, якщо $S \subset \mathbb{R}^n$ і вибірка X має *сумісну щільність*.

У випадку, коли вектор X містить всю наявну інформацію про *стохастичний експеримент*, для спрощення часто вважають, що *простір елементарних подій* збігається з *вибірковим простором*: $(\Omega, \mathfrak{F}) = (S, \Sigma)$, а окремі спостереження є елементарними подіями: $X(\omega) = \omega$.

Приклад. Побудова висновку про симетричність монети за результатами серії з 1000 підкидань. Якщо вибірка містить результати кожного з підкидань, то вибірковий простір має вигляд $S = \Omega = \{A, P\}^{1000}$. Для параметризації задачі позначимо $\theta \in \Theta = (0,1)$ імовірність аверсу при одному підкиданні. Тоді ймовірність $\mathbf{P}_\theta$ задає розподіл вектора індикаторів успіхів у схемі *випробувань Бернуллі* з параметрами $n = 1000, \ p = \theta$.

Нагадаємо таке важливе поняття з теорії міри.

Означення. *Міра μ на деякому вимірному просторі $(S, \ \Sigma)$* **абсолютно неперервна відносно міри** *λ, (позначення $\mu \ll \lambda$), якщо для довільної множини $B \in \Sigma$ із $\lambda(B) = 0$ випливає $\mu(B) = 0$.*

За *теоремою Радона – Нікодима* ця властивість еквівалентна існуванню вимірної *інтегровної за мірою* λ функції $f(x), x \in S$, такої, що

$$\mu(B) = \int_B f(x)\lambda(dx), \ \forall B \in \Sigma.$$

Функція $f = d\mu/d\lambda$ називається **щільністю міри** *μ відносно λ.*

У деяких розділах статистики використовується **умова підпорядкованості.** Вона полягає в тому, що розподіл вибірки є абсолютно неперервним (має щільність) відносно фіксованої міри λ на *вибірковому просторі*:

$$\mathbf{P}_\theta(X \in \cdot) \ll \lambda(\cdot), \ \forall \theta \in \Theta.$$

У будь-якому випадку для зліченного параметричного простору Θ така міра завжди існує. Дійсно, довільна міра зі зліченної множини ймовірнісних мір $(\mu_\theta, \theta \in \Theta)$ *абсолютно неперервна відносно міри* $\lambda = \sum_{\theta \in \Theta} 2^{-n(\theta)} \mu_\theta$, де $n(\theta)$ – номер елемента θ у послідовності Θ.

Приклад. Спостерігається результат підкидання несиметричної монети з невідомою ймовірністю аверса. У цьому випадку

$$\Omega = \{\text{А,Р}\}, \ \ \mathfrak{F} = 2^\Omega, \ S = \Omega, \ \Sigma = \mathfrak{F}, \ \ \lambda \text{ – точкова міра},$$

$$\mathbf{P}(\{\text{А}\}) = 1 - \mathbf{P}(\{\text{Р}\}) = \theta \in \Theta = [0, 1].$$

3.1.2. Функція вірогідності

Означення. *Нехай $X : \Omega \to S$ – вибірка зі значеннями у вимірному просторі $(S, \ \Sigma, \ \lambda)$, яка задовольняє умову підпорядкованості.* **Функцією вірогідності** *(або* **функцією правдоподібності***) вибірки називається сумісна щільність розподілу вибірки відносно міри λ у вибірковому просторі:*

$$L(x, \theta) \equiv \frac{d\mathbf{P}_\theta(X \in \cdot)}{d\lambda(\cdot)}(x),$$

тобто така вимірна за x функція, що для всіх $B \in \Sigma$ і всіх $\theta \in \Theta$

$$\mathbf{P}_\theta(X \in B) = \int_B L(x, \theta) \ \lambda(dx).$$

Для *абсолютно неперервної* вибірки X міра λ є *мірою Лебега*, функція вірогідності як функція x є *сумісною щільністю* розподілу:

$$\mathbf{P}_\theta(X \in B) = \int_B L(x,\theta)\, dx,$$

а для дискретної вибірки λ – *точкова міра*, функція вірогідності є *дискретним розподілом* імовірностей:

$$\mathbf{P}_\theta(X \in B) = \sum\nolimits_{x \in B \cap S} L(x,\theta).$$

Означення. Вибірковою функцією вірогідності (емпіричною функцією вірогідності) *називається випадкова величина, що отримується в результаті підстановки у функцію вірогідності замість аргумента $x \in S$ значення вибірки як випадкового вектора*

$$L(X,\theta) \equiv L(x,\theta) \mid_{\, x = X} .$$

3.1.3. Кратні вибірки

Часто у статистиці використовується схема багатократних спостережень.

Означення. *Вимірний простір* $(S,\ \Sigma,\ \lambda) = (R, \mathfrak{B}, \upsilon)^n$ *є* n-кратним прямим добутком *вимірного простору* $(R, \mathfrak{B}, \upsilon)$, *якщо*

$$S = R^n \equiv \{x = (x_1, ..., x_n),\ x_k \in R\},$$

$$\Sigma = \mathfrak{B} \otimes ... \otimes \mathfrak{B} \equiv\ \sigma[B_1 \times ... \times B_n,\ B_k \in \mathfrak{B},\ k = \overline{1,n}],$$

а міра $\lambda \equiv \upsilon \times ... \times\ \upsilon$ *визначається на прямокутниках як добуток*

$$\lambda(B_1 \times ... \times B_n) \equiv \upsilon(B_1)...\upsilon(B_n).$$

Наприклад, $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n), L_n) = (\mathbb{R}, \mathfrak{B}(\mathbb{R}), L_1)^n$, де L_n – n-вимірна *міра Лебега* (довжина, площа, об'єм...).

Означення. *Нехай вибірковий простір* $(S,\ \Sigma,\ \lambda)$ *є n-кратним прямим добутком* $(R, \mathfrak{B}, \upsilon)^n$. *Випадковий вектор* $X : \Omega \to S$ *називається* n-кратною вибіркою, *якщо* $X = (\xi_1, ..., \xi_n)$ *складається з незалежних у сукупності однаково розподілених випадкових величин* ξ_k, $k = \overline{1,n}$, *зі значеннями у просторі* $(R, \mathfrak{B}, \upsilon)$, *тобто*

$$\mathbf{P}_\theta(X \in B_1 \times ... \times B_n) = \prod\nolimits_{k=1}^{n} \mathbf{P}_\theta(\xi_1 \in B_k),\ \forall B_k \in \mathfrak{B},\ \theta \in \Theta.$$

Означення. *Число n називається* об'ємом вибірки X.

Якщо X – *кратна вибірка*, її *вибіркова функція вірогідності* позначається через $L_n(X,\theta)$, де n – об'єм вибірки.

Зауваження. Для кратної вибірки $X = (\xi_1, ..., \xi_n)$ множина всіх випадкових величин ξ, що мають однакову з ξ_1 функцію розподілу, називається **генеральною сукупністю** (популяцією). У зв'язку з цим вибірку X можна уявляти як результат n-кратного незалежного послідовного вибору представників з генеральної сукупності.

Означення. **Функцією вірогідності спостереження** *для кратної вибірки називається щільність розподілу величини ξ_1 відносно міри υ* :

$$f(y, \theta) = \frac{d\mathbf{P}_\theta(\xi_1 \in \cdot)}{d\upsilon(\cdot)}(y),$$

тобто така вимірна функція f, що

$$\mathbf{P}_\theta(\xi_k \in B) = \int_B f(y, \theta)\ \upsilon(dy), \quad \forall B \in \mathfrak{B}, \forall \theta \in \Theta.$$

Теорема (про функцію вірогідності кратної вибірки). *Нехай кратна вибірка $X = (\xi_1, ..., \xi_n)$ містить спостереження ξ_k зі значеннями в просторі $(R, \mathfrak{B}, \upsilon)$. Припустимо, що ξ_k мають функцію вірогідності спостережень $f(y, \theta)$. Тоді функція вірогідності всієї вибірки X дорівнює добуткові*

$$L_n(x, \theta) = \prod_{k=1}^n f(x_k, \theta),\ x = (x_1, ..., x_n).$$

Доведення є наслідком теореми *про критерій незалежності абсолютно неперервних величин*, оскільки вибірковий вектор $X = (\xi_1, ..., \xi_n)$ утворений саме незалежними однаково розподіленими величинами $\boxdot$

3.1.4. Статистики та оцінки

Означення. **Статистикою** *називається довільна вимірна функція від вибірки: $T = T(X)$, яка не містить значень невідомого параметра θ.*

Множина значень статистики є довільним *вимірним простором* $(C, \mathfrak{C})$. Найчастіше це евклідів простір: $(C, \mathfrak{C}) = (\mathbb{R}^m, \mathfrak{B}(\mathbb{R}^m))$.

Зауваження. Терміном "*статистика*" будемо одночасно визначати як саму функціональну залежність $T(x) : S \to C$ від вибірки, так і її значення $T(\omega) = T(X(\omega)) : \Omega \to C$, яке отримується після підстановки *вибірки* $x = X(\omega)$. Тлумачення випливатиме зі змісту відповідного аргумента.

Означення. **Оцінкою** *невідомого параметра $\theta \in \Theta$ називається будь-яка статистика зі значеннями у параметричному просторі Θ.*

Щоб підкреслити спеціальний характер оцінки, часто її зображають у вигляді $\widehat{\theta}$. Очевидно, оцінка є засобом для прогнозування, передбачення, оцінювання значення невідомого параметра θ на підставі спостережень X.

Означення. *Нехай* $\Theta \subset \mathbb{R}^d$. Інтервальною оцінкою *(або* надійним інтервалом*) невідомого параметра* θ *називається пара* $\mathbb{R}^d$*-значних статистик* $(\widehat{\theta}_1, \widehat{\theta}_2)$ *таких, що* $\widehat{\theta}_1 \le \widehat{\theta}_2$ *майже напевне.* Оцінкою тут є паралелепіпед $[\widehat{\theta}_1, \widehat{\theta}_2) = \{\theta \in \mathbb{R}^d : \widehat{\theta}_1 \le \theta < \widehat{\theta}_2\}$.

Іноді щодо невідомого значення θ досить вказати обмеження лише з одного боку. У цьому разі на відміну від попередньої інтервальної оцінки, яка називається двобічною, розглядають також однобічні оцінки: лівобічну інтервальну оцінку $[\widehat{\theta}, \infty)$ та правобічну інтервальну оцінку $(-\infty, \widehat{\theta})$.

Зауваження. Якщо при кожному n спостерігається *кратна вибірка* X об'єму n, а спосіб, яким утворена оцінка, один і той самий (не залежить від *об'єму вибірки* n), то поняття "оцінка" використовують також у широкому розумінні як "послідовність оцінок", що утворені за одним правилом при різних значеннях об'єму вибірки. Послідовності оцінок позначаються через $\widehat{\theta}_n$.

Приклад. *Вибіркове середнє* є оцінкою, що дорівнює середньому арифметичному спостережень, які утворюють вибірку. Однак це є послідовність оцінок, оскільки при кожному значенні *об'єму вибірки* є окрема статистика.

3.1.5. Властивості оцінок

Надалі для двох випадкових величин ξ, η запис $\xi \simeq \eta$ означає, що ці величини мають однакові *функції розподілу*,.отже, і однакові породжені *міри Лебега – Стілтьєса*. Символом $N(\mu, \sigma^2)$ позначатимемо випадкову величину з *нормальним розподілом* та середнім μ і дисперсією σ^2.

Якість тієї чи іншої *оцінки* потребує порівняльного аналізу. Для порівняння оцінок чи їх послідовностей будемо використовувати такі поняття.

Означення. *Оцінка* $\widehat{\theta}$ *називається* незміщеною, *якщо її математичне сподівання збігається з точним значенням* θ :

$$\mathbf{E}_\theta\, \widehat{\theta} = \theta, \quad \forall \theta \in \Theta.$$

Означення. *Оцінка* $\widehat{\theta}_n$ *називається* асимптотично незміщеною, *якщо має місце асимптотична збіжність середніх*

$$\mathbf{E}_\theta\, \widehat{\theta}_n \to \theta,\ n \to \infty, \quad \forall \theta \in \Theta.$$

Означення. *Оцінка* $\widehat{\theta}_n$ *називається* консистентною *(або* слушною*), якщо вона збігається за ймовірністю до істинного значення* θ :

$$\widehat{\theta}_n \to^{P_\theta} \theta,\ n \to \infty, \quad \forall \theta \in \Theta,$$

тобто $\mathbf{P}_\theta(|\widehat{\theta}_n - \theta| \ge \varepsilon) \to 0, n \to \infty,\ \forall \varepsilon > 0,\ \forall \theta \in \Theta.$

Означення. *Оцінка* $\widehat{\theta}_n$ *називається* строго консистентною, *якщо вона збігається з імовірністю 1 до істинного значення* θ :

$$\widehat{\theta}_n \to^{P_\theta 1} \theta,\ n \to \infty,\quad \forall \theta \in \Theta,$$

тобто $\mathbf{P}_\theta(\exists \lim_{n\to\infty} \widehat{\theta}_n = \theta) = 1,\ \forall \theta \in \Theta.$

Зауваження. Наведені властивості стосуються оцінок $\widehat{\theta}$ для значення невідомого параметру θ. У випадку, коли цей параметр – векторний, доцільно розглядати також оцінки $\widehat{\tau}$ для значень деякої функції $\tau = \tau(\theta)$ від параметра θ. Сформульовані вище означення поширюються також і на дану схему, якщо замінити θ на $\tau(\theta)$, а $\widehat{\theta}$ на $\widehat{\tau}$.

Означення. *Інтервальна оцінка* $(\widehat{\theta}_1, \widehat{\theta}_2)$ *невідомого параметра* θ *є* незміщеною оцінкою надійності p, *якщо для всіх* $\theta \in \Theta$

$$\mathbf{P}_\theta(\widehat{\theta}_1 \le \theta \le \widehat{\theta}_2) = p.$$

Означення. *Оцінка* $\widehat{\theta}_n$ *параметра* θ *називається* асимптотично нормальною, *якщо знайдеться числова нормуюча послідовність* $c_n = c_n(\theta)$ *така, що має місце слабка збіжність*

$$c_n(\widehat{\theta}_n - \theta) \to^{W_\theta} \zeta \simeq N(0,1),\ n \to \infty,\ \forall \theta \in \Theta.$$

За теоремою *про еквівалентність слабкої збіжності та в основному* асимптотична нормальність еквівалентна збіжності

$$\mathbf{P}_\theta(c_n(\widehat{\theta}_n - \theta) < x) \to \Phi(x), n \to \infty,\ \forall x \in \mathbb{R}, \forall \theta \in \Theta,$$

де Φ – *стандартна нормальна функція розподілу.*

Зауваження (про побудову інтервальної оцінки). Якщо оцінка $\widehat{\theta}_n$ є *асимптотично нормальною*, а $c_n \sim \sqrt{n}/\sigma$, то за теоремою *про добуток слабко збіжної послідовності зі збіжною*

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to^{W} \sigma\zeta \equiv \eta \simeq N(0, \sigma^2),\ n \to \infty.$$

У цьому разі величина σ^2 називається асимптотичною дисперсією оцінки $\widehat{\theta}_n$. Істинне її значення $\sigma^2 = \sigma^2(\theta)$ є відомою функцією параметра θ. З наведеної слабкої збіжності випливає, що розподіл нормованої величини $\sqrt{n}(\widehat{\theta}_n - \theta)/\sigma(\theta)$ наближається до розподілу *стандартної нормальної* величини ζ. Оберемо для заданого рівня $p \in (0,1)$ значення x_p так, щоб

$$\mathbf{P}(|\zeta| \le x_p) = p.$$

Для знаходження x_p досить чисельно розв'язати рівняння

$$\Phi(x_p) = (1+p)/2.$$

Наприклад, $x_{0.997} \approx 3$ за *правилом трьох сигма.*

Тоді при великих n наближено

$$\mathbf{P}_\theta\left(\sqrt{n}|\widehat{\theta}_n - \theta|/\sigma(\theta) \le x_p\right) \approx \mathbf{P}(|\zeta| \le x_p) = p.$$

Припустимо, що функція $\sigma(\theta)$ неперервна, а оцінка $\widehat{\theta}_n$ – *консистентна.* З цих припущень випливає збіжність *за ймовірністю* $\sigma(\widehat{\theta}_n) \to^P \sigma(\theta)$. Тому можна наближено замінити $\sigma(\theta)$ на $\sigma(\widehat{\theta}_n)$ під знаком імовірності і стверджувати (внаслідок теореми *про добуток слабко збіжної послідовності зі збіжною*), що при великих n подія $\{\sqrt{n}|\widehat{\theta}_n - \theta|/\sigma(\widehat{\theta}_n) \le x_p\}$ наближено теж має ймовірність p. Отже, властивості *асимптотичної нормальності* та *консистентності* дають можливість наближеної побудови інтервальних асимптотично *незміщених оцінок надійності р* для невідомого параметра, що мають вигляд

$$\mathbf{P}_\theta\left(\widehat{\theta}_n - x_p\sigma(\widehat{\theta}_n)/\sqrt{n} \le \theta \le \widehat{\theta}_n + x_p\sigma(\widehat{\theta}_n)/\sqrt{n}\right) \approx p, \ \forall\theta \in \Theta.$$

Означення. *Оцінка* $\widehat{\theta}_n$ *називається* локально незміщеною, локально консистентною *тоді, якщо відповідна властивість виконується лише для значень параметра* $\theta \in \Theta$ *із деякого околу істинного значення.*

3.2. Оцінювання ймовірності успіху у схемі Бернуллі

3.2.1. Частота успіхів та її властивості

Розглянемо *статистичний простір* $(\Omega,\ \mathfrak{F},\ (\mathbf{P}_\theta,\ \theta \in \Theta))$, в якому

$$\Omega = \{\omega = (\omega_1, ..., \omega_n),\ \omega_k \in \{0,1\}\} = \{0,1\}^n,\ \mathfrak{F} = 2^\Omega,$$

$$\Theta = [0,1],\ \mathbf{P}_\theta(\{\omega\}) = \theta^{\nu_n(\omega)}(1-\theta)^{n-\nu_n(\omega)},$$

де функція

$$\nu_n(\omega) = \nu_n((\omega_1, ..., \omega_n)) = |\{k : \omega_k = 1\}| = \sum\nolimits_{k=1}^n \omega_k$$

задає кількість успіхів (одиниць) в елементарній події ω.

Даний простір відповідає експерименту, в якому проводяться n *випробувань Бернуллі* з невідомою ймовірністю $p = \theta$ успіху в окремому випробуванні.

Нехай спостерігаються результати всіх випробувань, тобто вимірний простір $(S,\ \Sigma) = (\Omega, \mathfrak{F})$ і *кратна вибірка* має вигляд $X = (\chi_1, ..., \chi_n)$, де $\chi_k(\omega) = 1\!\!\mathrm{I}_{\{\omega_k\, =\, 1\}} = \omega_k$ – *індикатор успіху* в k-му випробуванні.

Величини $(\chi_1, ..., \chi_n)$ *незалежні в сукупності* та *однаково розподілені*,

$$\mathbf{P}_\theta(\chi_1 = 1) = \theta,\ \mathbf{P}_\theta(\chi_1 = 0) = 1 - \theta,\ \ \mathbf{E}_\theta\ \chi_1 = \theta.$$

Розглянемо *оцінку*, що збігається з відносною частотою успіхів

$$\widehat{\theta}_n = \frac{\nu_n}{n} = \frac{1}{n}\sum\nolimits_{k=1}^n \chi_k.$$

Теорема (про властивості відносної частоти). *Відносна частота успіху у схемі випробувань Бернуллі:*

(1) має біноміальний розподіл:

$$\mathbf{P}_\theta(\widehat{\theta}_n = k/n) = C_n^k \theta^k (1-\theta)^{n-k},\ k = \overline{0, n},$$

(2) є незміщеною оцінкою ймовірності успіху θ,

(3) є консистентною і строго консистентною оцінкою,

(4) є асимптотично нормальною з асимптотичною дисперсію $\theta(1-\theta)$.

Доведення

(1) Перше твердження є наслідком означення біноміального розподілу, оскільки чисельник ν_n в означенні $\widehat{\theta}_n$ є кількістю успіхів у n випробуваннях Бернуллі з імовірністю успіху θ.

(2) Незміщеність виводиться з *лінійності* математичного сподівання та з формули для математичного сподівання *індикаторної величини*

$$\mathbf{E}_\theta \widehat{\theta}_n = \frac{1}{n}\sum\nolimits_{k=1}^n \mathbf{E}\chi_k = \frac{1}{n}\sum\nolimits_{k=1}^n \theta = \theta,\ \forall\theta \in \Theta.$$

(3) Зауважимо, що частота успіху

$$\nu_n = \sum\nolimits_{k=1}^n \chi_k$$

є сумою *незалежних у сукупності однаково розподілених* величин. За лінійністю математичного сподівання та теоремою *про дисперсію суми незалежних величин*

$$\mathbf{E}_\theta \nu_n = n\theta,\ \mathbf{D}_\theta \nu_n = n\theta(1-\theta).$$

Звідси $\mathbf{D}_\theta \widehat{\theta}_n = \mathbf{E}_\theta(\widehat{\theta}_n - \theta)^2 = n\theta(1-\theta)/n^2 \to 0,\ \ n \to \infty$. Тому має місце збіжність $\widehat{\theta}_n \to^{L_2} \theta$, отже, $\widehat{\theta}_n \to^{P_\theta} \theta$ за теоремою *про співвідношення між різними видами збіжності*. Більше того, за *критерієм Колмогорова посиленого закону великих чисел* $\widehat{\theta}_n \to^{P_\theta 1} \mathbf{E}_\theta\ \chi_{_1} = \theta$.

(4) На підставі зображення пункту (3) із *класичної центральної граничної теореми* виводимо, що

$$\sqrt{n}(\widehat{\theta}_n - \theta)/\sqrt{\theta(1-\theta)} =$$
$$(\nu_n - \mathbf{E}\nu_n)/\sqrt{\mathbf{D}\nu_n} \to^W \zeta \simeq N(0,1),\ n \to \infty,$$

звідки

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to^W \sqrt{\theta(1-\theta)}\zeta = \eta \simeq N(0,\ \theta(1-\theta)),\ n \to \infty.$$

Отже, оцінка $\widehat{\theta}_n$ є *асимптотично нормальною* з дисперсією $\theta(1-\theta)$ $\boxdot$

3.2.2. Інтервальні оцінки ймовірності успіху

Властивість асимптотичної нормальності відносної частоти $\widehat{\theta}_n$ можна використати для побудови *інтервальних оцінок* параметра θ. Нехай $p \in (0,1)$ – деякий вірогідний рівень (наприклад, $p = 0.99$). Визначимо значення x_p так, щоб $\mathbf{P}(|\zeta| \le x_p) = p$, як це зроблено в розділі про властивості оцінок у зауваженні про побудову інтервальних оцінок.

(а) *Спеціальний метод.*

З *асимптотичної нормальності* виводимо, що подія

$$\left\{\sqrt{n}\left|\widehat{\theta}_n - \theta\right| / \sqrt{\theta(1-\theta)} \le x_p\right\}$$

при досить великих n наближено має ймовірність p. Розв'язуючи нерівність $n(\widehat{\theta}_n - \theta)^2 \le \theta(1-\theta)\, x_p^2$ відносно θ, робимо висновок, що з наперед заданою ймовірністю p невідомий параметр θ належить *надійному інтервалу* з кінцями

$$\left(n\widehat{\theta}_n + x_p^2/2 \pm x_p\sqrt{n\widehat{\theta}_n(1-\widehat{\theta}_n) + x_p^2/4}\right) / \ (n + x_p^2),$$

який є наближено асимптотично *незміщеною оцінкою надійності р.*

(б) *Загальний наближений метод.*

Замінимо множник $\theta(1-\theta)$ у нерівності $\sqrt{n}\left|\widehat{\theta}_n - \theta\right| \le x_p\sqrt{\theta(1-\theta)}$, яка відбувається при великих n з імовірністю p, на величину $\widehat{\theta}_n(1-\widehat{\theta}_n)$. Ця заміна вносить похибку другого порядку малості, оскільки оцінка $\widehat{\theta}_n$ є *консистентною*, а замінюваний вираз впливає лише на асимптотичну дисперсію. Тому наближено отримуємо асимптотично *незміщену оцінку надійності р*

$$\mathbf{P}_\theta\left(\widehat{\theta}_n - x_p\sqrt{\widehat{\theta}_n(1-\widehat{\theta}_n)/n} \le \theta \le \widehat{\theta}_n + x_p\sqrt{\widehat{\theta}_n(1-\widehat{\theta}_n)/n}\right) \approx p,\ \forall\theta \in \Theta.$$

Вправи

(1) Нехай $\widehat{\theta}_n$ – довільна незміщена консистентна оцінка параметра θ, а подія A_n не залежить від $\widehat{\theta}_n$ і $\mathbf{P}_\theta(A_n) = 1/n$. Довести, що оцінка $\widetilde{\theta}_n = \widehat{\theta}_n \mathbb{1}_{\overline{A_n}} + n^2 \mathbb{1}_{A_n}$ є консистентною, однак зміщена та асимптотично зміщена.

(2) Випадкова величина ζ_θ має розподіл Пуассона. (а) Довести асимптотичну нормальність: $(\zeta_\theta - \theta)/\sqrt{\theta} \to^W \zeta \simeq N(0,1), \theta \to \infty$. (б) Для заданого $p \in (0,1)$ побудувати наближений надійний інтервал рівня p для параметра θ, який має вигляд $(\zeta_\theta + x_p^2/2 \pm x_p \left(\zeta_\theta + x_p^2/4\right)^{1/2})$.

(3) Спостерігається величина ν, що має біноміальний розподіл з параметром p та невідомою кількістю випробувань n. Знайти надійний інтервал для n.

(4) Спостерігається кількість успіхів ν_n у перших n випробуваннях Бернуллі з невідомою ймовірністю успіхів p. Знайти надійний інтервал для числа успіхів ν_m у наступних m випробуваннях.

(5) Спостерігаються n випробувань Бернуллі з імовірністю успіху θ та відносною частотою $\widehat{\theta}_n$. Тоді $\sqrt{n}(\widehat{\theta}_n - \theta)/(\widehat{\theta}_n(1 - \widehat{\theta}_n))^{1/2} \to^W \zeta \simeq N(0,1), n \to \infty$.

(6) Спостерігається дві незалежні послідовності випробувань Бернуллі з імовірностями успіхів θ_i, та з n_i випробуваннями, $i = 1, 2$. Нехай $\widehat{\theta}_{n_i}$ – відносні частоти успіхів. Тоді $\sqrt{n_1}(\widehat{\theta}_{n_1} - \widehat{\theta}_{n_2} - \theta_1 + \theta_2) \to^W \zeta \simeq N(0,\ (\theta_1 - \theta_1^2) + \rho(\theta_2 - \theta_2^2))$ при $n_i \to \infty$ так, що $n_1/n_2 \to \rho$. Побудувати надійний інтервал для $\theta_1 - \theta_2$.

3.3. Емпірична функція розподілу

Розглянемо *статистичний простір*, в якому спостерігається *кратна вибірка* $X = (\xi_1, ..., \xi_n)$ зі спостереженнями $\xi_k,\ k = \overline{1,n}$, що мають невідому теоретичну *функцію розподілу* $F(x) = \mathbf{P}(\xi_k < x)$. У цьому випадку параметром θ статистичного простору є функція розподілу F. Для відображення цієї обставини у наступних декількох розділах імовірність та математичне сподівання будемо позначати через $\mathbf{P}$ та $\mathbf{E}$.

Означення. Емпіричною функцією розподілу *називається така параметрична сім'я статистик:*

$$\widehat{F}_n(x) = \frac{1}{n} \sum\nolimits_{k=1}^{n} \mathbb{1}_{\{\xi_k < x\}}\ ,\ x \in \mathbb{R}.$$

Функція $\widehat{F}_n(x)$ будується лише за значеннями вибірки та заданим аргументом x, є кусково-сталою та має прирости величини $1/n$ у точках $\xi_k,\ k = \overline{1,n}$. При кожній фіксованій елементарній події ω вона є *дискретною функцією розподілу* як функція аргументу x.

3.3.1. Загальні властивості

Теорема (про властивості емпіричної функції розподілу). *Для кожного $x \in \mathbb{R}$ значення $\widehat{F}_n(x)$ емпіричної функції розподілу:*

(1) має біноміальний розподіл:

$$\mathbf{P}\left(\widehat{F}_n(x) = k/n\right) = C_n^k F^k(x)(1-F(x))^{n-k},\ k = \overline{0,n},$$

(2) є незміщеною оцінкою значення теоретичної функції розподілу:

$$\mathbf{E}\widehat{F}_n(x) = F(x),$$

(3) є строго консистентною оцінкою для цього значення:

$$\widehat{F}_n(x) \to^{P1} F(x),\ n \to \infty,$$

(4) є асимптотично нормальною оцінкою з асимптотичною дисперсією $F(x)(1-F(x)$:

$$\sqrt{n}(\widehat{F}_n(x) - F(x)) \to^{W} \eta \simeq N(0,\ F(x)(1-F(x))),\ n \to \infty.$$

Доведення. Зафіксуємо x. Розглянемо послідовність із n випробувань, в яких k-м успіхом називається подія $\{\xi_k < x\}$. Оскільки величини ξ_k *незалежні в сукупності* та *однаково розподілені*, то дана послідовність є схемою *випробувань Бернуллі*, причому ймовірність успіху не залежить від номера випробування та дорівнюватиме $\theta = \mathbf{P}(\xi_k < x) = F(x)$. За означенням величина $\widehat{F}_n(x)$ збігатиметься з *відносною частотою успіхів*. Тому всі вказані властивості емпіричної функції розподілу випливають із наведеної вище теореми *про властивості відносної частоти* – частотної оцінки ймовірності успіху у випробуваннях Бернуллі $\boxdot$

3.3.2. Рівномірні за x властивості

Емпірична функція розподілу $\widehat{F}_n(x)$ є *випадковим процесом*: вона одночасно є функцією елементарної події $\omega \in \Omega$ та аргумента $x \in \mathbb{R}$.

Теорема (теорема Глівенка – Кантеллі). *Емпірична функція розподілу є рівномірно строго консистентною оцінкою функції розподілу F :*

$$\sup_{x\in\mathbb{R}} \left|\widehat{F}_n(x) - F(x)\right| \to^{P1} 0,\ n \to \infty.$$

Доведення. Зафіксуємо $m \in \mathbb{N}$ і визначимо при $0 \le k < m$ величини

$$x_{mk} = \sup(x \in \mathbb{R} : F(x) < k/m),$$

де $\sup\varnothing \equiv -\infty$, та визначимо $x_m = \infty$. Ця послідовність не спадає за k та у припущенні невиродженості F містить хоча б один скінченний елемент. У випадку, коли $F(x) = 1\!\!\mathrm{I}_{c<x}$, твердження теореми очевидне.

Зауважимо, що $F(x_{mk}) = F(x_{mk}-0) \le k/m$ та $F(x_{mk}+0) \ge k/m$ за умови $x_{mk} > -\infty$, оскільки з $F(x_{mk}+0) < k/m$ випливало б існування $x > x_{mk}$ таких, що $F(x) < k/m$.

Визначимо випадкову величину

$$d_{mn} = \max{}_{0\le k\le m} |\widehat{F}_n(x_{mk}) - F(x_{mk})|.$$

За теоремою *про властивості емпіричної функції розподілу* має місце збіжність $\widehat{F}_n(x_{mk}) - F(x_{mk}) \to^{P1} 0$ при $n\to\infty$ для кожного k. Тому $d_{mn} \to^{P1} 0$ при $n\to\infty$ для кожного фіксованого m.

Для кожного x існує k таке, що $x \in (x_{mk}, x_{m,k+1}]$. Якщо $x_{mk} > -\infty$, то

$$\widehat{F}_n(x) - F(x) \le \widehat{F}_n(x_{m,k+1}) - F(x_{mk}+0) =$$

$$\widehat{F}_n(x_{m,k+1}) - F(x_{m,k+1}) + F(x_{m,k+1}) - F(x_{mk}+0) \le$$

$$d_{mn} + (k+1)/m - k/m = d_{mn} + 1/m.$$

У випадку $x_{mk} = -\infty$ за означенням $F(x) \ge k/m$ для всіх x, звідки отримуємо заміною $x_{mk}+0$ на x таку ж саму оцінку.

Аналогічно доводимо відповідну оцінку знизу. Зважаючи на довільність x, з отриманих нерівностей виводимо, що

$$\sup{}_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \le d_{mn} + 1/m.$$

Отже, для довільного $m \ge 1$ з імовірністю 1 виконується нерівність

$$\overline{\lim}_{n\to\infty} \sup{}_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \le \overline{\lim}_{n\to\infty} d_{mn} + 1/m = 1/m,$$

звідки отримуємо при $m\to\infty$ твердження теореми $\boxdot$

Теорема (теорема Колмогорова про відхилення емпіричної функції розподілу). *Якщо теоретична функція розподілу F неперервна,*

(а) то розподіл ***статистики Колмогорова****:*

$$\widehat{\varkappa}_n \equiv \sqrt{n}\ \sup{}_{x\in\mathbb{R}} \left|\widehat{F}_n(x) - F(x)\right|$$

не залежить від вигляду невідомої функції F,

(б) і має місце слабка збіжність

$$\widehat{\varkappa}_n \to^{W} \varkappa,\ n\to\infty,$$

причому гранична величина має ***функцію розподілу Колмогорова****:*

$$\mathbf{P}(\varkappa < x) = K(x) \equiv \sum\nolimits_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2x^2),\ x > 0.$$

Доведення

(а) Для спрощення припустимо, що функція F строго монотонна. Тоді коректно визначена обернена до F функція

$$F^{(-1)}(u) = \sup(x : F(x) < u),\ 0 < u < 1,$$

причому $F(F^{(-1)}(u)) = u$.

Зробимо заміну змінної $x = F^{(-1)}(u), 0 < u < 1$, в означенні *емпіричної функції розподілу*

$$\widehat{F}_n(F^{(-1)}(u)) = \frac{1}{n}\sum\nolimits_{k=1}^{n} 1\!\!\mathrm{I}_{\left\{\xi_k\, <\, F^{(-1)}(u)\right\}} = \frac{1}{n}\sum\nolimits_{k=1}^{n} 1\!\!\mathrm{I}_{\{F(\xi_k)\, <\, u\}},$$

де за монотонністю та означенням оберненої функції використано тотожність $\{\xi_k < F^{(-1)}(u)\} = \{F(\xi_k) < u\}$. Величини $\alpha_k \equiv F(\xi_k)$ незалежні за теоремою *про перетворення незалежних величин*, та *рівномірно розподілені* на $[0,1]$:

$$\mathbf{P}(F(\xi_k) < u) = \mathbf{P}(\xi_k < F^{(-1)}(u)) = F(F^{(-1)}(u)) = u,\ \forall u \in [0,1].$$

З означення отримуємо $\{F^{(-1)}(u), 0 < u < 1\} = \{x : 0 < F(x) < 1\}$, а при $F(x) \in \{0,1\}$ виводимо рівність $\widehat{F}_n(x) = F(x)$ м.н. Тому заміна $x = F^{(-1)}(u)$ в означенні *статистики Колмогорова* веде до рівності м.н.

$$\widehat{\varkappa}_n = \sqrt{n}\ \sup{}_{u\in(0,1)} |\widehat{F}_n(F^{(-1)}(u)) - F(F^{(-1)}(u))| =$$
$$\sqrt{n} \sup{}_{u\in(0,1)} |\tfrac{1}{n}\textstyle\sum_{k=1}^{n}\ 1\!\!\mathrm{I}_{\{\alpha_k\, <\, u\}} - u|.$$

Отже, за теоремою *про обчислення ймовірностей і математичного сподівання функції від випадкового вектора* розподіл величини $\widehat{\varkappa}_n$ як функції від вектора $(\alpha_1, ..., \alpha_n)$ з незалежних рівномірно розподілених величин визначається однозначно і не залежить від F.

(б) Обчислення граничного розподілу *статистики Колмогорова* спирається на граничні теореми теорії випадкових процесів і не наводиться $\boxdot$

Теорема Колмогорова дає можливість обчислити інтервальну асимптотично *незміщену оцінку надійності p* для функції розподілу

$$\mathbf{P}(|\widehat{F}_n(x) - x_p\ /\sqrt{n} \le F(x) \le \widehat{F}_n(x) + x_p\ /\sqrt{n},\ \forall x \in \mathbb{R}|) \approx p,$$

де x_p – розв'язок рівняння $K(x_p) = p$.

Для більш точного оцінювання замість K слід використати формули для функцій розподілу статистик $\widehat{\varkappa}_n$, що були знайдені В.С. Королюком.

Вправи

(1) Для нестрого монотонних F існує неперервний зліва варіант $F^{(-1)}$. Використати його для доведення теореми Колмогорова у загальному випадку.

(2) Довести, що $K(x) = 1 - 2\sum_{k=1}^{+\infty}(-1)^{k-1}\exp(-2k^2x^2)$.

(3) Знайти розподіл статистик $\widehat{\varkappa}_1, \widehat{\varkappa}_2$.

(4) Вивести з теореми про закон повторного логарифму, що $\mathbf{P}(\overline{\lim}_{n\to\infty}\sqrt{2n/\ln\ln n}\,\sup_{x\in\mathbb{R}}|\widehat{F}_n(x) - F(x)| = 1) = 1$.

(5) Нехай $\widehat{F}_n(x)$ – емпірична функція розподілу для незалежних спостережень з рівномірним на інтервалі $[0,1]$ розподілом, $\alpha, \beta > 0, \alpha + \beta > 1$. Розглянемо ймовірності $p_n(\alpha,\beta) \equiv \mathbf{P}(\widehat{F}_n(x) < \alpha + \beta x,\ \forall x \in [0,1])$. Довести, що: (а) $p_n(\alpha,\beta) = 1 - ((1-\alpha)/\beta)^n$ при $\alpha \in [1 - 1/n, 1]$, (б) при $\alpha < 1 - 1/n$ $p_{n+1}(\alpha,\beta) = \int_{(1-\alpha)/\beta}^{1} p_n\left(\frac{n+1}{n}\alpha, \frac{n+1}{n}\beta t\right)(n+1)t^n dt$. (в) Обчислити $p_n(\alpha,\beta)$.

3.4. Варіаційний ряд. Квантилі

У процесі побудови графіка *емпіричної функції розподілу* ми стикаємося з задачею впорядкування точок її стрибків – вибіркових значень.

Означення. Варіаційним рядом *вибірки* $X = (\xi_1, ..., \xi_n)$ *називаються вибіркові значення* $\left\{\xi_{(k)},\ k = \overline{1,n}\right\} = \{\xi_k,\ k = \overline{1,n}\}$, *що впорядковані за зростанням:* $\xi_{(1)} \le \xi_{(2)} \le ... \le \xi_{(n)}$.

Означення. *k-ою* порядковою статистикою *називається k-й елемент $\xi_{(k)}$ варіаційного ряду.*

Наприклад, варіаційним рядом вибірки $\{3,2,5\}$ є $\{2,3,5\}$.

Для спрощення формулювань визначимо також $\xi_{(0)} \equiv -\infty, \xi_{(n+1)} \equiv \infty$.

Теорема (про однозначність визначення порядкових статистик). *Нехай вибірка $X = (\xi_1, ..., \xi_n)$ складається з незалежних у сукупності однаково розподілених величин, що мають неперервну функцію розподілу F. Тоді з імовірністю 1 всі нерівності в означенні варіаційного ряду є строгими:*

$$\xi_{(1)} < \xi_{(2)} < ... < \xi_{(n)}.$$

Доведення. За *напівадитивністю* ймовірності

$$\mathbf{P}\left(\bigcup\nolimits_{i\neq j}\{\xi_{(i)} = \xi_{(j)}\}\right) = \mathbf{P}\left(\bigcup\nolimits_{i\neq j}\{\xi_i = \xi_j\}\right) \le \sum\nolimits_{i\neq j}\mathbf{P}(\xi_i = \xi_j) = 0,$$

оскільки при $h > 0$ та $i \neq j$

$$\mathbf{P}(\xi_i = \xi_j) \le \mathbf{P}([\xi_i/h] = [\xi_j/h]) = \sum\nolimits_{n\in\mathbb{Z}}\mathbf{P}([\xi_i/h] = [\xi_j/h] = n) =$$
$$\sum\nolimits_{n\in\mathbb{Z}}\mathbf{P}(\xi_i \in [nh, nh+h),\ \xi_j \in [nh, nh+h)) =$$
$$\sum\nolimits_{n\in\mathbb{Z}}(F(nh+h) - F(nh))^2 \le \sup\nolimits_{n\in\mathbb{Z}}(F(nh+h) - F(nh)) \to 0,\ h \to 0 \ \boxdot$$

У даному розділі функція розподілу F вважається неперервною.

3.4.1. Розподіл порядкових статистик

Лема (про емпіричну функцію розподілу та порядкові статистики). *Справедливі такі співвідношення:*

$$\widehat{F}_n(x) = \sum\nolimits_{k=0}^{n} \frac{k}{n} 1\!\!1_{\{\xi_{(k)} < x \le \xi_{(k+1)}\}}, \quad \{\xi_{(k)} < x\} = \{n\widehat{F}_n(x) \ge k\}, \ \forall k = \overline{1, n}.$$

Доведення обох тотожностей полягає в тому, що за означенням *порядкових статистик* включення $x \in (\xi_{(k)}, \xi_{(k+1)}]$ та $\xi_{(k)} \in (-\infty, x)$ мають місце тоді і тільки тоді, коли інтервал $(-\infty, x)$ містить точно k чи відповідно, не менше ніж k порядкових статистик. Це еквівалентно наявності в $(-\infty, x)$ такої самої кількості спостережень, що дорівнює $n\widehat{F}_n(x)$ $\boxdot$

Теорема (про функцію розподілу порядкових статистик). *Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з неперервною функцією розподілу F окремих спостережень. Тоді для $k = \overline{1, n}$ справедливі рівності*

$$\mathbf{P}\left(\xi_{(k)} < x\right) = \sum\nolimits_{i=k}^{n} C_n^i \, F^i(x)(1 - F(x))^{n-i}, \ \forall x \in \mathbb{R}.$$

Доведення. Визначимо цілозначну випадкову величину $\nu_n(x) = n\widehat{F}_n(x)$. Як показано в розділі про *емпіричну функцію розподілу*, ця величина дорівнює кількості успіхів у n *випробуваннях Бернуллі* з імовірністю успіху $F(x)$, якщо k-й успіх інтерпретувати як подію $\{\xi_k < x\}$. Тому наведений вираз випливає з леми *про емпіричну функцію розподілу та порядкові статистики*, і з формули для *біноміального розподілу* величини $\nu_n(x)$:

$$\mathbf{P}(\xi_{(k)} < x) = \mathbf{P}(\nu_n(x) \ge k) = \textstyle\sum_{i=k}^{n} \mathbf{P}(\nu_n(x) = i) =$$
$$\textstyle\sum_{i=k}^{n} C_n^i \, F^i(x)(1 - F(x))^{n-i} \ \boxdot$$

Теорема (про емпіричний розподіл порядкових статистик). *Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з неперервною функцією розподілу $F(x)$ окремих спостережень, а величина ξ не залежить від $(\xi_1, ..., \xi_n)$ і має таку саму функцію розподілу. Тоді*

(а) $\mathbf{P}(\xi_{(k)} < \xi \le \xi_{(k+1)}) = 1/(n+1), \ k = \overline{0, n},$

(б) $\mathbf{P}(\xi \le \xi_{(k)}) = k/(n+1), \ k = \overline{0, n},$ *де* $\xi_{(0)} \equiv -\infty, \ \xi_{(n+1)} \equiv \infty.$

Зауваження. Дану теорему можна інтерпретувати так: *варіаційний ряд* об'єму n розбиває числову вісь на $n+1$ інтервал, які є рівноймовірними для наступного незалежного спостереження.

Доведення

(а) За формулою про *ймовірність вкладеної різниці* подій та за теоремою *про функцію розподілу порядкових статистик* для кожного x

$$\mathbf{P}(\xi_{(k)} < x \le \xi_{(k+1)}) = \mathbf{P}(\{\xi_{(k)} < x\} \backslash \{\xi_{(k+1)} < x\}) =$$
$$\mathbf{P}(\xi_{(k)} < x) - \mathbf{P}(\xi_{(k+1)} < x) = C_n^k \, F^k(x)(1 - F(x))^{n-k}.$$

Оскільки випадкові величини у парах ξ і $\xi_{(k)}, \xi$ і $\xi_{(k+1)}$ незалежні за теоремою *про векторні перетворення незалежних величин*, то внаслідок теореми *про математичне сподівання функції від незалежних величин* та теореми *про емпіричний розподіл порядкових статистик*

$$\mathbf{P}\left(\xi_{(k)} < \xi \le \xi_{(k+1)}\right) = \mathbf{P}(\xi_{(k)} < \xi) - \mathbf{P}\left(\xi_{(k+1)} < \xi\right) =$$
$$\textstyle\int_{-\infty}^{\infty} \left(\mathbf{P}(\xi_{(k)} < x) - \mathbf{P}\left(\xi_{(k+1)} < x\right)\right) dF_\xi(x) =$$
$$\textstyle\int_{-\infty}^{\infty} \mathbf{P}\left(\xi_{(k)} < x \le \xi_{(k+1)}\right) dF_\xi(x) = \int_{-\infty}^{\infty} C_n^k \, F^k(x)(1 - F(x))^{n-k} dF(x) =$$
$$\textstyle\int_0^1 C_n^k \, u^k(1-u)^{n-k} du = C_n^k \, B(k, n-k) =$$
$$C_n^k \, \Gamma(k+1)\Gamma(n-k+1)/\Gamma(k+1+n-k+1) = 1/(n+1),$$

де використана заміна змінної $u = F(x)$ та означення повної бета-функції.

(б) Дана рівність випливає з попереднього твердження та з формули

$$\mathbf{P}(\xi \le \xi_{(k)}) = \sum\nolimits_{i=0}^{k-1} \mathbf{P}(\xi_{(i)} < \xi \le \xi_{(i+1)}) = k/(n+1) \; \boxdot$$

3.4.2. Вектор рангів, сумісний розподіл порядкових статистик

Означення. *Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з варіаційним рядом $(\xi_{(1)} ... \xi_{(n)})$.* Рангом спостереження *ν_k називається номер k-го спостереження ξ_k у складі варіаційного ряду: $\xi_{(\nu_k)} = \xi_k$.* Вектором рангів *називається випадковий вектор $\nu = (\nu_1, ..., \nu_n)$, що складений з рангів спостережень, та задовольняє умову:*

$$(\xi_{(\nu_1)}, ..., \xi_{(\nu_n)}) = (\xi_1, ..., \xi_n).$$

Іншими словами, величина ν_k збігається з місцем k-го спостереження ξ_k у вибірці за порядком зростання, де найменший є першим. **Наприклад**, для вибірки $(3, 9, 2, 5)$ величина $\nu_1 = 2$, оскільки 1-е спостереження $\xi_1 = 3$ знаходиться на 2-му місці у варіаційному ряді $(2, 3, 5, 9)$.

Зауваження. За умови неперервності функції розподілу F *вектор рангів* визначений однозначно з імовірністю 1, оскільки всі значення порядкових статистик є різними *майже напевне* за теоремою *про однозначність визначення порядкових статистик*. Надалі будемо припускати, що ця умова виконана.

Очевидно, що вектор рангів набуває значень у просторі Π_n *усіх перестановок розмірності* n. Позначимо через π^- обернену до π перестановку. Оскільки $\nu_{\nu_k^-} = k$, то з $\xi_k = \xi_{(\nu_k)}$ випливає $\xi_{\nu_k^-} = \xi_{(k)}$, тобто значення ν_k^- інтерпретується як номер k-ї порядкової статистики $\xi_{(k)}$ у вибірці X. Отже, за означенням *порядкових статистик* справедлива тотожність

$$\{\nu = \pi^-\} = \{\xi_{\pi_1} < \xi_{\pi_2} < ... < \xi_{\pi_n}\}.$$

Теорема (про розподіл вектора рангів). *Вектор рангів ν кратної вибірки $X = (\xi_1, ..., \xi_n)$ рівномірно розподілений на Π_n :*

$$\mathbf{P}(\nu = \pi) = 1/n!, \ \forall \pi \in \Pi_n.$$

Доведення. Нехай $\pi = (\pi_1, ..., \pi_n) \in \Pi_n$.

За теоремою *про обчислення ймовірностей, пов'язаних із випадковим вектором*, імовірності у правій частині рівності

$$\mathbf{P}(\nu = \pi^-) = \mathbf{P}(\xi_{\pi_1} < \xi_{\pi_2} < ... < \xi_{\pi_n}) = \mathbf{P}(\xi_1 < \xi_2 < ... < \xi_n),$$

не залежать від π, оскільки однозначно визначаються *сумісними функціями розподілу* випадкових векторів $(\xi_{\pi_1}, ..., \xi_{\pi_n})$ та $(\xi_1, ..., \xi_n)$, які внаслідок *незалежності* та *однакової розподіленості* ξ_k є однаковими і дорівнюють добуткові:

$$\mathbf{P}(\xi_{\pi_1} < x_1, ...\xi_{\pi_n} < x_n) = F(x_1)...F(x_n) = \mathbf{P}(\xi_1 < x_1, ...\xi_n < x_n).$$

Очевидно, що множина всіх обернених перестановок $\{\pi^-, \pi \in \Pi_n\}$ збігається з Π_n, а їх кількість дорівнює $n!$, тому

$$1 = \sum\nolimits_{\pi \in \Pi_n} \mathbf{P}(\nu = \pi^-) = n! \mathbf{P}(\nu = \pi) \ \boxdot$$

Теорема (про сумісний розподіл порядкових статистик). *Припустимо, що функція розподілу спостережень F має щільність f. Тоді сумісна щільність вектора варіаційного ряду $(\xi_{(1)}, ..., \xi_{(n)})$ дорівнює*

$$f_{(\xi_{(1)}, ..., \xi_{(n)})}(x_1, ..., x_n) = n! \ f(x_1)...f(x_n) \mathbb{1}_{x_1 < x_2 < ... < x_n}.$$

Доведення. За властивостями вектора рангів

$$\mathbf{P}(\xi_{(1)} < x_1, ..., \xi_{(n)} < x_n) = \sum\nolimits_{\pi \in \Pi_n} \mathbf{P}(\xi_{(1)} < x_1, ..., \xi_{(n)} < x_n, \ \nu = \pi^-) =$$

$$\sum\nolimits_{\pi \in \Pi_n} \mathbf{P}(\xi_{\pi_1} < x_1, ..., \xi_{\pi_n} < x_n, \ \xi_{\pi_1} < \xi_{\pi_2} < ... < \xi_{\pi_n}) =$$

$$n! \ \mathbf{P}(\xi_1 < x_1, ..., \xi_n < x_n, \ \xi_1 < \xi_2 < ... < \xi_n) =$$

$$\int .. \int_{y_1 < x_1, ..., y_n < x_n} n! \ f(y_1)...f(y_n) \mathbb{1}_{y_1 < y_2 < .. < y_n} dy_1 ... dy_n,$$

де передостання рівність випливає з того, що внаслідок незалежності та однакової розподіленості спостережень $(\xi_1, ..., \xi_n)$ випадкові вектори $(\xi_{\pi_1}, ..., \xi_{\pi_n})$ та $(\xi_1, ..., \xi_n)$ мають однакові *сумісні функції розподілу*, а остання рівність є наслідком теореми *про обчислення ймовірностей, пов'язаних із неперервним вектором*, оскільки сумісна щільність вектора $(\xi_1, ..., \xi_n)$ дорівнює добуткові $f(y_1)...f(y_n)$ за теоремою *про критерій незалежності абсолютно неперервних величин*. За означенням *сумісна щільність* вектора *порядкових статистик* дорівнює підінтегральній функції в правій частині останньої рівності, яка збігається зі вказаною в формулюванні теореми ⊡

Вправи

(1) Знайти: (а) сумісну щільність порядкових статистик $\xi_{(j)}, \xi_{(k)}, j < k$, (б) функцію розподілу, математичне сподівання і дисперсію розмаху $\xi_{(n)} - \xi_{(1)}$.

(2) Спостереження у кратній вибірці мають щільність f та функцію розподілу F. Тоді щільність статистики $\xi_{(k)}$ дорівнює $nC_{n-1}^{k-1}F^{k-1}(x)f(x)(1-F(x))^{n-k}$.

(3) Довести, що для кратної вибірки X з неперервною функцією розподілу *статистика Колмогорова* $\widehat{\varkappa}_n = \sqrt{n}\ \sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)|$ дорівнює
$\widehat{\varkappa}_n = \sqrt{n} \max_{1\le k\le n} \max\left(k/n - F(\zeta_{(k)}),\ F(\zeta_{(k)}) - (k-1)/n\right)$.

(4) Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з неперервною функцією розподілу F спостережень. Довести, що статистики $(F(\xi_{(k)})/F(\xi_{(k+1)}))^k,\ 1 \le k < n$, незалежні, та знайти їх розподіл.

(5) Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з експоненційного розподілу $Exp(\theta)$, а $(\xi_{(k)}, k = \overline{1,n})$ – її варіаційний ряд, $\xi_{(0)} = 0$. (а) Сумісна щільність вектора $(\xi_{(k)}, k = \overline{1,r})$ дорівнює $\theta^r C_n^r r! \exp\left(-\theta\left[\sum_{k=1}^r x_k + (n-r)x_r\right]\right)$. (б) Величина $2\theta\left[\sum_{k=1}^r \xi_{(k)} + (n-r)\xi_{(r)}\right]$ має хі-квадрат розподіл з $2r$ ступенями свободи. (в) Випадкові величини $\eta_k = (n-k+1)(\xi_{(k)} - \xi_{(k-1)}), k = \overline{1,r}$ незалежні та експоненційно розподілені з параметром θ. (г) У класі лінійних незміщених оцінок для θ^{-1} від $\xi_{(i)}$, що мають вигляд $\sum_{i=1}^k c_i\xi_{(i)}$, найменшу дисперсію має оцінка $k^{-1}\sum_{i=1}^k \xi_{(i)} + k^{-1}(n-k)\xi_{(k)}$, а відповідна дисперсія дорівнює $1/k\theta^2$.

(6) Нехай $(\xi_n, n \ge 1)$ незалежні однаково розподілені невід'ємні величини з неперервною функцією розподілу. Довести, що $\mathbf{P}(\xi_n > \max_{1\le k<n} \xi_k) = 1/n$.

(7) Нехай $(\xi_n, n \ge 1)$ незалежні однаково розподілені величини з неперервною функцією розподілу, а $\xi_{(n-r+1)}$ – порядкові статистики для n-кратної вибірки. Визначимо випадковий номер $\nu_{n,r} = \min(k \ge 1 : \xi_{n+k} \ge \xi_{(n-r+1)})$ при $r \ge 1$. Довести, що: (а) $\mathbf{P}(\nu_{n,1} > k) = n/(n+k)$, (б) $\mathbf{P}(\nu_{n,r} > k) = C_n^r/C_{n+k}^r$, (в) $\lim_{n\to\infty} \mathbf{P}(\nu_{n,r} > nx) = (1+x)^{-r},\ x \ge 0$.

(8) Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з рівномірним на $[0, 1]$ розподілом спостережень. (а) Знайти сумісну щільність, математичні сподівання та дисперсії статистик $\xi_{(1)}, \xi_{(n)}$. (б) Вивести незалежність та функції розподілу спейсингів $\delta_k \equiv \xi_{(k+1)} - \xi_{(k)},\ 0 \le k < n$, де $\xi_{(0)} = 0, \xi_{(n+1)} = 1$. (в) Обчислити функцію розподілу величини $\min_{0 \le k \le n} \delta_k$. (г) Довести, що $\lim_{n\to\infty} \mathbf{P}(n\xi_{(k)} < x) = \Gamma_{k-1,1}(x)$ для всіх k, де $\Gamma_{k\lambda}$ – функція гама-розподілу з параметрами k, λ.

(9) Нехай $(\xi_n, n \ge 1)$ незалежні рівномірно розподілені на інтервалі $[0, 1]$ величини. Визначимо випадковий номер $\nu = \inf(k \ge 1 : \xi_k < \xi_{k+1})$. Довести при $x \in [0, 1], k \ge 1$ тотожність $\mathbf{P}(\xi_1 < x, \nu = k) = x^k/k! - x^{k+1}/(k+1)!$.

3.4.3. Теоретичні квантилі

При побудові *надійних інтервалів* широко застосовуються такі числові характеристики функцій розподілу.

Означення. *Теоретичним* **квантилем** *рівня $p \in (0, 1)$ для величини ξ з функцією розподілу F називається число x_p, що є розв'язком рівняння*

$$F(x_p) = p = \mathbf{P}(\xi < x_p).$$

Якщо функція F неперервна і строго монотонно зростає на $\mathbb{R}$, то квантиль будь-якого рівня визначений однозначно. При порушенні строгої монотонності природно визначати x_p як середину відрізка, що утворений розв'язками наведеного рівняння.

Частковими випадками квантилей є поняття медіани та квартилей.

Означення. *Теоретичною* **медіаною**, нижнім та верхнім **квартилем** *називаються відповідно квантилі*

$$m \equiv x_{1/2},\ q_l \equiv x_{1/4},\ q_u \equiv x_{3/4}.$$

Приклад. Для *стандартного нормального розподілу*

$$(x_{1/4},\ x_{1/2},\ x_{3/4}) \approx (-0.674,\ 0,\ +0.674).$$

Вправи

(1) Довести, що теоретична медіана $m = x_{1/2}$ є абсолютним центром положення інтегровної випадкової величини ξ, тобто

$$m = \arg\min_{a \in \mathbb{R}} \mathbf{E}\,|\xi - a|\,.$$

(2) Випадкові величини ξ_n, ξ мають однозначно визначені медіани m_n, m. Довести, що з $\xi_n \to^W \xi, n \to \infty$, випливає збіжність $m_n \to m$. Навести приклад, коли таке твердження не виконується для математичних сподівань.

3.4.4. Емпіричні квантилі

Враховуючи теорему *про емпіричний розподіл порядкових статистик*, згідно з якою $\mathbf{P}(\xi < \xi_{(k)}) = k/(n+1) \approx p$ при $k \sim np$, приходимо до вибіркових аналогів квантилів.

Означення. *Нехай* $X = (\xi_1, ..., \xi_n)$ *– кратна вибірка, а* $(\xi_{(k)},\ k = \overline{1,n})$ *– її варіаційний ряд.* Емпіричним квантилем (вибірковим квантилем) *рівня* $p \in (0,1)$ *називається статистика*

$$\zeta_{np} \equiv \xi_{([np])}.$$

Вибірковою медіаною *називається статистика*

$$\widehat{m}_n \equiv \begin{cases} \xi_{([n/2]+1)}\ , & \text{якщо } n \text{ – непарне,} \\ (\xi_{(n/2)} + \xi_{(n/2+1)})/2, & \text{якщо } n \text{ – парне.} \end{cases}$$

Нижній та верхній вибірковий квартиль *дорівнюють відповідно*

$$\widehat{q}_l \equiv \xi_{([n/4+1/2])}, \quad \widehat{q}_u \equiv \xi_{([3n/4+1/2])}.$$

Розмахом вибірки *називається відстань* $\xi_{(n)} - \xi_{(1)}$ *між* екстремальними статистиками – *найбільшим* $\xi_{(n)}$ *та найменшим* $\xi_{(1)}$ *спостереженнями.*

Зауваження. Для скороченого опису вибірки $X = (\xi_1, ..., \xi_n)$ використовується поняття ящик із вусами (whisker-box). Це п'ятірка статистик

$$\xi_{(1)} \le \widehat{q}_l \le \widehat{m}_n \le \widehat{q}_u \le \xi_{(n)},$$

які зображують у вигляді горизонтально розташованого ящика шириною $\widehat{q}_u - \widehat{q}_l$ (міжквартильний розмах) із виділеним центром $\widehat{m}_n$ та вусами зліва й справа шириною $\widehat{q}_l - \xi_{(1)}$ та $\xi_{(n)} - \widehat{q}_u$ відповідно.

3.4.5. Консистентність вибіркових квантилей

Теорема (про строгу консистентність вибіркових квантилей). *Нехай функція розподілу* $F(x)$ *неперервна і строго монотонно зростає. Тоді вибірковий квантиль* $\zeta_{n\alpha} = \xi_{([n\alpha])}$ *є строго консистентною оцінкою для теоретичного квантиля* x_α*. Зокрема, це справедливо для вибіркової медіани та вибіркових квартилей.*

Доведення. Нехай, як і вище, $\nu_n(x) = n\widehat{F}_n(x)$ – кількість спостережень, менших за x. Тоді за лемою *про емпіричну функцію розподілу та порядкові статистики* справедливі тотожності:

$$\{\zeta_{n\alpha} < x\} = \{\xi_{([n\alpha])} < x\} = \{\nu_n(x) \ge [n\alpha]\} = \{\widehat{F}_n(x) \ge [n\alpha]/n\}.$$

Оскільки при кожному $\varepsilon > 0$ внаслідок зауваження *про верхню границю величин та подій* мають місце включення:

$$\{\underline{\lim}_{n\to\infty}\zeta_{n\alpha} < x_\alpha - \varepsilon\} \subset \overline{\lim}_{n\to\infty}\{\zeta_{n\alpha} < x_\alpha - \varepsilon\},$$
$$\overline{\lim}_{n\to\infty}\{\eta_n \ge \varepsilon\} \subset \{\overline{\lim}_{n\to\infty}\eta_n \ge \varepsilon\},$$

то за *монотонністю* ймовірності

$$\mathbf{P}\left(\{\underline{\lim}_{n\to\infty}\zeta_{n\alpha} < x_\alpha - \varepsilon\}\right) \le \mathbf{P}\left(\overline{\lim}_{n\to\infty}\ \{\zeta_{n\alpha} < x_\alpha - \varepsilon\}\right) =$$
$$\mathbf{P}\left(\overline{\lim_{n\to\infty}}\left\{\widehat{F}_n(x_\alpha - \varepsilon) \ge [n\alpha]/n\right\}\right) = \mathbf{P}\left(\overline{\lim_{n\to\infty}}\left\{\widehat{F}_n(x_\alpha - \varepsilon)\ n/[n\alpha] \ge 1\right\}\right) \le$$
$$\mathbf{P}\left(\left\{\overline{\lim}_{n\to\infty}\widehat{F}_n(x_\alpha - \varepsilon)\ n/[n\alpha] \ge 1\right\}\right) = \mathbf{P}\left(F(x_\alpha - \varepsilon)/\alpha \ge 1\right) = 0.$$

Тут використана збіжність $\widehat{F}_n(x_\alpha - \varepsilon) \to F(x_\alpha - \varepsilon)$ *майже напевне* за властивістю *строгої консистентності емпіричної функції розподілу*, і збіжність $[n\alpha]/n \to \alpha$, та врахована нерівність $F(x_\alpha - \varepsilon) < \alpha = F(x_\alpha)$ за умови строгої монотонності F. Отже, $\underline{\lim}_{n\to\infty}\zeta_{n\alpha} \ge x_\alpha - \varepsilon$ майже напевне. З довільності $\varepsilon > 0$ звідси виводимо, що $\underline{\lim}_{n\to\infty}\ \zeta_{n\alpha} \ge x_\alpha$ майже напевне. Аналогічно доводиться нерівність $\overline{\lim}_{n\to\infty}\ \zeta_{n\alpha} \le x_\alpha$ майже напевне $\boxdot$

3.4.6. Асимптотична нормальність вибіркових квантилей

Теорема (про асимптотичну нормальність вибіркових квантилей). *Нехай функція розподілу $F(x)$ має щільність $f(x)$ та квантиль x_α рівня $\alpha \in (0,1)$, а функція f неперервна і додатна в точці x_α. Тоді вибірковий квантиль $\zeta_{n\alpha} \equiv \xi_{([n\alpha])}$ є асимптотично нормальною оцінкою для x_α :*

$$\sqrt{n}(\zeta_{n\alpha} - x_\alpha) \to^W\ \eta \simeq N(0,\sigma^2),\ n \to \infty,$$

де асимптотична дисперсія $\sigma^2 = \alpha(1-\alpha)\ /\ f^2(x_\alpha)$.

Доведення. Позначимо $\beta_n = \sqrt{n}(\zeta_{n\alpha} - x_\alpha)\ /\ \sigma$.

Твердження теореми випливатиме з теореми *про добуток слабко збіжної послідовності зі збіжною*, якщо довести, що

$$\beta_n \to^W \beta \simeq N(0,1).$$

Нехай, як і вище,

$$\nu_n(x) = n\widehat{F}_n(x) = \textstyle\sum_{k=1}^n \mathrm{1\!I}_{\{\xi_k < x\}}.$$

За лемою *про емпіричну функцію розподілу та порядкові статистики* має місце тотожність $\{\xi_{(k)} < x\} = \{\nu_n(x) \ge k\}$. Тому

$$\mathbf{P}(\beta_n < y) = \mathbf{P}(\zeta_{n\alpha} < x_\alpha + y\sigma/\sqrt{n}) =$$

$$\mathbf{P}(\xi_{([n\alpha])} < x_\alpha + y\sigma/\sqrt{n}) = \mathbf{P}(\xi_{([n\alpha])} < z_n) = \mathbf{P}(\nu_n(z_n) \ge [n\alpha]),$$
$$\text{де} \quad z_n \equiv x_\alpha + y\sigma/\sqrt{n} = x_\alpha + o(1), n \to \infty.$$

Послідовність $\nu_n(z_n)$ є сумою *незалежних у сукупності*, *однаково розподілених* індикаторних випадкових величин $\chi_{nk} \equiv 1\!\!1_{\{\xi_k < z_n\}}$, що утворюють *загальну послідовність серій* випадкових величин, однаково розподілених у кожній серії та обмежених. Тому вони задовольняють умови *центральної граничної теореми Ляпунова для загальних серій* згідно з її наслідком. Отже, має місце *асимптотична нормальність*

$$\gamma_n \equiv (\nu_n(z_n) - m_n)/s_n \to^W \beta \simeq N(0,1),$$

де $m_n \equiv \mathbf{E}\nu_n(z_n) = nF(z_n), \ \ s_n^2 \equiv \mathbf{D}\nu_n(z_n) = nF(z_n)(1 - F(z_n))$.

З формули Тейлора та диференційовності F у точці x_α отримуємо

$$F(z_n) = F(x_\alpha) + f(x_\alpha)\ y\sigma/\sqrt{n} + o(1/\sqrt{n}) =$$
$$= \alpha + y\sqrt{\alpha(1-\alpha)/n} + o(1/\sqrt{n}) \to \alpha,\ n \to \infty,$$

звідки виводимо асимптотичні зображення:

$$m_n = n\alpha + y\sqrt{n\alpha(1-\alpha)} + o(\sqrt{n}), \quad s_n^2 = n\alpha(1-\alpha) + o(n),\ n \to \infty.$$

Зокрема, має місце збіжність $y_n \equiv ([n\alpha] - m_n)/s_n \to -y,\ n \to \infty$.

З наведених зображень виводимо граничне співвідношення

$$\mathbf{P}(\beta_n < y) = \mathbf{P}(\nu_n(z_n) \ge [n\alpha]) = \mathbf{P}(\gamma_n \ge ([n\alpha] - m_n)/s_n) =$$
$$\mathbf{P}(\gamma_n - y_n \ge 0) \to_{n\to\infty} \mathbf{P}(\beta + y \ge 0) = \mathbf{P}(\beta \ge -y) = \mathbf{P}(\beta < y).$$

Остання рівність випливає з симетрії нормального розподілу. Отже, має місце збіжність *в основному* $\beta_n \to^O \beta$, і за теоремою *про еквівалентність слабкої збіжності та в основному* $\beta_n \to^W \beta$, що доводить теорему ⊡

Вправи

(1) Нехай $(\xi_1, ..., \xi_n)$ – кратна вибірка з неперервною функцією розподілу F. Знайти граничний розподіл при $n \to \infty$ величин $n\sqrt{F(\xi_{(1)})(1 - F(\xi_{(n)}))}$.

(2) Нехай $(\xi_1, ..., \xi_n)$ – кратна вибірка з рівномірним на $[0,1]$ розподілом. Довести асимптотичну нормальність:,$\mathbf{P}(\widehat{m}_n - 1/2 < x/\sqrt{8n}) \to \Phi(x), n \to \infty$.

(3) Нехай $(\xi_1, ..., \xi_n)$ – кратна вибірка з нормальним розподілом спостережень $N(\theta, 1)$. Знайти граничний розподіл величин $\sqrt{n}(\widehat{m}_n - \theta), n \to \infty$, для вибіркової медіани $\widehat{m}_n$. Як співвідносяться асимптотичні ефективності медіани та вибіркового середнього при оцінювання θ ?

3.5. Вибіркові моменти. Метод моментів

Велику групу *статистик* утворюють вибіркові моменти, які за *законом великих чисел* є природними оцінками для теоретичних моментів – математичних сподівань степеневих функцій від випадкової величини.

3.5.1. Теоретичні та вибіркові моменти

Означення. *Нехай ξ – випадкова величина. Її (нецентральним)* **теоретичним моментом** *порядку $k \in \mathbb{N}$ називається число*

$$\mu_k \equiv \mathbf{E}\xi^k,$$

за умови інтегрованості величини ξ^k.

Центральним теоретичним моментом *порядку k називається число*

$$\mu_k^0 \equiv \mathbf{E}(\xi - \mu)^k,$$

де $\mu \equiv \mu_1$ – математичне сподівання, $\mu_2^0 = \sigma^2$ – дисперсія ξ.

Зауваження. Значення центральних моментів використовуються в теорії розподілів для означення таких спеціальних характеристик:

$k_v = \sigma \ / \ \mu_1$ – **коефіцієнт варіації** (дорівнює 1 для показникового розподілу),

$k_s = 3(\mu_1 - x_{1/2})/\sigma$– **коефіцієнт скошеності** (нульовий для симетричних розподілів),

$k_a = \mu_3^0 \ /\sigma^3$ – **коефіцієнт асиметрії** (аналогічно),

$k_e = \mu_4^0 \ /\sigma^4 - 3$ – **коефіцієнт ексцесу** (нульовий для нормальних спостережень).

Всі наведені характеристики є безрозмірними та відображають певні особливості форми відповідного розподілу.

Означення. *Нехай $X = (\xi_1, ..., \xi_n)$ – вибірка. Її (нецентральним)* **вибірковим моментом** *порядку k називається статистика*

$$\widehat{\mu}_{kn} \equiv \frac{1}{n} \sum\nolimits_{i=1}^{n} \xi_i^k.$$

Центральним вибірковим моментом *порядку k називається статистика*

$$\widehat{\mu}_{kn}^0 \equiv \frac{1}{n} \sum\nolimits_{i=1}^{n} (\xi_i - \widehat{\mu}_n)^k \,,$$

де $\widehat{\mu}_n \equiv \widehat{\mu}_{1n}$ – **вибіркове середнє**, *а,* $\widehat{\mu}_{2n}^0 = \widehat{\sigma}_n^2$ – **вибіркова дисперсія**.

Зауваження. Важливою властивістю центральних моментів є **інваріантність відносно зсувів** – вони не змінюються при одночасному зсуві всіх спостережень на сталу:

$$\widehat{\mu}_{kn}^{0} = \frac{1}{n}\sum\nolimits_{i=1}^{n}(\xi_i - \frac{1}{n}\sum\nolimits_{j=1}^{n}\xi_j)^k = \frac{1}{n}\sum\nolimits_{i=1}^{n}(\xi_i - c - \frac{1}{n}\sum\nolimits_{j=1}^{n}(\xi_j - c))^k.$$

Зауваження. Як і в теоремі *про властивості дисперсії*, для *вибіркової дисперсії* має місце тотожність

$$\widehat{\sigma}_n^2 = \widehat{\mu}_{2,n} - (\widehat{\mu}_n)^2.$$

Дійсно, $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \widehat{\mu}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}\xi_i^2 - 2(\widehat{\mu}_n)^2 + (\widehat{\mu}_n)^2$.

Теорема (про моменти вибіркових моментів). *Нехай* $(\xi_1, ..., \xi_n)$ – *кратна вибірка. Тоді*

$$\mathbf{E}\widehat{\mu}_{kn} = \mu_k, \ \mathbf{D}\widehat{\mu}_{kn} = \frac{1}{n}(\mu_{2k} - \mu_k^2).$$

Зокрема, $\mathbf{E}\widehat{\mu}_n = \mu$, $\mathbf{D}\widehat{\mu}_n = \sigma^2 / n$. *Крім того,*

$$\mathbf{E}\widehat{\sigma}_n^2 = \frac{n-1}{n}\sigma^2, \quad \mathbf{D}\widehat{\sigma}_n^2 = \frac{n-1}{n^3}((n-1)\mu_4^0 - (n-3)\sigma^4).$$

Доведення. *Незміщеність* нецентральних моментів є очевидним наслідком *лінійності* математичного сподівання. Вираз для їх *дисперсій* випливає з *незалежності в сукупності* і *однакової розподіленості* степеневих функцій $(\xi_i^k, i = \overline{1,n})$ та з теореми *про дисперсію суми незалежних величин*:

$$\mathbf{D}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n}\xi_i^k\right) = \frac{1}{n^2}\sum\nolimits_{i=1}^{n}\mathbf{D}(\xi_i^k) = \frac{1}{n^2}n\mathbf{D}(\xi_1^k) = \frac{1}{n}(\mathbf{E}\xi_1^{2k} - (\mathbf{E}\xi_1^k)^2).$$

Середнє для *вибіркової дисперсії* обчислюється з урахуванням останнього зауваження шляхом піднесення до квадрату під знаком суми:

$$\mathbf{E}\widehat{\sigma}_n^2 = \mathbf{E}(\widehat{\mu}_{n2} - \widehat{\mu}_n^2) = \mathbf{E}\widehat{\mu}_{n2} - \mathbf{E}\,\widehat{\mu}_n^2 =$$
$$\mu_2 - \left(\mathbf{E}\sum\nolimits_{i=1}^{n}\xi_i^2 + \mathbf{E}\sum\nolimits_{i\neq j}\xi_i\xi_j\right) / n^2 =$$
$$\mu_2 - \mu_2/n - \mu^2(n-1)/n = (\mu_2 - \mu^2)(n-1)/n = \sigma^2(n-1)/n.$$

При обчисленні дисперсії $\widehat{\sigma}_n^2$ можна вважати, що $\mu = 0$, оскільки оцінка $\widehat{\sigma}_n^2$ *інваріантна відносно зсувів*, зокрема, не змінюється при відніманні від спостережень їх спільного середнього μ. Тому:

$$n^4\mathbf{E}\left(\widehat{\sigma}_n^2\right)^2 = \mathbf{E}\left(n\sum\nolimits_k \xi_k^2 - (\sum\nolimits_k \xi_k)^2\right)^2 = \mathbf{E}\left((n-1)\sum\nolimits_k \xi_k^2 - 2\sum\nolimits_{i<j}\xi_i\xi_j\right)^2 =$$
$$(n-1)^2\mathbf{E}\left(\sum\nolimits_k \xi_k^2\right)^2 + 4\mathbf{E}(\sum\nolimits_{i<j}\xi_i\xi_j)^2 - 4(n-1)\mathbf{E}(\sum\nolimits_k \xi_k^2\ \sum\nolimits_{i<j}\xi_i\xi_j) =$$

$$(n-1)^2\left(n\,\mu_4^0+n(n-1)\sigma^4\right)+2n(n-1)\sigma^4-0=$$
$$(n-1)\left[(n-1)\mu_4^0+(n^2-2n+3)\sigma^4\right],$$
$$\mathbf{D}\widehat{\sigma}_n^2=\mathbf{E}\left(\widehat{\sigma}_n^2\right)^2-\left(\mathbf{E}\widehat{\sigma}_n^2\right)^2=\mathbf{E}\left(\widehat{\sigma}_n^2\right)^2-\sigma^4(n-1)^2/n^2=$$
$$(n-1)\left((n-1)\mu_4^0-(n-3)\sigma^4\right)/\;n^3,$$

де використане припущення $\mathbf{E}\xi_j=0$ ⊡

З теореми *про моменти вибіркових моментів* випливає, що *вибіркова дисперсія* $\widehat{\sigma}_n^2$ є зміщеною оцінкою для теоретичної дисперсії σ^2. Тому часто використовують її скоригований *незміщений* варіант.

Означення. Нормованою вибірковою дисперсією є *статистика*

$$\widehat{s}_n^2\equiv\frac{1}{n-1}\sum\nolimits_{i=1}^{n}(\xi_i-\widehat{\mu}_n)^2=\frac{n}{n-1}\widehat{\sigma}_n^2\,,$$

що є незміщеною оцінкою дисперсії: $\mathbf{E}\widehat{s}_n^2=\frac{n}{n-1}\mathbf{E}\widehat{\sigma}_n^2=\sigma^2$ ⊡

Теорема (про асимптотичні властивості вибіркових моментів). *Нехай* $X=(\xi_1,...,\xi_n)$ *– кратна вибірка.*

(а) Якщо $\mathbf{E}\left|\xi_1\right|^k<\infty$, *то нецентральний вибірковий момент* $\widehat{\mu}_{kn}$ *є незміщеною та строго консистентною оцінкою теоретичного моменту* μ_k.

(б) Якщо $\mathbf{E}\xi_1^{2k}<\infty$, *то оцінки* $\widehat{\mu}_{kn}$ *є асимптотично нормальними:*

$$\sqrt{n}(\,\widehat{\mu}_{kn}-\mu_k)\to^{W}\eta\simeq N(0,\;\mu_{2k}-\mu_k^2).$$

Доведення. Випадкові величини $\zeta_i\equiv\xi_i^k$ є незалежними за теоремою *про перетворення незалежних величин* та *однаково розподіленими* за теоремою *про збереження однакової розподіленості*, оскільки отримані застосуванням однієї (степеневої порядку k) функції до величин з однаковими розподілами. Умова (а) гарантує справедливість *критерію Колмогорова посиленого закону великих чисел*, що і доводить (а).

Для доведення (б) застосуємо *класичну центральну граничну теорему*. Скінченність математичних сподівань та дисперсій, а також формула для дисперсії доданка наведені в теоремі *про моменти вибіркових моментів*. Тому *асимптотична нормальність* є наслідком вказаної теореми ⊡

Вибіркові моменти можна розглядати одночасно у вигляді випадкового вектора вибіркових моментів наперед заданої розмірності d:

$$\widehat{\mu}_n^{(d)}\equiv(\widehat{\mu}_{kn},\;k=\overline{1,d}),\qquad\mathbf{E}\widehat{\mu}_n^{(d)}=\mu^{(d)}\equiv(\mu_k,\;k=\overline{1,d}).$$

Теорема (про асимптотику вектора вибіркових моментів). *Нехай* $X=(\xi_1,...,\xi_n)$ *– кратна вибірка.*

(а) Якщо $\mathbf{E}\left|\xi_1\right|^d<\infty$, *то векторна статистика* $\widehat{\mu}_n^{(d)}$ *є строго консистентною оцінкою для вектора* $\mu^{(d)}$.

(б) Якщо $\mathbf{E}\xi_1^{2d} < \infty$, *то вектор вибіркових моментів є асимптотично нормальним з коваріаційною матрицею* $V^{(d)} = (\mu_{k+l} - \mu_k\,\mu_l,\ k,l = \overline{1,d})$:

$$\sqrt{n}\left(\widehat{\mu}_n^{(d)} - \mu^{(d)}\right) \to^W \eta \simeq N_d(0,\ V^{(d)}),\ n \to \infty,$$

Доведення

(а) З теореми *про асимптотичні властивості вибіркових моментів* виводимо збіжність $\widehat{\mu}_{kn} \to^{P1} \mu_k, n \to \infty$, для всіх $k = \overline{1,d},\ \theta \in \Theta$. Тому

$$\mathbf{P}\left(\lim \widehat{\mu}_n^{(d)} = \mu^{(d)}(\theta)\right) = \mathbf{P}\left(\cap_{k=1}^d \{\lim \widehat{\mu}_{kn} = \mu_k\}\right) = 1.$$

(б) Розглянемо послідовність випадкових векторів

$$\gamma_j \equiv \left(\xi_j^k - \mu_k,\ k = \overline{1,d}\right),\quad j \ge 1.$$

Вони є незалежними за теоремою *про перетворення незалежних величин*, та *однаково розподіленими*, оскільки отримані з однаково розподілених величин застосуванням однієї функції. Дана послідовність задовольняє умови *класичної центральної граничної теореми для випадкових векторів* із параметрами $m = 0$, $V = V^{(d)}$, тому що

$$\mathbf{E}\gamma_j = 0,\quad Cov(\gamma_j) = (\mathbf{E}(\xi_j^k - \mu_k)(\xi_j^l - \mu_l), k,l = \overline{1,d}) =$$
$$((\mu_{k+l} - \mu_k\mu_l),\ k,l = \overline{1,d}) = V^{(d)}.$$

Отже, зі вказаної теореми отримуємо при $n \to \infty$

$$\sqrt{n}(\widehat{\mu}_n^{(d)} - \mu^{(d)}) = \frac{1}{\sqrt{n}}\sum_{j=1}^n \gamma_j = \eta_n \to^W \eta \simeq N_d(0,\ V^{(d)})\ \boxdot$$

Вправи

(1) Довести, що вибіркові моменти та емпірична функція розподілу $\widehat{F}_n$ пов'язані такими співвідношеннями:

$\widehat{\mu}_{kn} = \int_{-\infty}^{\infty} x^k d\widehat{F}_n(x),\quad \widehat{\mu}_{kn}^0 = \int_{-\infty}^{\infty} (x - \widehat{\mu}_n)^k\, d\widehat{F}_n(x).$

(2) Визначити вибіркові аналоги коефіцієнтів варіації, асиметрії, скошеності та ексцесу як відповідних функцій від вибіркових моментів.

(3) Обчислити коефіцієнти асиметрії та ексцесу для стандартних дискретних та абсолютно неперервних розподілів.

(4) Для стандартної нормальної випадкової величини $\xi \simeq N(0,1)$ довести інтегруванням за частинами тотожність $\mu_k = (k-1)\mu_{k-2}$ та рівності

$\mu_{2k} = 2^k\pi^{-1/2}\Gamma(k + 1/2) = 1 \cdot 3 \cdot ...(2k-1) \equiv (2k-1)!!.$

(5) Для $\xi \simeq \Gamma(\lambda, \alpha)$ обчислити $\mu_k = \lambda^{-k}\alpha(\alpha+1)...(\alpha+k-1)$.

(6) Довести асимптотичне зображення для дисперсії

$\mathbf{D}(\widehat{\mu}_{kn}^0) = (\mu_{2k} - \mu_k^2 + k\mu_{k-1}(k\mu_{k-1}\mu_2 - 2\mu_{k+1}))/n\ +\ O(1/n^2),\ n \to \infty.$

(7) Довести, що вибіркове середнє для кратної вибірки зі щільністю Коші $1/\pi(1+(x-\theta)^2)$ не є консистентною оцінкою центра симетрії θ, оскільки розподіл цієї статистики не залежить від n.

(8) Нехай $(\xi_1, ..., \xi_n)$ – кратна вибірка з неперервною функцією розподілу F, варіаційним рядом $(\xi_{(1)}, ..., \xi_{(n)})$, а $\alpha \in (0, 1/2)$. Для усунення впливу можливих випадкових збурень із занадто великими чи малими значеннями використовують робастний (стійкий до збурень) аналог вибіркового середнього, що має вигляд $\widehat{\mu}_{n\alpha} \equiv (n(1-2\alpha))^{-1} \sum_{k=[n\alpha]+1}^{[n-n\alpha]} \xi_{(k)}$. Довести (а) рівність $\widehat{\mu}_{n,0} = \widehat{\mu}_n$, (б) збіжність $\widehat{\mu}_{n\alpha} \to^{P1} \int_{x_\alpha}^{x_{1-\alpha}} y dF(y), n \to \infty$, де x_p – квантиль рівня p для F.

(9) Нехай $(\xi_1, ..., \xi_n)$ – кратна вибірка з показниковим розподілом $Exp(\theta)$. Знайти оцінку методу моментів для через (а) перший вибірковий момент, (б) другий вибірковий момент, (в) перші два вибіркові моменти. (г) Знайти таку оцінку для ймовірності $\mathbf{P}_\theta(\xi_1 \geq 1)$.

(10) Довести такі співвідношення між центральними та нецентральними моментами: $\mu_2^0 = \mu_2 - \mu_1^2$, $\mu_3^0 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^2$, $\mu_n^0 = \sum_{k=0}^n C_n^k \mu_{n-k}(-\mu_1)^k$.

(11) Для покращення властивостей оцінок застосовують метод розщеплень (bootstrap). Для довільних оцінок $T_n = T_n(\xi_1, ..., \xi_n)$ обирають оцінку вигляду $T_n^* = n^{-1} \sum_{k=1}^n T_{n-1}(\xi_1, ...\xi_{k-1}, \xi_{k+1}, ..., \xi_n)$. Довести, що застосування методу розщеплень до статистик $\widehat{\sigma}_n^2$ при $n > 2$ призводить до статистик $\widehat{s}_n^2$.

3.5.2. Метод моментів

Метод моментів є спеціальним методом оцінювання невідомих параметрів, який спирається на асимптотичні властивості *вибіркових моментів*.

Припустимо, що параметричний простір є d-вимірним: $\Theta \subset \mathbb{R}^d$.

Оскільки розподіл вибірки $X = (\xi_1, ..., \xi_n)$ відомий повністю при заданому значенні $\theta \in \Theta$, то повністю відомими є функції

$$\mu_k(\theta) \equiv \mathbf{E}_\theta\, \xi_1^k, \ \theta \in \Theta.$$

Розглянемо векторну функцію

$$\mu^{(d)}(\theta) \equiv (\mu_k(\theta), \ k = \overline{1, d}) : \quad \Theta \to \mathbb{R}^d.$$

Припустимо, що існує неперервне відображення $T_d(\mu) : \mathbb{R}^d \to \Theta$, яке є оберненим до $\mu^{(d)}(\theta)$, тобто

$$T_d(\mu^{(d)}(\theta)) = \theta, \ \forall \theta \in \Theta.$$

Ця умова виконується, зокрема, за теоремою про обернене відображення з курсу математичного аналізу, якщо функція $\mu^{(d)}(\theta)$ неперервно диференційовна, якобіан $\det\left|\frac{d}{d\theta}\mu^{(d)}(\theta_0)\right| \neq 0$ для деякого $\theta_0 \in \Theta$ і простір Θ звужено до деякого околу точки θ_0.

Означення. Оцінкою методу моментів *параметра θ називаєтьcя така статистика від вектора вибіркових моментів $\widehat{\mu}_n^{(d)} = (\widehat{\mu}_{nk},\ k = \overline{1,d})$, що містить значення перших d вибіркових моментів:*

$$\widehat{\theta}_n \equiv T_d(\widehat{\mu}_n^{(d)}),$$

де $T_d : \mathbb{R}^d \to \Theta$ – обернена функція до вектора моментів $\mu^{(d)}(\theta)$.

Зауваження. З означення оберненої функції випливає, що оцінка методу моментів $\widehat{\theta}_n$ є єдиним розв'язком системи **рівнянь методу моментів**

$$\mu^{(d)}(\widehat{\theta}_n) = \widehat{\mu}_n^{(d)}.$$

Зауваження. У деяких випадках кількість дійсно залежних від θ координат вектора $\mu^{(d)}(\theta)$ може бути меншою за d, тому відображення $T_d(\mu)$ не існує – наприклад, при рівномірному на $[-\theta, \theta]$ розподілу спостережень та $d = 1$. У цьому разі необхідно збільшити розмірність d вектора $\mu^{(d)}(\theta)$.

Теорема (про строгу консистентність оцінок методу моментів). *Якщо $X = (\xi_1, ..., \xi_n)$ – кратна вибірка, $\mathbf{E}_\theta \left|\xi_1\right|^d < \infty$, і функція T_d неперервна, то оцінка методу моментів є строго консистентною оцінкою параметра θ.*

Доведення. За теоремою *про асимптотику вектора вибіркових моментів* $\widehat{\mu}_n^{(d)} \to^{P_\theta 1} \mu^{(d)}(\theta), \forall\theta \in \Theta$. Тому з неперервності та означення T_d маємо

$$\widehat{\theta}_n = T_d(\widehat{\mu}_n^{(d)}) \to^{P_\theta 1} T_d(\mu^{(d)}(\theta)) = \theta,\ n \to \infty,\ \forall\theta \in \Theta \ \boxdot$$

Зауваження. Якщо $\mathbf{E}_\theta \xi_1^{2d} < \infty$ і функція T_d неперервно диференційовна, то можна довести, що *оцінка методу моментів* є *асимптотично нормальною* з матрицею $t_d = \frac{\partial}{\partial\mu} T_d(\mu)\mid_{\mu=\mu^{(d)}(\theta)}$:

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to^W \eta \simeq N_d(0,\ t_d V^{(d)} t_d'),\ n \to \infty.$$

Приклади

1. *Оцінка параметрів гама-розподілу*. Нехай $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка* з *гама-розподілом* $\xi_1 \simeq \Gamma(\lambda, \alpha)$, невідомий параметр $\theta = (\lambda, \alpha), \dim\theta = 2$. Тоді теоретичні моменти мають вигляд $\mu_1(\theta) = \alpha/\lambda$, $\mu_2(\theta) = \alpha/\lambda^2 + (\alpha/\lambda)^2$.

Оцінку методу моментів $(\widehat{\alpha}, \widehat{\lambda})$ знаходимо з системи рівнянь

$$\widehat{\alpha}/\widehat{\lambda} = \widehat{\mu}_n, \quad \widehat{\alpha}/(\widehat{\lambda})^2 + (\widehat{\alpha}/\widehat{\lambda})^2 = \widehat{\mu}_{2n} = \widehat{\sigma}_n^2 + (\widehat{\mu}_n)^2,$$

звідки дістанемо

$$\widehat{\lambda} = \widehat{\mu}_n / \ \widehat{\sigma}_n^2, \quad \widehat{\alpha} = (\widehat{\mu}_n)^2 / \ \widehat{\sigma}_n^2.$$

2. *Оцінка параметрів рівномірного розподілу*. Нехай $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка* з *рівномірним розподілом* $\xi_1 \simeq U(a, b)$, невідомий параметр $\theta = (a, b), \dim\theta = 2$. Тоді $\mu_1(\theta) = (a+b)/2, \ \ \mu_2(\theta) = \mu_1^2(\theta) + (b-a)^2/12$.

Оцінку методу моментів $(\widehat{a}, \widehat{b})$ знаходимо з системи

$$(\widehat{a} + \widehat{b})/2 = \widehat{\mu}_n, \ (\widehat{b} - \widehat{a})^2/12 = \widehat{\sigma}_n^2,$$

$$\widehat{b} = \widehat{\mu}_n + \sqrt{3}\widehat{\sigma}_n, \ \widehat{a} = \widehat{\mu}_n - \sqrt{3}\widehat{\sigma}_n.$$

3. *Оцінка параметрів логнормального розподілу*. Випадкова величина ξ має *логнормальний розподіл*, позначення $\xi \simeq LN(\mu, \sigma^2)$, якщо її логарифм має нормальний розподіл: $\ln\xi \simeq N(\mu, \sigma^2)$. Нехай $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка* з логнормальним розподілом спостережень: $\xi_1 \simeq LN(\mu, \sigma^2)$. Для оцінки параметрів методом моментів можна скористатися перетворенням вибірки за допомогою логарифмічної функції. За теоремою *про перетворення незалежних величин* статистика $Y = (\eta_1, ..., \eta_n)$ з координатами $\eta_k = \ln\xi_k, k = \overline{1, n}$, є *кратною вибіркою,* а її елементи мають *нормальний розподіл*. За теоремою *про інтерпретацію параметрів нормального розподілу* вибіркові середнє та дисперсія вектора Y :

$$\widehat{\mu}_n = \frac{1}{n}\sum\nolimits_{k=1}^{n} \ln\xi_k, \ \ \widehat{\sigma}_n^2 = \frac{1}{n}\sum\nolimits_{k=1}^{n} (\ln\xi_k - \widehat{\mu}_n)^2$$

є *строго консистентними* оцінками параметрів μ, σ^2.

Зауважимо, що одночасно з логарифмічним у прикладній статистиці широко застосовується ціла шкала первинних трансформацій:

$$g_\alpha(x) = \begin{cases} (x^\alpha - 1)/\alpha, & \alpha > 0, \\ \ln x, & \alpha = 0. \end{cases}$$

Параметр α підбирається згідно з властивостями розподілу спостережень.

Вправи

(1) Знайти щільність, математичне сподівання та дисперсію випадкової величини з логнормальним розподілом.

(2) Для кратної вибірки $X = (\xi_1, ..., \xi_n)$ з $\mu_4 < \infty$ довести асимптотичну нормальність: $\sqrt{n}(\widehat{s}_n^2 - \sigma^2) \to^W N(0, \mu_4^0 - \sigma^4), n \to \infty$.

(3) Нехай при F_n – функція розподілу з моментами $(\mu_k(n), k \geq 0)$, і при кожному k існує границя $m_k = \lim_{n\to\infty} \mu_k(n)$. Довести, що (а) послідовність $(m_k, k \geq 0)$ утворена моментами деякої функції розподілу F. (б) Якщо F – нормальна функція розподілу, то $F_n \to^W F, n \to \infty$.

(4) Нехай випадкова величина ξ набуває значень на відрізку $[0, 1]$. (а) Довести, що відповідна послідовність нецентральних моментів $(\mu_n, n \geq 0)$ з $\mu_0 \equiv 1$ є цілком монотонною, тобто $(-1)^k\Delta^k\mu_n \geq 0$ для всіх $k \geq 1, n \geq 0$, де різницевий

оператор визначається як $\Delta x_n \equiv x_{n+1} - x_n, n \geq 0$. (б) Вивести з теореми Бернштейна, що $\sum_{k<nx} C_n^k (-1)^{n-k} \Delta^{n-k} \mu_k \to^O \mathbf{P}(\xi < x),\ n \to \infty$ (в) Застосувати формулу (б) до послідовностей $\mu_n = p^n$, $1/(n+1)$, $2/(n+2)$.

3.6. Незміщені оптимальні оцінки

У межах задачі оцінювання невідомого параметра θ важливу роль відіграє умова *незміщеності*, оскільки вона сприяє задовільній якості оцінки навіть при невеликих об'ємах вибірки.

Часто у випадку багатовимірного параметра треба оцінити не всі його складові, а лише деякі. Тому розглянемо задачу незміщеного оцінювання деякої функції від невідомого параметра.

Означення. *Статистика $T = T(X)$ називається* незміщеною оцінкою *значення функції $\tau = \tau(\theta)$ від параметра $\theta \in \Theta$, якщо* $\dim T = \dim \tau$ *та*

$$\mathbf{E}_\theta T = \tau(\theta), \quad \forall \theta \in \Theta.$$

Приклад. Для *кратної вибірки $X = (\xi_1, ..., \xi_n)$ вибіркова дисперсія* $\widehat{\sigma}_n^2$ є *незміщеною оцінкою* функції $\frac{n-1}{n}\sigma^2$ від дисперсії σ^2, а *нормована вибіркова дисперсія* $\widehat{s}_n^2$ є незміщеною оцінкою σ^2. Так само статистики ξ_1 та $(\xi_1 - \xi_2)^2/2$ є незміщеними оцінками середнього та дисперсії, хоча і не надто ефективними.

3.6.1. Оптимальні оцінки

Якщо існують незміщені оцінки, то їх можна порівнювати між собою за величиною середнього квадратичного відхилення від значення, яке оцінюється, та одночасно збігається з середнім цих оцінок.

Позначимо через

$$\Gamma_\tau = \{T = T(X) : \mathbf{E}_\theta T = \tau(\theta), \quad \forall \theta \in \Theta\}$$

клас усіх *незміщених оцінок* дійсної параметричної функції $\tau(\theta) : \Theta \to \mathbb{R}$.

Означення. *Статистика $T^* \in \Gamma_\tau$ називається* оптимальною *незміщеною оцінкою значення $\tau(\theta)$, або ж* незміщеною оцінкою з рівномірно найменшою дисперсією, *якщо вона має найменшу скінченну дисперсію у класі всіх незміщених оцінок:*

$$\mathbf{D}_\theta T^* \leq \mathbf{D}_\theta T, \quad \forall T \in \Gamma_\tau,\ \forall \theta \in \Theta.$$

Теорема (про єдиність оптимальної оцінки). *Якщо T_1, T_2 – дві оптимальні оцінки для значення $\tau = \tau(\theta)$, то $T_1 = T_2$ майже напевне.*

Доведення. Позначимо $\sigma^2(\theta) = \mathbf{D}_\theta T_1 = \mathbf{D}_\theta T_2$. Розглянемо таку *оцінку*: $T = (T_1 + T_2)/2$. Оскільки $\mathbf{E}_\theta T = (\mathbf{E}_\theta T_1 + \mathbf{E}_\theta T_2)/2 = \tau$, то $T \in \Gamma_\tau$, і

$$\sigma^2(\theta) = \mathbf{D}_\theta T_1 \le \mathbf{D}_\theta T = \mathbf{E}_\theta (T_1 - \tau + T_2 - \tau)^2 \ / \ 4 =$$
$$(\mathbf{D}_\theta T_1 + \mathbf{D}_\theta T_2 + 2\,\mathbf{E}(T_1 - \tau)(T_2 - \tau)) \ / \ 4 \le$$
$$(2\sigma^2(\theta) + 2\,\sqrt{\mathbf{D}_\theta T_1\ \mathbf{D}_\theta T_2}) \ / \ 4 = \sigma^2(\theta),$$

де враховані оптимальність T_1 та *нерівність Коші* у останній нерівності. Отже, за означенням *оптимальної* оцінки $\sigma^2(\theta) = \mathbf{D}_\theta T$, і нерівність Коші перетворюється на рівність. За зауваженням про *рівність у нерівності Коші* існує невипадкова стала bтака, що $T_2 - \tau = b(T_1 - \tau)$ *м.н.* Тому

$$\sigma^2(\theta) = \mathrm{Cov}(T_1, T_2) = b\ \mathrm{Cov}(T_1, T_1) = b\mathbf{D}_\theta T_1 = b\sigma^2(\theta).$$

Звідси $b = 1$ та $T_1 = T_2$ *м.н* $\boxdot$

Приклади

1. *Оцінювання ймовірності успіху за одним спостереженням з геометричним розподілом.* Нехай вибірка $X = \xi$ містить одну випадкову величину, що має *геометричний розподіл* $G(\theta)$ на $\mathbb{N}$ із невідомою ймовірністю успіху $\theta \in (0,1)$. Для довільної статистики $T = T(X)$ умова незміщеності при оцінюванні функції $\tau = \tau(\theta)$ має вигляд

$$\mathbf{E}_\theta T(X) = \sum\nolimits_{n \ge 1} (1-\theta)^{n-1}\theta\ T(n) = \tau(\theta),\ \ \forall \theta \in (0,1).$$

Нехай $\tau(\theta) = \theta$. Тоді з єдиності розкладу функції 1 у ряд Тейлора виводимо, що існує єдина незміщена оцінка для θ, до того ж не дуже змістовна: $T(X) = 1$ при $X = 1$ та $T(X) = 0$ для $X > 1$.

Якщо ж $\tau(\theta) = \theta/(1-\theta)$, то незміщених оцінок взагалі не існує, оскільки ця функція не припускає розклад у збіжний ряд Тейлора при $\theta \in (0,1)$.

2. *Оцінювання дисперсії у загальній нормальній моделі*. Нехай вибірка $X = (\xi_1, ..., \xi_n)$ складається з *незалежних у сукупності* величин із *нормальним розподілом* $\xi_k \simeq N(\mu, \sigma^2)$. Як було з'ясовано в теоремі *про моменти вибіркових моментів*, *нормована вибіркова дисперсія*

$$\widehat{s}^2_n = \frac{1}{n-1}\sum\nolimits_{i=1}^{n} (\xi_i - \widehat{\mu}_n)^2 = \frac{n}{n-1}\widehat{\sigma}^2_n$$

є *незміщеною* оцінкою невідомої дисперсії σ^2. З тієї ж теореми виводимо, що відповідна середньоквадратична похибка дорівнює

$$\mathbf{D}_\theta \widehat{s}^2_n = \mathbf{D}_\theta \left(\frac{n}{n-1}\widehat{\sigma}^2_n\right)^2 = \frac{1}{n(n-1)}((n-1)\mu^0_4 - (n-3)\sigma^4) =$$
$$((n-1)3\sigma^4 - (n-3)\sigma^4) \ / \ n(n-1) = 2\sigma^4 \ / \ (n-1),$$

де врахована тотожність $\mu_4^0 = 3\sigma^4$ для моментів нормального розподілу.

Вправа. Отримати останнє співвідношення диференціюванням характеристичної функції нормального розподілу.

Одночасно розглянемо зміщену оцінку для σ^2 вигляду $\widehat{\tau}_n^2 = t\ \widehat{s}_n^2$, для деякої сталої t. Обчислимо середньоквадратичну похибку

$$\mathbf{E}_\theta(\widehat{\tau}_n^2 - \sigma^2)^2 = \mathbf{E}_\theta(t(\widehat{s}_n^2 - \sigma^2) - (1-t)\sigma^2)^2 =$$

$$t^2\mathbf{D}_\theta \widehat{s}_n^2 + (1-t)^2\sigma^4 = (2t^2/(n-1) + (1-t)^2)\sigma^4.$$

Вираз у правій частині набуває найменшого значення при $t = \frac{n-1}{n+1} \neq 1$. Отже, незміщена оцінка $\widehat{s}_n^2$ з погляду середнього квадратичного відхилення від оцінюваного значення σ^2 гірша за зміщену оцінку $\widehat{\tau}_n^2$. Це не дивно – адже мінімум у класі всіх оцінок часто є меншим від мінімуму у підкласі.

Не зважаючи на негативний зміст наведених прикладів, у багатьох схемах незміщені оцінки все-таки є досить ефективними. Зокрема, у випадку відносно невеликого об'єму вибірки, коли на середнє квадратичне відхилення не варто зважати з причини його надмірних значень.

3.7. Нерівність Крамера – Рао, ефективність

Розглянемо *кратну вибірку* $X = (\xi_1, ..., \xi_n)$ з *функцією вірогідності*

$$L(x,\theta) = \prod_{k=1}^n f(x_k,\theta),\ x = (x_1,...,x_n) \in S,\ \theta \in \Theta,$$

де $f(x,\theta)$ – щільність одного спостереження.

У даному розділі припускатимемо, що *параметричний простір* евклідів: $\Theta \subset \mathbb{R}^d$, а функція вірогідності $L(x,\theta)$ – диференційовна за θ.

3.7.1. Функція впливу

Означення. **Функцією впливу,** *або* **функцією внеску,** *вибірки X називається частинна похідна за параметром θ від логарифма вибіркової функції вірогідності:*

$$U(X,\theta) \equiv \frac{\partial}{\partial\theta} \ln L(X,\theta).$$

У випадку кратної вибірки **функцією впливу спостереження** ξ *називається похідна за θ від логарифма функції вірогідності спостереження:*

$$u(\xi,\theta) \equiv \frac{\partial}{\partial\theta} \ln f(\xi,\theta).$$

Зауваження. Якщо параметр $\theta \in \Theta \subset \mathbb{R}$ скалярний, то функція впливу є дійсною функцією від випадкової величини, тобто *випадковою величиною*. Якщо ж $\Theta \subset \mathbb{R}^d$, то похідну слід розуміти як вектор, складений із частинних похідних (градієнт), а функція впливу є d-вимірним *випадковим вектором*:

$$U(X,\theta) \equiv (U_k(X,\theta), k = \overline{1,d}), \quad U_k(X,\theta) \equiv \frac{\partial}{\partial \theta_k} \ln L(X,\theta).$$

Важливе значення та назва "функція впливу" пояснюється таким міркуванням. Якщо припустити, що розподіл вибірки взагалі не залежить від параметра θ (параметр не впливає на розподіл), то функція впливу буде тотожнім нулем. Одночасно зауважимо, що якраз у цьому випадку марно намагатись оцінити параметр на основі спостережень, чий розподіл не залежить від значення такого параметра. Тому відносно "малі" значення функції впливу вказують на брак інформації щодо параметра в наявних спостереженнях.

Теорема (про функцію впливу кратної вибірки). *Для кратної вибірки $X = (\xi_1, ..., \xi_n)$ вибіркова функція впливу дорівнює сумі функцій впливу спостережень, які її утворюють:*

$$U(X,\theta) = \sum\nolimits_{k=1}^{n} u(\xi_k,\theta).$$

Доведення випливає з теореми *про функцію вірогідності кратної вибірки* та лінійності частинної похідної:

$$\frac{\partial}{\partial \theta} \ln L(X,\theta) = \frac{\partial}{\partial \theta} \ln \prod\nolimits_{k=1}^{n} f(\xi_k,\theta) =$$

$$\frac{\partial}{\partial \theta} \sum\nolimits_{k=1}^{n} \ln f(\xi_k,\theta) = \sum\nolimits_{k=1}^{n} \frac{\partial}{\partial \theta} \ln f(\xi_k,\theta) = \sum\nolimits_{k=1}^{n} u(\xi_k,\theta) \boxdot$$

Приклади

1. *Пуассонівська вибірка.* Для кратної вибірки з *розподілом Пуассона* $\Pi(\lambda)$ та невідомим параметром $\theta = \lambda$

$$u(y,\theta) = \tfrac{\partial}{\partial \theta} \ln(\theta^y \exp(-\theta)/y!) = y/\theta - 1, \ y \in \mathbb{Z}_+,$$
$$U(X,\theta) = \textstyle\sum_{k=1}^{n} (\xi_k/\theta - 1) = n\,(\widehat{\mu}_n/\theta - 1)\,.$$

2. *Вибірка з гама-розподілу.* Для кратної вибірки з *гама-розподілом* $\Gamma(\lambda,\alpha)$ та невідомим параметром $\theta = (\lambda,\alpha)$

$$\ln f(y,\theta) = \alpha \ln \lambda + (\alpha - 1) \ln y - \lambda y - \ln \Gamma(\alpha),$$
$$u_1(y,\theta) = \alpha/\lambda - y,$$

$$u_2(y,\theta) = \ln\lambda + \ln y - \psi(\alpha), \quad \psi(\alpha) = \tfrac{\partial}{\partial\alpha}\ln\Gamma(\alpha),$$
$$U_1(X,\theta) = n\left(\alpha/\lambda - \widehat{\mu}_n\right),$$
$$U_2(X,\theta) = n\ln\lambda + \ln\textstyle\prod_{k=1}^{n}\xi_k - n\psi(\alpha).$$

3.7.2. Умови регулярності

У даному розділі будемо припускати, що виконані певні умови регулярності на *функцію вірогідності* вибірки. Ці умови ми не будемо точно деталізувати, однак позначимо їх.

1. Множина тих значень вибірки X, для яких функція вірогідності $L(X,\theta)$ додатна, не залежить від θ.

2. Функція вірогідності $L(X,\theta)$ двічі неперервно диференційовна за θ.

3. *Функція впливу* $U(X,\theta)$ – ненульова та інтегровна у квадраті, тобто: $0 < \mathbf{E}_\theta U^2(X,\theta) < \infty,\ \forall\theta\in\Theta$.

4. Знак похідної за параметром θ можна внести під знак інтегралів вигляду $\int_S g(x,\theta)L(x,\theta)\lambda(dx)$ з функцією вірогідності $L(x,\theta)$ для певних функцій g, що спричиняється умовою 1 і збіжністю інтеграла від похідної.

3.7.3. Властивості функції впливу, інформація за Фішером

Теорема (про центрованість функції впливу). *За умов регулярності функція впливу центрована:*

$$\mathbf{E}_\theta\, U(X,\theta) = 0, \quad \forall\theta\in\Theta.$$

Доведення. За означенням *функції вірогідності*

$$\mathbf{P}_\theta(X\in S) = 1 = \int_S L(x,\theta)\ \lambda(dx),$$

оскільки вибірка X набуває значень у *вибірковому просторі* S, а $L(x,\theta)$ – її *сумісна щільність* розподілу відносно міри λ.

Позначимо $S_0 = \{x\in S : L(x,\theta) > 0\}$ вимірну підмножину *вибіркового простору* (S,Σ), яка не залежить від θ за *умовами регулярності*. Використовуючи ці умови та властивості *інтеграла Лебега* від нульової функції, з наведеної вище тотожності дістанемо

$$0 = \frac{\partial}{\partial\theta}\,1 = \frac{\partial}{\partial\theta}\int_S L(x,\theta)\ \lambda(dx) = \frac{\partial}{\partial\theta}\int_{S_0} L(x,\theta)\ \lambda(dx) =$$
$$\int_{S_0}\frac{\partial}{\partial\theta}L(x,\theta)\ \lambda(dx) = \int_{S_0}\left(\frac{\partial}{\partial\theta}\ln L(x,\theta)\right)L(x,\theta)\ \lambda(dx) =$$

$$\int_S \left(\frac{\partial}{\partial\theta}\ln L(x,\theta)\right) L(x,\theta)\ \lambda(dx) = \mathbf{E}_\theta U(X,\theta),$$

де остання рівність випливає з теореми *про обчислення математичного сподівання функції від неперервної величини*, і враховано тотожність

$$\frac{\partial}{\partial\theta}L(x,\theta) = \left(\frac{\partial}{\partial\theta}\ln L(x,\theta)\right) L(x,\theta) = U(x,\theta)L(x,\theta)$$

на множині S_0, де $L(x,\theta) > 0$ ⊡

Означення. *Нехай параметр $\theta \in \Theta \subset \mathbb{R}$ – скалярний.* Інформацією за Фішером *у вибірці X називається функція*

$$I(\theta) \equiv \mathbf{D}_\theta U(X,\theta) = \mathbf{E}_\theta U^2(X,\theta).$$

Друга рівність в означенні випливає з теореми *про центрованість функції впливу* та з властивостей *дисперсії*.

Теорема (про обчислення інформації за Фішером). *За умов регулярності справедлива тотожність*

$$I(\theta) = -\mathbf{E}_\theta\frac{\partial^2}{\partial\theta^2}\ln L(X,\theta) = -\mathbf{E}_\theta\frac{\partial}{\partial\theta}U(X,\theta), \quad \forall\theta\in\Theta.$$

Доведення. Визначимо $S_0 = \{x \in S : L(x,\theta) > 0\}$, продиференціюємо тотожність із теореми *про центрованість функції впливу* та використаємо умови регулярності:

$$0 = \frac{\partial}{\partial\theta}\mathbf{E}_\theta U(X,\theta) = \frac{\partial}{\partial\theta}\int_S U(x,\theta)L(x,\theta)\lambda(dx) =$$

$$\frac{\partial}{\partial\theta}\int_{S_0} U(x,\theta)L(x,\theta)\lambda(dx) = \int_{S_0}\frac{\partial}{\partial\theta}(U(x,\theta)L(x,\theta))\lambda(dx) =$$

$$\int_{S_0}\left(\frac{\partial}{\partial\theta}U(x,\theta)\right)L(x,\theta)\ \lambda(dx) + \int_{S_0}U(x,\theta)\frac{\partial}{\partial\theta}L(x,\theta)\lambda(dx) =$$

$$\int_S\left(\frac{\partial}{\partial\theta}U(x,\theta)\right)L(x,\theta)\ \lambda(dx) +$$

$$\int_{S_0}U(x,\theta)\left(\frac{\partial}{\partial\theta}\ln L(x,\theta)\right)L(x,\theta)\ \lambda(dx) =$$

$$\mathbf{E}_\theta\frac{\partial}{\partial\theta}U(X,\theta) + \mathbf{E}_\theta U^2(X,\theta) = \mathbf{E}_\theta\frac{\partial^2}{\partial\theta^2}\ln L(X,\theta) + I(\theta) \ \boxdot$$

Теорема (про адитивність інформації за Фішером). *Нехай для кратної вибірки $X = (\xi_1, ..., \xi_n)$ функція $i(\theta) \equiv \mathbf{D}_\theta u(\xi_1,\theta)$ задає **інформацію за Фішером в одному спостереженні**. Тоді повна інформація за Фішером дорівнює:*

$$I(\theta) = n\, i(\theta).$$

Доведення. Оскільки величини $u(\xi_k,\theta)$ незалежні за теоремою *про перетворення незалежних величин*, то внаслідок теорем *про функцію впливу кратної вибірки* та *про дисперсію суми незалежних величин*

$$I(\theta)=\mathbf{D}_\theta U(X,\theta)=\mathbf{D}_\theta \sum\nolimits_{k=1}^n u(\xi_k,\theta)=\sum\nolimits_{k=1}^n \mathbf{D}_\theta u(\xi_k,\theta)=n\,i(\theta),$$

де використана також *однакова розподіленість* цих величин $\boxdot$

Приклади

1. *Схема Бернуллі*. Нехай вибірка $X=(\chi_1,...,\chi_n)$ містить результати *випробувань Бернуллі* з невідомою ймовірністю успіху θ, де χ_k – індикатор успіху в k-му випробуванні. *Функція вірогідності* одного спостереження є щільністю відносно точкової міри і має вигляд

$$f(y,\theta)=\theta 1\!\!\mathrm{I}_{y=1}+(1-\theta)1\!\!\mathrm{I}_{y=0}=\theta^y(1-\theta)^{1-y},\ y\in\{0,1\},$$

теоретична *функція впливу* дорівнює

$$u(y,\theta)=\frac{\partial}{\partial\theta}\ln f(y,\theta)=\frac{y}{\theta}-\frac{1-y}{1-\theta}=\frac{y-\theta}{\theta(1-\theta)},$$

функції вірогідності та впливу всієї вибірки

$$L(X,\theta)=\prod\nolimits_{k=1}^n \theta^{\chi_k}(1-\theta)^{1-\chi_k}=\theta^{\nu_n}(1-\theta)^{n-\nu_n},\ \nu_n=\sum\nolimits_{k=1}^n \chi_k,$$

$$U(X,\theta)=\sum\nolimits_{k=1}^n u(\chi_k,\theta)=n(\widehat{\theta}_n-\theta)/\theta(1-\theta),\ \ \widehat{\theta}_n=\nu_n/n,$$

функції *інформації за Фішером* мають вигляд

$$i(\theta)=\mathbf{E}_\theta\left(\frac{\chi_1-\theta}{\theta(1-\theta)}\right)^2=\frac{\mathbf{D}_\theta\chi_1}{\theta^2(1-\theta)^2}=\frac{1}{\theta(1-\theta)},\ I(\theta)=\frac{n}{\theta(1-\theta)}.$$

2. *Показниковий розподіл*. Нехай $X=(\xi_1,...,\xi_n)$ *кратна вибірка* з *показниковим розподілом:* $\xi_k\simeq Exp(\theta)$. *Функції впливу* дорівнюють:

$$u(y,\theta)=\frac{\partial}{\partial\theta}(\ln\theta-\theta y)=1/\theta-y,\ \ U(X,\theta)=\sum\nolimits_{k=1}^n u(\xi_k,\theta)=n/\theta-n\widehat{\mu}_n,$$

а *інформація за Фішером* у спостереженні та у вибірці має вигляд

$$i(\theta)=\mathbf{E}_\theta(1/\theta-\xi_1)^2=\mathbf{D}_\theta\xi_1=1/\theta^2,\ \ I(\theta)=n/\theta^2.$$

3.7.4. Нерівність Крамера – Рао

Розглянемо задачу оцінювання значення дійсної параметричної функції $\tau(\theta)$ у класі Γ_τ *незміщених* її оцінок.

Теорема (про нерівність та критерій Крамера – Рао). *Нехай параметр θ є скалярним:* $\Theta\subset\mathbb{R}$.

(а) Якщо $T = T(X) \in \Gamma_\tau$ – довільна незміщена оцінка значення $\tau(\theta)$, і виконуються умови регулярності, то при всіх $\theta \in \Theta$ має місце **нерівність Крамера – Рао**

$$\mathbf{E}_\theta(T-\tau)^2 \equiv \mathbf{D}_\theta T \geq \frac{\tau_\theta^2(\theta)}{I(\theta)},$$

де $\tau_\theta(\theta) = \frac{d}{d\theta}\tau(\theta)$, $I(\theta)$ – інформація за Фішером у вибірці X.

(б) Рівність у нерівності (а) виконується тоді й тільки тоді, коли оцінка T є лінійною функцією від функції впливу вибірки:

$$T(X) - \tau(\theta) = c(\theta)U(X,\theta) \text{ м.н.}, \ \forall\theta \in \Theta,$$

для деякої дійсної $c(\theta)$. Ця стала дорівнює $c(\theta) = \tau_\theta(\theta)/\, I(\theta)$.

Означення. *Оцінка $T \in \Gamma_\tau$ називається* ефективною оцінкою *параметричної функції $\tau(\theta)$, якщо нерівність Крамера – Рао для неї є рівністю, тобто у випадку, коли ця оцінка має найменше можливе значення дисперсії у класі Γ_τ всіх незміщених оцінок.*

Твердження (б) дає критерій ефективності Крамера – Рао.

Зауваження. З означення та єдиності оптимальної оцінки випливає, що *ефективна оцінка є оптимальною оцінкою* свого математичного сподівання, тобто критерій Крамера – Рао є достатньою умовою *оптимальності* для регулярних спостережень.

Зауваження. *Нерівність Крамера – Рао* для дисперсії $\mathbf{D}_\theta T$ має місце також і для зміщених оцінок T, якщо в правій частині τ_θ^2 замінити на $(\tau_\theta + b_\theta)^2$, де $b_\theta(\theta) = \frac{d}{d\theta}b(\theta)$, $b(\theta) = \mathbf{E}_\theta T - \tau(\theta)$. Дійсно, у даному випадку оцінка T є незміщеною для значення $\tau + b$.

Доведення теореми.

За означенням *незміщеності* та за теоремою *про обчислення математичного сподівання функції від випадкової величини*

$$\mathbf{E}_\theta T(X) = \int_S T(x)L(x,\theta)\ \lambda(dx) = \tau(\theta),\ \forall\theta \in \Theta.$$

За *умови регулярності* з урахуванням означення *функції впливу* та теореми *про обчислення математичного сподівання функції від випадкової величини* знайдемо

$$\tau_\theta(\theta) = \frac{\partial}{\partial\theta}\int_S T(x)L(x,\theta)\ \lambda(dx) = \int_S T(x)\frac{\partial}{\partial\theta}L(x,\theta)\ \lambda(dx) =$$

$$\int_S T(x)U(x,\theta)L(x,\theta)\ \lambda(dx) =$$

$$\mathbf{E}_\theta T(X)U(X,\theta) = \mathbf{E}_\theta(T(X)-\tau(\theta))U(X,\theta),$$

де використана теорема *про центрованість функції впливу* та тотожність для похідної: $\partial L/\partial\theta = UL$ з доведення цієї теореми.

Згідно з *нерівністю Коші* з отриманої тотожності виводимо

$$\tau_\theta^2(\theta) = (\mathbf{E}_\theta(T(X) - \tau(\theta))U(X,\theta))^2 \le$$

$$\mathbf{E}_\theta(T(X) - \tau(\theta))^2\, \mathbf{E}_\theta U^2(X,\theta) = \mathbf{D}_\theta T \cdot I(\theta),$$

за означенням $I(\theta)$. Це доводить твердження (а).

За зауваженням про *рівність у нерівності Коші* ця нерівність перетворюється на рівність тоді і тільки тоді, коли множники під знаком математичного сподівання є лінійно пов'язаними *майже напевне*, тобто $T(X) - \tau(\theta) = c(\theta)U(X,\theta)$ м.н. для сталої $c(\theta)$ при кожному θ. Звідси виводимо твердження (б) теореми. Вираз для значення відповідної сталої знаходимо після підстановки тотожності критерію рівності в отриману вище тотожність:

$$\tau_\theta(\theta) = \mathbf{E}_\theta(T(X) - \tau(\theta))U(X,\theta) = \mathbf{E}_\theta c(\theta)U^2(X,\theta) = c(\theta)I(\theta) \ \boxdot$$

Зауваження. У випадку, коли $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка*, *інформація за Фішером* у знаменнику дорівнює $I(\theta) = n\ i(\theta)$, отже *нерівність Крамера – Рао* набуває вигляду

$$\mathbf{D}_\theta(\sqrt{n}(T - \tau)) = n\ \mathbf{D}_\theta T \ge \frac{\tau_\theta^2(\theta)}{i(\theta)} \equiv \widetilde{\sigma}^2_{opt}.$$

Означення. *Якщо послідовність оцінок $T_n \in \Gamma_\tau$ є асимптотично нормальною з асимптотичною дисперсією σ_a^2, тобто*

$$\sqrt{n}(T_n - \tau) \to^W \eta \simeq N(0, \sigma_a^2),$$

причому $\sigma_a^2 = \widetilde{\sigma}^2_{opt}$ збігається з найменшим можливим значенням у нерівності Крамера – Рао, то оцінка (T_n) називається асимптотично ефективною.

3.7.5. Ефективна оцінка у схемі Бернуллі

Розглянемо задачу оцінювання, в якій проводяться n *випробувань Бернуллі* з невідомою ймовірністю $p = \theta$ успіху в окремому випробуванні. Припустимо, що спостерігається вибірка $X = (\chi_1, ..., \chi_n)$, де χ_k – індикатор k-го успіху. Розподіл вибірки абсолютно неперервний відносно точкової міри, логарифмічна *функція вірогідності* має вигляд

$$\ln L(X,\theta) = \nu_n(X) \ln\theta + (n - \nu_n(X)) \ln(1 - \theta),$$

де $\nu_n(X) = \sum_{k=1}^{n} \chi_k$ – загальна кількість успіхів. Звідси

$$U(X,\theta) = \frac{\nu_n(X)}{\theta} - \frac{n - \nu_n(X)}{1-\theta} = \frac{\nu_n(X)}{\theta(1-\theta)} - \frac{n}{1-\theta},$$

функція *інформації за Фішером* дорівнює

$$I(\theta) = \mathbf{D}_\theta \nu_n(X)/\theta^2(1-\theta)^2 = n \;/\; \theta(1-\theta),$$

нижня границя для дисперсії *незміщеної оцінки* параметра θ має вигляд

$$\mathbf{D}_\theta T \geq \theta(1-\theta) \;/\; n.$$

За *критерієм ефективності Крамера – Рао* знаходимо ефективну оцінку

$$T^* = \theta + c\left(\frac{\nu_n(X)}{\theta(1-\theta)} - \frac{n}{1-\theta}\right).$$

Для того, щоб права частина була *статистикою* (не залежала від θ), є єдиний спосіб обрати сталу: $c = \theta(1-\theta)/n$. При такому виборі *нерівність Крамера – Рао* перетворюється на рівність.

Отже, у схемі Бернуллі *оптимальною незміщеною* оцінкою ймовірності успіху θ є *відносна частота успіхів* $T^* = \nu_n(X)/n = \frac{1}{n}\sum_{k=1}^{n} \chi_k$.

Вправи

(1) Довести, що при $|\Theta| > 1$ в класі всіх оцінок параметра θ не існує такої, що має рівномірно найменше середньоквадратичне відхилення.

(2) У схемі випробувань Бернуллі розглянемо такі функції від частоти успіхів: $T_{\alpha\beta} = (\nu_n(X) + \alpha)/(n+\beta),\ \alpha, \beta \in \mathbb{R}$. (а) Обчислити квадратичне відхилення $T_{\alpha\beta}$ від θ. (б) Довести, що при $\alpha = \sqrt{n}/2$ та $\beta = \sqrt{n}$ це відхилення не залежить від θ і дорівнює $(2\sqrt{n}+2)^{-2}$. (в) Порівняти його з таким же відхиленням для T^*.

(3) Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з експоненційного розподілу $Exp(\theta)$. Довести, що оптимальна незміщена оцінка для значення функції $\exp(-\theta x)$ дорівнює $(\max(0, 1 - x/S_n))^n$, де $S_n = \xi_1 + ... + \xi_n$.

3.7.6. Нерівність Крамера – Рао для векторного параметра

Означення. *Нехай параметр* $\theta \in \Theta \subset \mathbb{R}^d$ *– векторний, тоді функція впливу також є вектором-стовпчиком*

$$U(X,\theta) \equiv \left(U_i(X,\theta) \equiv \frac{\partial}{\partial\theta_i} \ln L(X,\theta),\ i = \overline{1,d}\right).$$

Інформаційною матрицею за Фішером *вибірки X називається коваріаційна матриця векторної функції впливу:*

$$I(\theta) \equiv \mathrm{Cov}_\theta\, U(X,\theta) = \mathbf{E}_\theta U(X,\theta)U'(X,\theta) =$$
$$\left(\mathrm{Cov}(U_i(X,\theta),\ U_j(X,\theta)),\ i,j = \overline{1,d}\right).$$

Як і у скалярному випадку, матриця інформації для кратних вибірок має властивість адитивності та центрованості. Доведення цих властивостей проводиться цілком аналогічно, тому не наводиться.

Приклад. Розглянемо *кратну вибірку* $X = (\xi_1, ..., \xi_n)$ із нормальних спостережень $\xi_k \simeq N(\mu,\sigma^2)$, невідомий параметр $\theta = (\mu,\sigma^2)$, $\dim\theta = 2$. Не зважаючи на форму запису, вважатимемо σ^2 незалежною змінною. Логарифм *функції вірогідності* вибірки має такий вигляд

$$\ln L(X,\theta) = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum\nolimits_{k=1}^{n}(\xi_k - \widehat{\mu}_n + \widehat{\mu}_n - \mu)^2 =$$
$$-\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}\left(\widehat{\sigma}_n^2 + (\widehat{\mu}_n - \mu)^2\right).$$

Звідси знаходимо координати *функції впливу*

$$U_1(X,\theta) = \frac{n(\widehat{\mu}_n - \mu)}{\sigma^2},$$
$$U_2(X,\theta) = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}\left(\widehat{\sigma}_n^2 + (\widehat{\mu}_n - \mu)^2\right).$$

За теоремою *про моменти вибіркових моментів* обчислюємо

$$I_{11}(\theta) = \frac{n^2}{\sigma^4}\mathbf{D}_\theta\widehat{\mu}_n = \frac{n}{\sigma^2}.$$

Вправа. Знайдіть інші елементи інформаційної матриці, якщо відомо, що статистики $\widehat{\mu}_n$ і $\widehat{\sigma}_n^2$ незалежні.

Теорема (нерівність Крамера – Рао для векторного параметра). *Припустимо, що $\dim\Theta = d > 1$, та інформаційна матриця за Фішером $I(\theta)$ є невиродженою.*

(а) Нехай $T = T(X) \in \Gamma_\tau$ – незміщена оцінка скалярної функції $\tau(\theta)$ і виконані умови регулярності. Тоді має місце нерівність

$$\mathbf{E}_\theta(T-\tau)^2 = \mathbf{D}_\theta T \ge \tau'_\theta I^{-1}(\theta)\tau_\theta,\ \forall\theta\in\Theta,$$

де $\tau_\theta(\theta) = \frac{d}{d\theta}\tau(\theta)$ – вектор градієнта, τ'_θ – спряжений вектор-рядок.

(б) Рівність у нерівності (а) має місце тоді й тільки тоді, коли T є лінійною функцією від функції впливу вибірки, тобто для деякої сталої $c(\theta) \in \mathbb{R}^d$:

$$T(X) - \tau(\theta) = c'(\theta)U(X,\theta)\ \text{м.н.},\quad \forall\theta\in\Theta.$$

Зауваження. Твердження (а) теореми залишається справедливим і для виродженої матриці $I(\theta)$, якщо замінити квадратичну форму в правій частині (а) на такий вираз:

$$B(I(\theta),\tau_\theta) = \sup_{b\in\mathbb{R}^d:\ b'I(\theta)b>0}(b'\tau_\theta)^2\ /\ b'I(\theta)b.$$

Доведення. Як і у скалярному випадку, з умов *нормованості* функції вірогідності та *незміщеності* на підставі умов регулярності доводимо тотожності

$$\mathbf{E}_\theta U(X,\theta)=0,\ \mathbf{E}_\theta(T(X)-\tau(\theta))U(X,\theta)=\tau_\theta(\theta).$$

Оберемо довільне $b\in\mathbb{R}^d$. Скалярно помножимо другу тотожність на b і скористаємося *нерівністю Коші* та теоремою *про дисперсію лінійної форми від випадкового вектора*:

$$(b'\tau_\theta)^2 = (b'\mathbf{E}_\theta(T-\tau)U(X,\theta))^2 = (\mathbf{E}_\theta(T-\tau)b'U(X,\theta))^2 \le$$

$$\mathbf{E}_\theta(T-\tau)^2\ \mathbf{E}_\theta(b'U(X,\theta))^2 = \mathbf{D}_\theta T\ \mathbf{D}_\theta(b'U(X,\theta)) = \mathbf{D}_\theta T\ \ b'I(\theta)b.$$

Звідси виводимо нерівність

$$\mathbf{D}_\theta T \ge (b'\tau_\theta)^2\ /\ b'I(\theta)b = \tau'_\theta I^{-1}(\theta)\tau_\theta,$$

де остання рівність досягається при $b=I^{-1}(\theta)\tau_\theta$. Умова рівності випливає з зауваження про *рівність у нерівності Коші.*

Твердження зауваження для виродженої інформаційної матриці дістанемо після переходу до верхньої межі за b в останній нерівності $\boxdot$

3.7.7. Ефективні оцінки параметрів нормальних спостережень

Оцінювання середнього при відомій дисперсії

Нехай $X=(\xi_1,...,\xi_n)$ – *кратна вибірка* з *нормальним розподілом* спостережень: $\xi_k \simeq N(\mu,\sigma^2)$, де *дисперсія* σ^2 відома, а середнє $\mu=\theta$ треба оцінити. Логарифмічна *функція вірогідності* має вигляд

$$\ln L(X,\theta) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum\nolimits_{k=1}^n(\xi_k-\mu)^2 =$$

$$-\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}(\widehat{\mu}_{2n}-2\widehat{\mu}_n\mu+\mu^2).$$

де $\widehat{\mu}_n$ – *вибіркове середнє*, $\widehat{\mu}_{2n}$ – другий нецентральний *вибірковий момент*. Звідси знаходимо *функцію впливу*

$$U(X,\theta)=\frac{n}{\sigma^2}(\widehat{\mu}_n-\mu).$$

За *критерієм ефективності Крамера – Рао оптимальна* оцінка μ повинна задовольняти рівняння

$$T^* - \mu = c\frac{n}{\sigma^2}(\widehat{\mu}_n - \mu) = \widehat{\mu}_n - \mu,$$

що має місце при виборі $c = \sigma^2/n$. Отже, вибіркове середнє $T^* = \widehat{\mu}_n$ є *оптимальною* оцінкою середнього при відомій дисперсії.

Оцінювання дисперсії при відомому середньому

Розглянемо нормальну *кратну вибірку* $X = (\xi_1, ..., \xi_n)$, $\xi_k \simeq N(\mu, \sigma^2)$, де відоме середнє μ, а *дисперсія* $\sigma^2 = \theta$ оцінюється. У цьому випадку *функція впливу* має вигляд

$$U(X, \theta) = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}(\widehat{\mu}_{2n} - 2\widehat{\mu}_n\mu + \mu^2).$$

Вибором множника $c = 2\sigma^4/n$ забезпечується рівність $T - \sigma^2 = cU$ у *критерії ефективності Крамера – Рао*, а *оптимальна* оцінка дорівнює *вибірковій дисперсії* з урахуванням відомого середнього

$$T^* = \widetilde{\sigma}_n^2 = \widehat{\mu}_{2n} - 2\widehat{\mu}_n\mu + \mu^2 = \frac{1}{n}\sum\nolimits_{k=1}^{n}(\xi_k - \mu)^2.$$

Оцінювання середнього при невідомій дисперсії

Розглянемо нормальну *кратну вибірку* $X = (\xi_1, ..., \xi_n)$, $\xi_k \simeq N(\mu, \sigma^2)$, де невідомі як μ, так і σ^2, параметр $\theta = (\mu, \sigma^2)$, а оцінюється середнє $\mu = \tau(\theta)$. У цьому випадку координати *функції впливу* мають вигляд

$$U_1(X, \theta) = \frac{n}{\sigma^2}(\widehat{\mu}_n - \mu),$$

$$U_2(X, \theta) = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}(\widehat{\mu}_{2,n} - 2\widehat{\mu}_n\mu + \mu^2).$$

Оберемо векторну сталу $c = (c_1, c_2) = (\sigma^2/n, 0)$. Тоді у критерії (б) *нерівності Крамера – Рао для векторного параметра*

$$T^* - \mu = c_1U_1(X, \theta) + c_2U_2(X, \theta) = \widehat{\mu}_n - \mu$$

має місце рівність для $T^* = \widehat{\mu}_n$, тобто *вибіркове середнє* є *оптимальною* оцінкою і при невідомій *дисперсії*.

3.7.8. Ефективні оцінки для експоненційної моделі

Інтегруванням рівняння $\frac{\partial}{\partial\theta}\ln L(X,\theta) = (T(X) - \tau(\theta))/c(\theta)$ з *критерію ефективності Крамера-Рао* отримуємо **експоненційну модель розподілів**.

Нехай $\Theta \subset \mathbb{R}^d$, $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка*, причому щільність одного спостереження має вигляд

$$f(y,\theta) = \exp(b'(\theta)a(y) + c(y) + d(\theta)),$$

де a, b – векторні функції однакової розмірності, а функції b, d – диференційовні за векторним параметром θ. За теоремою *про функцію вірогідності кратної вибірки*

$$\ln L(X,\theta) = nb'(\theta)A(X) + C(X) + nd(\theta),$$

$$A(X) \equiv \frac{1}{n}\sum\nolimits_{k=1}^{n} a(\xi_k),\ C(X) \equiv \sum\nolimits_{k=1}^{n} c(\xi_k),$$

тому з урахуванням позначення $\beta(\theta) \equiv \partial b(\theta)/\partial\theta$:

$$U(X,\theta) = n\beta'(\theta)A(X) + n\partial d(\theta)/\partial\theta.$$

Припустимо, що виконані *умови регулярності* (основна умова автоматично випливає з додатності експоненційної функції на всій числовій осі). За теоремою *про центрованість функції впливу* звідси отримуємо

$$U(X,\theta) = n\beta'(\theta)(A(X) - \alpha(\theta)), \quad \alpha(\theta) \equiv \mathbf{E}_\theta A(X) = \mathbf{E}_\theta a(\xi_1).$$

Оскільки вектор $A(X)$ є сумою *незалежних у сукупності*, *однаково розподілених* векторів, то *інформаційна матриця за Фішером* дорівнює

$$I(\theta) = n^2\beta'(\theta)\mathrm{Cov}_\theta A(X)\beta(\theta) = n\beta'(\theta)V(\theta)\beta(\theta), \quad V(\theta) \equiv \mathrm{Cov}_\theta a(\xi_1).$$

Рівність у *критерії ефективності Крамера – Рао* для оцінювання невідомої функції $\tau(\theta)$ має вигляд

$$T(X) - \tau(\theta) = nc'(\theta)\beta'(\theta)(A(X) - \alpha(\theta)),$$

та виконується тоді й тільки тоді, коли $nc'(\theta)\beta'(\theta) = \gamma'$ для всіх θ та деякого сталого вектора γ. Очевидно, що для існування відповідної функції $c(\theta)$ достатньо невиродженості матриці $\beta(\theta)$, $\forall\theta$. У цьому випадку *ефективна оцінка* (одночасно *оптимальна*) значення $\tau(\theta)$ існує та дорівнює

$$T^*(X) = \gamma' A(X) = \frac{1}{n}\sum\nolimits_{k=1}^{n} \gamma' a(\xi_k),$$

причому $\tau(\theta) = \gamma'\alpha(\theta) = \mathbf{E}_\theta T^*(X) = \mathbf{E}_\theta\gamma' a(\xi_1)$. *Дисперсія* цієї оцінки дорівнює $\mathbf{D}_\theta T^*(X) = \mathbf{D}_\theta\gamma' a(\xi_1)/n = \gamma' V(\theta)\gamma/n$.

Вправи

(1) Розподіли: біноміальний, Пуассона, негативний біноміальний, експоненційний, багатовимірний нормальний підпорядковуються експоненційній моделі.

(2) Для векторної функції $\tau(\theta)$ в умовах попередньої теореми та для її незміщеної оцінки $T = T(X) \in \Gamma_\tau$ матриця $\mathrm{Cov}_\theta(T(X) - \tau(\theta)) - \tau'_\theta I^{-1}(\theta)\tau_\theta$ невід'ємно визначена, зокрема, має невід'ємні діагональні елементи.

(3) Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з логістичною щільністю спостережень $f(y,\theta) = \exp(\theta - y)(1 + \exp(\theta - y))^{-2}, y \in \mathbb{R}, \theta \in \mathbb{R}$. Довести, що: (а) вибіркове середнє $\widehat{\mu}_n$ є незміщеною оцінкою для θ з дисперсією $\pi^2/3n$, (б) функція інформації дорівнює $I_n(\theta) = n/3$, тобто вибіркове середнє у даному разі не є ефективною оцінкою, хоча має непогану ефективність $9/\pi^2$.

3.8. Достатні статистики та оптимальність

Означення. *Статистика $T = T(X)$ називається* достатньою статистикою *для вибірки X, якщо для деяких вимірних невід'ємних функцій g, h, j вибіркова функція вірогідності має вигляд*

$$L(X,\theta) = g(T(X),\theta)\ h(X)\ j(\theta) \ \ \text{м.н.}\ \forall\theta \in \Theta.$$

Зауважимо, що на розмірність достатньої статистики обмеження не накладаються, однак при виборі цієї статистики доцільно мінімізувати її розмірність. В кожному разі, достатні статистики напевне існують. Наприклад, статистика $T(X) = X$ завжди є достатньою: досить обрати $g = L, h \equiv 1, j \equiv 1$.

Якщо достатня статистика S має вигляд $S(X) = f(T(X))$ для вимірної функції f, то за означенням статистика T – також достатня.

Означення. *Достатня статистика $S = S(X)$ називається мінімальною достатньою, якщо для кожної достатньої статистики T знайдеться невипадкова вимірна функція f така, що $S = f(T)$ м.н.*

3.8.1. Приклади достатніх статистик

1. *Пуассонівська вибірка*. Якщо X – *n-кратна вибірка* з *розподілом Пуассона,* $\xi_1 \simeq \Pi(\lambda)$, то

$$\ln L(X,\theta) = \ln \prod_{k=1}^{n} \frac{\lambda^{\xi_k}}{\xi_k!} \exp(-\lambda) = (\ln\lambda) \sum_{k=1}^{n} \xi_k - \sum_{k=1}^{n} \ln(\xi_k!) - n\lambda,$$

отже, існує скалярна достатня статистика $T(X) = \widehat{\mu}_n \equiv \frac{1}{n}\sum_{k=1}^{n} \xi_k$.

2. *Рівномірна вибірка*. Якщо X – n-*кратна вибірка* з $\xi_1 \simeq U(a,b)$, то

$$L(X,\theta) = (b-a)^{-n}\ \mathbb{1}_{\{a\ <\ \min \xi_k,\ \max \xi_k\ <\ b\}}.$$

Отже, достатньою є двовимірна статистика

$$T(X) = (\min_{1\le k\le n} \xi_k,\ \max_{1\le k\le n} \xi_k) = \left(\xi_{(1)}, \xi_{(n)}\right).$$

3. *Нормальна вибірка*. Якщо X – n-*кратна вибірка* з $\xi_1 \simeq N(\mu,\sigma^2)$, то

$$\ln L(X,\theta) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}\left(\widehat{\mu}_{2n} - 2\widehat{\mu}_n\mu + \mu^2\right).$$

Пара $(\widehat{\mu}_n,\ \widehat{\mu}_{2n})$ утворює *достатню статистику*, так само як і $(\widehat{\mu}_n,\ \widehat{\sigma}^2_n)$ – оскільки між цими парами статистик є ізоморфізм.

3.8.2. Умовний розподіл вибірки

Теорема (про критерій достатності статистики). *Статистика* $T(X)$ *є достатньою статистикою тоді й тільки тоді, коли умовний розподіл*

$$\mathbf{P}_\theta(X = x \mid T(X) = t) = l(x,t)$$

не залежить від θ.

Доведення проведемо в припущенні, що вибірка є дискретним випадковим вектором. В цьому випадку *функція вірогідності* є *дискретним розподілом* вибірки: $\mathbf{P}_\theta(X = x) = L(x,\theta),\ \forall x\in S, \forall\theta\in\Theta$.

Достатність. Для заданого x визначимо $t = T(x)$. Тоді за означенням $l(x,t)$ як *умовної ймовірності*

$$L(x,\theta) = \mathbf{P}_\theta(X = x) = \mathbf{P}_\theta(X = x, T(X) = t) =$$

$$\mathbf{P}_\theta(T(X) = t) l(x,t) = g(t,\theta) l(x,t) = g(T(x),\theta) l(x,T(x)),$$

що доводить зображення у означенні достатньої статистики.

Необхідність. Зауважимо, що вказана *умовна ймовірність* дорівнює нулю та не залежить від θ, якщо $T(x)\ne t$. Отже, можна припустити, що $T(x) = t$. За означенням *умовної ймовірності*

$$\mathbf{P}_\theta(X = x \mid T(X) = t) = \frac{\mathbf{P}_\theta(X = x,\ T(X) = t)}{\mathbf{P}_\theta(T(X) = t)} =$$

$$\frac{\mathbf{P}_\theta(X = x,\ T(x) = t)}{\mathbf{P}_\theta(T(X) = t)} = L(x,\theta)\ /\sum\nolimits_{y:\ T(y)=t} L(y,\theta) =$$

$$g(T(x),\theta)\ h(x)\ j(\theta)\ /\sum\nolimits_{y:\ T(y)=t} g(T(y),\theta)\ h(y)\ j(\theta) =$$

$$g(t,\theta) h(x)\ /\sum\nolimits_{y:\ T(y)=t} g(t,\theta)\ h(y) = h(x)\ /\sum\nolimits_{y:\ T(y)=t} h(y) = l(x,t),$$

оскільки має місце зображення $\{T(X) = t\} = \cup_{y:\ T(y)=t}\{X = y\}$ $\boxdot$

3.8.3. Достатність та оптимальність

Теорема (про дисперсію функції від достатньої статистики). *Нехай $T = T(X)$ – достатня статистика, а $T_0 \in \Gamma_\tau$ – довільна незміщена оцінка функції $\tau(\theta)$. Тоді знайдеться функція $T_1(X) = H_0(T(X))$ від достатньої статистики $T(X)$, яка також є незміщеною оцінкою для $\tau(\theta)$ та має не більшу дисперсію, ніж T_0, при кожному $\theta \in \Theta$.*

Доведення. Нехай $T_0 \in \Gamma_\tau$ – довільна незміщена оцінка. Розглянемо функцію, що задає умовне сподівання T_0:

$$H_0(t) = \mathbf{E}_\theta(T_0(X) \mid T(X) = t) \equiv \frac{\mathbf{E}_\theta\left(T_0(X)\ 1\!\!1_{\{T(X)\,=\,t\}}\right)}{\mathbf{P}_\theta(T(X) = t)} =$$

$$\mathbf{E}_\theta \sum\nolimits_{x\in S} T_0(x)\ 1\!\!1_{\{X=x,\ T(X)=t\}} / \ \mathbf{P}_\theta(T(X) = t) =$$

$$\sum\nolimits_{x\in S} T_0(x)\ \mathbf{P}_\theta(X = x \mid T(X) = t) = \sum\nolimits_{x\in S} T_0(x)\ l(x,t),$$

та за теоремою *про критерій достатності статистики* не залежить від θ.

Визначимо *статистику*

$$T_1(X) = H_0(T(X)).$$

Вона є незміщеною, оскільки за *формулою повної ймовірності* та за означенням функцій $H_0(t)$ і $l(x,t)$

$$\mathbf{E}_\theta T_1(X) = \mathbf{E}_\theta H_0(T(X)) = \sum\nolimits_t H_0(t)\ \mathbf{P}_\theta(T(X) = t) =$$

$$\sum\nolimits_t \left(\sum\nolimits_{x\in S} T_0(x)\ l(x,t)\right)\ \mathbf{P}_\theta(T(X) = t) =$$

$$\sum\nolimits_{x\in S} T_0(x) \sum\nolimits_t \mathbf{P}_\theta(X = x \mid T(X) = t)\ \mathbf{P}_\theta(T(X) = t) =$$

$$\sum\nolimits_{x\in S} T_0(x) \mathbf{P}_\theta(X = x) = \mathbf{E}_\theta T_0(X) = \tau(\theta),\ \forall\theta \in \Theta.$$

Далі, обчислимо *дисперсію*

$$\mathbf{D}_\theta T_0 = \mathbf{E}_\theta(T_0 - H_0(T) + H_0(T) - \tau(\theta))^2 = \mathbf{E}_\theta(T_0 - H_0(T))^2 +$$

$$2\mathbf{E}_\theta(T_0 - H_0(T))(H_0(T) - \tau(\theta)) + \mathbf{D}_\theta H_0(T) \geq \mathbf{D}_\theta H_0(T) = \mathbf{D}_\theta T_1,$$

оскільки за означенням $H_0(t)$ математичне сподівання добутку – нульове:

$$\mathbf{E}_\theta(T_0 - H_0(T))(H_0(T) - \tau(\theta)) =$$

$$\sum\nolimits_t \mathbf{E}_\theta\left(T_0 - H_0(T)\right)\ \left(H_0(T) - \tau(\theta)\right)\ 1\!\!1_{\{T\,=\,t\}} =$$

$$\sum\nolimits_t \mathbf{E}_\theta\left(T_0 - H_0(t)\right)\ \left(H_0(t) - \tau(\theta)\right)\ 1\!\!1_{\{T\,=\,t\}} =$$

$$\sum\nolimits_t \left(\mathbf{E}_\theta T_0\ 1\!\!1_{\{T\,=\,t\}} - H_0(t)\mathbf{P}_\theta(T = t)\right)\ \left(H_0(t) - \tau(\theta)\right) =$$

$$\sum\nolimits_t \left(\mathbf{E}_\theta(T_0/T = t) - H_0(t)\right)\ \mathbf{P}_\theta(T = t)\ \left(H_0(t) - \tau(\theta)\right) = 0.$$

Отже, $\mathbf{D}_\theta T_0 \ge \mathbf{D}_\theta T_1$ для довільної *незміщеної* оцінки T_0 ⊡

Теорема (теорема Рао – Блекуела – Колмогорова). *Якщо існує оптимальна оцінка $T^* \in \Gamma_\tau$, а $T = T(X)$ – деяка достатня статистика, то T^* є вимірною функцією від цієї статистики: $T^* = H^*(T)$ м.н.*

Доведення. Оберемо в теоремі *про дисперсію функції від достатньої статистики* незміщену оцінку $T_0 \equiv T^*$, звідки $\mathbf{D}_\theta T^* \ge \mathbf{D}_\theta H_0(T)$ для *вимірної* функції H_0, що побудована в цій теоремі, причому $H_0(T) \in \Gamma_\tau$. Однак за означенням *оптимальної* оцінки $\mathbf{D}_\theta H_0(T) \ge \mathbf{D}_\theta T^*$. Отже, обидві нерівності є рівностями для всіх θ і кожна з оцінок $T^*, H_0(T)$ є оптимальною. З теореми *про єдиність оптимальної оцінки* виводимо: $T^* = H^*(T)$ *м.н.* при $H^* \equiv H_0$ ⊡

Означення. *Статистика $T = T(X)$ називається* **повною** *статистикою, якщо для кожної вимірної функції $\varphi(\cdot)$ з тотожностей*

$$\mathbf{E}_\theta\, \varphi(T) = 0, \quad \forall\theta \in \Theta,$$

випливає, що $\varphi(T) = 0$ майже напевне.

Теорема (про оптимальність повної достатньої статистики). *Якщо існує повна достатня статистика, то довільна вимірна функція від неї є оптимальною оцінкою свого математичного сподівання.*

Доведення. Нехай $T = T(X)$ – повна достатня статистика, а g – вимірна функція. Визначимо функцію $\tau(\theta) = \mathbf{E}_\theta g(T(X))$. Якщо $T_0 \in \Gamma_\tau$ – довільна *незміщена* оцінка для значення $\tau = \tau(\theta)$, то внаслідок теореми *про дисперсію функції від достатньої статистики* знайдеться вимірна функція H_0 (що визначається через T_0) така, що статистика $T_1 \equiv H_0(T)$ знову є незміщеною, причому $\mathbf{D}_\theta T_1 \le \mathbf{D}_\theta T_0$, $\forall\theta \in \Theta$. За умовою незміщеності $\mathbf{E}_\theta T_1 = \mathbf{E}_\theta H_0(T) = \tau(\theta)$, де за означенням $\tau(\theta) = \mathbf{E}_\theta g(T)$. Звідси $\mathbf{E}_\theta(H_0(T) - g(T)) = 0$, $\forall\theta \in \Theta$. Тому з *повноти* T випливає $H_0(T) - g(T) = 0$ *м.н.*, тобто $T_1 \equiv H_0(T) = g(T)$ *м.н.*, причому функція g вже не залежить від вибору T_0. Отже, $\mathbf{D}_\theta g(T) \le \mathbf{D}_\theta T_0$, $\forall\theta \in \Theta$, $\forall T_0 \in \Gamma_\tau$, тобто $g(T)$ – *оптимальна незміщена* оцінка для $\tau(\theta)$ ⊡

Приклад. *Рівномірна на* $[0, \theta]$ *вибірка* (*існування суперефективної оптимальної оцінки в нерегулярній моделі*). Нехай X – *кратна вибірка* з *рівномірним розподілом* $\xi_k \simeq U(0, \theta)$, $k = \overline{1, n}$, $\theta \in \Theta = \mathbb{R}_+$. Як зроблено вище в прикладі *достатніх статистик*, перевіряємо, що в даному випадку статистика $T(X) = \max \xi_k$ є достатньою. Її *функція розподілу* дорівнює $\mathbf{P}_\theta(T < x) = (x/\theta)^n$, $0 \le x \le \theta$. Ця статистика є *повною*, оскільки з $\mathbf{E}_\theta \varphi(T) = \theta^{-n} \int_0^\theta \varphi(x) n x^{n-1} dx = 0$, $\forall\theta > 0$, випливає, що $\varphi(x) = 0$ майже для всіх $x \ge 0$ та $\varphi(T) = 0$ *м.н.*

Отже, *статистика* $T(X) = \max \xi_k$ є *оптимальною незміщеною* оцінкою свого математичного сподівання $\mathbf{E}_\theta\ T(X) = \tau(\theta) = \frac{n\theta}{n+1}$. Слід зазначити, що *дисперсія* цієї оцінки дорівнює **(вправа)**

$$\mathbf{D}_\theta T = \theta^2 n/(n+2)(n+1)^2 = O(1/n^2),\ n \to \infty,$$

і має суттєво вищий порядок малості, ніж гарантує *нерівність Крамера – Рао*. Цей факт пов'язаний з тим, що в даній схемі не виконується *умова регулярності*, тому критерій Крамера – Рао не може бути застосований.

Вправи

(1) Статистика $\varphi(T)$ для достатньої статистики T та бієкції φ є достатньою.

(2) Для наведених прикладів знайти умовний розподіл X за умови $T = t$.

(3) Для кратної нормальної вибірки з невідомими середнім та дисперсією (μ, σ^2) (а) довести, що вибіркові середнє та дисперсія утворюють повну достатню статистику, (б) знайти оптимальну оцінку для $\Phi(-\mu/\sigma)$.

(4) Для n-кратної нормальної вибірки з невідомими середнім θ та одиничною дисперсією довести, що оптимальна оцінка для значення $\theta^k, k \in \mathbb{N}$, має вигляд $(-1)^k n^{-k/2} H_k(-\sqrt{n}\widehat{\mu}_n)$, де $H_k(x) \equiv (-1)^k \exp(x^2/2)\ d^k/dx^k \exp(-x^2/2)$.

(5) Для n-кратної вибірки з (а) геометричним розподілом $G(\theta)$, (б) показниковим розподілом $Exp(\theta)$ вибіркова сума $\widehat{S}_n$ є повною достатньою статистикою.

(6) Для кратної вибірки $X = (\xi_1, ..., \xi_n)$ із гама-розподілом $\xi_k \simeq \Gamma(\lambda, \alpha)$ пара статистик $(\sum_{k=1}^n \xi_k,\ \sum_{k=1}^n \ln \xi_k)$ утворює повну достатню статистику. Знайти звідси оптимальні оцінки для (а) α/λ, (б) $\lambda/(\alpha - 1/n)$, (б) α^k при заданих k та $\lambda = 1$.

(7) Для (а) біноміальної, (б) пуассонівської кратної вибірки з параметром θ вибіркова сума $\widehat{S}_n$ є повною достатньою статистикою. Зокрема, за умови (б) оптимальною оцінкою для полінома $\sum_{k=1}^m c_k \theta^k$ є статистика $\sum_{k=1}^m c_k \pi_k(\widehat{S}_n) n^{-k}$, де $\pi_k(s) \equiv s(s-1)...(s-k+1)$.

(8) Для кратної вибірки X із щільністю $f(y, \theta) = (\frac{2}{3} 1\!\!\mathrm{I}_{y \in [0,\theta)} + \frac{4}{3} 1\!\!\mathrm{I}_{y \in [\theta, 2\theta)})/2\theta$ існує одновимірна достатня статистика. Знайти її.

(9) Для кратної вибірки X із неперервною функцією розподілу спостережень її варіаційний ряд є повною достатньою статистикою.

(10) Довести, що для кратної вибірки X оптимальна оцінка завжди є симетричною функцією спостережень. Для цього довести, що дисперсія симетризованої оцінки не перевищує дисперсії вихідної незміщеної оцінки.

(11) Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з степеневим розподілом спостережень: $\mathbf{P}_\theta(\xi_1 = y) = a(y)\theta^y / f(\theta), y \in \mathbb{Z}_+$, де $\theta \in \Theta = (0, r) \subset \mathbb{R}_+$, причому $f(\theta) = \sum_{y \geq 0} a(y)\theta^y < \infty$ для всіх $\theta < r$. Довести, що: (а) статистика $T = \sum_{k=1}^n \xi_k$ є достатньою, (б) розподіл T має аналогічний степеневий вигляд,

(в) статистика T є повною. (г) Знайти оптимальну оцінку для $\tau(\theta) = \theta^k$.

(12) Вибірка $X = (\xi_1, ..., \xi_n)$ містить незалежні $\xi_j \simeq N(\alpha + \beta c_j, \sigma^2)$. Довести, що статистика $\left(\sum_{j=1}^n \xi_j, \sum_{j=1}^n c_j \xi_j, \sum_{j=1}^n \xi_j^2\right)$ є повною достатньою.

(13) Для кратної вибірки X із щільністю Коші $f(y) = (\pi(1 + (y - \theta)^2))^{-1}$ єдиною достатньою статистикою є X.

(14) Довести, що статистика S є мінімальною достатньою статистикою для вибірки X тоді й тільки тоді, коли $\sigma[S] = \cap_{T \text{ – достатня}} \sigma[T]$. Вивести звідси існування мінімальної достатньої статистики.

(15) Довести, що статистика S є мінімальною достатньою статистикою для вибірки X, якщо для довільної достатньої статистики T статистика S є достатньою для T. Знайти мінімальну достатню статистику: (а) у схемі Бернуллі, для щільностей спостережень (б) $2(1 - x/\theta)^+$, (в) $\exp(-x + \theta - \exp(-x + \theta))$.

3.9. Оцінки максимальної вірогідності

Визначення оцінки максимальної вірогідності ґрунтується на **принципі максимальної вірогідності** – "*те, що спостерігається, є найбільш імовірним серед усіх можливих альтернатив*".

Надалі будемо припускати, що вибірка X задовольняє *умову підпорядкованості* її розподілу деякій мірі у *вибірковому просторі*. За такої умови повністю визначена *вибіркова функція вірогідності* $L(X,\ \theta)$. Значення цієї функції і дають критерій "найбільшої вірогідності".

Означення. Оцінкою максимальної вірогідності (ОМВ, *EML*) *параметра θ за вибіркою X називається статистика, що максимізує вибіркову функцію вірогідності* $L(X, \theta)$:

$$\widehat{\theta} = \widehat{\theta}(X) \equiv \arg\max{}_{\theta \in \Theta} L(X,\ \theta),$$

тобто це така статистика $\widehat{\theta} = \widehat{\theta}(X)$, що задовольняє умову:

$$L(X,\ \theta) \leq L(X,\ \widehat{\theta}), \quad \forall \theta \in \Theta.$$

Для кратної вибірки ОМВ позначається як $\widehat{\theta}_n$, де n – об'єм вибірки.

Іноді простішим є обчислення ОМВ з еквівалентного означення

$$\widehat{\theta} = \widehat{\theta}(X) \equiv \arg\max{}_{\theta \in \Theta}\ \ln L(X,\ \theta),$$

що збігається з основним через монотонність логарифмічної функції.

Необхідною умовою існування *ОМВ* є припущення існування точки максимуму. ОМВ визначена однозначно за умови єдиності цієї точки.

Теорема (про рівняння максимальної вірогідності). *Якщо параметр є векторним:* $\Theta \subset \mathbb{R}^d$, *максимум в означенні оцінки максимальної вірогідності досягається всередині параметричної множини, а функція вірогідності диференційовна, то ОМВ задовольняє* ***рівняння максимальної вірогідності****:*

$$\frac{\partial}{\partial\theta} \ln L(X,\theta) \mid_{\theta = \widehat{\theta}} = 0,$$

тобто ОМВ є коренем функції впливу $U(X,\theta)$:

$$U(X,\widehat{\theta}) = 0.$$

Якщо $\dim\Theta = d > 1$, *то рівняння максимуму вірогідності є векторним, і перетворюється на* ***систему рівнянь максимальної вірогідності*** *для кожної координати вектора-градієнта.*

Зауваження. Для практичного обчислення *ОМВ* як кореня *рівняння максимальної вірогідності* $U(X,\widehat{\theta}) = 0$ часто використовують **метод Ньютона – Рафсона**, який складається з двох етапів.

(1) Обирають довільну *консистентну* оцінку $\widehat{\theta}^{(0)} = \widehat{\theta}^{(0)}(X)$.

(2) Рекурентно при $m \ge 0$ обчислюють $\widehat{\theta}^{(m+1)} = \widehat{\theta}^{(m)} - U(X,\widehat{\theta}^{(m)})/I(\widehat{\theta}^{(m)})$. Тут $I(\theta)$ – функція *інформації за Фішером.*

За певних умов послідовність $\widehat{\theta}^{(m)}$ м.н. збігається до *ОМВ*.

3.9.1. Співвідношення з ефективними оцінками. Інваріантність

Теорема (про співвідношення ОМВ з ефективними оцінками).

(а) Якщо параметр θ *– скалярний і існує його ефективна оцінка* $\widehat{\theta}^*$, *то існує оцінка максимальної вірогідності* $\widehat{\theta}$, *причому* $\widehat{\theta} = \widehat{\theta}^*$.

(б) Якщо T *– достатня статистика і існує ОМВ* $\widehat{\theta}$, *то остання є вимірною функцією від* $T : \widehat{\theta} = H(T)$.

Доведення

(а) За теоремою *про нерівність та критерій Крамера – Рао* (а саме, за критерієм (б) існування $\widehat{\theta}^*$) похідна за аргументом θ від логарифмічної *функції вірогідності* має вигляд

$$\frac{\partial}{\partial\theta} \ln L(X,\theta) \equiv U(X,\theta) = (\widehat{\theta}^* - \theta)/c(\theta),$$

де функція $c(\theta) = \tau_\theta / I(\theta) = 1/I(\theta)$ невід'ємна. Тому логарифмічна функція вірогідності зростає при $\theta < \widehat{\theta}^*$ і спадає при $\theta > \widehat{\theta}^*$. Отже, найбільшим є її значення при $\theta = \widehat{\theta}^*$ і $\widehat{\theta}^*$ є *оцінкою максимальної вірогідності.*

(б) Твердження випливає з факторизації функції вірогідності за означенням *достатньої статистики*:

$$\widehat{\theta} = arg\max{}_{\theta\in\Theta}\ g(T(X),\theta)\ h(X)\ j(\theta) =$$

$$arg\max{}_{\theta\in\Theta}\ g(T(X),\theta)\ j(\theta) = H(T(X)),$$

оскільки точка максимуму за θ не змінюється при множенні функції на додатну сталу $h(X)$, а вираз для функції під другим знаком максимуму залежить лише від θ та $T(X)$ ⊡

Теорема (про інваріантність оцінки максимальної вірогідності). *Нехай існує оцінка максимальної вірогідності $\widehat{\theta}$ параметра θ, а функція $q : \Theta \to Q$ є взаємно-однозначною. Тоді оцінка $q(\widehat{\theta})$ є оцінкою максимальної вірогідності для значення функції $q(\theta)$.*

Доведення. Позначимо через $\theta(q)$ функцію, обернену до $q(\theta)$.

Функція вірогідності через параметр q дорівнює $L_q(X,q) = L(X,\theta(q))$ і за умови однозначності задовольняє тотожність

$$L_q(X,q(\theta)) = L(X,\theta(q(\theta))) = L(X,\theta),\ \forall\theta\in\Theta.$$

Оскільки множина $\{q(\theta),\theta\in\Theta\} = Q$, то з означення *ОМВ* дістаємо

$$L_q(X,q(\theta)) = L(X,\theta) \le L(X,\widehat{\theta}) = L_q(X,q(\widehat{\theta})),$$

$$L_q(X,q) \le \max{}_{\theta\in\Theta}\ L_q(X,q(\theta)) \le L_q(X,q(\widehat{\theta})),$$

Дані нерівності є рівностями при $q = q(\widehat{\theta})$. Отже, $q(\widehat{\theta})$ є ОМВ для $q(\theta)$ ⊡

3.9.2. Приклади обчислення оцінок максимальної вірогідності

1. *Схема Бернуллі:*

$$\arg\max\ln L(X,\theta) = \arg\max\ln\theta^{\nu_n(X)}(1-\theta)^{n-\nu_n(X)} =$$

$$\arg\max(\nu_n(X)\ln\theta + (n-\nu_n(X))\ln(1-\theta)) = \nu_n(X)/n = \widehat{\theta}_n.$$

2. *Пуассонівська вибірка:*

$$\ln L(X,\theta) = \ln\prod\nolimits_{k=1}^n \frac{\theta^{\xi_k}}{\xi_k!}e^{-\theta} = \ln\theta\sum\nolimits_{k=1}^n \xi_k - n\theta + \ln h(X),$$

$$\widehat{\theta}_n = \arg\max\ln L(X,\theta) = \frac{1}{n}\sum\nolimits_{k=1}^n \xi_k = \widehat{\mu}_n.$$

3. *Показникова вибірка* (зміщеність *ОМВ*):

$$\ln L(X,\theta) = \ln \prod\nolimits_{k=1}^{n} \theta \exp(-\theta\xi_k) = n \ln\theta - \theta \sum\nolimits_{k=1}^{n} \xi_k,$$

$$\widehat{\theta}_n = \arg\max \ln L(X,\theta) = n \;/ \sum\nolimits_{k=1}^{n} \xi_k = 1 \;/\; \widehat{\mu}_n.$$

Оскільки сума $\sum_{k=1}^{n} \xi_k$ має *розподіл Ерланга* з параметрами n, θ, то

$$\mathbf{E}_\theta \widehat{\theta}_n = \int_0^\infty \frac{n}{x} \theta \frac{(\theta x)^{n-1}}{(n-1)!} \exp(-\theta x) dx = \theta \frac{n}{n-1} \neq \theta \;\forall n \geq 1 \;\forall \theta > 0.$$

Отже, у загальному випадку не можна розраховувати на *незміщеність оцінок максимальної вірогідності*.

4. *Нормальна вибірка*. Як показано вище в розділі про оцінювання середнього для нормальних спостережень,

$$\ln L(X,\theta) + \frac{n}{2} \ln 2\pi = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2\sigma^2} (\widehat{\mu}_{2n} - 2\widehat{\mu}_n \mu + \mu^2) =$$

$$-\frac{n}{2} \ln \sigma^2 - \frac{n}{2\sigma^2} \left(\widehat{\mu}_{2n} - (\widehat{\mu}_n)^2 + (\widehat{\mu}_n - \mu)^2\right) =$$

$$-\frac{n}{2} \ln \sigma^2 - \frac{n}{2\sigma^2} (\widehat{\sigma}^2_n + (\widehat{\mu}_n - \mu)^2) \leq -\frac{n}{2} \ln \sigma^2 - \frac{n}{2\sigma^2} \widehat{\sigma}^2_n =$$

$$-\frac{n}{2} \left(\ln \sigma^2 + \frac{1}{\sigma^2} \widehat{\sigma}^2_n \right) \leq -\frac{n}{2} \left(\ln \widehat{\sigma}^2_n + \frac{1}{\widehat{\sigma}^2_n} \widehat{\sigma}^2_n \right) = -\frac{n}{2} (\ln \widehat{\sigma}^2_n + 1),$$

оскільки функція $\ln s \; + \frac{a}{s}$ набуває найменшого значення при $s = a$. Дані нерівності є рівностями при $\mu = \widehat{\mu}_n, \;\sigma^2 = \widehat{\sigma}^2_n \equiv n^{-1} \sum_{k=1}^{n} (\xi_k - \widehat{\mu}_n)^2$. Тому *оцінка максимальної вірогідності* параметра $\theta = (\mu, \sigma^2)$ дорівнює

$$\widehat{\theta}_n \equiv \arg\max \ln L(X,\theta) = (\widehat{\mu}_n, \;\widehat{\sigma}^2_n).$$

Зокрема, за теоремою *про інваріантність оцінки максимальної вірогідності* ОМВ для границь *надійного інтервалу* з *правила трьох сигма* дорівнюють

$$\widehat{\mu \pm 3\sigma} = \widehat{\mu}_n \pm 3\sqrt{\widehat{\sigma}^2_n}.$$

Зауваження. Метод максимальної вірогідності можна застосовувати також для спостережень, що не є кратними вибірками. Наприклад, у випадку **цензурування даних** для k вимірювань з загальної кількості n можуть бути відомі точні значення $(\xi_1, ..., \xi_k)$, а решта $n-k$ вимірювань могла б привести до перевищення шкали приладу – тобто відомою є тільки інформація щодо $\{\xi_j > x\}, j = \overline{k+1, n}$, при заданому рівні цензурування x. В такому разі функція вірогідності матиме вигляд

$$L(X,\theta) = \left(\prod\nolimits_{j=1}^{k} f(\xi_j, \theta) \right) \left(\int_x^\infty f(y,\theta) dy \right)^{n-k}.$$

Вправи

(1) Довести, що ОМВ центра симетрії θ розподілу, що має щільність Лапласа: $(1/2)\exp(-|x-\theta|)$, збігається з вибірковою медіаною.

(2) Знайти ОМВ параметрів кратної вибірки з логнормальним розподілом.

(3) Кратна вибірка X утворена спостереженнями з розподілом $N(\theta, 2\theta)$. Довести, що ОМВ для θ має вигляд $\sqrt{1 + \sum_{k=1}^{n} \xi_k^2 \, /n} - 1$ і є консистентною.

(4) Обчислити ОМВ значення функції $\tau(\theta) = \Phi(-\mu/\sigma) \equiv \mathbf{P}_\theta(\xi_1 < 0)$ для нормальної кратної вибірки $X = (\xi_1, ..., \xi_n)$, $\xi_k \simeq N(\mu, \sigma^2)$ із невідомим параметром $\theta = (\mu, \sigma^2)$. Знайти асимптотичний розподіл для цієї оцінки після її центрування на $\tau(\theta)$ та нормування на $\sqrt{n}$.

(5) Популяція бегемотів містить невідому кількість N особин. Випадково обрані n з них були помічені та відпущені. З числа m наступних випадково відловлених особин виявилось k помічених. Знайти ОМВ для N.

(6) Знайти ОМВ для невідомої кількості спостережень n за біноміальною одноелементною вибіркою $X \simeq B(n, p)$.

(7) Функція вірогідності $L(X, \theta)$ неперервна за θ, а ОМВ $\widehat{\theta}_n$ визначена однозначно та є достатньою статистикою. Довести, що $\widehat{\theta}_n$ – мінімальна достатня.

(8) Довести, що ОМВ для вибірки з рівномірним розподілом $U(\theta, 1+\theta)$ утворюють цілий інтервал.

(9) Спостерігаються випадкові величини $\xi_k \simeq N(\theta_k, 1), k = \overline{1,2}$. (а) знайти ОМВ для $\theta = (\theta_1, \theta_2)$, (б) знайти вказану ОМВ у припущенні, що $\theta_1 \le \theta_2$.

(10) Вибірка має умовний розподіл Пуассона $\mathbf{P}(\xi = n \mid \xi \ge 1), n \in \mathbb{N}$, де $\xi \simeq \Pi(\lambda)$. Довести, що ОМВ для λ збігається з оцінкою методу моментів.

(11) Незалежні кратні вибірки $X = (\xi_1, ..., \xi_n), Y = (\zeta_1, ..., \zeta_m)$ мають розподіли $N(\mu_1, \sigma^2)$ та $N(\mu_2, \sigma^2)$ відповідно. Знайти ОМВ для $\theta = (\mu_1, \mu_2, \sigma^2)$.

(12) Нехай $(\xi_1, ..., \xi_n)$ – кратна вибірка розміру n з розподілу зі щільністю $f(y, \theta) = \alpha\varphi((y-\mu)/\sigma)/\sigma + (1-\alpha)\varphi(y-\mu)$, де φ – стандартна нормальна щільність розподілу. Довести, що ОМВ для параметра $\theta = (\mu, \sigma^2)$ не існує.

3.9.3. Умови консистентності ОМВ

Нехай $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка* зі щільністю $f(y, \theta)$ одного спостереження, тобто $\mathbf{P}_\theta(\xi_1 \in B) = \int_B f(y, \theta)\nu(dy), \forall B \in \mathfrak{B}$.

У даному розділі через θ_0 позначимо істинне значення параметра θ.

Розглянемо такі **умови консистентності ОМВ**:

(c1) Для всіх $\theta \ne \theta_0$ відношення $f(y, \theta)/f(y, \theta_0)$ не дорівнює майже

скрізь одиниці на просторі значень спостережень $y \in R$ – тобто

$$\nu\left(\{y : f(y,\theta) \neq f(y,\theta_0)\}\right) > 0,\ \forall\theta \neq \theta_0.$$

(c2) Параметрична множина Θ є компактною підмножиною $\mathbb{R}^d$.

(c3) При кожному y функція $f(y,\cdot)$ є строго додатною та неперервно диференційовною за θ, причому величини $\ln f(\xi_1,\theta)$ та $\partial \ln f(\xi_1,\theta)/\partial\theta$ мажоруються *інтегровною* випадковою величиною η :

$$|\ln f(\xi_1,\theta)| \leq \eta,\ |\partial \ln f(\xi_1,\theta)/\partial\theta| \leq \eta,\ \forall\theta \in \Theta, \quad \mathbf{E}_{\theta_0}\eta < \infty.$$

3.9.4. Інформація за Кульбаком

Означення. *Функцією* інформації за Кульбаком *називається функція*

$$I(\theta,\theta_0) = -\,\mathbf{E}_{\theta_0} \ln \frac{f(\xi_1,\theta)}{f(\xi_1,\theta_0)}.$$

Теорема (про властивості інформації за Кульбаком). *За умов консистентності ОМВ інформація за Кульбаком має такі властивості.*

(а) $I(\theta,\theta_0) \geq 0$,

(б) $I(\theta,\theta_0) = 0$ *тоді й тільки тоді, коли* $\theta = \theta_0$,

(в) $I(\theta,\theta_0)$ *неперервна за* θ.

Доведення. Розглянемо випадкову величину

$$\zeta = f(\xi_1,\theta)\ /\ f(\xi_1,\theta_0).$$

За формулою з теореми *про обчислення математичного сподівання функції від випадкової величини* та з означення щільності $f(y,\theta)$ обчислимо

$$\mathbf{E}_{\theta_0}\zeta = \int_R \frac{f(y,\theta)}{f(y,\theta_0)} f(y,\theta_0)\nu(dy) = \int_R f(y,\theta)\nu(dy) = \mathbf{P}_\theta(\xi_1 \in R) = 1.$$

З опуклості функції $\ln x$ виводимо для функції $h(x) \equiv x - 1 - \ln x$ елементарну нерівності $h(x) \geq 0$ при $x \in \mathbb{R}_+$ та $h(x) > 0$ при $x \neq 1$. Звідки за *невід'ємністю* математичного сподівання

$$I(\theta,\theta_0) = -\,\mathbf{E}_{\theta_0} \ln \zeta = \mathbf{E}_{\theta_0}(\zeta - 1 - \ln \zeta) = \mathbf{E}_{\theta_0} h(\zeta) \geq 0.$$

За властивістю *позитивності* математичного сподівання від невід'ємної ненульової м.н. випадкової величини $h(\zeta)$ *інформація за Кульбаком* $I(\theta,\theta_0)$ дорівнює нулю тоді й тільки тоді, коли $h(\zeta) = 0$ *м.н.* Останнє рівняння еквівалентне $\zeta = 1$ м.н., тобто умові $\mathbf{P}_{\theta_0}(\xi_1 \in \{y : f(y,\theta) \neq f(y,\theta_0)\}) = 0$. Тому $f(y,\theta) = f(y,\theta_0)$ майже для всіх y за мірою ν на просторі значень спостережень. За умовою консистентності *(c1)* остання властивість еквівалентна рівності $\theta = \theta_0$.

Неперервність $I(\theta, \theta_0)$ випливає з неперервності щільності спостережень та з *теореми Лебега про мажоровану збіжність*:

$$\lim_{n\to\infty} I(\theta_n, \theta_0) = -\lim_{n\to\infty} \mathbf{E}_{\theta_0} \ln \frac{f(\xi_1, \theta_n)}{f(\xi_1, \theta_0)} = -\mathbf{E}_{\theta_0} \lim_{n\to\infty} \ln \frac{f(\xi_1, \theta_n)}{f(\xi_1, \theta_0)} = I(\theta, \theta_0),$$

при $\theta_n \to \theta$, оскільки величини під знаком математичного сподівання за умов консистентності не перевищують *інтегровну* величину 2η $\boxdot$

3.9.5. Консистентність ОМВ

Теорема (про строгу консистентність ОМВ). *За умов консистентності ОМВ (c1-c3) оцінка максимальної вірогідності $\widehat{\theta}_n$ параметра θ за кратною вибіркою X існує і є строго консистентною.*

Доведення. Існування *ОМВ* є очевидним наслідком теореми *про функцію вірогідності кратної вибірки*, звідки випливає неперервність за θ функції вірогідності, та компактності параметричної множини Θ.

Для доведення консистентності позначимо

$$\chi_k(\theta) = \ln \frac{f(\xi_k, \theta_0)}{f(\xi_k, \theta)},\ \chi_k(U) = \inf\nolimits_{\theta\in U}\ \chi_k(\theta),\ k = \overline{1, n},$$

де $U \subset \Theta$ – деяка параметрична підмножина.

Зауважимо, що величини $(\chi_k(\theta))$, $(\chi_k(U))$ *незалежні в сукупності* та *однаково розподілені* при різних k всередині кожної послідовності за теоремами *про перетворення незалежних величин* та *про обчислення розподілу функції від випадкової величини*, і за означенням

$$I(\theta, \theta_0) = \mathbf{E}_{\theta_0} \chi_1(\theta) \geq 0.$$

Крім того, з умов *(c2)-(c3)* випливає, що

$$\chi_1(U) = \chi_1(\widetilde{\theta}(U)) \to^{P1} \chi_1(\theta), \quad U \downarrow \{\theta\},$$

оскільки при $\theta \in U$

$$|\chi_1(U) - \chi_1(\theta)| \leq (\mathrm{diam} U) \sup_{\theta\in U} |\partial \ln f(\xi_1, \theta)/\partial\theta| \leq \eta \ \mathrm{diam} U.$$

Тому за *теоремою Лебега про мажоровану збіжність*

$$\mathbf{E}_{\theta_0}\ \chi_1(U) \to \mathbf{E}_{\theta_0}\ \chi_1(\theta) = I(\theta, \theta_0) > 0,\ U \downarrow \{\theta\},\ \forall\theta \neq \theta_0.$$

Більше того, з умови мажорованості похідної виводимо, що

$$0 \leq \mathbf{E}_{\theta_0}\ \chi_1(\theta) - \mathbf{E}_{\theta_0}\ \chi_1(U) \leq \mathbf{E}_{\theta_0}\ |\chi_1(U) - \chi_1(\theta)| \leq$$
$$(\mathrm{diam} U)\ \mathbf{E}_{\theta_0} \sup\nolimits_{\theta\in U}\ |\partial \ln f(\xi_1, \theta)/\partial\theta| \leq K \ \mathrm{diam} U,$$

де стала $K = \mathbf{E}_{\theta_0}\eta < \infty$ за умовою *(c3)*.

Нехай $\varepsilon > 0$. Визначимо компактну внаслідок *(c2)* множину

$$\Theta(\varepsilon) \equiv \Theta \cap \{\theta : |\theta - \theta_0| \geq \varepsilon\}.$$

З неперервності функції *інформації за Кульбаком* та з компактності $\Theta(\varepsilon)$ виводимо додатність величини

$$\alpha \equiv \inf\ _{\theta\in\Theta(\varepsilon)} I(\theta, \theta_0) > 0.$$

Дійсно, при $\alpha = 0$ у множині $\Theta(\varepsilon)$ знайдеться точка θ (часткова границя для мінімізуючої послідовності), в якій $I(\theta, \theta_0) = 0$, що суперечить твердженню (б) теореми *про властивості інформації за Кульбаком*.

Оскільки множина $\Theta(\varepsilon)$ – компактна, то за теоремою про відкрите покриття компакту для довільного $\delta > 0$ знайдеться її скінченне покриття $U(\delta) = (U_j(\delta),\ j = \overline{1, m}),\ \ \Theta(\varepsilon) = \cup_{j=1}^m U_j(\delta)$, із діаметром $\mathrm{diam} U_j(\delta) \leq \delta$. Оберемо $\delta < \alpha/2K$. Тоді за наведеною вище нерівністю та за означенням нижньої межі при виборі довільних $\theta_j \in U_j$

$$\mathbf{E}_{\theta_0}\ \chi_1(U_j(\delta)) \geq \mathbf{E}_{\theta_0}\ \chi_1(\theta_j) - |\mathbf{E}_{\theta_0}\chi_1(U_j(\delta)) - \mathbf{E}_{\theta_0}\chi_1(\theta_j)| \geq$$
$$I(\theta_j, \theta_0) - K\ \mathrm{diam} U_j(\delta) \geq \alpha - K\delta \geq \alpha/2 > 0, \quad \forall j = \overline{1, m}.$$

З виконання строгих нерівностей $L(X, \theta_0) > L(X, \theta)$ для всіх $\theta \in \Theta(\varepsilon)$ випливає, що максимум функції $L(X, \theta)$ за аргументом θ досягається поза множиною $\Theta(\varepsilon)$, адже $\theta_0 \notin \Theta(\varepsilon)$. Тому з теореми *про функцію вірогідності кратної вибірки* та з означення величин $\chi_k(\theta)$ виводимо, що

$$\left\{|\widehat{\theta}_n - \theta_0|\ < \varepsilon\right\} = \{\arg\max_{\theta\in\Theta} L(X, \theta) \notin \Theta(\varepsilon)\} \supset \left\{\inf_{\theta\in\Theta(\varepsilon)} \tfrac{L(X,\theta_0)}{L(X,\theta)} > 1\right\} =$$
$$\left\{\inf_{\theta\in\Theta(\varepsilon)} \ln \tfrac{L(X,\theta_0)}{L(X,\theta)} > 0\right\} = \left\{\inf_{\theta\in\Theta(\varepsilon)} \textstyle\sum_{k=1}^n \ln \tfrac{f(\xi_k,\theta_0)}{f(\xi_k,\theta)} > 0\right\} =$$
$$\left\{\inf_{\theta\in\Theta(\varepsilon)} \textstyle\sum_{k=1}^n \chi_k(\theta) > 0\right\} = \left\{\min_{1\leq j\leq .m} \inf\ _{\theta\in U_j(\delta)} \textstyle\sum_{k=1}^n \chi_k(\theta) > 0\right\}.$$

Отже, за *монотонністю* ймовірності

$$\mathbf{P}_{\theta_0}\left(|\widehat{\theta}_n - \theta_0| < \varepsilon\right) \geq \mathbf{P}_{\theta_0}\left(\min_{1\leq j\leq m} \inf_{\theta\,\in\,U_j(\delta)} \textstyle\sum_{k=1}^n \chi_k(\theta) > 0\right) \geq$$
$$\mathbf{P}_{\theta_0}\left(\min_{1\leq j\leq .m} \textstyle\sum_{k=1}^n \chi_k(U_j(\delta)) > 0\right) =$$
$$\mathbf{P}_{\theta_0}\left(\bigcap_{j=1}^m \left\{\tfrac{1}{n}\textstyle\sum_{k=1}^n \chi_k(U_j(\delta)) > 0\right\}\right) \to 1,\ n \to \infty,$$

оскільки за *критерієм Колмогорова посиленого закону великих чисел* при кожному j має місце збіжність

$$\frac{1}{n}\sum\nolimits_{k=1}^{n}\chi_k(U_j(\delta)) \to^{P1} \mathbf{E}_{\theta_0}\ \chi_1(U_j(\delta)) \geq \alpha/2 > 0,\ n\to\infty,$$

а скінченний переріз подій одиничної ймовірності має ймовірність 1.

Отже, $\mathbf{P}_{\theta_0}\left(|\widehat{\theta}_n-\theta_0| \ <\varepsilon\right)\to 1, n\to\infty, \forall\varepsilon>0$, і *ОМВ* – консистентна.

Застосувавши наведені вище міркування до подій

$$\underline{\lim}_{n\to\infty}\left\{|\widehat{\theta}_n-\theta_0| \ <\varepsilon\right\}\supset \underline{\lim}_{n\to\infty}\bigcap\nolimits_{j=1}^{m}\left\{\frac{1}{n}\sum\nolimits_{k=1}^{n}\chi_k(U_j(\delta))>0\right\},$$

аналогічно отримуємо нерівність

$$\mathbf{P}_{\theta_0}\left(\underline{\lim}_{n\to\infty}\left\{|\widehat{\theta}_n-\theta_0| \ <\varepsilon\right\}\right)\geq$$

$$\mathbf{P}_{\theta_0}\left(\underline{\lim}_{n\to\infty}\bigcap\nolimits_{j=1}^{m}\left\{\tfrac{1}{n}\sum\nolimits_{k=1}^{n}\chi_k(U_j(\delta))>0\right\}\right),$$

де права частина дорівнює 1 за *критерієм Колмогорова посиленого закону великих чисел*. Оскільки при кожному $\varepsilon>0$

$$\left\{\overline{\lim}_{n\to\infty}|\widehat{\theta}_n-\theta_0| \ \geq 2\varepsilon\right\}\subset \overline{\lim}_{n\to\infty}\left\{|\widehat{\theta}_n-\theta_0| \ \geq\varepsilon\right\},$$

де подія справа має нульову ймовірність, то $\widehat{\theta}_n \to^{P1} \theta_0$, що і доводить *строгу консистентність* ОМВ ⊡

3.9.6. Асимптотична нормальність і ефективність ОМВ

У даному розділі припускатимемо, що параметрична множина $\Theta\subset\mathbb{R}^1$.

Теорема (про асимптотичну нормальність ОМВ).

Нехай $X=(\xi_1,...,\xi_n)$ – кратна вибірка, щільність одного спостереження $f(y,\theta)$ задовольняє умови регулярності, умови консистентності ОМВ (c1-c3) та тричі неперервно диференційовна за θ, причому відповідні похідні мажоруються за модулем інтегровною величиною η :

$$\left|\partial^{(i)}\ln f(\xi_1,\theta)/\partial^{(i)}\theta\right|\leq\eta,\quad i=\overline{1,3},\quad \mathbf{E}_{\theta_0}\eta<\infty.$$

Якщо ОМВ $\widehat{\theta}_n$ невідомого параметра θ_0 є розв'язком рівняння максимальної вірогідності, то ця оцінка є асимптотично нормальною:

$$\sqrt{n}(\widehat{\theta}_n-\theta_0)\to^{W}\zeta\simeq N(0,\ 1/i(\theta_0)),\ n\to\infty,$$

де асимптотична дисперсія визначається кількістю інформації за Фішером, що міститься в одному спостереженні:

$$i(\theta) \equiv \mathbf{E}_\theta \left(\frac{\partial}{\partial \theta} \ln f(\xi_1, \theta) \right)^2 .$$

Зауваження. За теоремою *про адитивність.інформації за Фішером* повна інформація за Фішером у вибірці дорівнює $I_n(\theta) = ni(\theta)$, тому з даної теореми випливає *асимптотична ефективність* ОМВ, оскільки $1/ni(\theta_0) = 1/I_n(\theta_0)$ збігається з найменшою можливою межею для дисперсій незміщених оцінок параметра θ в теоремі *про нерівність та критерій Крамера – Рао.*

Доведення. Розглянемо при $i = \overline{1,3}$ такі випадкові величини:

$$U^{(i)}(X, \theta) = \partial^{(i)} \ln L_n(X, \theta) / \partial^{(i)} \theta,$$

$$\zeta_k^{(i)}(\theta) = u^{(i)}(\xi_k, \theta), \text{ де } u^{(i)}(y, \theta) = \partial^{(i)} \ln f(y, \theta) / \partial^{(i)} \theta.$$

За теоремою *про функцію вірогідності кратної вибірки* має місце адитивність функції впливу та її похідних:

$$U^{(i)}(X, \theta) = \sum\nolimits_{k=1}^{n} \zeta_k^{(i)}(\theta), \quad \forall i = \overline{1,3},$$

де доданки у сумах є *незалежними у сукупності* та *однаково розподіленими* випадковими величинами, як борелєві функції від незалежних та однаково розподілених величин ξ_k. За зробленим у теоремі припущенням перші доданки мажоруються інтегровною величиною η.

Зауважимо, що за теоремою *про рівняння максимальної вірогідності ОМВ* є коренем функції впливу $U = U^{(1)}$. Тому розкладом у ряд Тейлора отримуємо

$$0 = U(X, \widehat{\theta}_n) \equiv U^{(1)}(X, \widehat{\theta}_n) = U^{(1)}(X, \theta_0) +$$
$$(\widehat{\theta}_n - \theta_0) U^{(2)}(X, \theta_0) + \frac{1}{2} (\widehat{\theta}_n - \theta_0)^2 U^{(3)}(X, \widetilde{\theta}_n),$$

де $\widetilde{\theta}_n \in [\widehat{\theta}_n, \theta_0]$ і $\widetilde{\theta}_n \to^{P1} \theta_0, n \to \infty$, внаслідок теореми *про строгу конзистентність ОМВ*. Розв'язуючи останню тотожність відносно $\widehat{\theta}_n - \theta_0$, отримуємо таку рівність для $\widehat{\theta}_n$:

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = - \frac{\alpha_n}{\beta_n + (\widehat{\theta}_n - \theta_0)\, \widetilde{\gamma}_n},$$

де

$$\alpha_n = \frac{1}{\sqrt{n}} U^{(1)}(X, \theta_0) = \frac{1}{\sqrt{n}} \sum\nolimits_{k=1}^{n} \zeta_k^{(1)}(\theta_0),$$

$$\beta_n = \frac{1}{n}\sum\nolimits_{k=1}^{n} \zeta_k^{(2)}(\theta_0),\ \widetilde{\gamma}_n = \gamma_n(\widetilde{\theta}_n),\ \gamma_n(\theta) = \frac{1}{2n}\sum\nolimits_{k=1}^{n} \zeta_k^{(3)}(\theta).$$

За теоремою *про центрованість функції впливу* та *про обчислення інформації за Фішером* (в умовах регулярності)

$$\mathbf{E}_{\theta_0}\zeta_k^{(1)}(\theta_0) = \mathbf{E}_{\theta_0}u(\xi_1, \theta_0) = 0,$$

$$\mathbf{D}_{\theta_0}\zeta_k^{(1)}(\theta_0) = \mathbf{D}_{\theta_0}u(\xi_1, \theta_0) = i(\theta_0).$$

Оскільки величини $\zeta_k^{(1)}(\theta_0)$ *незалежні у сукупності*, *однаково розподілені* та *квадратично інтегровні* за умовами регулярності, то за *класичною центральною граничною теоремою*

$$\alpha_n \to^W \alpha \simeq N(0,\ i(\theta_0)),\ n \to \infty.$$

Далі, з умов регулярності та з теореми *про обчислення інформації за Фішером* випливає тотожність

$$\mathbf{E}_{\theta_0}\zeta_k^{(2)}(\theta_0) = \mathbf{E}_{\theta_0}\partial u(\xi_1, \theta_0)/\partial\theta_0 = -\mathbf{E}_{\theta_0}(u(\xi_1, \theta_0))^2 = -i(\theta_0),$$

до того ж абсолютна інтегровність вказаних величин є наслідком умови мажорованості.

Тому за *критерієм Колмогорова посиленого закону великих чисел*

$$\beta_n \to^{P1} -i(\theta_0),\ n \to \infty.$$

Нарешті, з умови мажорованості третіх похідних логарифмічної функції вірогідності: $\left|\zeta_1^{(3)}(\theta)\right| = \left|u^{(3)}(\xi_1, \theta)\right| \le \eta,\ \mathbf{E}_{\theta_0}\eta < \infty$, та з *критерію Колмогорова посиленого закону великих чисел* виводимо, що при кожному θ послідовність $\gamma_n(\theta)$ збігається м.н., а тому обмежена м.н. Оскільки функція $\gamma_n(\theta)$ неперервна за θ, і $\widetilde{\theta}_n \to^{P1} \theta_0 \in \Theta$, де Θ – компакт, то послідовність $\widetilde{\gamma}_n = \gamma_n(\widetilde{\theta}_n)$ також обмежена майже напевне. Тому з теореми *про строгу консистентність ОМВ* випливає збіжність

$$(\widehat{\theta}_n - \theta_0)\,\widetilde{\gamma}_n \to^{P1} 0,\ n \to \infty.$$

Тому знаменник у наведеному зображенні для $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ збігається:

$$\beta_n + (\widehat{\theta}_n - \theta_0)\,\widetilde{\gamma}_n \to^{P1} -i(\theta_0),\ n \to \infty.$$

За теоремою *про добуток слабко збіжної послідовності зі збіжною*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to^W -\frac{\alpha}{-i(\theta_0)} \simeq \frac{1}{i(\theta_0)}N(0, i(\theta_0)) \simeq N(0,\ i^{-1}(\theta_0)),\ n \to \infty \ \boxdot$$

Вправи

(1) Обчислити інформацію за Кульбаком для кратних вибірок (а) з нормальним розподілом $N(\mu, \sigma^2)$, (б) гама-розподілом $\Gamma(\lambda, \alpha)$, (г) біноміальним розподілом з параметрами n, p, при невідомих параметрах.

(2) Для кратної вибірки з гама-розподілом $\Gamma(1, \theta)$ спостережень знайти ОМВ та довести її асимптотичну нормальність і асимптотичну ефективність. Знайти оцінку методу моментів та довести, що її ефективність менша за 1.

(3) Знайти оцінку максимальної вірогідності для кратної вибірки з щільністю спостережень $\exp(-x + \theta - \exp(-x + \theta)),\ x, \theta \in \mathbb{R}$.

(4) Довести, що середнє квадратичне відхилення оцінки максимальної вірогідності $\widehat{\sigma}_n^2$ від параметра σ^2 для кратної вибірки з розподілом $N(\mu, \sigma^2)$ та невідомими μ, σ^2 строго більше за таке ж відхилення оцінки $n\widehat{\sigma}_n^2/(n+1)$. Порівняти це відхилення з дисперсією незміщеної оцінки $\widehat{s}_n^2$.

(5) Знайти інформацію за Кульбаком для кратної нормальної вибірки.

3.10. Статистики від нормальних вибірок

Особливість нормальних спостережень полягає у тому, що *вибіркове середнє* та *вибіркова дисперсія* незалежні й мають відомі розподіли.

Усі вектори надалі є векторами-стовпчиками, однак у друкованому вигляді їх координати будемо виписувати в рядок. Скалярний добуток векторів α, β дорівнює $\alpha'\beta$, де α' – спряжений (транспонований) до α вектор-рядок. Нагадаємо, шо матриця U називається ортонормованою, якщо $U^{-1} = U'$.

3.10.1. Стандартна нормальна вибірка

Теорема (про вибіркові моменти стандартної нормальної вибірки). *Нехай вибірка $\zeta = (\zeta_1, ..., \zeta_n)$ є стандартним нормальним вектором. Тоді її перші два вибіркові моменти*

$$\widehat{\mu}_n = \frac{1}{n} \sum\nolimits_{k=1}^{n} \zeta_k, \ \ \widehat{s}_n^2 = \frac{1}{n-1} \sum\nolimits_{k=1}^{n} (\zeta_k - \widehat{\mu}_n)^2$$

є незалежними випадковими величинами, причому

$$\sqrt{n}\ \widehat{\mu}_n \simeq N(0,1), \quad \text{та} \quad (n-1)\ \widehat{s}_n^2 \simeq \chi_{n-1}^2,$$

де величина χ_{n-1}^2 має хі-квадрат розподіл з $n-1$ ступенем свободи.

Доведення. Визначимо вектор $q = (1/\sqrt{n},\ k = \overline{1,n})$, що має одиничну евклідову норму. Нехай $(q_1, ..., q_{n-1})$ – доповнення вектора q до ортонормованого базису в $\mathbb{R}^n$, що існує за теоремою Грама – Шмідта. Позначимо $U = (q_1, ..., q_{n-1}, q)'$ *ортонормовану* матрицю з рядками $q_1', ..., q_{n-1}', q'$, що породжена цим базисом. Зауважимо, що за ортонормованістю значення $Uq = (0, ..., 0, 1) \equiv e$ є одиничним вектором. Крім того, за теоремою *про лінійні перетворення нормальних векторів* вектор $\eta = U\zeta$ знову є *стандартним нормальним вектором.*

Оскільки $\widehat{\mu}_n = \zeta' q/\sqrt{n}$ та $\zeta = U'\eta$, то виконуються рівності

$$(n-1)\widehat{s}_n^2 = \sum\nolimits_{k=1}^{n} (\zeta_k - \widehat{\mu}_n)^2 = \sum\nolimits_{k=1}^{n} \zeta_k^2 - n\,\widehat{\mu}_n^2 = \zeta'\zeta - (\zeta' q)(q'\zeta) =$$

$$\zeta'(I - q\cdot q')\zeta = \eta' U(I - q\cdot q')U'\eta = \eta'(I - e\cdot e')\eta = \sum\nolimits_{k=1}^{n-1} \eta_k^2 \simeq \chi_{n-1}^2$$

за означенням *хі-квадрат розподілу*, оскільки $(\eta_k, k = \overline{1,n})$ є *незалежними у сукупності* величинами зі *стандартним нормальним* розподілом внаслідок означення *стандартного нормального вектора* $\eta = (\eta_1, ..., \eta_n)$.

Крім того, за теоремою *про векторні перетворення незалежних величин* статистика χ_{n-1}^2 не залежить від $\eta_n = q'\zeta = \sqrt{n}\widehat{\mu}_n$ через незалежність (η_k).

Нарешті, середнє $\widehat{\mu}_n$ як лінійне перетворення ζ за теоремою *про лінійні перетворення нормальних векторів* має нормальний розподіл, із середнім 0 і дисперсією $\mathbf{D}\widehat{\mu}_n = n/n^2 = 1/n$ ⊡

3.10.2. Загальна нормальна вибірка

Теорема (про вибіркові середнє та дисперсію нормальної вибірки). *Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з нормальним розподілом спостережень: $\xi_k \simeq N(\mu, \sigma^2)$. Тоді статистики $\widehat{\mu}_n$ і $\widehat{s}_n^2$ незалежні та*

$$\sqrt{n}\ \frac{\widehat{\mu}_n - \mu}{\sigma} \simeq N(0,1), \quad \frac{(n-1)\widehat{s}_n^2}{\sigma^2} \simeq \chi_{n-1}^2,$$

$$n\,\frac{\widehat{\sigma}_n^2}{\sigma^2} = \frac{1}{\sigma^2}\sum\nolimits_{k=1}^{n} (\xi_k - \mu)^2 \simeq \chi_n^2.$$

Зауваження. Припущення нормальності кратної вибірки X необхідне і достатнє для незалежності вибіркових середнього та дисперсії.

Доведення. Центруванням та нормуванням спостережень ξ_k визначимо $\zeta_k = (\xi_k - \mu)/\sigma,\ k = \overline{1,n}$. Величини ζ_k є незалежними стандартними нормальними за означенням *нормального розподілу* та за теоремою *про перетворення незалежних величин*. Тому вектор $\zeta = (\zeta_k, k = \overline{1,n})$ є *стандартним нормальним вектором*. Оскільки перші дві статистики в твердженні теореми збігаються відповідно з нормованими *вибірковим середнім* та *вибірковою дисперсією* для вибірки ζ, то перше твердження теореми є наслідком теореми *про вибіркові моменти стандартної нормальної вибірки*. Третя статистика збігається з сумою $\sum_{k=1}^{n}\zeta_k^2$ і має *хі-квадрат розподіл* за означенням $\boxdot$

3.10.3. Статистики Стьюдента та Фішера

Означення. *Випадкова величина τ_n має* t-розподіл, *або* розподіл Стьюдента, *з n ступенями свободи, якщо її можна подати у вигляді*

$$\tau_n = \frac{\zeta}{\sqrt{\chi_n^2 \,/\, n}},$$

де $\zeta \simeq N(0,1)$ – стандартна нормальна величина, а χ_n^2 – незалежна від неї величина з хі-квадрат розподілом та n ступенями свободи.

Означення. *Випадкова величина $\phi_{n,m}$ має* розподіл Фішера, *або ж* розподіл Снедекора – Фішера, *з n, m ступенями свободи, якщо її можна подати у вигляді*

$$\phi_{n,m} = \frac{\chi_n^2 \,/\, n}{\chi_m^2 \,/\, m},$$

де χ_n^2, χ_m^2 – незалежні величини з хі-квадрат розподілом та n, m ступенями свободи відповідно.

Теорема (про статистику Стьюдента від нормальної вибірки). *Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з нормальним розподілом $N(\mu, \sigma^2)$. Тоді статистика*

$$\tau_{n-1} = \sqrt{n}\ \frac{\widehat{\mu}_n - \mu}{\widehat{s}_n}$$

має розподіл Стьюдента з $n-1$ ступенем свободи.

Доведення. За теоремою *про вибіркові середнє та дисперсію нормальної вибірки* після ділення на σ чисельник $\sqrt{n}(\widehat{\mu}_n - \mu)/\sigma$ має *стандартний нормальний розподіл*, і не залежить від знаменника

$$\widehat{s}_n/\sigma = \sqrt{(n-1)\widehat{s}_n^2 \ /\sigma^2(n-1)} \simeq \sqrt{\chi_{n-1}^2 \,/\, (n-1)}.$$

Тому весь дріб τ_{n-1} має *розподіл Стьюдента* $\boxdot$

Вправи

(1) Довести, що для кратних нормальних спостережень вибіркові асиметрія та ексцес $\widehat{\mu}_{3n}^0/\widehat{s}_n^{3/2}$, $\widehat{\mu}_{4n}^0/\widehat{s}_n^2$ не залежать від вибіркових середнього та дисперсії.

(2) Довести збіжність: $(\chi_n^2 - n)/\sqrt{2n} \to^W \zeta \simeq N(0,1),\ n \to \infty$.

(3) Довести, що $\sqrt{2\chi_n^2} - \sqrt{2n-1} \to^W \zeta \simeq N(0,1),\ n \to \infty$.

(4) Вивести апроксимацію Вілсона
$\mathbf{P}(\chi_n^2 < x) \to \mathbf{P}\left(\zeta < [(x/n)^{1/3} - 1 + 2/9n](9n/2)^{1/2}\right), n \to \infty.$

(5) Знайти щільність τ_n, виходячи з виразу для гама-щільності::
$f_{\tau_n}(x) = c_n(1 + x^2/n)^{-(n+1)/2},\ \ c_n = (\pi n)^{-1/2}\Gamma((n+1)/2)/\Gamma(n/2),$
обчислити $\mathbf{E}\tau_n = 0,\ \mathbf{D}\tau_n = n/(n-2), n > 2$. Зокрема, τ_1 має розподіл Коші.

(6) Довести, що $\tau_n \to^W \zeta \simeq N(0,1),\ n \to \infty$.

(7) Довести, що щільність $\phi_{n,m}$ має вигляд
$f_{nm}(x) = c_{nm}x^{n/2-1}\left(1 + \frac{n}{m}x\right)^{-(n+m)/2},\ c_{nm} = (\frac{n}{m})^{n/2}B(\frac{n}{2}, \frac{m}{2}).$

3.11. Інтервальні оцінки параметрів нормальних спостережень

Розглянемо лише двобічні *інтервальні оцінки* параметрів кратної нормальної вибірки $X = (\xi_1, ..., \xi_n),\ \xi_k \simeq N(\mu, \sigma^2)$, однобічні оцінки виводяться аналогічно. Побудова **надійних інтервалів для нормальних спостережень** ґрунтується на властивостях вибіркових моментів для нормальних виборок.

Позначимо через x_α двобічний *квантиль* вірогідного рівня $1-\alpha$ для *стандартного нормального розподілу*:

$$\mathbf{P}(|\zeta| \le x_\alpha) = 1 - \alpha,\ \Phi(x_\alpha) = 1 - \alpha/2.$$

Зокрема, при побудові інтервальних оцінок можна скористатись тим, що при $1 - \alpha = 0.95$ *надійний інтервал* для середнього $N(0,1)$ розподілу дорівнює:

$(-1.96,\ 1.96)$ – у симетричному варіанті,
$(-\infty,\ 1.645)$ – у правобічному варіанті,
$(-1.645,\ \infty)$ – у лівобічному варіанті,
$(-1.88,\ 2.05)$ – у правозкошеному варіанті,
$(-2.05,\ 1.88)$ – у лівозкошеному варіанті.

Аналогічний зміст мають *квантилі* $y_{n\alpha}$ для *розподілу Стьюдента*, та квантилі $z_{n\alpha}$ для *хі-квадрат розподілу*:

$$\mathbf{P}(|\tau_n| \le y_{n\alpha}) = 1 - \alpha,\ \ \mathbf{P}(\chi_n^2 \le z_{n\alpha}) = 1 - \alpha/2.$$

Доведення наведених нижче оцінок є наслідком теореми *про вибіркові середнє та дисперсію нормальної вибірки*, та означення відповідних *квантилей.*

3.11.1. Оцінка середнього при відомій дисперсії

Оскільки при відомій *дисперсії* σ^2 статистика $\zeta = \sqrt{n}(\widehat{\mu}_n - \mu)/\sigma \simeq N(0,1)$ є стандартною нормальною, то

$$\mathbf{P}\left(\widehat{\mu}_n - \frac{\sigma}{\sqrt{n}}x_\alpha \le \mu \le \widehat{\mu}_n + \frac{\sigma}{\sqrt{n}}x_\alpha\right) = \mathbf{P}(|\zeta| \le x_\alpha) = 1 - \alpha.$$

3.11.2. Оцінка дисперсії при відомому середньому

У випадку відомого середнього μ для відповідно модифікованої вибіркової дисперсії $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^n(\xi_k - \mu)^2$ зі співвідношення $n\widehat{\sigma}_n^2/\sigma^2 \simeq \chi_n^2$ виводимо, що

$$\mathbf{P}\left(n\widehat{\sigma}_n^2 \,/\, z_{n\alpha} \le \sigma^2 \le n\widehat{\sigma}_n^2 \,/\, z_{n,2-\alpha}\right) =$$
$$\mathbf{P}\left(z_{n,2-\alpha} \le \chi_n^2 \le z_{n\alpha}\right) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha.$$

3.11.3. Оцінка середнього при невідомій дисперсії

За теоремою *про статистику Стьюдента від нормальної вибірки* статистика $\tau_n = \sqrt{n}(\widehat{\mu}_n - \mu) \,/\, \widehat{s}_n$ має *розподіл Стьюдента* з $n-1$ ступенем свободи. Тому

$$\mathbf{P}\left(\widehat{\mu}_n - \frac{\widehat{s}_n}{\sqrt{n}}y_{n-1,\alpha} \le \mu \le \widehat{\mu}_n + \frac{\widehat{s}_n}{\sqrt{n}}y_{n-1,\alpha}\right) = \mathbf{P}\left(|\tau_n| \le y_{n-1,\alpha}\right) = 1 - \alpha.$$

3.11.4. Оцінка дисперсії при невідомому середньому

Оскільки $(n-1)\,\widehat{s}_n^2/\sigma^2 \simeq \chi_{n-1}^2$, то

$$\mathbf{P}\left((n-1)\widehat{s}_n^2 \,/\, z_{n-1,\alpha} \le \sigma^2 \le (n-1)\widehat{s}_n^2 \,/\, z_{n-1,2-\alpha}\right) =$$
$$\mathbf{P}\left(z_{n-1,2-\alpha} \le \chi_{n-1}^2 \le z_{n-1,\alpha}\right) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha.$$

3.11.5. Загальний метод опорних величин

Для довільно розподіленої вибірки X при побудові *надійних інтервалів* використовують метод **опорних величин**. Він полягає у знаходженні опорної випадкової величини вигляду $g(X,\theta)$, що має такі властивості:

(1) вона є функцією вибірки X та невідомого параметра θ,

(2) має повністю відомий розподіл,

(3) монотонно залежить від θ.

Виходячи з відомого розподілу, знаходять такі значення g_1, g_2, що

$$\mathbf{P}_\theta(g_1 < g(X,\theta) < g_2) = 1 - \alpha, \ \forall\theta \in \Theta.$$

Далі, за монотонністю розв'язують нерівності так, щоб

$$\{g(X,\theta) < g_2\} = \{\theta < \widehat{\theta}_2(X)\},$$

$$\{g(X,\theta) > g_1\} = \{\theta > \widehat{\theta}_1(X)\}.$$

Отже, $(\widehat{\theta}_1, \widehat{\theta}_2)$ – шуканий *надійний інтервал* для θ рівня $1-\alpha$.

У випадку, коли розподіл опорної величини лише асимптотично наближається до відомого, то відповідні асимптотичні надійні інтервали будують, виходячи при великих об'ємах вибірки n з наближених рівностей

$$\mathbf{P}_\theta(g_1 < g_n(X,\theta) < g_2) \approx 1-\alpha.$$

Вправи

(1) Припустимо, що для статистики $T = T(X)$ імовірність $\mathbf{P}_\theta(T < x)$ монотонно залежить від θ при кожному x. Нехай для заданого рівня α функції θ_i визначено з рівнянь $\mathbf{P}_{\theta_1(x)}(T < x) = \alpha/2$, $\mathbf{P}_{\theta_2(x)}(T \geq x) = \alpha/2$. Довести, що: (а) $(\theta_1(T(X)), \theta_2(T(X)))$ є надійним інтервалом рівня $1-\alpha$ для θ, (б) вказане припущення виконується для біноміального та пуассонівського розподілу.

(2) Побудувати надійний інтервал заданого рівня для параметра кратної вибірки з (а) показниковим розподілом $Exp(\theta)$, (б) рівномірним на $(\theta, 2\theta)$ розподілом, (в) розподілом Пуассона.

(3) Випадкова величина ξ_λ має розподіл Пуассона з параметром λ. Побудувати асимптотичний надійний інтервал для λ за спостереженням ξ_λ, виходячи з асимптотичної нормальності величини $(\xi_\lambda - \lambda)\,\lambda^{-1/2}$ при $\lambda \to \infty$.

(5) Нехай $(\xi_1, ..., \xi_n)$ – кратна нормальна вибірка з розподілом $\xi_k \simeq N(\mu, \sigma^2)$, $\Phi_{\mu\sigma}$ – нормальна функція розподілу з параметрами μ, σ^2. Довести, що: (а) для кожного $t > 0$ розподіл приросту $\Delta_t = \Phi_{\mu\sigma}(\widehat{\mu}_n + t\widehat{\sigma}_n) - \Phi_{\mu\sigma}(\widehat{\mu}_n - t\widehat{\sigma}_n)$ залежить лише від t, (б) якщо $t = y_{n-1,\alpha}\sqrt{1+1/n}$, то $\mathbf{E}\Delta_t = \alpha$. (в) має місце збіжність $\sqrt{n}(\widehat{\mu}_n/\widehat{s}_n - \kappa)/\sqrt{1+\kappa^2/2} \to^W N(0,1), n \to \infty$, де $\kappa = \mu/\sigma$.

3.12. Перевірка статистичних гіпотез

На відміну від задач статистичного оцінювання, задача перевірки статистичної гіпотези полягає у формуванні дихотомічного висновку щодо відповідності наявних спостережень (вибірки) певним припущенням про їх розподіл, тобто про властивості *ймовірності* $\mathbf{P}_\theta$. Останні визначаються істинним значенням параметра θ. Тут $(\mathbf{P}_\theta, \theta \in \Theta)$ – параметрична сім'я розподілів з означення *статистичного простору*.

Статистичною **гіпотезою** називається довільне *припущення про розподіл вибірки*, яка спостерігається у стохастичному експерименті. Оскільки цей розподіл вважається повністю відомим при істинному значенні параметра θ, статистичні гіпотези часто формулюють у вигляді

$$H_0 : \theta \in \Theta_0,$$

де $\Theta_0 \subset \Theta$ – певна параметрична підмножина. Якщо параметричний простір Θ ототожнити з множиною всіх можливих розподілів вибірки, то будь-яка гіпотеза може бути зображена у наведеному вигляді.

Зокрема, розглядають гіпотези про те, що:
(1) значення параметра дорівнює заданому,
(2) значення параметра перевищує задане,
(3) розподіл спостережень збігається з заданим,
(4) розподіл спостережень належить заданому класу розподілів,
(5) групи спостережень є однорідними (мають однакові розподіли),
(6) групи спостережень є незалежними,
(7) спостереження є випадковими.

Як ми побачимо пізніше, статистичні властивості (зокрема, якість) висновків щодо розподілу вибірки суттєво залежать не тільки від вигляду статистичної гіпотези, а й від множини значень параметра, що не задовольняють цю гіпотезу. Треба мати на увазі, що вказана множина значень не завжди збігається з доповненням параметричної множини Θ_0.

Статистичною **альтернативою** називають таке припущення про розподіл вибірки, яке вважається виконаним у випадку, коли не справджується основна статистична *гіпотеза*. Статистичні альтернативи мають вигляд

$$H_1 : \theta \in \Theta_1,$$

де $\Theta_1 \subset \Theta$. Якщо $\Theta_1 = \Theta \setminus \Theta_0$, то альтернатива може не формулюватися явно. Якщо ж має місце строге включення $\Theta_0 \cup \Theta_1 \subset \Theta$, то основна гіпотеза H_0, на відміну від альтернативної, називається **нульовою гіпотезою**.

Означення. *Статистична гіпотеза* $H_0 : \theta \in \Theta_0$ *називається* **простою гіпотезою,** *якщо множина* $\Theta_0 = \{\theta_0\}$ *є одноелементною, причому розподіл* $\mathbf{P}_{\theta_0}(X \in \cdot)$ *відомий повністю.*

Приклади

1. *Нормальна вибірка з відомою дисперсією*. Невідомим параметром *нормального розподілу* є середнє: $\theta = \mu$, *дисперсія* σ^2 вважається відомою. При фіксації μ щільність спостережень відома, тому гіпотеза $H_0 : \theta = \mu_0$ – проста.

2. *Нормальна вибірка з невідомою дисперсією*. Якщо параметр $\theta = (\mu, \sigma^2)$, то при виконанні гіпотези $H_0 : \mu = \mu_0$ вибіркові щільності можуть мати різні дисперсії, тому ця гіпотеза не є простою.

3.12.1. Статистика критерію, критична область

Абстрактно критерій для перевірки H_0 проти H_1 на підставі вибірки X можна визначити як функцію $\delta(X) : S \to \{H_0, H_1\}$, що для кожного вектора X приймає висновок на користь однієї з гіпотез.

Однак на практиці статистичний висновок робиться на підставі розгляду значення певної функції від вибірки – **статистики критерію.** Ця функція $\widehat{\varkappa}(X) : S \to D$ є довільною *статистикою* (*вимірною* функцією від вибірки) зі значеннями в деякому вимірному просторі D. Відзначимо, що на відміну від *оцінки*, статистика критерію може містити значення параметрів – у випадку, коли гіпотеза висувається саме щодо цих значень.

Для визначення результату статистичного висновку щодо якісної (дихотомічної) властивості розподілу спостережень на основі значення статистики $\widehat{\varkappa}$ досить розбити множину D на дві частини: $D = D_0 \cup D_1$, та

(0) у випадку включення $\widehat{\varkappa} \in D_0$ – приймати *нульову гіпотезу*,

(1) при $\widehat{\varkappa} \in D_1$ – приймати *альтернативу* і відхиляти нульову гіпотезу.

Множина D_1, на якій відхиляється нульова гіпотеза, називається **критичною областю** статистики критерію. Очевидно, що довільний алгоритм дихотомічного вибору на підставі значення вибірки X можна подати у наведеному вигляді, обираючи, наприклад, $D = \{0, 1\}$, $D_0 = \{0\}$, $D_1 = \{1\}$ та відповідно конструюючи D-значну *статистику критерію*.

На практиці частіше обирають $D = \mathbb{R}$, $D_0 = (-\infty, x_0)$, $D_1 = [x_0, \infty)$. У цьому випадку статистика критерію є числовою величиною, а x_0 визначає **критичний рівень** *статистики критерію*. Нульова гіпотеза відкидається за умови перевищення статистикою критерію критичного рівня: $\widehat{\varkappa} \geq x_0$.

У загальному випадку з *критичною областю* статистики D_1 можна пов'язати **критичну область вибірки**

$$W = \{x \in S : \widehat{\varkappa}(x) \in D_1\} = \{x \in S : \delta(x) = H_1\}.$$

При потраплянні вибіркового вектора X у *критичну область* нульова гіпотеза відкидається, у іншому випадку відкидається альтернатива.

Означення. *Статистичним* **критерієм** (*статистичним* **тестом**) *називається пара* $(\widehat{\varkappa}(X), D_1)$, *що утворена D-значною статистикою критерію* $\widehat{\varkappa}(X)$ *та її критичною областю* $D_1 \subset D$.

Алгоритм перевірки статистичної *гіпотези* H_0 проти *альтернативи* H_1 за допомогою *критерію* $(\widehat{\varkappa}, D_1)$ виконується в два етапи:

(1) обчислюють значення $\widehat{\varkappa} = \widehat{\varkappa}(X)$,

(2) перевіряють включення $\widehat{\varkappa} \in D_1$, що еквівалентне $X \in W$,

(2.1) якщо воно справджується, то *нульова гіпотеза* H_0 на підставі спостережень X відкидається і приймається альтернатива H_1,

(2.0) якщо ж це включення не справджується, то нульова гіпотеза H_0 на підставі спостережень X не може бути відкинута, отже, приймається, а *альтернатива* H_1 відкидається.

Зауваження. У зв'язку зі статистичним характером перевірки гіпотез зауважимо таке. Перевірка гіпотези ні в якому разі не є доведенням справедливості чи несправедливості припущення гіпотези. Невдача у відхиленні H_0 означає лише, ще немає досить вагомих свідчень для відхилення H_0 – це і є зміст висновку про те, що ця гіпотеза приймається. Про справедливість припущення можна казати лише при наявності багатосторонніх свідчень на його користь, як статистичних, так і інших.

3.12.2. Рівень та потужність критерію

Наведені вище означення *критерію* перевірки гіпотези є суто технічними і безпосередньо не пов'язані з якістю статистичного висновку. Для визначення показників якості зауважимо, що *алгоритм перевірки* гіпотези спричиняє лише одне з двох можливих рішень: прийняття або відхилення *нульової гіпотези*. Кожне з цих рішень може призвести до похибки.

Означення. Похибкою першого роду *статистичного критерію називається відхилення нульової гіпотези за умови, що вона справджується.*

Похибкою другого роду *називається прийняття нульової гіпотези за умови, коли справджується альтернатива.*

Вказані помилкові рішення пов'язані з відповідними подіями. Їх імовірності називаються ймовірностями похибок першого та другого роду.

Означення. *Нехай нульова гіпотеза має вигляд* $H_0 : \theta \in \Theta_0$, *а альтернатива –* $H_1 : \theta \in \Theta_1$. Вірогідним рівнем (*або* критичним рівнем) *статистичного критерія* $(\widehat{\varkappa}, D_1)$ *називається функція від* θ, *що задає ймовірності похибок першого роду:*

$$\mathbf{P}_\theta(\widehat{\varkappa}(X) \in D_1) = \mathbf{P}_\theta(X \in W),\ \theta \in \Theta_0.$$

Потужністю *критерію називається ймовірність відсутності похибки другого роду (тобто ймовірність правильного – альтернативного – висновку при альтернативі), що задається функцією:*

$$1 - \mathbf{P}_\theta(\widehat{\varkappa}(X) \in D_0) = \mathbf{P}_\theta(\widehat{\varkappa}(X) \in D_1) = \mathbf{P}_\theta(X \in W),\ \theta \in \Theta_1.$$

Зауваження. Якщо нульова гіпотеза є *простою гіпотезою*, то *вірогідний рівень* задається одним числом: $\mathbf{P}_{\theta_0}(X \in W)$, оскільки при $\theta = \theta_0$ розподіл вибірки відомий. Це число називають **Р-значенням критерію**.

Імовірність $\mathbf{P}_\theta(X \in W)$ попадання вибірки у критичну область як функція від усіх значень параметра $\theta \in \Theta_0 \cup \Theta_1$ називається **оперативною характеристикою** критерію.

Маючи на меті одночасну мінімізацію *ймовірностей похибок першого та другого роду*, для створення якісного критерію треба відшукати таку статистику $\widehat{\varkappa}(X)$, яка була б "чутливою" до істинного значення параметра:

– при виконанні нульової гіпотези набувала переважно значень із множини D_0 (наприклад, "досить помірних" значень) – що зменшує ймовірність *похибки першого роду*,

– при виконанні альтернативи набувала переважно значень у доповненні $D_1 = D \setminus D_0$ (наприклад, "надмірно великих значень") – що зменшує ймовірність *похибки другого роду*.

Очевидно, що задача побудови оптимального статистичного критерію зводиться до одночасної мінімізації *ймовірностей похибок першого та другого роду*, або ж до мінімізації його *вірогідного рівня* при максимізації *потужності*. Ця задача є внутрішньо суперечливою, оскільки і рівень, і потужність (як значення однієї *оперативної характеристики* на різних множинах) одночасно монотонно (у напрямку зростання) залежать від *критичної області вибірки W*. Зокрема, при $W = \varnothing$ імовірність похибки першого роду нульова, другого роду дорівнює одиниці, а потужність – нульова. При $W = S$ має місце протилежне: похибка першого роду і потужність одночасно одиничні.

Тому задачу відшукання оптимального критерію зводять до задачі умовної оптимізації, тобто знаходженню найбільшої потужності при обмеженні похибки першого роду.

Альтернативним до викладеного є **Байесовський підхід** до побудови оптимального критерію. Він полягає у постулюванні апріорного розподілу між нульовою гіпотезою та альтернативою, причому оптимальність зводиться до мінімізації середньої похибки, що отримується відповідним усередненням *імовірностей похибок першого та другого роду*.

Ще одну альтернативу становить **мінімаксний підхід**, згідно з яким оптимальним є критерій, що мінімізує найбільшу з імовірностей похибок першого та другого роду.

Означення. *Критерій* $(\widehat{\varkappa}, D_1)$ *має* **вірогідний рівень** α, *якщо ймовір-*

ності похибок першого роду не перевищують α :

$$\mathbf{P}_\theta(\widehat{\varkappa}(X) \in D_1) = \mathbf{P}_\theta(X \in W) \leq \alpha, \ \forall\theta \in \Theta_0,$$

і хоча б одна з цих нерівностей є рівністю (хоча б для одного θ).

Зауваження. Якщо *нульова гіпотеза* є *простою гіпотезою*, то в означенні рівня критерію обов'язково має місце рівність. Дійсно, в такому випадку маємо єдине значення параметру і одну нерівність, яка за означенням перетворюється на рівність.

Означення. *Критерій* $(\widehat{\varkappa}, D_1)$ *є* **незміщеним критерієм**, *якщо його потужність не менша за його рівень* α :

$$\mathbf{P}_\theta(\widehat{\varkappa}(X) \in D_1) = \mathbf{P}_\theta(X \in W) \geq \alpha, \ \forall\theta \in \Theta_1.$$

Зауваження. Як і у випадку статистичних оцінок, послідовність критеріїв $(\widehat{\varkappa}_n, D_1)$, що побудовані для кожного n за *кратною вибіркою* об'єму n,за одним алгоритмом, також будемо називати критерієм.

Означення. *Критерій* $(\widehat{\varkappa}_n, D_1)$ *вірогідного рівня* α *є* **консистентним** *критерієм, якщо його потужність прямує до одиниці:*

$$\mathbf{P}_\theta(\widehat{\varkappa}_n \in D_1) \to 1, \ n \to \infty, \ \forall\theta \in \Theta_1.$$

Очевидно, що консистентні критерії є асимптотичними розв'язками умовної задачі оптимізації критерію за рівнем та потужністю.

3.13. Непараметричні критерії для функції розподілу

Розглянемо *кратну вибірку* $X = (\xi_1, ..., \xi_n)$, в якій окремі спостереження мають невідому неперервну *функцію розподілу* $F(x) = \mathbf{P}(\xi_1 < x)$. З формального погляду дана схема відповідає параметричному просторові Θ, що містить усі функції розподілу, а нульова гіпотеза щодо значення параметра $\theta = F$ для заданої функції розподілу F має вигляд

$$H_0 : \mathbf{P}(\xi_1 < x) = F(x), \forall x \in \mathbb{R}, \ X \text{ – кратна вибірка.}$$

3.13.1. Критерій узгодженості Колмогорова

Перевіряється **гіпотеза узгодженості** кратної вибірки з заданою функцією розподілу F, тобто припущення H_0 про те, що функція розподілу окремих спостережень збігається з наперед заданою неперервною функцією розподілу F. Як альтернативу будемо розглядати клас усіх неперервних функцій розподілу, відмінних від F :

$$H_1 : \mathbf{P}(\xi_1 < x) = G(x), \forall x \in \mathbb{R},\ G \neq F, G \in C(\mathbb{R}),\ X \text{ – кратна вибірка.}$$

Критерій можна вважати непараметричним, адже невідомою є вся функція розподілу, а не окремі параметри.

Побудова критерію ґрунтується на *статистиці Колмогорова*:

$$\widehat{\varkappa}_n(X) = \sqrt{n}\ \sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F(x) \right|,$$

де $\widehat{F}_n(x)$ – *емпірична функція розподілу*:

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{k=1}^{n} 1\!\!\mathrm{I}_{\{\xi_k < x\}}.$$

З вигляду статистики $\widehat{\varkappa}_n$ виводимо, що критичними для H_0 є її великі значення, принаймні при значних об'ємах вибірки n.

За *теоремою Колмогорова про відхилення емпіричної функції розподілу* (що наведена вище у розділі про *емпіричну функцію розподілу*) розподіл статистики $\widehat{\varkappa}_n(X)$ за *нульової гіпотези* не залежить від невідомої функції розподілу F та має місце збіжність *в основному*

$$\mathbf{P}(\widehat{\varkappa}_n < x) \to \mathbf{P}(\varkappa < x) = K(x),\ n \to \infty,$$

де *функція розподілу Колмогорова* $K(x)$ повністю відома:

$$K(x) \equiv \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2 x^2).$$

Критерій Колмогорова задається парою $(\widehat{\varkappa}_n,\ [x_\alpha, \infty))$, де *критичний рівень* x_α знаходиться за *вірогідним рівнем* α з умови $K(x_\alpha) = 1 - \alpha$. *Нульова гіпотеза* про узгодженість відхиляється, якщо $\widehat{\varkappa}_n \geq x_\alpha$. Вірогідний рівень критерію наближено (і тим точніше, чим більший обсяг вибірки) дорівнює

$$\mathbf{P}(\widehat{\varkappa}_n \geq x_\alpha) \approx \mathbf{P}(\varkappa \geq x_\alpha) = 1 - K(x_\alpha) = \alpha.$$

Отже, для скінченних n *критерій Колмогорова* є наближеним. Його можна зробити точним, якщо для побудови критичного рівня замість $K(x)$ використати точне значення функції розподілу статистики $\widehat{\varkappa}_n$.

Для дослідження *потужності* критерію припустимо, що розглядається *нульова гіпотеза* $H_0 : \theta = F$, де F – гіпотетична функція розподілу спостережень, а істинна функція розподілу відповідає *альтернативі* $H_1 : \theta = \theta_1 \equiv G$. Тоді *потужність* критерію дорівнює

$$\mathbf{P}_{\theta_1}\left(\sqrt{n}\sup_{x\,\in\,\mathbb{R}}\left|\widehat{F}_n(x)-F(x)\right|\geq x_\alpha\right)=$$

$$\mathbf{P}_{\theta_1}\left(\sqrt{n}\sup_{x\,\in\,\mathbb{R}}\left|\widehat{F}_n(x)-G(x)+G(x)-F(x)\right|\geq x_\alpha\right)\geq$$

$$\mathbf{P}_{\theta_1}(\sqrt{n}\Delta-\widetilde{\varkappa}_n\geq x_\alpha),\quad \text{де } \Delta=\sup_{x\,\in\,\mathbb{R}}|F(x)-G(x)|>0,$$

$$\widetilde{\varkappa}_n=\sqrt{n}\ \sup_{x\,\in\,\mathbb{R}}\left|\widehat{F}_n(x)-G(x)\right|\to^W\varkappa,\ n\to\infty,$$

за теоремою Колмогорова у припущенні H_1, та $\mathbf{P}_{\theta_1}(\sqrt{n}\Delta-\widetilde{\varkappa}_n\geq x_\alpha)\to 1$, $n\to\infty$. Отже, потужність *критерію Колмогорова* прямує до одиниці при кожній простій альтернативі. Критерій є *консистентним.*

Вправи

(1) Довести, що $\mathbf{P}(\varkappa_n<c_n)\to 1$, якщо $\varkappa_n\to^W\varkappa$, і $c_n\to\infty, n\to\infty$.

(2) Довести що в умовах теореми Колмогорова про відхилення емпіричної функції розподілу за нульової гіпотези розподіл статистики омега-квадрат

$\widehat{\omega}_n^2=n\int_{-\infty}^{\infty}(\widehat{F}_n(x)-F(x))^2dF(x)$

не залежить від вигляду теоретичної функції розподілу та обчислюється через варіаційний ряд: $n\widehat{\omega}_n^2=(12n)^{-1}+\sum_{k=1}^n(F(\xi_{(k)})-(2k-1)/2n)^2$.

(3) Нехай $\widehat{F}_n(x)$ – емпірична функція розподілу для вибірки з неперервною функцією розподілу F одного спостереження, $\alpha>0$, а функція $w(t)$ невід'ємна. Довести, що розподіл наступних статистик не залежить від F :

$S_{nw\alpha}=\sup_{x\in\mathbb{R}}w(F(x))|\widehat{F}_n(x)-F(x)|^\alpha,$
$T_{nw\alpha}=\sup_{x\in\mathbb{R}}w(\widehat{F}_n(x))|\widehat{F}_n(x)-F(x)|^\alpha,$
$U_{nw\alpha}=\int_{\mathbb{R}}w(F(x))|\widehat{F}_n(x)-F(x)|^\alpha F(dx),$
$V_{nw\alpha}=\int_{\mathbb{R}}w(\widehat{F}_n(x))|\widehat{F}_n(x)-F(x)|^\alpha\widehat{F}(dx).$

3.13.2. Критерій однорідності Смірнова

Одночасно з вибіркою X розглянемо *кратну вибірку* $Y=(\eta_1,...,\eta_m)$ об'ємом m, що не залежить від X, із неперервною функцією розподілу окремих спостережень $G(y)=\mathbf{P}(\eta_1<y)$. Параметризуючи розподіл повного вектора спостережень (X,Y) як пару функцій розподілу $\theta=(F,G)$, розглянемо складну нульову гіпотезу однорідності

$$H_0: F=G,\ F \text{ – неперервна},\ X \text{ і } Y \text{ – незалежні, кратні}$$

проти альтернативи

$$H_1: F\neq G,\ F,G \text{ – неперервні},\ X \text{ і } Y \text{ – незалежні, кратні.}$$

Позначимо *емпіричну функцію розподілу* другої вибірки

$$\widehat{G}_m(y) = \frac{1}{m}\sum\nolimits_{k=1}^{m} 1\!\!\mathrm{I}_{\{\eta_k\,<\,y\}}.$$

Теорема (теорема Смірнова про відхилення емпіричних функцій розподілу). *За гіпотези H_0 розподіл статистики Смірнова:*

$$\widehat{\varkappa}_{nm} = \sqrt{\frac{mn}{m+n}}\sup{}_{x\in\mathbb{R}}\ \left|\widehat{F}_n(x) - \widehat{G}_m(x)\right|$$

не залежить від вигляду невідомої функції розподілу $F = G$, і для всіх $x \in \mathbb{R}$ має місце збіжність

$$\mathbf{P}(\widehat{\varkappa}_{nm} < x) \to \mathbf{P}(\varkappa < x) = K(x),\ n, m \to \infty,$$

де функція $K(x)$ та сама, що й у теоремі Колмогорова про відхилення емпіричної функції розподілу.

Доведення не наводиться.

Критерій Смірнова задається парою $(\widehat{\varkappa}_{nm},\ [x_\alpha, \infty))$, де *критичний рівень* x_α визначається за *вірогідним рівнем* α умовою $K(x_\alpha) = 1 - \alpha$. Тут вигляд критичної області обумовлений характером статистики $\widehat{\varkappa}_{nm}$. Рівень критерію наближено дорівнює

$$\mathbf{P}(\widehat{\varkappa}_{nm} \ge x_\alpha) \approx \mathbf{P}(\varkappa \ge x_\alpha) = 1 - K(x_\alpha) = \alpha.$$

Для оцінки *потужності* припустимо, що виконується альтернатива H_1 та позначимо $b_{nm} = \sqrt{nm/(n+m)}$. Потужність критерію дорівнює

$$\mathbf{P}_\theta\left(b_{nm}\sup\nolimits_{x\in\mathbb{R}}|\widehat{F}_n(x) - \widehat{G}_m(x)|\ \ge x_\alpha\right) =$$

$$\mathbf{P}_\theta\left(b_{nm}\sup\nolimits_{x\in\mathbb{R}}|\widehat{F}_n(x) - F(x) + G(x) - \widehat{G}_m(x) + F(x) - G(x)|\ \ge x_\alpha\right) \ge$$

$$\mathbf{P}_\theta\left(b_{nm}\Delta - \widetilde{\varkappa}_n - \widetilde{\varkappa}_m \ge x_\alpha\right),\ \text{де}\ \Delta = \sup\nolimits_{x\in\mathbb{R}}|F(x) - G(x)| > 0,$$

$$\widetilde{\varkappa}_n = \sqrt{n}\ \sup\nolimits_{x\in\mathbb{R}}|\widehat{F}_n(x) - F(x)|\ \to^W \varkappa',$$

$$\widetilde{\varkappa}_m = \sqrt{m}\ \sup\nolimits_{x\in\mathbb{R}}|\widehat{G}_m(x) - G(x)|\ \to^W \varkappa''$$

за *теоремою Колмогорова про відхилення емпіричної функції розподілу*. Тому $b_{nm}\Delta - \widetilde{\varkappa}_n - \widetilde{\varkappa}_m \to^P \infty$, отже, потужність критерію Смірнова прямує до одиниці при кожній простій альтернативі. Критерій є *консистентним*.

3.14. Деякі рангові критерії

Значна кількість критеріїв, що стійкі до випадкових збурень, будується з використанням рангових статистик.

Означення. Ранговими статистиками *вибірки $X = (\xi_1, ..., \xi_n)$ називаються координати вектора рангів $\nu = (\nu_1, ..., \nu_n)$, де значення ν_k задає номер k-го спостереження ξ_k у варіаційному ряді* $(\xi_{(k)}, k = \overline{1,n})$:

$$\xi_k = \xi_{(\nu_k)}, \quad (\xi_1, ..., \xi_n) = (\xi_{(\nu_1)}, ..., \xi_{(\nu_n)}).$$

Приклад. Нехай спостерігається вибірка $X = (2, 7, 4, 9, 3, 5)$. Її варіаційний ряд має вигляд $(2, 3, 4, 5, 7, 9)$, *вектор рангів* збігається з перестановкою $(1, 5, 3, 6, 2, 4)$, а вибіркове середнє дорівнює 5. Якщо вибірка випадково збурена (наприклад, через неякісну передачу інформації) до вигляду $X = (2, 7, 4, 9, 3, 5000)$, то вектор рангів зазнає не дуже істотних змін лише до $(1, 4, 3, 5, 2, 6)$ – на відміну від середнього 837.5 замість 5.

Як показано вище у теоремі *про розподіл вектора рангів*, для *кратної вибірки* з неперервною функцією розподілу вектор рангів ν *рівномірно розподілений* на множині всіх перестановок Π_n порядку n.

Зауваження. У випадку, коли функція розподілу спостережень (ξ_k) не є неперервною (наприклад, коли ці спостереження є цілозначними), не можна виключити, що серед них знайдуться однакові. У цьому разі ранги визначають так, щоб вони були однаковими для однакових спостережень, однак щоб сума таких рангів не змінилася. Тому при виконанні подій $\xi_{n_1} = ... = \xi_{n_k} = x$, та $\xi_{(r-1)} < \xi_{(r)} = ... = \xi_{(r+k-1)} = x < \xi_{(r+k)}$ обирають

$$\nu_{n_1} = \nu_{n_2} = ... = \nu_{n_k} = (r + (r + 1) + .. + (r + k - 1))/k = r + (k - 1)/2.$$

Наприклад, для вибірки $X = (2, 7, 2, 5)$ з повтореннями *вектор рангів* дорівнює $\nu = (3/2, 4, 3/2, 3)$.

3.14.1. Критерій однорідності Вілкоксона

Розглянемо *кратну вибірку* $X = (\xi_1, ..., \xi_n)$, в якій окремі спостереження мають невідому неперервну *функцію розподілу* $F(x) = \mathbf{P}(\xi_1 < x)$ і одночасно незалежну кратну вибірку $Y = (\eta_1, ..., \eta_m)$ зі спільною неперервною функцією розподілу $G(y) = \mathbf{P}(\eta_1 < y)$. Для перевірки нульової *гіпотези однорідності*

$$H_0 : G = F, \quad X \text{ та } Y \text{ незалежні, кратні вибірки,}$$

використаємо об'єднану вибірку

$$Z = (\xi_1, ..., \xi_n, \ \eta_1, ..., \eta_m),$$

та позначимо через ν_j ранг j-го спостереження у вибірці Z. За нульової гіпотези вибірка Z є *кратною вибіркою* об'єма $N = n + m$, тому вектор ν *рівномірно розподілений* на множині перестановок Π_N порядку N.

Визначимо **статистику Вілкоксона**

$$S_{nm} = \sum\nolimits_{j=1}^{n} \nu_j,$$

яка задає сумарний ранг вибірки X в об'єднаній вибірці Z.

За *нульової гіпотези* H_0 розподіл статистики S_{nm} однозначно визначається лише значеннями n, m. Це дає можливість знайти *критичний рівень* статистики для забезпечення заданого *вірогідного рівня* α. Вигляд *критичної області* визначається альтернативою.

Нехай *альтернативою* є від'ємний зсув вибірки Y відносно X на $\Delta > 0$:

$$H_1 : G(y) = F(y + \Delta),\ \forall y \in \mathbb{R}, \quad X \text{ та } Y \text{ незалежні, кратні.}$$

У цьому разі $\eta_1 \simeq \xi_1 - \Delta$, отже, значення статистики S_{nm} будуть переважно більшими при виконанні альтернативи порівняно з *нульовою гіпотезою*, оскільки елементи вибірки X отримуватимуть переважно більші ранги. Тому критичними для нульової гіпотези слід вважати великі значення статистики. Отже, *критична область* повинна мати вигляд $[x_{nm}(\alpha), \infty)$, де критичний рівень $x_{nm}(\alpha)$ визначається умовою $\mathbf{P}(S_{mn} \ge x_{nm}(\alpha)) \approx \alpha$. Знаходження рівня можливе з використанням таблиць або спеціальних комп'ютерних програм.

3.14.2. Зауваження щодо рандомізованих критеріїв

Оскільки статистика S_{mn} має дискретний розподіл, то її *функція розподілу* набуває зліченну множину значень, і для довільного $\alpha \in (0, 1)$ неможливо підібрати критичне значення $x_{nm}(\alpha)$ так, щоб за нульової гіпотези критерій мав точний *вірогідний рівень* α.

Для забезпечення точного рівня застосовують **процедуру рандомізації.** Нехай $\alpha_1 \le \alpha \le \alpha_2$ – найближчі до α імовірності, для яких рівняння

$$\mathbf{P}(S_{mn} \ge x_i) = \alpha_i, \quad i = 1, 2,$$

мають точні розв'язки x_i. Згенеруємо незалежну від S_{mn} випадкову величину $\upsilon \in \{1, 2\}$ таку, що $\mathbf{P}(\upsilon = 1) = \varepsilon$, $\mathbf{P}(\upsilon = 2) = 1 - \varepsilon$. Визначимо випадковий (рандомізований) *критичний рівень* $x(\alpha) = x_\upsilon \equiv x_1 1\!\!\mathrm{I}_{\{\upsilon=1\}} + x_2 1\!\!\mathrm{I}_{\{\upsilon=2\}}$. Тоді за *формулою повної ймовірності*

$$\mathbf{P}(S_{mn} \ge x(\alpha)) = \varepsilon \mathbf{P}(S_{mn} \ge x_1) + (1 - \varepsilon)\mathbf{P}(S_{mn} \ge x_2) =$$

$$\varepsilon\alpha_1 + (1 - \varepsilon)\alpha_2 = \alpha,$$

якщо обрати $\varepsilon = (\alpha_2 - \alpha)/(\alpha_2 - \alpha_1) \in [0, 1]$. Отже, **рандомізована критична область** $[x_\upsilon, \infty)$ має точний вірогідний рівень α.

3.14.3. Асимптотична нормальність статистики Вілкоксона

При великих значеннях об'єму вибірки можна скористатись *асимптотичною нормальністю статистики Вілкоксона*. Для її обґрунтування обчислимо моменти вектора рангів з урахуванням теореми *про розподіл вектора рангів* та *нульової гіпотези*:

$$\mathbf{E}\nu_k = \sum\nolimits_{\pi\in\Pi_N} \pi_k/N! = \sum\nolimits_{r=1}^{N}\sum\nolimits_{\pi:\pi_k=r} r/N! = N(N+1)/2N = (N+1)/2.$$

Аналогічно,

$$\mathbf{E}\nu_k^2 = \sum\nolimits_{r=1}^{N}\sum\nolimits_{\pi:\pi_k=r} r^2/N! = (N+1)(2N+1)/6,\ \mathbf{D}\nu_k = (N^2-1)/12.$$

Нарешті, при $k \neq j$

$$\mathbf{E}\nu_k\nu_j = \sum\nolimits_{r\neq s}^{N}\sum\nolimits_{\pi:\pi_k=r,\pi_j=s} rs/N! = (N+1)(3N+2)/12,$$

$$\mathsf{Cov}(\nu_k,\nu_j) = \mathbf{E}\nu_k\nu_j - \mathbf{E}\nu_k\mathbf{E}\nu_j = -(N+1)/12.$$

Зауважимо, що коефіцієнт кореляції рангів дорівнює

$$\rho(\nu_k,\nu_j) = \mathsf{Cov}(\nu_k,\nu_j)/\sqrt{\mathbf{D}\nu_k\mathbf{D}\nu_j} = -1/(N-1) \to 0, N \to \infty,$$

тому ранги різних спостережень асимптотично некорельовані. До того ж вони однаково рівномірно розподілені на $\overline{1..N}$. Отже, є підстави очікувати, що за *нульової гіпотези* для сум рангів виконується *класична центральна гранична теорема,* тобто $(S_{nm} - \mathbf{E}S_{nm})/\sqrt{\mathbf{D}S_{nm}} \to^W \zeta \simeq N(0,1), n,m \to \infty$.

Обчислимо

$$\mathbf{E}S_{nm} = \sum\nolimits_{j=1}^{n}\mathbf{E}\nu_j = n(N+1)/2 = n(n+m+1)/2,$$

$$\mathbf{D}S_{nm} = \mathbf{E}S_{nm}^2 - (\mathbf{E}S_{nm})^2 =$$
$$\sum\nolimits_{j=1}^{n}\mathbf{E}\nu_j^2 + \sum\nolimits_{i\neq j=1}^{n}\mathbf{E}\nu_j\nu_i - (\mathbf{E}S_{nm})^2 =$$
$$n\ (N+1)(2N+1)/6 + n(n-1)(N+1)(3N+2)/12 - (\mathbf{E}S_{nm})^2 =$$
$$n(N+1)(N-n)/12 = nm(n+m+1)/12.$$

Отже, за нульової гіпотези має місце *слабка збіжність* при $n,m \to \infty$ центрованої та нормованої статистики Вілкоксона:

$$\varkappa_{nm} = \frac{S_{nm} - n(n+m+1)/2}{\sqrt{nm(n+m+1)/12}} \to^W \zeta \simeq N(0,1).$$

Звідси за припущення критичності великих значень статистики S_{nm} знаходимо *критичну область* для статистики S_{nm} так, щоб *вірогідний рівень* критерію наближено збігався з заданим:

$$x_{nm}(\alpha) = n(n+m+1)/2 + x_\alpha \sqrt{nm(n+m+1)/12},$$

де x_α – *квантиль* рівня $1-\alpha$ *стандартного нормального розподілу*.

Якщо виконується *альтернатива* $H_1 : F \neq G$, то можна підрахувати

$$\mathbf{E}S_{nm} = \sum\nolimits_{j=1}^{n} \mathbf{E}\left(1 + \sum\nolimits_{1\le k\neq j}^{n} 1\!\!\mathrm{I}_{\{\xi_j \,\geq\, \xi_k\}} + \sum\nolimits_{k=1}^{m} 1\!\!\mathrm{I}_{\{\xi_j \,\geq\, \eta_k\}}\right) =$$
$$n\,(1 + (n-1)/2 + m\,\mathbf{P}(\xi_1 \geq \eta_1)) =$$
$$n(n+m+1)/2 - nm\,\delta, \quad \text{де } \delta = \mathbf{P}(\xi_1 < \eta_1) - 1/2.$$

За умови, що альтернативою є припущення $\delta \equiv \int F(x)dG(x) - 1/2 < 0$, звідси виводимо, що

$$\varkappa_{nm} = \frac{S_{nm} - n(n+m+1)/2}{\sqrt{nm(n+m+1)/12}} = \frac{S_{nm} - \mathbf{E}S_{nm} - nm\,\delta}{\sqrt{nm(n+m+1)/12}} \to^{P} \infty,\ n,m\to\infty,$$

тобто критерій є *консистентним* проти альтернативи з $\delta < 0$.

Вправи

(1) Довести, що $S_{mn} = n(n+1)/2 + \sum_{i=1}^{n}\sum_{j=1}^{m} 1\!\!\mathrm{I}_{\{\xi_i<\eta_j\}}$.

(2) Обчислити з означення $\varkappa_{nm}$, що за нульової гіпотези

$\mathbf{E}(\varkappa_{nm})^{2r-1} = 0$, $\mathbf{E}(\varkappa_{nm})^{2r} \to (2r-1)!!, n\to\infty$,

при всіх $r \geq 1$, та довести асимптотичну нормальність $\varkappa_{nm}$.

(3) Розподіл статистики $\omega^2_{nm} = \int_{\mathbb{R}} (\widehat{F}_n(x) - \widehat{G}_m(x))^2 d\widehat{H}_{n+m}(x)$ за нульової гіпотези про однорідність не залежить від функції розподілу F, де $\widehat{F}, \widehat{G}, \widehat{H}$ – емпіричні функції розподілу вибірок та об'єднаної вибірки. Довести, що

$$\omega^2_{nm} = (mn)^{-1}\left[1/6 + m^{-1}\sum\nolimits_{i=1}^{n}(\nu_i - i)^2 + n^{-1}\sum\nolimits_{j=1}^{m}(\tau_j - j)^2\right] - 2/3,$$

де ν_i, τ_j – ранги виборок X, Y у об'єднаній вибірці.

3.14.4. Критерій незалежності Спірмена

Розглянемо дві *кратні вибірки* $X = (\xi_1, ..., \xi_n)$ та $Y = (\eta_1, ..., \eta_n)$ однакового об'єму n. Позначимо через $\nu = (\nu_1, ..., \nu_n)$ та $\tau = (\tau_1, ..., \tau_n)$ – *вектори рангів* цих вибірок. При виконанні гіпотези H_0 про *незалежність* X і Y та неперервність спільної функції розподілу вектори ν і τ *незалежні в сукупності* та *рівномірно розподілені* на множині Π_n перестановок порядку n. Зокрема, для математичних сподівань, дисперсій та коваріацій рангів справедливі наведені у попередньому розділі формули з заміною об'єму вибірки N на n :

$$\mathbf{E}\nu_k = \mathbf{E}\tau_k = (n+1)/2,\ \mathbf{D}\nu_k = (n^2-1)/12,$$
$$\mathrm{Cov}(\nu_k, \nu_j) = -(n+1)/12, \quad j \neq k.$$

Використаємо для перевірки H_0 статистику, що визначає *вибірковий коефіцієнт кореляції* між *векторами рангів*:

$$\widehat{\rho}_n = \sum\nolimits_{k=1}^{n} (\nu_k - \mathbf{E}\nu_k)(\tau_k - \mathbf{E}\tau_k) \ / \ \widehat{\sigma}_\nu \ \widehat{\sigma}_\tau,$$
$$\widehat{\sigma}^2_\nu = \sum\nolimits_{k=1}^{n} (\nu_k - \mathbf{E}\nu_k)^2,\ \ \widehat{\sigma}^2_\tau = \sum\nolimits_{k=1}^{n} (\tau_k - \mathbf{E}\tau_k)^2.$$

За *нерівністю Коші* статистика $\widehat{\rho}_n$ набуває значень із відрізку $[-1, 1]$. При повній тотожності векторів рангів $\nu = \tau$, що свідчить про повну позитивну залежність між X і Y, маємо $\widehat{\rho}_n = 1$, а при повній протилежності $\widehat{\rho}_n = -1$. Отже, великі за абсолютною величиною значення статистики $\widehat{\rho}_n$ вказують на залежність *векторів рангів* та відповідних вибірок.

З використанням теореми *про розподіл вектора рангів* можна підрахувати **(Вправа)**, що

$$\widehat{\sigma}^2_\nu = \widehat{\sigma}^2_\tau = n(n^2-1)/12.$$

Звідси знаходимо еквівалентне зображення для статистики Спірмена

$$\widehat{\rho}_n = 12\ c_n \sum\nolimits_{k=1}^{n} (\nu_k - (n+1)/2)(\tau_k - (n+1)/2) =$$
$$1 - 6\ c_n \sum\nolimits_{k=1}^{n} (\nu_k - \tau_k)^2.$$

де $c_n = 1/n(n^2-1)$. За нульової гіпотези

$$\mathbf{E}\widehat{\rho}_n = 12\ c_n \sum\nolimits_{k=1}^{n} \mathbf{E}(\nu_k - (n+1)/2)(\tau_k - (n+1)/2) = 0,$$

$$\mathbf{D}\widehat{\rho}_n = \mathbf{E}\widehat{\rho}^2_n = 144\ c^2_n \sum\nolimits_{i,j=1}^{n} \mathrm{Cov}(\nu_i,\ \nu_j)\ \mathrm{Cov}(\tau_i,\ \tau_j) = \frac{1}{n-1}.$$

Як відзначено вище, критичними для гіпотези незалежності є великі за модулем значення $\widehat{\rho}_n$. Тому *критична область* критерію про незалежність має вигляд $D_1 = \{\rho : |\rho| \geq r_\alpha\}$. Для забезпечення рівня α за *нульової гіпотези* при малих об'ємах n використовують табульовані значення *квантилей* розподілу статистики Спірмена, а при великих n – *асимптотичну нормальність*

$$\sqrt{n-1}\, \widehat{\rho}_n \to^W \zeta \simeq N(0, 1),$$

звідки наближено $r_\alpha = x_{\alpha/2}/\sqrt{n-1}$, де $x_{\alpha/2}$ – *квантиль* рівня $1 - \alpha/2$ *стандартного нормального розподілу*.

Вправи

(1) Перевірити рівність $\mathbf{D}\widehat{\rho}_n = 1/(n-1)$ на підставі наведених вище виразів для коваріацій рангових статистик.

(2) Довести асимптотичну нормальність $\widehat{\rho}_n$ обчисленням моментів $\mathbf{E}\widehat{\rho}_n^{2r}$ при $r \geq 1$ та їх границь при $n \to \infty$.

3.15. Критерій хі-квадрат Пірсона у поліноміальній схемі Бернуллі

Критерій хі-квадрат – найбільш універсальний з відомих критеріїв і може бути застосований для великої кількості статистичних моделей.

Користуючись групуванням спостережень, або ж підрахунком емпіричних частот для певної групи подій, часто вибірковий вектор можна звести до вигляду $X = (\zeta_1, ..., \zeta_n)$, де окремі спостереження *незалежні у сукупності*, *однаково розподілені* та набувають k різних значень $\{x_1, ..., x_k\}$:

$$\mathbf{P}_\theta(\zeta_1 = x_i) = p_i > 0,\ i = \overline{1,k},\ \sum\nolimits_{i=1}^{k} p_i = 1.$$

Така схема випробувань є узагальненням біноміальної схеми *випробувань Бернуллі*, де $k = 2$, і називається *поліноміальною схемою випробувань*. Параметром у даній схемі виступає невідомий *дискретний розподіл* $\theta = (p_i, i = \overline{1,k})$.

Аналогом *відносної частоти успіхів* для поліноміальної схеми є вектор **емпіричних частот**: $\widehat{\nu}_n = (\widehat{\nu}_{n1}, ..., \widehat{\nu}_{nk})$, де величина

$$\widehat{\nu}_{ni} = \sum\nolimits_{j=1}^{n} 1\!\!\mathrm{I}_{\{\zeta_j\, =\, x_i\}} = \left|\{j : \zeta_j = x_i\}\right|,\ i = \overline{1,k},$$

є кількістю тих випробувань із числа n, що привели до значення x_i.

Оскільки вектор параметрів $\theta = (p_i, i = \overline{1,k})$ *кратної* вибірки X містить розподіл спостережень, то її *функція вірогідності* має вигляд

$$L(X,\theta) = \prod\nolimits_{j=1}^{n} \left(\sum\nolimits_{i=1}^{k} p_i 1\!\!\mathrm{I}_{\{\zeta_j = x_i\}}\right) = \prod\nolimits_{j=1}^{n} \left(\prod\nolimits_{i=1}^{k} p_i^{1\!\!\mathrm{I}_{\{\zeta_j = x_i\}}}\right) = \prod\nolimits_{i=1}^{k} p_i^{\widehat{\nu}_{ni}},$$

а вектор **відносних емпіричних частот** $\widehat{\nu}_n/n$ збігається з *оцінкою максимальної вірогідності* параметра θ :

$$\arg\max\nolimits_{\theta\in\Theta} L(X,\theta) = \arg\max\nolimits_{\theta\in\Theta} \sum\nolimits_{i=1}^{k} \widehat{\nu}_{ni} \ln p_i = (\widehat{\nu}_{ni}/n, i = \overline{1,k}) = \widehat{\nu}_n/n.$$

Для обчислення умовного максимуму на множині можливих розподілів $\{\theta = (p_i) : p_i > 0,\ \sum p_i = 1\}$ тут застосовано метод множників Лагранжа.

Якщо інтерпретувати при кожному фіксованому i подію $\{\zeta_j = x_i\}$ як j-й успіх у послідовності з n *випробувань Бернуллі*, а решту значень ζ_j – як неуспіх, прийдемо до висновку, що величина $\widehat{\nu}_{ni}$ є відповідною кількістю успіхів у n випробуваннях Бернуллі. Тому вона має *біноміальний розподіл* із параметрами n і p_i, зокрема,

$$\mathbf{E}_\theta \widehat{\nu}_{ni} = np_i, \ \mathbf{D}_\theta \widehat{\nu}_{ni} = np_i(1 - p_i),$$

а *відносна частота успіхів* $\widehat{\nu}_{ni}/n$ є *строго консистентною асимптотично нормальною* оцінкою ймовірності p_i за теоремою *про властивості відносної частоти*. Оскільки переріз подій одиничної ймовірності має ймовірність 1, то вектор $\widehat{\nu}_n/n$ є *строго консистентною* оцінкою для дискретного розподілу θ :

$$\mathbf{P}_\theta \left(\lim_{n\to\infty} \widehat{\nu}_n/n = \theta \right) = \mathbf{P}_\theta \left(\bigcap\nolimits_{i=1}^k \{\lim_{n\to\infty} \widehat{\nu}_{ni}/n = p_i\} \right) = 1.$$

Емпіричні частоти залежні, причому випадковий вектор $\widehat{\nu}_n \in \mathbb{R}^k$ лежить у $(k-1)$-вимірному підпросторі: $\sum_{i=1}^k \widehat{\nu}_{ni} = n$.

3.15.1. Статистика хі-квадрат

Теорема (теорема Пірсона про асимптотику статистики хі – квадрат). *Нехай $\widehat{\nu}_n = (\widehat{\nu}_{n1}, ..., \widehat{\nu}_{nk})$ – вектор емпіричних частот у схемі з n поліноміальними випробуваннями $X = (\zeta_1, ..., \zeta_n)$ та розподілом імовірностей $\theta = (p_1, ..., p_k) > 0$ окремих спостережень. Тоді* ***статистика хі-квадрат***

$$\widehat{\chi}^2_{k-1}(n) \equiv \sum\nolimits_{i=1}^k \frac{(\widehat{\nu}_{ni} - np_i)^2}{np_i}$$

має граничний хі-квадрат розподіл із $k-1$ ступенями свободи, що не залежить від параметра θ :

$$\widehat{\chi}^2_{k-1}(n) \to^W \chi^2_{k-1}, \ n \to \infty.$$

Зауваження. В означення *статистики хі-квадрат* входять емпіричні частоти $\widehat{\nu}_{ni}$, які спостерігаються статистиком, та їх очікувані (середні) значення $np_i = \mathbf{E}\widehat{\nu}_{ni}$. Тому корисна мнемонічна форма для обчислення статистики:

$$\chi^2 = \sum \frac{(O-E)^2}{E},$$

де O означає Observed (частота, що спостерігається), а E – Expected (частота, що очікується).

Доведення. Розглянемо k-вимірний випадковий вектор

$$\eta(n) = \left((\widehat{\nu}_{ni} - np_i) \ / \ \sqrt{np_i}, \ i = \overline{1,k}\right)$$

та для кожного $j = \overline{1,n}$ незалежні *однаково розподілені* вектори

$$\gamma_j = \left((1\!\!\mathrm{I}_{\{\xi_j \,=\, x_i\}} - p_i) \ / \ \sqrt{p_i}, \ i = \overline{1,k}\right).$$

За означенням *емпіричних частот* $\widehat{\nu}_{ni}$ та вектора $\eta(n)$

$$\eta(n) = \left(\sum\nolimits_{j=1}^{n} (1\!\!\mathrm{I}_{\{\xi_j \,=\, x_i\}} - p_i) \ / \ \sqrt{np_i}, \ i = \overline{1,k}\right) = \sum\nolimits_{j=1}^{n} \gamma_j \ / \ \sqrt{n},$$

$$\widehat{\chi}^2_{k-1}(n) = \eta^2(n) = \eta'(n)\eta(n).$$

Обчислимо середнє та *коваріацію* одного доданку в сумі для $\eta(n)$:

$$\mathbf{E}_\theta \gamma_1 = ((\mathbf{P}_\theta(\xi_1 = x_i) - p_i) \ / \sqrt{p_i}, \ i = \overline{1,k}) = 0,$$

$$\mathrm{Cov}(\gamma_1) = \left(\mathbf{E}_\theta (1\!\!\mathrm{I}_{\{\xi_1 \,=\, x_i\}} - p_i)(1\!\!\mathrm{I}_{\{\xi_1 \,=\, x_l\}} - p_l) \ / \sqrt{p_i p_l}, \ i,l = \overline{1,k}\right) =$$

$$\left((p_i \delta_{il} - p_i p_l) / \sqrt{p_i p_l}, \ i,l = \overline{1,k}\right) = \left((\delta_{il} - \sqrt{p_i p_l}), \ i,l = \overline{1,k}\right) = I - q \cdot q',$$

де I – одинична матриця, а матриця $q \cdot q' = (\sqrt{p_i p_l}, \ i,l = \overline{1,k})$ розміру $k \times k$ утворена декартовим добутком вектора-стовпчика $q \equiv (\sqrt{p_i}, \ i = \overline{1,k})'$ на свій транспонований вектор-рядок.

Оскільки вектори γ_j незалежні і однаково розподілені, то за *класичною центральною граничною теоремою для випадкових векторів* має місце *слабка збіжність* при $n \to \infty$:

$$\eta(n) = \sum\nolimits_{j=1}^{n} \gamma_j \ / \ \sqrt{n} \to^W \xi \simeq N_k(0, \ I - q \cdot q').$$

З неперервності квадратичної функції та з означення *слабкої збіжності* векторів звідси випливає *слабка збіжність* випадкових величин

$$\widehat{\chi}^2_{k-1}(n) = \eta^2(n) \to^W \xi^2, \ n \to \infty.$$

За означенням вектор $q = (\sqrt{p_i}, \ i = \overline{1,k})'$ має одиничну норму. Нехай $(q_1, ..., q_{k-1})$ – доповнення q до ортонормованого базису в $\mathbb{R}^k$, а *ортонормована* матриця $U \equiv (q_1, ..., q_{k-1}, q)'$. Тоді $Uq = e = (0, ..., 0, 1)'$.

Розглянемо випадковий вектор $\beta = U\xi$. За теоремою *про лінійні перетворення нормальних векторів* вектор $\beta = (\beta_1, ..., \beta_k)'$ є нормальним. Оскільки за теоремою *про коваріаційну матрицю лінійного перетворення*

$$\mathbf{E}_\theta \beta = 0, \quad \mathrm{Cov}(\beta) = U\mathrm{Cov}(\xi)U' = U(I - q \cdot q')U' =$$

$$UU' - \ Uq \cdot (Uq)' = I - e \cdot e' = (\delta_{ij} 1\!\!\mathrm{I}_{i<k}, \ i,j = \overline{1,k}),$$

то випадкові величини $(\beta_1, ..., \beta_{k-1})$ є *незалежними у сукупності стандартними нормальними* величинами, а $\beta_k = 0$, оскільки $\mathbf{D}\beta_k = \mathrm{Cov}(\beta)_{kk} = 0$. Тоді за означенням *хі-квадрат розподілу* $\xi^2 \simeq \chi^2_{k-1}$, оскільки:

$$\xi^2 = \xi'\xi = (U'\beta)'U'\beta = \ \beta'UU'\beta = \beta^2 = \sum\nolimits_{i=1}^{k-1} \beta_i^2 \ \simeq \chi^2_{k-1}.$$

Отже, $\widehat{\chi}^2_{k-1}(n) \to^W \xi^2 \simeq \chi^2_{k-1}$ $\boxdot$

3.15.2. Критерій хі-квадрат для простих гіпотез

Припустимо, що *статистичний простір* відповідає наведеній вище *поліноміальній схемі випробувань*, причому невідомим параметром є розподіл результату одного випробування: $\theta = (p_1, ..., p_k)$. Розглянемо задачу перевірки *простої гіпотези* H_0, яка полягає в тому, що цей розподіл збігається з наперед заданим розподілом: $H_0 : \theta = (p_1, ..., p_k)$.

Як статистику критерію оберемо *статистику хі-квадрат*

$$\widehat{\chi}^2_{k-1}(n) = \sum\nolimits_{i=1}^{k} \frac{(\widehat{\nu}_{ni} - np_i)^2}{np_i} = n \sum\nolimits_{i=1}^{k} \frac{(\widehat{\nu}_{ni}/n - p_i)^2}{p_i},$$

що визначена в *теоремі Пірсона про асимптотику статистики хі-квадрат*. Зауважимо, що за *теоремою Бореля про асимптотику частоти успіху* має місце збіжність $\widehat{\nu}_{ni}/n \to^{P1} \mathbf{P}(\zeta_1 = x_i), n \to \infty$. Якщо гіпотеза H_0 не виконується, тобто остання гранична ймовірність не дорівнює p_i при певному i, то має місце збіжність $\widehat{\chi}^2_{k-1}(n) \to^{P1} \infty, n \to \infty$. Отже, критичними для гіпотези H_0 є великі значення статистики $\widehat{\chi}^2_{k-1}(n)$.

Тому **критерій узгодженості хі-квадрат** для перевірки гіпотези H_0 має вигляд $(\widehat{\chi}^2_{k-1}(n),\ [x_\alpha, \infty))$, де *критичний рівень* x_α знаходиться з умови $\mathbf{P}(\chi^2_{k-1} < x_\alpha) = 1 - \alpha$. З теореми Пірсона випливає, що рівень критерію при великих об'ємах вибірки наближено дорівнює α.

Відоме емпіричне (позаматематичне) правило: *при обмеженій кількості спостережень критерію хі-квадрат можна довіряти*, коли *на кожний рівень припадає не менше 5 (6, 12, 20, ...) спостережень* (тобто $\widehat{\nu}_{ni} \geq 5, 6, 12, 20, ..)$.

Більш помірний варіант необхідних застережень полягає у тому, що всі очікувані частоти не менші за 2, і не менше 80% з них не менші 5.

Критерій хі-квадрат є *конзистентним* для кожної простої альтернативи. Дійсно, якщо гіпотетичне значення $\theta = (\widetilde{p}_i, i = \overline{1,k})$ не дорівнює істинному $(p_i,\ i = \overline{1,k})$, то

$$\widetilde{\chi}^2_{k-1}(n) = \sum\nolimits_{i=1}^{k} (\widehat{\nu}_{ni} - n\widetilde{p}_i)^2 / n\widetilde{p}_i = \sum\nolimits_{i=1}^{k} (np_i - n\widetilde{p}_i + \widehat{\nu}_{ni} - np_i)^2 / n\widetilde{p}_i \geq$$

$$\sum\nolimits_{i=1}^{k}((np_i - n\widetilde{p}_i)^2/2 - (\widehat{\nu}_{ni} - np_i)^2)/n\widetilde{p}_i \geq n\Delta - c\widehat{\chi}^2_{k-1}(n),$$

де $\Delta = \sum_{i=1}^{k}(p_i - \widetilde{p}_i)^2/2\ \widetilde{p}_i > 0, \quad c = \max_i(p_i/\widetilde{p}_i)$, а *статистика хі-квадрат*

$$\widehat{\chi}^2_{k-1}(n) = \sum\nolimits_{i=1}^{k}(\widehat{\nu}_{ni} - np_i)^2/np_i$$

слабко збігається за припущенням. Тому $\widetilde{\chi}^2_{k-1}(n) \to^P \infty,\ n \to \infty$, і для кожного критичного рівня x_α *потужність* $\mathbf{P}_\theta(\widetilde{\chi}^2_{k-1}(n) \geq x_\alpha) \to 1, n \to \infty$. Отже, критерій є консистентним $\boxdot$

Приклад. Для перевірки гіпотези про симетрію гральної кості її підкидають 600 разів. Побудуємо таблицю зі стовпчиками, які відповідають номерам результатів випробувань, та з рядками, що містять такі дані:

O – *емпіричні частоти*, що спостерігаються (Observed values),

E – відповідні очікувані частоти (Expected values),

$O - E$ – різниці емпіричних та очікуваних частот,

$(O - E)^2\ /\ E$ – відповідний доданок у сумі хі-квадрат:

Грань	1	2	3	4	5	6	Σ
O	104	108	97	103	94	94	600
E	100	100	100	100	100	100	600
$O - E$	4	8	-3	3	-6	-6	0
$(O - E)^2/E$	.16	.64	.09	.09	.36	.36	1.70

За таблицями знаходимо, що $\mathbf{P}(\chi^2_5 \geq 15.09) \approx 0.01$. Значення статистики менше за критичне: $1.70 < 15.09$, тому альтернатива про асиметрію повинна бути відкинута на рівні 0.01, і приймається нульова гіпотеза про симетрію.

Вправи

(1) Нижче наведені частоти цифр $\overline{0,9}$ у 10002 десяткових знаках числа $\pi - 3$

0	1	2	3	4	5	6	7	8	9
968	1026	1021	974	1014	1046	1021	970	948	1014

За допомогою критерію хі-квадрат на рівні 0.01 перевірити гіпотезу про рівномірну розподіленість десяткових цифр числа π.

(2) Знайти сумісний розподіл та його генератрису для вектора емпіричних частот $\widehat{\nu}_n = (\widehat{\nu}_{n1}, ..., \widehat{\nu}_{nk})$ у поліноміальній схемі Бернуллі.

(3) Довести, що статистика хі-квадрат у поліноміальній схемі Бернуллі дорівнює $\widehat{\chi}^2_{k-1}(n) = \sum_{i=1}^{k}\widehat{\nu}^2_{ni}/np_i - n$.

(4) Нехай $\widehat{\nu}_{nij}$ – кількість спостережень j-го результату до першої появи i-го результату у поліноміальній схемі Бернуллі. (а) Знайти генератрису, математичні сподівання та дисперсії $\widehat{\nu}_{nij}$. (б) Знайти сумісну генератрису $(\widehat{\nu}_{nij}, j = \overline{1, i-1})$.

3.15.3. Групування, гістограма

Значна кількість статистичних випробувань зводиться до *поліноміальної схеми випробувань*. Для цього застосовують **групування спостережень.**

Для прикладу розглянемо *кратну вибірку* $X = (\xi_1, ..., \xi_n)$.

Нехай $H = (H_i,\ i = \overline{1,k})$ – довільне вимірне *розбиття* простору значень окремих спостережень ξ_j. Визначимо на цьому просторі вимірну функцію $g(y) = \sum_{i=1}^{k} i\ 1\!\!\mathrm{I}_{y \in H_i}$, що набуває k різних значень. **Групованою вибіркою** за розбиттям H називається вектор

$$\zeta = g(X) = (\zeta_j = g(\xi_j),\ j = \overline{1,n}).$$

Випадкові величини $(\zeta_j,\ j = \overline{1,n})$ *незалежні у сукупності*, *однаково розподілені* і набувають k різних значень, тобто $\zeta = (\zeta_1, ..., \zeta_n)$ є вибіркою спостережень із *поліноміальної схеми випробувань*.

Теоретичні частоти обчислюється через розподіл спостережень:

$$p_i = \mathbf{P}_\theta(\xi_1 \ \in\ H_i),\ i = \overline{1,k}.$$

Відповідний вектор *емпіричних частот* $\widehat{\nu}_n = (\widehat{\nu}_{n1}, ..., \widehat{\nu}_{nk})$ для групованої вибірки (ζ_j) має вигляд

$$\widehat{\nu}_{ni} = \sum\nolimits_{j=1}^{n} 1\!\!\mathrm{I}_{\{\zeta_j = i\}} = \sum\nolimits_{j=1}^{n} 1\!\!\mathrm{I}_{\{\xi_j \ \in\ H_i\}} = \left|\{j : \xi_j \ \in\ H_i\}\right|,\ i = \overline{1,k},$$

і називається **гістограмою** вибірки $X = (\xi_j,\ j = \overline{1,n})$ при розбитті H. Гістограми часто зображують у вигляді ряду вертикальних стовпчиків, висоти яких пропорційні відповідним емпіричним частотам

5					■■	
4				■■	■■	
3	■■			■■	■■	■■
2	■■		■■	■■	■■	■■
1	■■	■■	■■	■■	■■	■■
0	H_1	H_2	H_3	H_4	H_5	H_6

Існує більш адекватний візуальний спосіб групування даних, що називається **графіком стебла та листя** (stem and leaf plot). Для його побудови дані масштабують так, щоб вони належали відрізку $[0.0, ..., 9.9]$, а потім у стовпчик з номером, що дорівнює цілій частині числа, записують його першу після коми цифру. Кількість цифр у стовпчику відображає частоту потрапляння даних у відповідний інтервал, а склад цифр свідчить про розподіл даних всередині інтервалів групування. Наприклад, вибірка 21, 27, 172, 233, 251, 259, 414, 542, 584, 590, 633, 691 зображується у вигляді:

		5			9				
2		5			8	9			
2	7	3		1	4	3			
—	—	—	—	—	—	—	—	—	—
0	1	2	3	4	5	6	7	8	9

3.15.4. Критерій хі-квадрат узгодженості з функцією розподілу

Для перевірки узгодженості *кратної вибірки* $X = (\xi_1, ..., \xi_n)$ із заданою *функцією розподілу* F одного спостереження числову вісь розбивають на k несумісних інтервалів: $\mathbb{R} = \cup_{i=1}^{k} \Delta_i$, та обчислюють *гістограму*

$$\widehat{\nu}_{ni} = \sum\nolimits_{j=1}^{n} 1\!\!\mathrm{I}_{\{\xi_j \,\in\, \Delta_i\}} = \left|\{j : \xi_j \;\in\; \Delta_i\}\right|, \quad i = \overline{1, k}.$$

Статистика критерію дорівнює

$$\widehat{\chi}_{k-1}^{2}(n) = \sum\nolimits_{i=1}^{k} \frac{(\widehat{\nu}_{ni} - nF(\Delta_i))^2}{nF(\Delta_i)}$$

і асимптотично має *хі-квадрат розподіл* χ_{k-1}^{2}. Теоретичні ймовірності дорівнюють приростам функції F на інтервалах: $F(\Delta_i) = \mathbf{P}(\xi_j \in \Delta_i)$.

Критерій узгодженості хі-квадрат із функцією розподілу має вигляд $(\widehat{\chi}_{k-1}^{2}(n),\ [z_\alpha, \infty))$, де значення *критичного рівня* z_α визначається з умови $\mathbf{P}(\chi_{k-1}^{2} < z_\alpha) = 1 - \alpha$. З *теореми Пірсона про асимптотику статистики хі-квадрат* випливає, що *вірогідний рівень* критерію $\mathbf{P}(\widehat{\chi}_{k-1}^{2}(n) \geq z_\alpha)$ наближено (при великих об'ємах вибірки) дорівнює α.

Зауваження. Слід мати на увазі, що при застосуванні групування відбувається підміна формулювання *нульової гіпотези*. Так, у задачі про узгодженість вибірки з заданою функцією розподілу нульова гіпотеза після групування зводиться до твердження, що ймовірності потрапляння одного спостереження в інтервали розбиття збігаються з наперед заданими. Останню умову задовольняють багато *різних* теоретичних функцій розподілу. Для подолання даної обставини слід збільшувати кількість k інтервалів групування. З іншого боку, при завеликих k кількість спостережень в деяких інтервалах може бути неприйнятно малою. Тому на практиці рекомендують обирати $k \approx c \ln n$.

3.15.5. Критерій хі-квадрат для складних гіпотез

Розглянемо *поліноміальну схему випробувань*, в якій розподіл спостереження залежить від значення векторного параметра $\theta \in \Theta \subset \mathbb{R}^d$:

$$\mathbf{P}_\theta(\zeta_1 = x_i) = p_i(\theta),\ i = \overline{1,k}.$$

Функції $p_i(\theta)$ вважаються повністю відомими і утворюють при кожному θ *дискретний розподіл*: $p_i(\theta) > 0$, $\sum_{i=1}^k p_i(\theta) = 1$. Як і вище, результати спостережень представлені вектором *емпіричних частот* $\widehat{\nu}_n = (\widehat{\nu}_{n1}, ..., \widehat{\nu}_{nk})$.

Розглянемо випадок, коли параметр θ не тільки невідомий, але й відсутні будь-які припущення про його значення. Тоді спочатку доведеться висунути якесь припущення щодо значення θ на підставі спостережень (тобто фактично оцінити це значення), а потім перевірити статистичну *гіпотезу* про відповідність істинного параметра його оцінці. Очевидно, що використання наведеного вище критерію хі-квадрат стає некоректним, оскільки підстановка замість значення θ його оцінки змінює розподіл *статистики хі-квадрат* $\widehat{\chi}^2_{k-1}(n)$ і, зокрема, деформує рівень критерію. Як виявляється, для розв'язання проблеми існує необхідна модифікація критерію Пірсона.

Теорема (про асимптотику модифікованої статистики хі- квадрат). *Припустимо, що у поліноміальній схемі випробувань з k значеннями у одному випробуванні виконуються умови:*

(а) $\inf_{\theta,\, i} p_i(\theta) > 0$,

(б) $p_i(\theta) \in C^2(\Theta)$, $\forall i = \overline{1,k}$,

(в) $rang\big(\partial p_i(\theta)/\partial\theta,\ i = \overline{1,k}\big) = d \equiv \dim\Theta < k$.

Якщо оцінка $\widehat{\theta}_n$ є оцінкою максимальної вірогідності за емпіричними частотами $\widehat{\nu}_n = (\widehat{\nu}_{n1}, ..., \widehat{\nu}_{nk})$, тобто

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} L(\widehat{\nu}_n, \theta) = \arg\max_{\theta\in\Theta} \prod_{i=1}^k p_i^{\widehat{\nu}_{ni}}(\theta),$$

*то **модифікована статистика хі-квадрат***

$$\widehat{\chi}^2\left(\widehat{\theta}_n\right) \equiv \sum_{i=1}^k \frac{(\widehat{\nu}_{ni} - np_i(\widehat{\theta}_n))^2}{np_i(\widehat{\theta}_n)}$$

слабко збігається до величини з хі-квадрат розподілом і кількістю ступенів свободи, що скоригована на число оцінених параметрів:

$$\widehat{\chi}^2\left(\widehat{\theta}_n\right) \to^W \chi^2_{k-1-d},\ n \to \infty.$$

Доведення теореми не наводиться.

Як і вище, критичними для гіпотези H_0 є великі значення статистики.

Модифікований критерій узгодженості хі-квадрат задається парою $(\widehat{\chi}^2_{k-d-1}(n),\ [z_\alpha, \infty))$, де *критичний рівень* z_α обчислюється з рівняння $\mathbf{P}(\chi^2_{k-d-1} < z_\alpha) = 1 - \alpha$. З теореми *про асимптотику модифікованої статистики хі-квадрат* випливає, що *вірогідний рівень* критерію наближено (при великих об'ємах вибірки) дорівнює α.

3.15.6. Критерій хі-квадрат однорідності

Розглянемо статистичні випробування, що складаються із m серій поліноміальних випробувань $X_l = (\zeta_{lj},\ j = \overline{1, n_l})$ по n_l спостережень у l-й серії, $l = \overline{1, m}$. Окреме випробування в кожній серії може закінчитися одним із k результатів $\{x_1, ..., x_k\}$. Спостереження зображені подвійним вектором *емпіричних частот*

$$\widehat{\nu}_n = (\widehat{\nu}_l(n_l), l = \overline{1, m}) = ((\widehat{\nu}_{li},\ i = \overline{1, k}),\ l = \overline{1, m}),$$

$$\widehat{\nu}_{li} = \sum\nolimits_{j=1}^{n_s} \mathrm{I\!I}_{\{\zeta_{lj} = x_i\}} = \left|\{j : \zeta_{lj} = x_i\}\right|, \quad n = \sum\nolimits_{l=1}^{m} n_l.$$

Усі випробування *незалежні у сукупності*. Нульова гіпотеза полягає в тому, що розподіли всіх спостережень однакові, однак припущення щодо спільного розподілу не формулюються. Якби був відомий спільний розподіл спостережень: $\theta = (p_i, i = \overline{1, k})$, то для перевірки нульової *простої гіпотези*

$$H_0 : \mathbf{P}_\theta(\zeta_{l1} = x_i) = p_i,\ \forall i = \overline{1, k},\ \forall l = \overline{1, m},$$

можна було б використати при $n_l \to \infty$, $l = \overline{1, m}$, *статистику хі-квадрат*

$$\widehat{\chi}^2(\theta) = \sum\nolimits_{l=1}^{m} \sum\nolimits_{i=1}^{k} \frac{(\widehat{\nu}_{li} - n_l p_i)^2}{n_l p_i} \to^W \sum\nolimits_{l=1}^{m} \chi^2_{k-1,\ l} \simeq \chi^2_{mk-m}.$$

Остання рівність тут випливає з адитивності *хі-квадрат розподілів* відносно додавання незалежних величин, що є наслідком теореми *про інваріантність гама-розподілів відносно згортки*. Якщо розглянути об'єднану вибірку $(\zeta_{lj}, j = \overline{1, n_l}, l = \overline{1, m})$ без припущення однорідності, тобто при залежності розподілу (p_i) також від номера серії l, то число ступенів свободи $mk - m$ можна інтерпретувати як справжню розмірність вектора розподілів $(\mathbf{P}_\theta(\zeta_{l1} = x_i), i = \overline{1, k}, l = \overline{1, m})$, оскільки суми ймовірностей для кожної з m серій дорівнюють 1.

При невідомому розподілі вибірки, тобто за *нульової гіпотези*

$$H_0' : \mathbf{P}_\theta(\zeta_{lj} = x_i) = \mathbf{P}_\theta(\zeta_{11} = x_i),\ \forall i = \overline{1, k},\ \forall l = \overline{1, m},\ \forall \theta,$$

для цього розподілу $\theta = (p_i, i = \overline{1,k})$ необхідно спочатку побудувати *оцінку максимальної вірогідності*

$$\widehat{\theta}_n = \arg\max\nolimits_{\theta\in\Theta} \prod\nolimits_{l=1}^{m} \prod\nolimits_{i=1}^{k} p_i^{\widehat{\nu}_{li}} = \arg\max\nolimits_{\theta\in\Theta} \prod\nolimits_{i=1}^{k} p_i^{\widehat{\nu}_{\cdot i}} =$$

$$\arg\max\nolimits_{\theta\in\Theta} \sum\nolimits_{i=1}^{k} \widehat{\nu}_{\cdot i} \ln p_i = (\widehat{\nu}_{\bullet i} \;/\; n,\ i = \overline{1,k}),$$

$$\text{де } \ \widehat{\nu}_{\bullet i} = \sum\nolimits_{l=1}^{m} \widehat{\nu}_{li},\ n = \sum\nolimits_{l=1}^{m} n_l.$$

Зауважимо, що фактична розмірність параметричного простору має вигляд $d \equiv \dim\Theta = k - 1$, оскільки сума координат k-вимірного вектора θ завжди дорівнює одиниці.

Отже, **модифікована статистика хі-квадрат для перевірки однорідності** має вигляд

$$\widehat{\chi}^2\left(\widehat{\theta}_n\right) = \sum\nolimits_{l=1}^{m} \sum\nolimits_{i=1}^{k} \frac{(\widehat{\nu}_{li} - n_l \widehat{\nu}_{\bullet i}/n)^2}{n_l \widehat{\nu}_{\bullet i}/n}$$

і за теоремою *про асимптотику модифікованої статистики хі-квадрат* асимптотична кількість ступенів свободи дорівнює зменшеній на $k-1$ фактичної кількості параметрів $mk - m$:

$$\widehat{\chi}^2\left(\widehat{\theta}_n\right) \to^{W} \chi^2_{mk-m-(k-1)} = \chi^2_{(m-1)(k-1)},\ n \to \infty.$$

Критерій хі-квадрат однорідності має вигляд $\left(\widehat{\chi}^2(\widehat{\theta}_n),\ [z_\alpha, \infty)\right)$, де *критичний рівень* z_α визначається умовою $\mathbf{P}(\chi^2_{(m-1)(k-1)} < z_\alpha) = 1 - \alpha$. З теореми *про асимптотику модифікованої статистики хі-квадрат* випливає, що *вірогідний рівень* критерію наближено (при великих n) дорівнює α.

3.15.7. Критерій хі-квадрат незалежності

Розглянемо *кратну вибірку*, що містить двовимірні спостереження (пари спостережень):

$$X = ((\xi_j, \eta_j),\ j = \overline{1,n}).$$

Сумісний розподіл окремих пар спостережень є невідомим, а нульова гіпотеза полягає в тому, що спостереження в парі є незалежними. За допомогою групування можна звести задачу до випадку, коли множина вибіркових значень є скінченною: $\xi_1 \in \{x_1, ..., x_k\}$, $\eta_1 \in \{y_1, ..., y_m\}$. У цьому випадку нульова *проста гіпотеза* формулюється у вигляді

$$H_0 : \mathbf{P}_\theta(\xi_1 = x_i, \eta_1 = y_l) = p_i q_l,\ \forall i = \overline{1,k},\ \forall l = \overline{1,m},$$

де вектор розподілу координат $\theta = (p_i, i = \overline{1,k},\ \ q_l, l = \overline{1,m})$, є невідомим параметром і підлягає оцінці. Якби ці розподіли були відомими, то критерій хі-квадрат містив би сумісні *емпіричні частоти*

$$\nu_{il} = \sum\nolimits_{j=1}^{n} 1\!\!1_{\{\xi_j = x_i,\, \eta_j = y_l\}} = \left|\{j : \xi_j = x_i,\ \eta_j = y_l\}\right|$$

і мав використовувати *статистику хі-квадрат*

$$\widehat{\chi}^2_{km-1}(\theta) = \sum\nolimits_{i=1}^{k} \sum\nolimits_{l=1}^{m} \frac{(\nu_{il} - np_i q_l)^2}{np_i q_l}.$$

Таблиця з частотами $(\nu_{il},\ i = \overline{1,k},\ l = \overline{1,m})$ називається **таблицею спряженості** факторів, що впливають на значення спостережень у рядках та стовпчиках.

При розподілі $\theta = (p_i, i = \overline{1,k},\ q_l, l = \overline{1,m})$ нульова гіпотеза має вигляд

$$H_0' : \mathbf{P}_\theta(\xi_1 = x_i,\ \eta_1 = y_l) =$$
$$\mathbf{P}_\theta(\xi_1 = x_i)\mathbf{P}_\theta(\eta_1 = y_l),\ \forall i = \overline{1,k},\ \forall l = \overline{1,m},$$

а *оцінка максимальної вірогідності* для цього розподілу дорівнює

$$\widehat{\theta}_n = \arg\max\nolimits_{\theta\in\Theta} \prod\nolimits_{i=1}^{k} \prod\nolimits_{l=1}^{m} (p_i q_l)^{\nu_{il}} = \arg\max\nolimits_{\theta\in\Theta} \left(\prod\nolimits_{i=1}^{k} p_i^{\widehat{\nu}_{i\bullet}}\right) \left(\prod\nolimits_{l=1}^{m} q_l^{\widehat{\nu}_{\bullet l}}\right)$$
$$= \arg\max_{\theta\in\Theta} \left(\sum\nolimits_{i=1}^{k} \widehat{\nu}_{i\bullet} \ln p_i + \sum\nolimits_{l=1}^{m} \widehat{\nu}_{\bullet l} \ln q_l\right) =$$
$$(\widehat{\nu}_{i\bullet}/n,\ i = \overline{1,k},\ \ \widehat{\nu}_{\bullet l}/n,\ l = \overline{1,m}),\ \text{де}\ \ \widehat{\nu}_{i\bullet} = \sum\nolimits_{l=1}^{m} \widehat{\nu}_{il},\ \widehat{\nu}_{\bullet l} = \sum\nolimits_{i=1}^{k} \widehat{\nu}_{il}.$$

Отже, **модифікована статистика хі-квадрат для перевірки незалежності** парних спостережень має вигляд

$$\widehat{\chi}^2(\widehat{\theta}_n) \equiv \sum\nolimits_{i=1}^{k} \sum\nolimits_{l=1}^{m} \frac{(\widehat{\nu}_{il} - \widehat{\nu}_{i\bullet}\widehat{\nu}_{\bullet l}/n)^2}{\widehat{\nu}_{i\bullet}\widehat{\nu}_{\bullet l}/n}.$$

Кількість параметрів, що були оцінені, дорівнює $\dim\Theta = k - 1 + m - 1$. Тому з теореми *про асимптотику модифікованої статистики хі-квадрат* робимо висновок, що за нульової гіпотези

$$\widehat{\chi}^2(\widehat{\theta}_n) \to^W \chi^2_{mk-1-(m-1+k-1)} = \chi^2_{(m-1)(k-1)},\ n \to \infty.$$

Критерій хі-квадрат незалежності парних спостережень має вигляд $(\widehat{\chi}^2(\widehat{\theta}_n),\ [z_{(m-1)(k-1),\alpha}, \infty))$, де *критичний рівень* $z_{(m-1)(k-1),\alpha}$ визначається за рівнем α умовою $\mathbf{P}(\chi^2_{(m-1)(k-1)} < z_{(m-1)(k-1),\alpha}) = 1 - \alpha$. З теореми *про асимптотику модифікованої статистики хі-квадрат* робимо висновок, що *вірогідний рівень* даного критерію наближено (при великих об'ємах вибірки) дорівнює α.

Зауваження. У випадку, коли $m = k = 2$, тобто статистика хі-квадрат має 1 ступінь свободи, всі доданки у сумі є пропорційними. У цьому разі рекомендується використання **поправки Йетса на неперервність**, що полягає у заміні статистики на $\widehat{\chi}^2(\widehat{\theta}_n) \equiv (\widehat{\nu}_{11} - \widehat{\nu}_{1\bullet}\widehat{\nu}_{\bullet 1}/n - 1/2)^2 \;/\; (\widehat{\nu}_{1\bullet}\widehat{\nu}_{\bullet 1}\widehat{\nu}_{2\bullet}\widehat{\nu}_{\bullet 2}/n^3)$.

Приклад. Соціологи провели опитування трьох груп населення: з середньою, професійною та вищою освітою щодо необхідності проведення економічних реформ. Результати зображено у *таблиці спряженості*

	Середня	Професійна	Вища	Σ
Так	40	75	35	150
Ні	160	225	65	450
Σ	200	300	100	600

Результат обробки таблиці за критерієм хі-квадрат має вигляд:

Висновок	Освіта	O	E	$O-E$	$(O-E)^2/E$
Так	С	40	$50 = 200 \cdot 150/600$	-10	2
Так	П	75	$75 = 300 \cdot 150/600$	0	0
Так	В	35	$25 = 100 \cdot 150/600$	10	4
Ні	С	160	$150 = 200 \cdot 450/600$	10	0.67
Ні	П	225	$225 = 300 \cdot 450/600$	0	0
Ні	В	65	$75 = 100 \cdot 450/600$	-10	1.33
Σ		600	600	0	8.00

Оскільки значення статистики $8.00 > 5.99 = z_{2,\ 0.05}$, то нульова гіпотеза про незалежність фактора освіти від думки щодо необхідності реформ відкидається на рівні 0.05 – тобто приймається альтернатива про залежність (у даному випадку – позитивну залежність від рівня освіти).

Вправа. Довести, що у 2×2 таблиці спряженості умовний розподіл частоти $\widehat{\nu}_{11}$ за умови, що фіксовані $\widehat{\nu}_{\bullet j}$ та $\widehat{\nu}_{i\bullet}$, є гіпергеометричним. Знайти його. Побудувати точний **критерій Фішера перевірки незалежності**, якщо критичними є великі відхиленні у обидва боки величини $\widehat{\nu}_{11}$ від $\widehat{\nu}_{\bullet 1}\widehat{\nu}_{1\bullet}\;/n$.

3.16. Перевірка гіпотез про параметри нормальних спостережень

Критерії узгодженості для нормальних виборок ґрунтуються на спеціальних властивостях вибіркових моментів, які використовувались вище при побудові *надійних інтервалів для нормальних спостережень.*

Відзначимо, що існує прямий зв'язок між перевіркою гіпотез про параметри та побудовою надійних інтервалів для них. А саме, *надійний інтервал* для невідомого параметра можна розглядати для множину прийнятних значень для його гіпотетичного значення, а доповнення цього інтервалу – як відповідну критичну область для нульової гіпотези. Наприклад, більшість наведених вище *надійних інтервалів для нормальних спостережень* вигляду $[\widehat{\theta}_1, \widehat{\theta}_1]$ побудовані так, що $\{\widehat{\theta}_1 < \theta < \widehat{\theta}_2\} = \{|\widehat{\varkappa}(X, \theta)| < x_a\}$ для деякої *опорної величини* $\widehat{\varkappa}(X, \theta)$. Тут критичне значення x_a обране з умови $\mathbf{P}(\widehat{\theta}_1 < \theta < \widehat{\theta}_2) = 1 - \alpha$. Тоді $(|\widehat{\varkappa}(X, \theta_0)|, [x_a, \infty))$ є критерієм рівня α для перевірки гіпотези $H_0 : \theta = \theta_0$.

Нехай $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка* з нормальним розподілом спостережень $N(\mu, \sigma^2)$. Наведені критерії містять статистики: $\widehat{\mu}_n$ – *вибіркове середнє*, $\widehat{\sigma}_n^2$ – *вибіркова дисперсія*, та $\widehat{s}_n^2$ – *нормована вибіркова дисперсія*.

Нижче x_α означає двобічний *квантиль* вірогідного рівня $1 - \alpha$ для *стандартного нормального розподілу*: $\mathbf{P}(|\zeta| < x_\alpha) = 1 - \alpha$. Аналогічний зміст мають $y_{n\alpha}$ для *розподілу Стьюдента* з n ступенями свободи, та $z_{n\alpha}$ – для *хі-квадрат розподілу* з n ступенями свободи.

3.16.1. Перевірка гіпотез про параметри однієї вибірки

Гіпотеза про середнє при відомій дисперсії. Нехай *дисперсія* σ^2 відома. Випадкова величина

$$\widehat{\varkappa}_n(\mu, \sigma) = \sqrt{n}\ (\widehat{\mu}_n - \mu)/\sigma$$

має *стандартний нормальний розподіл* за теоремою *про вибіркові середнє та дисперсію нормальної вибірки*. Тому двобічний критерій вигляду $(|\widehat{\varkappa}_n(\mu_0, \sigma)|,\ [x_\alpha, \infty))$ для перевірки гіпотези $H_0 : \mu = \mu_0$ має *вірогідний рівень* α. Критерій є *консистентним*, оскільки при $\mu_0 \neq \mu$ статистика

$$\widehat{\varkappa}_n(\mu_0, \sigma) = \sqrt{n}(\widehat{\mu}_n - \mu + \mu - \mu_0)/\sigma \to^P \pm\infty,\ n \to \infty.$$

Гіпотеза про дисперсію при відомому середньому. При відомому середньому μ вибіркова дисперсія $\widehat{\sigma}_n^2$ є *статистикою*, а величина

$$\widehat{\chi}_n^2(\sigma) = n\widehat{\sigma}_n^2\ /\ \sigma^2 = \sum\nolimits_{k=1}^{n} (\xi_k - \mu)^2\ /\ \sigma^2$$

має *хі-квадрат розподіл* із n ступенями свободи за теоремою *про вибіркові середнє та дисперсію нормальної вибірки*. Тому критерій $(\widehat{\chi}_n^2(\sigma_0),\ [z_{n\alpha}, \infty))$ для перевірки гіпотези $H_0 : \sigma = \sigma_0$ має *вірогідний рівень* α.

Гіпотеза про середнє при невідомій дисперсії. За теоремою *про статистику Стьюдента від нормальної вибірки* величина

$$\widehat{\tau}_{n-1}(\mu) = \sqrt{n}\ \frac{\widehat{\mu}_n - \mu}{\widehat{s}_n}$$

має *розподіл Стьюдента* з $n-1$ ступенем свободи. Тому критерій вигляду $(|\widehat{\tau}_{n-1}(\mu_0)|\,,\ [y_{n-1,\alpha},\infty))$ для перевірки гіпотези $H_0:\mu=\mu_0$ має *вірогідний рівень* α. Критерій є *консистентним.*

Приклад. 10 випадково обраних студентів третього курсу показали такі результати тестування на 100-бальних випробуваннях на початку і в кінці навчального року:

N	1	2	3	4	5	6	7	8	9	10
Вересень	80	72	61	74	85	72	93	87	98	65
Липень	85	73	60	75	96	80	87	82	98	70
Різниця	5	1	−1	1	11	8	6	−5	0	5

Для перевірки гіпотези про наявне поліпшення рівня знань обчислимо для різниць $\widehat{\mu}_{10}=3.1, \widehat{s}_{10}=4.75$, та значення $\widehat{\tau}_9(0)=\sqrt{10}(3.1-0)/4.75$ $=2.1$. Оскільки $2.1 > y_{\ 10-1,\ 0.01}=1.83$, то на рівні 0.01 гіпотеза про поліпшення не може бути відхилена – приймається.

Гіпотеза про дисперсію при невідомому середньому. За теоремою *про вибіркові середнє та дисперсію нормальної вибірки* величина

$$\widehat{\chi}^2_{n-1}(\sigma)=(n-1)\widehat{s}^2_n\ /\ \sigma^2$$

має *хі-квадрат розподіл* із $n-1$ ступенем свободи. Відповідно критерій вигляду $\left(\widehat{\chi}^2_{n-1}(\sigma_0),\ [z_{n-1,\alpha},\infty)\right)$ має *вірогідний рівень* α.

3.16.2. Перевірка гіпотез про параметри двох вибірок

Гіпотеза про різницю середніх при відомих дисперсіях. Нехай спостерігаються дві незалежні кратні нормальні вибірки $X=(\xi_1,...,\xi_n)$ та $Y=(\eta_1,...,\eta_m)$, що мають розподіли

$$\xi_1\simeq N(\mu_1,\sigma_1^2),\quad \eta_1\simeq N(\mu_2,\sigma_2^2)$$

з невідомими середніми μ_k та відомими дисперсіями σ_k^2.

Нульова гіпотеза формулюється як

$$H_0:\mu_2-\mu_1=\Delta,$$

де величина Δ відома. Часто розглядають випадок, коли $\Delta=0$.

Нехай $\widehat{\mu}_n$, $\widehat{\mu}_m$ – відповідні *вибіркові середні* для X і Y. Величини $\widehat{\mu}_n$, $\widehat{\mu}_m$ незалежні за теоремою *про векторні перетворення незалежних величин*, причому $\widehat{\mu}_n-\mu_1\simeq N(0,\sigma_1^2/n)$, $\widehat{\mu}_m-\mu_2\simeq N(0,\sigma_2^2/m)$ за теоремою *про вибіркові середнє та дисперсію нормальної вибірки.* Тому за нульової гіпотези

$$\widehat{\mu}_m - \widehat{\mu}_n - \Delta \simeq N(0, \sigma_1^2/n + \sigma_2^2/m),$$

за теоремою *про нормальність суми незалежних нормальних векторів.*

Отже, статистика критерію

$$\widehat{\zeta}_{nm} = (\widehat{\mu}_m - \widehat{\mu}_n - \Delta)\ /\sqrt{\sigma_1^2/n + \sigma_2^2/m}$$

має за нульової гіпотези *стандартний нормальний розподіл*, а його *критична область* відповідає великим її значенням.

Критерій $(\left|\widehat{\zeta}_{nm}\right|,\ [x_\alpha, \infty))$ для перевірки гіпотези про різницю середніх має *вірогідний рівень* α та є *консистентним.*

Гіпотеза про різницю середніх при невідомій дисперсії. Нехай, як і вище, спостерігаються незалежні кратні нормальні вибірки $X = (\xi_1, ..., \xi_n)$ та $Y = (\eta_1, ..., \eta_m)$ з розподілами

$$\xi_1 \simeq N(\mu_1, \sigma^2), \quad \eta_1 \simeq N(\mu_2, \sigma^2)$$

з невідомими середніми μ_k та однаковими невідомими дисперсіями σ^2.

Нульова гіпотеза має вигляд

$$H_0 : \mu_2 - \mu_1 = \Delta,$$

де величина Δ відома, зокрема, можливо, $\Delta = 0$.

Нехай $\widehat{\mu}_n$, $\widehat{\mu}_m$ – відповідні *вибіркові середні* для X та Y, а $\widehat{s}_n^2$, $\widehat{s}_m^2$ – *нормовані вибіркові дисперсії.* Тоді з незалежності вибірок X та Y і з теорем *про векторні перетворення незалежних величин* та *про вибіркові середнє та дисперсію нормальної вибірки* випливає, що величини $\widehat{\mu}_n$, $\widehat{\mu}_m$, $\widehat{s}_n^2$, $\widehat{s}_m^2$ *незалежні у сукупності.* Тому випадкова величина

$$\sqrt{\frac{nm}{n+m}}\,(\widehat{\mu}_m - \widehat{\mu}_n - \Delta)\,/\sigma,$$

що має нульове середнє та одиничну дисперсію, є нормально розподіленою і до того ж не залежить від суми

$$((n-1)\widehat{s}_n^2 + (m-1)\widehat{s}_m^2)\ /\ \sigma^2,$$

яка має *хі-квадрат розподіл* із $n+m-2$ ступенями свободи за означенням.

Отже, за нульової гіпотези H_0 внаслідок означення *розподілу Стьюдента* статистика, що утворена відношенням:

$$\widehat{\tau}_{n+m-2} = \sqrt{\frac{nm}{n+m}}\ \frac{\widehat{\mu}_m - \widehat{\mu}_n - \Delta}{\sqrt{((n-1)\widehat{s}_n^2 + (m-1)\widehat{s}_m^2)\,/(n+m-2)}}$$

має розподіл Стьюдента з $n+m-2$ ступенями свободи.

Критерій Стьюдента $(|\widehat{\tau}_{n+m-2}|\,,\ [y_{n+m-2,\alpha},\infty))$ для перевірки гіпотези про різницю середніх має *вірогідний рівень* α. Критерій є *консистентним*.

Гіпотеза про відношення дисперсій при відомих середніх. Одночасно спостерігаються незалежні кратні нормальні вибірки $X=(\xi_1,...,\xi_n)$ та $Y=(\eta_1,...,\eta_m)$ з розподілами $\xi_1 \simeq N(\mu_1,\sigma_1^2)$ і $\eta_1 \simeq N(\mu_2,\sigma_2^2)$ з відомими середніми μ_k та невідомими дисперсіями σ_k^2.

Нульова гіпотеза формулюється як

$$H_0 : \sigma_2^2 = \rho\sigma_1^2,$$

де відношення ρ відоме (зокрема $\rho = 1$). Тоді за гіпотези H_0 статистика

$$\widehat{\phi}_{n,m} = \frac{\widehat{\sigma}_n^2\ /\ \sigma_1^2}{\widehat{\sigma}_m^2\ /\ \sigma_2^2} = \rho\frac{\widehat{\sigma}_n^2}{\widehat{\sigma}_m^2}$$

має *розподіл Фішера* з n,m ступенями свободи.

Відповідно критерій $(\widehat{\phi}_{n,m},\ [0,w_{\alpha/2}/\rho]\cup[w_{1-\alpha/2}/\rho,\infty))$ має рівень α, де границі w_α *критичної області* обрані з умови $\mathbf{P}(\phi_{n,m} < w_\alpha) = \alpha$.

Гіпотеза про відношення дисперсій при невідомих середніх. При невідомих середніх слід обрати *нормовані вибіркові дисперсії*. У цьому випадку статистика

$$\widehat{\phi}_{n-1,m-1} = \frac{\widehat{s}_n^2\ /\ \sigma_1^2}{\widehat{s}_m^2\ /\ \sigma_2^2} = \rho\frac{\widehat{s}_n^2}{\widehat{s}_m^2}$$

має *розподіл Фішера* з $n-1, m-1$ ступенями свободи. Тому критерій вигляду $(\widehat{\phi}_{n-1,m-1},\ [0,w_{\alpha/2}]\cup[w_{1-\alpha/2},\infty))$ має рівень α.

Іноді на практиці для перевірки гіпотези про рівність середніх одночасно використовують критерій про відношення дисперсій (для обґрунтування припущення про рівність дисперсій), а потім вже критерій про різницю середніх, що оснований на такому припущенні. Зауважимо, що в такому випадку статистики критеріїв є залежними, тому вірогідний рівень комбінованого критерію не є передбачуваним.

3.16.3. Кореляційний аналіз

Розглянемо *кратну вибірку* $X=(\xi_1,...,\xi_n)$ з двомірним нормальним розподілом спостережень: $\xi_j = (\xi_{1j},\xi_{2j}) \simeq N_2(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ з невідомими середніми μ_k, дисперсіями σ_k^2 та коефіцієнтом кореляції ρ. Ці параметри однозначно визначають розподіл спостережень за теоремою *про щільність нормального розподілу на площині*.

Нехай $\widehat{\mu}_{kn}$, $\widehat{s}^2_{kn}$ вибіркові середнє та дисперсія для вектора $(\xi_{k1}, ..., \xi_{kn})$, $k = 1, 2$. Визначимо **вибірковий коефіцієнт кореляції**

$$\widehat{\rho}_n = \frac{1}{(n-1)\widehat{s}_{1n}\widehat{s}_{2n}} \sum\nolimits_{j=1}^{n} (\xi_{1j} - \widehat{\mu}_{1n})(\xi_{2j} - \widehat{\mu}_{2n})$$

З *критерію Колмогорова посиленого закону великих чисел* випливає, що $\widehat{\rho}_n \to^{P1} \rho, n \to \infty$ (**Вправа**).

За нульової гіпотези $H_9 : \rho = 0$ координати вектора ξ_1 незалежні згідно з теоремою *про незалежність координат нормального вектора.*

Тому для перевірки незалежності критичною областю слід обрати множину відносно великих значень статистики $\widehat{\rho}_n$.

Можна довести, що статистика

$$\widehat{\tau}_{n-2} = \sqrt{n-2}\ \widehat{\rho}_n(1 - \widehat{\rho}^2_n)^{-1/2}$$

має розподіл Стьюдента з $n - 2$ ступенями свободи за гіпотези H_0. Тому критерій $(\widehat{\rho}^2_n,\ y^2_{n-2,\alpha}/(n-2+y^2_{n-2,\alpha}))$ для перевірки незалежності має рівень α. Критерій є конзистентним.

Для перевірки гіпотези про різницю середніх $H_0 : \mu_1 - \mu_2 = \Delta$ можна скористатись статистикою

$$\widehat{\tau}_{n-2} = \sqrt{n-1} \frac{\widehat{\mu}_{1n} - \widehat{\mu}_{2n} - \Delta}{\sqrt{\widehat{s}^2_{1n} + \widehat{s}^2_{2n} - 2\widehat{\rho}_n\widehat{s}_{1n}\widehat{s}_{2n}}},$$

що має при виконанні гіпотези H_0 розподіл Стьюдента з $n - 1$ ступенем свободи (**Вправа**). Критичними є великі значення статистики.

Для розв'язання задачі *інтервального оцінювання* коефіцієнта ρ використовується асимптотична нормальність

$$\sqrt{n}(\widehat{\rho}_n - \rho) \to^W \zeta \simeq N(0, (1 - \rho^2)^2), n \to \infty.$$

У випадку сильної залежності, коли $|\rho|$ близький до 1, використовується перетворення Фішера

$$z_n = \frac{1}{2} \ln \frac{1 + \widehat{\rho}_n}{1 - \widehat{\rho}_n} \simeq N\left(\frac{1}{2} \ln \frac{1 + \rho}{1 - \rho} + 2\frac{\rho}{n-1}, \frac{1}{n-3}\right), n \to \infty.$$

3.16.4. Асимптотичні критерії

На практиці всі наведені критерії використовуються також для вибірок, які не мають *нормального розподілу* (наприклад, для цілозначних спостережень). Правомірність такого використання може бути обґрунтована виходячи з того, що *вибіркові середні* та дисперсії все ж таки наближено мають відповідні нормальні та пов'язані з нормальним розподіли, оскільки вони є сумами незалежних (чи умовно незалежних) випадкових величин. Останні внаслідок центральної граничної теореми є *асимптотично нормальними.* Для ілюстрації цього положення зазначимо, що деякі датчики випадкових чисел для моделювання *стандартної нормальної* випадкової величини використовують алгоритм: $\zeta \simeq \alpha_1 + ... + \alpha_{12} - 6$, де α_i – незалежні величини з *рівномірним розподілом* на $[0, 1]$.

Приклади

(а) Якщо вибірка утворена результатами n *випробувань Бернуллі* з невідомою ймовірністю успіху θ, то для перевірки гіпотези $\theta = p$ можна скористатись асимптотичною нормальністю *відносної частоти успіху* $\widehat{\theta}_n$:

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to^W \zeta \simeq N(0, \theta(1-\theta)) \text{ при } n \to \infty,$$

як це зроблено у розділі про оцінку ймовірності успіху для побудови відповідних надійних інтервалів.

(б) Нехай спостерігається кратна вибірка $X = (\xi_1, ..., \xi_n)$ з розподілом Пуассона $\Pi(\lambda)$. Тоді для вибіркового середнього за *класичною центральною граничною теоремою*

$$(\widehat{\mu}_n - \lambda)(\lambda/n)^{-1/2} \to^W \zeta \simeq N(0, 1) \text{ при } n \to \infty.$$

Дану збіжність можна використати як для перевірки гіпотез щодо значення λ, так і для побудови відповідних *надійних інтервалів*.

(в) Одночасно спостерігаються незалежні кратні вибірки $X = (\xi_1, ..., \xi_n)$ та $Y = (\eta_1, ..., \eta_m)$, з розподілами Пуассона $\Pi(\lambda)$ та $\Pi(\mu)$ відповідно. Тоді з аналогічних підстав

$$(\widehat{\mu}_n - \widehat{\mu}_m - \lambda + \mu)(\lambda/n + \mu/m)^{-1/2} \to^W \zeta \simeq N(0, 1), n, m \to \infty,$$

що дозволяє як перевірити гіпотезу щодо різниці $\lambda - \mu$, так і побудувати надійні інтервали для неї.

Вправа. Спостерігаються дві незалежні схеми випробувань Бернуллі з n_i спостереженнями, ймовірностями успіху θ_i,та частотами успіхів $\widehat{\theta}_i$ $i = 1, 2$. Тоді: $\left(\widehat{\theta}_1 - \widehat{\theta}_2 - \theta_1 + \theta_2\right) / \sqrt{\theta_1(1-\theta_1)/n_1 + \theta_2(1-\theta_2)/n_2} \to^W \zeta \simeq N(0, 1)$, при $n_i \to \infty$, а за умови $H_0 : \theta_1 = \theta_2$ при $\widehat{\theta} = (n_1\widehat{\theta}_1 + n_2\widehat{\theta}_2)/(n_1 + n_2)$ також $\left(\widehat{\theta}_1 - \widehat{\theta}_2\right) / \sqrt{\widehat{\theta}(1-\widehat{\theta})(1/n_1 + 1/n_2)/n_1n_2} \to^W \zeta \simeq N(0, 1)$. Довести та використати ці твердження для перевірки H_0.

3.17. Найбільш потужні критерії, лема Неймана – Пірсона

Розглянемо задачу перевірки статистичної *гіпотези* $H_0 : \theta \in \Theta_0$ проти альтернативи $H_1 : \theta \in \Theta_1$ на підставі *вибірки* X зі значеннями у *вибірковому просторі* $(S,\ \Sigma,\ \lambda)$. Як вже відомо, кожен критерій перевірки гіпотези однозначно задається **вибірковою критичною областю** $W \in \Sigma$: гіпотеза H_0 відкидається, якщо $X \in W$, і не відкидається (приймається), якщо $X \notin W$. Вибіркова критична область визначається через *статистичний критерій* як така підмножина вибіркового простору:

$$W = \{x \in S : \widehat{\varkappa}(x) \in D_1\},$$

де $\widehat{\varkappa}(X) = \widehat{\varkappa}$ – *статистика критерію*, а D_1 – її *критична область*. Нагадаємо, що *вірогідний рівень* критерію визначається найбільшою з імовірностей *похибок першого роду* $\mathbf{P}_\theta(X \in W),\ \theta \in \Theta_0$.

Означення. *Критерій із критичною областю W^* є* **найбільш потужним** *рівня α, якщо його вірогідний рівень дорівнює α, причому довільний критерій із критичною областю W рівня α має не більшу потужність, ніж W^*:*

$$\mathbf{P}_\theta(X \in W^*) \geq \mathbf{P}_\theta(X \in W),\ \forall \theta \in \Theta_1.$$

3.17.1. Критерій відношення вірогідностей

Задача відшукання *найбільш потужного* критерію не завжди має розв'язок, оскільки часто неможливо максимізувати значення потужності одночасно при декількох значеннях параметра. Однак у випадку простих гіпотез та альтернатив *найбільш потужний* критерій існує. Припустимо, що *нульова гіпотеза* та її *альтернатива* є *простими гіпотезами*:

$$H_i : \theta = \theta_i,\ i = 0, 1.$$

Позначимо через $F_i(B) = \mathbf{P}_{\theta_i}(X \in B)$, $B \in \Sigma$, відповідні розподіли вибірок. Ці розподіли відомі повністю за означенням простої гіпотези.

За умови *абсолютної неперервності* $F_1 \ll F_0$ існує вимірна *щільність міри* $l_{01}(x)$ така, що

$$F_1(B) = \int_B l_{01}(x) dF_0(x)$$

для всіх вимірних множин $B \in \Sigma$ *вибіркового простору* S. За означенням *функції вірогідності* міри $F_i(B)$ мають щільності $L(x, \theta_i)$ відносно фіксованої міри λ на Σ :

$$F_i(B) = \int_B L(x, \theta_i)\lambda(dx),\ \forall B \in \Sigma,\ i = 0, 1.$$

Тому за теоремою про заміну змінної статистика $l_{01}(X)$ збігається з емпіричним **відношенням вірогідностей**

$$l_{01}(X) = \frac{L(X, \theta_1)}{L(X, \theta_0)}.$$

Теорема (лема Неймана – Пірсона). *Нехай нульова гіпотеза та альтернатива є простими гіпотезами і* $F_1 \ll F_0$. *Якщо для даного вірогідного рівня* $\alpha \in (0, 1)$ *існує стала* $l_\alpha > 0$, *така, що*

$$\mathbf{P}_{\theta_0}(l_{01}(X) \geq l_\alpha) = \alpha,$$

то ***критерій відношення вірогідностей*** $(l_{01}(X), [l_\alpha, \infty))$, *що має критичну область вибірки*

$$W^* = \{x \in S : l_{01}(x) \geq l_\alpha\},$$

є найбільш потужним рівня α, *та незміщеним критерієм.*

Зауваження. Функція

$$\mathbf{P}_{\theta_0}(l_{01}(X) \geq l)\ =\ F_0(\{x \in S : l_{01}(x) \geq l\})$$

не зростає за l і набуває значень із відрізка $[0, 1]$. Тому умова теореми щодо існування l_α виконується, якщо випадкова величина $l_{01}(X)$ є *абсолютно неперервною*. Якщо ж для даного рівня α умова існування точного критичного рівня не виконана, то можна провести *процедуру рандомізації*, як це описано вище в розділі про рангові критерії.

Доведення. Розглянемо критерій із *критичною областю* W рівня α :

$$\mathbf{P}_{\theta_0}(X \in W) = F_0(W) \leq \alpha.$$

Обчислимо його *потужність* з урахуванням означення щільності l_{01}:

$$\mathbf{P}_{\theta_1}(X \in W) = F_1(W) = F_1(W \setminus W^*) + F_1(W \cap W^*) =$$

$$F_1(W^*) + F_1(W \setminus W^*) - F_1(W^* \setminus W) =$$

$$F_1(W^*) + \textstyle\int_{W \setminus W^*} l_{01}(x)F_0(dx) - \int_{W^* \setminus W} l_{01}(x)F_0(dx) \leq$$

$$F_1(W^*) + l_\alpha F_0(W \setminus W^*) - l_\alpha F_0(W^* \setminus W) =$$

$$F_1(W^*) + l_\alpha F_0(W) - l_\alpha F_0(W^*) \leq F_1(W^*) + l_\alpha \alpha - l_\alpha \alpha = F_1(W^*),$$

де справедливість передостанньої нерівності є наслідком означення області W^*, згідно з яким $l_{01}(x) < l_\alpha$ при $x \notin W^*$, та $l_{01}(x) \geq l_\alpha$ при $x \in W^*$, а остання нерівність випливає з вибору l_α, оскільки вірогідний рівень W^* дорівнює $F_0(W^*) = \alpha$.

Для доведення *незміщеності критерію* припустимо, що $l_\alpha \geq 1$ у означенні W^*. Тоді $F_1(W^*) = \int_{W^*} l_{01}(x) F_0(dx) \geq l_\alpha F_0(W^*) \geq F_0(W^*) = \alpha$.

Нехай тепер $l_\alpha < 1$. Тоді $F_1(\overline{W^*}) = \int_{\overline{W^*}} l_{01}(x) F_0(dx) \leq l_\alpha F_0(\overline{W^*}) \leq F_0(\overline{W^*})$. Оскільки $F_i(\overline{W^*}) = 1 - F_i(W^*)$, звідси отримуємо $F_1(W^*) \geq \alpha$ $\boxdot$

3.17.2. Приклад критерію відношення вірогідностей

Нехай $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка* з *нормальним розподілом* спостережень: $\xi_k \simeq N(\mu, \sigma^2)$. Припустимо, що дисперсія σ^2 відома, а нульова та альтернативна гіпотези щодо середнього мають вигляд

$$H_i : \mu = \mu_i, \ i = 0, 1.$$

Для визначеності будемо вважати, що $\mu_1 > \mu_0$.

Відповідні *функції вірогідностей* є строго додатними, а емпіричне *відношення вірогідностей* дорівнює

$$l_{01}(X) = \frac{L_n(X, \theta_1)}{L_n(X, \theta_0)} = \frac{(2\pi\sigma^2)^{-n/2} \exp(-\sum_{k=1}^n (\xi_k - \mu_1)^2 / 2\sigma^2)}{(2\pi\sigma^2)^{-n/2} \exp(-\sum_{k=1}^n (\xi_k - \mu_0)^2 / 2\sigma^2)} =$$

$$\exp\left(\frac{n}{\sigma^2}(\mu_1 - \mu_0)\, \widehat{\mu}_n - \frac{n}{2\sigma^2}(\mu_1^2 - \mu_0^2)\right),$$

де $\widehat{\mu}_n$ – *вибіркове середнє*. Нерівність $l_{01}(X) \geq l_\alpha$ еквівалентна нерівності

$$\frac{\sqrt{n}}{\sigma}(\widehat{\mu}_n - \mu_0) \geq \frac{\sigma}{\sqrt{n}(\mu_1 - \mu_0)} \ln l_\alpha + \frac{\sqrt{n}}{2\sigma}(\mu_1 - \mu_0).$$

За H_0 унаслідок теореми *про вибіркові середнє та дисперсію нормальної вибірки* ліва частина має *стандартний нормальний розподіл*. Тому, позначивши через x_α праву частину останньої нерівності, для довільного рівня α з рівняння $\Phi(x_\alpha) = 1 - \alpha$ знаходимо *критичний рівень* x_α та обчислюємо

$$l_\alpha = \exp\left(\frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma}(x_\alpha - \frac{\sqrt{n}}{2\sigma}(\mu_1 - \mu_0))\right).$$

Отже, *найбільш потужний* критерій має критичну область

$$\left\{\frac{\sqrt{n}}{\sigma}(\widehat{\mu}_n - \mu_0) \geq x_\alpha\right\}.$$

Зауважимо, що в умовах альтернативи H_1 *вибіркове середнє* має розподіл $\widehat{\mu}_n \simeq N(\mu_1, \sigma^2/n)$. Тому *потужність* критерію дорівнює

$$\mathbf{P}_{\theta_1}\left(\frac{\sqrt{n}}{\sigma}(\widehat{\mu}_n-\mu_0)\geq x_\alpha\right)=\mathbf{P}_{\theta_1}\left(\frac{\sqrt{n}}{\sigma}(\widehat{\mu}_n-\mu_1)-\sqrt{n}(\mu_0-\mu_1)/\sigma\geq x_\alpha\right)=$$
$$1-\Phi(x_\alpha+\sqrt{n}(\mu_0-\mu_1)/\sigma)=\Phi(-x_\alpha+\sqrt{n}(\mu_1-\mu_0)/\sigma)\to 1,$$

при $n\to\infty$, отже, критерій є *консистентним*

Нехай $\alpha,\beta\in(0,1)$ наперед задані сталі. Знайдемо такий мінімальний об'єм вибірки n, щоб критерій рівня α мав потужність не меншу за $1-\beta$. Для цього необхідно і достатньо, щоб права частина останнього співвідношення була не меншою за $1-\beta$, тобто $-x_\alpha+\sqrt{n}(\mu_1-\mu_0)/\sigma\geq x_\beta$, де $\Phi(x_\beta)=1-\beta$. Звідси знаходимо

$$n\geq n(\alpha,\beta)\equiv\sigma^2(x_\alpha+x_\beta)^2\ /\ (\mu_1-\mu_0)^2.$$

Отже, кількість спостережень зростає зі зменшенням *імовірностей похибок першого та другого роду* α,β, пропорційна *дисперсії* спостережень та обернено пропорційна квадрату різниці між гіпотетичними середніми.

Вправи

(1) Знайти критерій відношення вірогідностей у нормальної схеми для перевірки гіпотез $H_i:\mu=\mu_i,\sigma=\sigma_i,i=0,1$, з відомими дисперсіями σ_i^2.

(2) Розглядаються дві незалежні кратні вибірки об'єму n з нормальними розподілами спостережень, що мають невідомі середні μ_k та відомі дисперсії σ_k^2, $k=1,2$. Довести, що мінімальне число спостережень для перевірки гіпотези $H_0:\mu_1-\mu_2=0$ проти альтернативи $H_0:\mu_1-\mu_2=\delta$ при похибках α,β першого та другого роду, дорівнює $n_{\alpha\beta}=(\sigma_1^2+\sigma_2^2)(x_\alpha+x_\beta)^2\ /\ \delta^2$, де x_α,x_β – квантилі рівнів α,β для стандартного нормального розподілу.

(3) Нехай $(\xi_1,\ldots,\xi_n)$ – кратна вибірка з функцією розподілу спостережень F. Побудувати критерій відношення вірогідностей для перевірки нульової гіпотези $F(x)=1-\exp(-x)$ проти альтернативи $F(x)=1-\exp(-x^\theta),x\geq 0,\theta>1$.

(4) Нехай $X=(\xi_1,...,\xi_n)$ – кратна вибірка, а $f(y,\theta)$ – функція вірогідності одного спостереження. (а) Довести зображення $\ln l_{01}(X)=\sum_{k=1}^n\eta_k$, де випадкові величини $\eta_k=\ln(f(\xi_k,\theta_1)/f(\xi_k,\theta_0))$ незалежні та однаково розподілені. (б) $\mathbf{E}_{\theta_0}\eta_1=-I(\theta_1,\theta_0)$, де $I(\theta_1,\theta_0)$ – інформація за Кульбаком. (в) Довести, що при $n\to\infty$ для критичного рівня у лемі Неймана-Пірсона виконується зображення $\ln l_\alpha+nI(\theta_1,\theta_0)\sim x_\alpha\sigma_0\sqrt{n}$, де $\sigma_0^2=\mathbf{D}_{\theta_0}\eta_1,\ \Phi(x_\alpha)=1-\alpha$. (г) Вивести асимптотичне зображення для ймовірності похибки другого роду.

(5) Нормальні випадкові величини ξ_k обчислюються з рекурентної системи рівнянь $\xi_k=\theta\xi_{k-1}+\varepsilon_k,\ \ k=\overline{1,n}$, де похибки незалежні та $\varepsilon_k\simeq N(0,\sigma^2)$, і $\xi_0=0$. Довести, що сумісна щільність вектора $X=(\xi_1,...,\xi_n)$ при $x=(x_k)$ дорівнює $L(x,\theta)=(2\pi\sigma^2)^{-n/2}\exp(-\sum_{k=1}^n(x_k-\theta x_{k-1})^2/2\sigma^2)$. де $x_0=0$.

Довести, що критерій відношення вірогідностей для перевірки $H_0 : \theta = 0$ проти $H_1 : \theta \neq 0$ оснований на статистиці $\left(\sum_{k=1}^{n-1} \xi_k \xi_{k+1}\right)^2 / \left(\sum_{k=1}^{n-1} \xi_k^2\right)$.

(6) Спостерігається кратна вибірка $X = ((\xi_{1k}, \xi_{2k}), k = \overline{1,n})$ з двовимірного нормального розподілу $N_2(0, 0, \sigma_1^2, \sigma_2^2, \rho)$. Довести, що статистика критерію відношення вірогідностей для перевірки гіпотези $H_0 : \rho = 0$ проти альтернативи $H_1 : \rho \neq 0$ є функцією від $|\tau|$, де $\tau = \left(\sum_{k=1}^{n} \xi_{1k}\xi_{2k}\right) / \sqrt{\left(\sum_{k=1}^{n} \xi_{1k}^2\right) \left(\sum_{k=1}^{n} \xi_{2k}^2\right)}$.

(7) Обчислити критерій відношення вірогідностей для перевірки простих гіпотез про значення параметра для кратної вибірки з (а) пуассоновим, (б) геометричним, (в) показниковим розподілом. Знайти потужність критерію, перевірити його незміщеність та конзистентність.

(8) Довести, що оптимальні (а) Байесовський, та (б) мінімаксний критерії перевірки простих гіпотез основані на статистиці відношення вірогідностей.

3.17.3. Рівномірно найбільш потужні критерії з монотонним відношенням вірогідностей

Найбільш потужні критерії існують також і для складних гіпотез. Для їх побудови розглянемо таке узагальнення поняття критерію.

Означення. Рандомізованим критерієм *перевірки статистичної гіпотези називається пара* $(\widehat{\varkappa}(X),\ \pi(t))$, *де* $\widehat{\varkappa}(X)$ *– статистика критерію, а функція* $\pi(t) \in [0,1]$ *задає ймовірність відхилити нульову гіпотезу при значенні статистики* $\widehat{\varkappa}(X) = t$.

При реалізації такого критерію у випадку, коли $\pi(t) = 0$, приймається нульова гіпотеза, при $\pi(t) = 1$ приймається альтернатива, а при $\pi(t) \in (0,1)$ для прийняття рішення додатково проводиться незалежне *випробування Бернуллі*, щоб із ймовірністю $\pi(t)$ відхилити нульову гіпотезу, та з імовірністю $1 - \pi(t)$ її прийняти.

Означення. *Критерій* $(\widehat{\varkappa}(X),\ \pi(t),\ t_0)$ *є* простим рандомізованим критерієм *із критичним рівнем* t_0, *якщо статистика* $\widehat{\varkappa}(X) = t$ *скалярна, функція* $\pi(t) = 0$ *при* $t < t_0$, $\pi(t) = 1$ *при* $t > t_0$, *а значення* $\pi(t_0) \in [0,1]$.

Даний критерій приймає нульову гіпотезу при $t < t_0$, приймає альтернативу при $t > t_0$, та при $t = t_0$, висновок розігрується з імовірністю $\pi(t_0)$ на користь альтернативи.

Вірогідний рівень рандомізованого критерію задається значеннями функції $\mathbf{E}_\theta \pi(\widehat{\varkappa})$ при $\theta \in \Theta_0$, а *потужність* – значеннями $\mathbf{E}_\theta \pi(\widehat{\varkappa})$ при $\theta \in \Theta_1$. Ці значення обчислюються за теоремою *про обчислення математичного* *сподівання функції від випадкової величини*

$$\mathbf{E}_\theta \pi(\widehat{\varkappa}) = \int \pi(t)\mathbf{P}_\theta(\widehat{\varkappa} \in dt) = \int_S \pi(\widehat{\varkappa}(x))\mathbf{P}_\theta(X \in dx).$$

Очевидно, що звичайний *статистичний критерій* $(\widehat{\varkappa}(X),\ [t_0, \infty))$ є частковим випадком *простого рандомізованого критерію* $(\widehat{\varkappa}(X),\ \pi(t),\ t_0)$: для нього $\pi(t_0) = 1$. Звідси, зокрема, випливає, що *найбільш потужний* критерій у класі всіх рандомізованих критеріїв рівня α має не меншу потужність, ніж будь-який звичайний критерій того ж рівня.

Наступне твердження відрізняє рандомізовані критерії від звичайних.

Лема (про побудову рандомізованого критерію). *Для довільних скалярної статистики $\widehat{\varkappa}$ та рівня $\alpha \in (0, 1)$ існує та однозначно визначений простий рандомізований критерій $(\widehat{\varkappa}, \pi(t), t_\alpha)$ з деяким t_α такий, що його вірогідний рівень при заданому θ збігається з α.*

Доведення. За наведеною вище формулою рівняння для вказаного вірогідного рівня має такий вигляд:

$$\mathbf{E}_\theta \pi(\widehat{\varkappa}) \equiv 0 \cdot \mathbf{P}_\theta(\widehat{\varkappa} < t_\alpha) + \pi(t_\alpha) \cdot \mathbf{P}_\theta(\widehat{\varkappa} = t_\alpha) + 1 \cdot \mathbf{P}_\theta(\widehat{\varkappa} > t_\alpha) = \alpha.$$

Звідси, зокрема, виводимо $\mathbf{P}_\theta(\widehat{\varkappa} > t_\alpha) \le \alpha$. Визначимо

$$t_\alpha = \sup(t : \mathbf{P}_\theta(\widehat{\varkappa} > t) \le \alpha).$$

Якщо t_α є точкою неперервності функції розподілу випадкової величини $\widehat{\varkappa}$, то $\mathbf{P}_\theta(\widehat{\varkappa} = t_\alpha) = 0$ і $\mathbf{P}_\theta(\widehat{\varkappa} > t_\alpha) = \alpha$, інакше значення t_α можна було б збільшити. Тому при $\pi(t_\alpha) = 1$ маємо $\mathbf{E}_\theta \pi(\widehat{\varkappa}) = \alpha$.

У випадку інтервалу сталості функції розподілу $\widehat{\varkappa}$ будь-який належний вибір t_α призводить до еквівалентного критерію, тому можна зупинитись на єдиному найменшому значенні t_α.

Якщо ж $\mathbf{P}_\theta(\widehat{\varkappa} = t_\alpha) > 0$, то досить обрати

$$\pi(t_\alpha) = (\alpha - \mathbf{P}_\theta(\widehat{\varkappa} > t_\alpha)) \ / \ \mathbf{P}_\theta(\widehat{\varkappa} = t_\alpha).$$

Зауважимо, що у першому випадку *простий рандомізований критерій* із критичним рівнем α збігається зі звичайним $(\widehat{\varkappa}(X), [t_\alpha, \infty))$. Отже, у випадку неперервності розподілу статистики $\widehat{\varkappa}$ прості рандомізовані критерії збігаються зі звичайними ⊡

Нагадаємо, що розподіл вибірки X за виконання гіпотез $H_i : \theta = \theta_i$, $i = 0, 1$, позначається через $F_i(B) = \mathbf{P}_{\theta_i}(X \in B)$, $B \in \Sigma$, а статистика $l_{01}(X)$ збігається з емпіричним *відношенням вірогідностей*

$$l_{01}(X) = \frac{L(X, \theta_1)}{L(X, \theta_0)}.$$

Теорема (теорема Неймана – Пірсона для рандомізованих критеріїв). *Нехай нульова гіпотеза H_0 та альтернатива H_1 є простими гіпотезами і $F_1 \ll F_0$, тобто коректно визначено відношення вірогідностей $l_{01}(X)$. Тоді для кожного рівня $\alpha \in (0,1)$ існує* ***простий рандомізований критерій відношення вірогідностей*** $(l_{01}(X), \pi^*(t), t^*_\alpha)$, *що є найбільш потужним рівня α. Його критичний рівень t^*_α згідно з лемою про побудову рандомізованого критерію однозначно знаходиться з умови*

$$\mathbf{E}_{\theta_0}\pi^*(l_{01}(X)) = \alpha.$$

Крім того, потужність цього критерію не менша за рівень α :

$$\mathbf{E}_{\theta_1}\pi^*(l_{01}(X)) \geq \alpha.$$

Доведення теореми практично не відрізняється від доведення *леми Неймана – Пірсона*. Нехай $(T(X),\ \pi(t))$ – довільний *рандомізований критерій* рівня α. Позначимо

$$S^{\pm} = \{x \in S : \pi^*(l_{01}(x)) \gtrless \pi(T(x))\}.$$

Тоді різниця *потужностей* дорівнює

$$\mathbf{E}_{\theta_1}\pi^*(l_{01}(X)) - \mathbf{E}_{\theta_1}\pi(T(X)) = \int_{S^+ \,\cup\, S^-}(\pi^*(l_{01}(x)) - \pi(T(x)))F_1(dx) =$$
$$\int_{S^+ \,\cup\, S^-}(\pi^*(l_{01}(x)) - \pi(T(x)))\ l_{01}(x)F_0(dx) \geq$$
$$\int_{S^+ \,\cup\, S^-}(\pi^*(l_{01}(x)) - \pi(T(x)))\ t^*_\alpha F_0(dx) =$$
$$t^*_\alpha\,(\mathbf{E}_{\theta_0}\pi^*(l_{01}(X)) - \mathbf{E}_{\theta_0}\pi(T(X))) = t^*_\alpha(\alpha - \mathbf{E}_{\theta_0}\pi(T(X)) \geq 0,$$

оскільки *вірогідний рівень* критерію $(T(X),\ \pi(t))$ не більший за α. Передостання нерівність є наслідком означення *простого рандомізованого критерію відношення вірогідностей*: з кожного включення $x \in S^{\pm}$ випливає відповідно $\pi^*(l_{01}(x)) = 1$ або $\pi^*(l_{01}(x)) = 0$ для всіх $l_{01}(x) = t \neq t^*_\alpha$, оскільки $\pi^*(t) \in \{0,1\}$ за означенням для таких t. За цим же означенням з останніх співвідношень виводимо, що $l_{01}(x)) \geq t^*_\alpha$ та $l_{01}(x)) \leq t^*_\alpha$ відповідно. Тому для всіх $x \in S^{\pm}$ нерівності $\pi^*(l_{01}(x)) - \pi(T(x)) \gtrless 0$ та $l_{01}(x)) \gtrless t^*_\alpha$, виконуються одночасно, що забезпечує вказану нерівність.

Твердження щодо *потужності* випливає зі вже доведеної властивості найбільшої потужності. Дійсно, завжди визначений тривіальний рандомізований критерій рівня α, для якого $\pi(t) \equiv \alpha$. Оскільки його потужність теж дорівнює α, то потужність *критерію відношення вірогідностей* не менша за α $\boxdot$

Розглянемо тепер задачу перевірки складних гіпотез. Для складних гіпотез *найбільш потужні* критерії називають **рівномірно найбільш потужними** критеріями, оскільки нерівність для потужності має виконуватись рівномірно за $\theta \in \Theta_1$.

Припустимо, що $\Theta \subset \mathbb{R}$. Нехай для всіх пар $\theta_0, \theta_1 \in \Theta$ коректно визначені *відношення вірогідностей* $l_{01}(X) = L(X, \theta_1)/L(X, \theta_0)$.

Означення. *Вибірка X має* **монотонне відношення вірогідностей,** *якщо існує така скалярна статистика $T(X)$, що для кожної пари параметрів $\theta_0 < \theta_1 \in \Theta$ знайдеться строго монотонно зростаюча за $t \in \mathbb{R}$ функція $g_{01}(t)$, для якої має місце тотожність*

$$l_{01}(X) \equiv \frac{L(X, \theta_1)}{L(X, \theta_0)} = g_{01}(T(X)).$$

Теорема (про рівномірно найбільш потужний критерій). *Нехай вибірка має монотонне відношення вірогідностей зі статистикою $T(X)$. Тоді для кожного вірогідного рівня $\alpha \in (0, 1)$ простий рандомізований критерій $(T(X),\ \pi^*(t),\ t^*_\alpha)$, що однозначно визначається з рівняння $\mathbf{E}_{\theta_0}\pi^*(T(X)) = \alpha$, є рівномірно найбільш потужним критерієм рівня α для перевірки гіпотези $H_0 : \theta = \theta_0$ проти складної альтернативи $H_1 : \theta > \theta_0$.*

Доведення. Нехай $\theta_1 > \theta_0$. Розглянемо задачу перевірки *простої гіпотези* $H_0 : \theta = \theta_0$ проти простої альтернативи $H'_1 : \theta = \theta_1$. Нехай $(l_{01}(X),\ \pi_1(l), l^*_{1\alpha}))$ – *простий рандомізований критерій відношення вірогідностей,* що за *теоремою Неймана – Пірсона для рандомізованих критеріїв* є найбільш потужним рівня α для перевірки H_0 проти H'_1.

Визначимо $t^*_\alpha = \inf(t : g_{01}(t) \geq l^*_{1\alpha})$. За означенням *монотонного відношення вірогідностей* внаслідок строгої монотонності g_{01} для кожного з двох випадків $\gtreqqless$ виконуються рівності

$$\{l_{01}(X) \gtreqqless l^*_{1\alpha}\} = \{g_{01}(T(X)) \gtreqqless l^*_{1\alpha}\} = \{T(X) \gtreqqless t^*_\alpha\}.$$

Звідси та з леми *про побудову рандомізованого критерію* виводимо рівності $\alpha = \mathbf{E}_{\theta_0}\pi_1(l_{01}(X)) = \mathbf{E}_{\theta_0}\pi^*(T(X))$, якщо обрати $\pi^*(t^*_\alpha) = \pi_1(l^*_{1\alpha})$. Отже, простий рандомізований критерій $(T(X),\ \pi^*(t),\ t^*_\alpha)$ має рівень α та еквівалентний *найбільш потужному* критерію для перевірки H_0 проти H'_1. Оскільки за згаданою лемою критичний рівень t^*_α та ймовірність $\pi^*(t)$ однозначно визначаються за рівнем α та статистикою $T(X)$, то побудований критерій не залежить від вибору $\theta_1 > \theta_0$.

Нехай $(\widehat{\varkappa}(X),\ \pi(t))$ *рандомізований критерій* рівня α для перевірки H_0, а $\theta_1 > \theta_0$.Тоді, зокрема, $\mathbf{E}_{\theta_0}\pi(\widehat{\varkappa}(X)) \leq \alpha$ і внаслідок доведеної вище властивості найбільшої потужності критерію $(T(X), \pi^*(t), t^*_\alpha)$ отримуємо

$$\mathbf{E}_{\theta_1}\pi(\widehat{\varkappa}(X)) \le \mathbf{E}_{\theta_1}\pi^*(T(X)),$$

тому при кожному альтернативному значенні параметра $\theta = \theta_1 > \theta_0$ потужність критерію $(\widehat{\varkappa}(X),\ \pi(t))$ не перевищує потужність простого критерію вигляду $(T(X), \pi^*(t), t^*_\alpha)$ $\boxdot$

Приклад. Нехай $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка* з *нормальним розподілом* спостережень: $\xi_k \simeq N(\mu, \sigma^2)$ при відомій дисперсії σ^2 та невідомому середньому $\theta = \mu$. Як показано у прикладі *критерію відношення вірогідностей*, це відношення дорівнює

$$l_{01}(X) = \exp\left(\frac{n}{\sigma^2}(\mu_1 - \mu_0)\widehat{\mu}_n - \frac{n}{2\sigma^2}(\mu_1^2 - \mu_0^2)\right),$$

і є зростаючою функцією відносно статистики $\widehat{\mu}_n$ для всіх $\mu_0 < \mu_1$. Крім того, розподіл статистики $\widehat{\mu}_n$ неперервний, тому *простий рандомізований критерій* збігається зі звичайним.

Отже, *рівномірно найбільш потужним* критерієм для перевірки гіпотези $H_0 : \mu = \mu_0$ проти $H_1 : \mu > \mu_0$ є звичайний критерій $(\widehat{\mu}_n,\ [t_\alpha, \infty))$.

Вправи

(1) Сформулювати умови застосування теореми про рівномірно найбільш потужний критерій для кратної вибірки з *експоненційної моделі розподілів*.

(2) Нехай $X = (\xi_{ik}, i = \overline{1, n_k}, k = 1, 2)$ – дві незалежні кратні нормальні вибірки, $\xi_{ik} \simeq N(\mu_k, \sigma_k^2)$, з невідомими σ_k^2. Побудувати рівномірно найбільш потужний критерій відношення вірогідностей для перевірки (а) гіпотези про середні $H_0 : \mu_1 \le \mu_2$ проти альтернативи $H_1 : \mu_1 > \mu_2$, у припущенні $\sigma_1^2 = \sigma_2^2$, (б) гіпотези $H_0 : \sigma_1^2 \le \sigma_2^2$ проти $H_1 : \sigma_1^2 > \sigma_2^2$ при відомих середніх.

(3) Знайти рівномірно найбільш потужний критерій відношення вірогідностей для перевірки гіпотези про ймовірність успіху для вибірки з біноміального розподілу.

(4) Знайти рівномірно найбільш потужний критерій для перевірки гіпотези $H_0 : \theta = \theta_0$ проти $H_1 : \theta > \theta_0$ для вибірки з рівномірного розподілу $U(0, \theta)$.

(5) Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка зі щільністю оберненого нормального розподілу $f(y, \theta) = (\theta/2\pi y^3)^{1/2} \exp(-\theta y/2 + \theta - \theta/y), y > 0$. Знайти критерій Неймана – Пірсона для перевірки гіпотези $H_0 : \theta = \theta_0$ проти альтернативи $H_1 : \theta < \theta_0$, з використанням нормальної апроксимації.

(6) Кратна вибірка має монотонне відношення вірогідностей тоді й тільки тоді, коли щільність спостереження $f(y, \theta)$ є строго одновершинною.

(7) Спостерігається вибірка $\xi_1, \ldots, \xi_n$ з незалежних спостережень, що мають розподіли Пуассона: $\xi_j \simeq \Pi(\lambda_j)$, де $\lambda_j = a + bt_j$ з невідомими (a, b) та

додатними різними t_j. Перевірити гіпотезу $H_0 : b = 0$ за критерієм відношення вірогідностей.

(8) Нехай X – кратна вибірка з нормальним розподілом $N(\mu, \sigma^2)$ спостережень та невідомим параметром $\theta = (\mu, \sigma^2)$. Побудувати найбільш потужний критерій для перевірки нульової гіпотези $H_0 : \mu = 0$, що спирається на статистику $T = (\sup_{\theta \in \Theta_0} L(X, \theta))/(\sup_{\theta \in \Theta} L(X, \theta))$.

3.17.4. Поняття про послідовний аналіз

У наведеному вище прикладі *критерію відношення вірогідностей* статистичний висновок щодо справедливості гіпотез формулюється одномоментно відразу після отримання вектора спостережень $X = (\xi_1, ..., \xi_n)$. Однак проведення спостережень часто є процесом, що відбувається в часі, до того ж небезкоштовним. Наприклад, такими є геологорозвідувальні роботи, де кожне спостереження пов'язане з бурінням нової свердловини. З цієї точки зору доцільним було б розширення правил формування статистичних висновків, що дозволило б приймати певні рішення після отримання кожного чергового спостереження. Дану ідею запропонував та розвинув А. Вальд у своїх працях, що створили основу теорії **послідовного статистичного аналізу**.

Для ілюстрації цієї теорії розглянемо задачу перевірки *простих гіпотез* $H_i : \theta = \theta_i, i = 0, 1$ для *кратної вибірки* $X = (\xi_1, ..., \xi_n)$. Позначимо через $f(y, \theta)$ щільність спостереження. За теоремою *про функцію вірогідності кратної вибірки* логарифм *відношення вірогідностей* дорівнює

$$\ln l_{01}(X) = \ln \frac{L_n(X, \theta_1)}{L_n(X, \theta_0)} = \sum\nolimits_{j=1}^{n} h(\xi_j) \equiv S_n, \;\; h(y) \equiv \ln \frac{f(y, \theta_1)}{f(y, \theta_0)},$$

де доданки $h(\xi_j)$ *незалежні в сукупності* та *однаково розподілені*.

У теоремі *про властивості інформації за Кульбаком* доведено, що за *умов консистентності ОМВ* $\mathbf{E}_{\theta_0} h(\xi_1) < 0$. Аналогічно доводиться, що $\mathbf{E}_{\theta_1} h(\xi_1) > 0$.

Найбільш потужний критерій відношення вірогідностей має таку критичну область: $\{l_{01}(X) \geq l_\alpha\} = \{S_n \geq \ln l_\alpha\}$.

Нехай для кожного $k = \overline{1, n}$ задано сталі $c_{0k} < 0 < c_{1k}$. Позначимо через $S_k = \sum_{j=1}^{k} h(\xi_j)$. Розглянемо такий узагальнений **послідовний критерій відношення вірогідностей**:

(0) якщо $\{S_1 \in (c_{01}, c_{11}), ..., S_{k-1} \in (c_{0,k-1}, c_{1,k-1}),\; S_k \leq c_{0k}\}$ для деякого $k \leq n$, то спостереження припиняються в такий момент k, і приймається *нульова гіпотеза* H_0,

(1) якщо $\{S_1 \in (c_{01}, c_{11}), ..., S_{k-1} \in (c_{0,k-1}, c_{1,k-1}),\ S_k \geq c_{1k}\}$ для деякого $k \leq n$, то спостереження припиняються в такий момент k, і приймається *альтернативна* гіпотеза H_1,

(n) інакше приймається рішення на користь H_0.

Вибір знаків сталих обґрунтовується *критерієм Колмогорова посиленого закону великих чисел*, внаслідок якого $S_k \to^{P1} -\infty, k \to \infty$ за нульової гіпотези та $S_k \to^{P1} \infty, k \to \infty$ за альтернативи, відповідно до знаку математичного сподівання одного доданку. Звідси випливає також, що ймовірність події в (n) прямує до нуля при $n \to \infty$.

Якщо формально обрати $c_{0k} = -\infty, k \leq n,\ c_{1k} = \infty, k < n$, і $c_{1n} = \ln l_\alpha$, то отримаємо *критерій відношення вірогідностей* як частковий випадок сформульованого *послідовного критерію відношення вірогідностей.*

На відміну від *критерію відношення вірогідностей,* послідовний критерій спирається на випадкову кількість спостережень

$$\tau = \min(k \leq n : S_k \notin (c_{0k}, c_{1k})).$$

Унаслідок доведеного вище включення мінімальна середня кількість випробувань у всьому класі послідовних критеріїв із заданими ймовірностями похибок α, β першого та другого роду не більша за відповідну кількість для простого *критерію відношення вірогідностей*: $\inf \mathbf{E}\tau \leq n$. Отже, використання послідовного критерію може призвести до зменшення кількості спостережень при збереженні якості критерію.

Для уточнення властивостей послідовного критерію припустимо, що сталі $c_{0k} = c_0, c_{1k} = c_1$ не залежать від k. Позначимо через A_{0k} подію в умові (0) критерію, через A_{1k} – в умові (1) та через A_n – у (n). За означенням всі ці події *попарно несумісні* та утворюють *повну групу подій.* Нехай випадковий вектор $X_k = (\xi_1, ..., \xi_k)$ містить перші k спостережень. Його *функція вірогідності* дорівнює $L_k(X_k, \theta) = \prod_{j=1}^k f(\xi_j, \theta)$, а *відношення вірогідностей* має вигляд

$$L_k(X_k, \theta_1)\ /\ L_k(X_k, \theta_0) = \exp\left(\sum\nolimits_{j=1}^k h(\xi_j)\right) = \exp(S_k).$$

Остання величина не перевищує $C_0 \equiv \exp(c_0) < 1$ на події A_{0k}, і не менша за $C_1 \equiv \exp(c_1) > 1$ на події A_{1k}. Позначимо через B_{ik} борелєві підмножини $\mathbb{R}^k$, що отримуються з A_{ik} заміною величин ξ_j на j-ті координати y_j вектора $x_k = (y_1, ..., y_k) \in \mathbb{R}^k$. Тоді $A_{ik} = \{X_k \in B_{ik}\}$. За означенням

$$L_k(x_k, \theta_1)\ /\ L_k(x_k, \theta_0) = \exp\left(\sum\nolimits_{j=1}^k h(y_j)\right) \geq \exp(c_1) = C_1,\ \forall x_k \in B_{1k},$$

$$L_k(x_k, \theta_1)\ /\ L_k(x_k, \theta_0) \leq \exp(c_0) = C_0,\ \forall x_k \in B_{0k}.$$

Звідси

$$\mathbf{P}(A_{1k} \mid H_0) = \int_{B_{1k}} L(x_k, \theta_0)\mu_k(dx_k) \le \int_{B_{1k}} \tfrac{L(x_k,\theta_1)}{C_1}\mu_k(dx_k) = \tfrac{\mathbf{P}(A_{1k}|H_1)}{C_1},$$
$$\mathbf{P}(A_{0k} \mid H_1) = \int_{B_{1k}} L(x_k, \theta_1)\mu_k(dx_k) \le \int_{B_{1k}} C_0 L(x_k, \theta_0)\mu_k(dx_k) =$$
$$C_0\mathbf{P}(A_{1k} \mid H_0).$$

Нехай α, β – *імовірності похибок першого та другого роду* для побудованого вище послідовного критерію. Тоді

$$\alpha = \mathbf{P}(\cup_{k=1}^n A_{1k} \mid H_0) = \textstyle\sum_{k=1}^n \mathbf{P}(A_{1k} \mid H_0) \le \sum_{k=1}^n \mathbf{P}(A_{1k} \mid H_1)/C_1 =$$
$$(1 - \textstyle\sum_{k=1}^n \mathbf{P}(A_{0k} \mid H_1) - \mathbf{P}(A_n \mid H_1))/C_1 = (1-\beta)/C_1,$$
$$\beta = \textstyle\sum_{k=1}^n \mathbf{P}(A_{0k} \mid H_1) + \mathbf{P}(A_n \mid H_1) \le$$
$$\textstyle\sum_{k=1}^n C_0\mathbf{P}(A_{1k} \mid H_0) + C_0\mathbf{P}(A_n \mid H_0) = (1-\alpha)C_0.$$

Отже, критичні значення $c_i = \ln C_i$ та похибки в *послідовному критерії відношення вірогідностей* пов'язані системою нерівностей

$$\alpha \le (1-\beta)/C_1, \quad \beta \le (1-\alpha)C_0.$$

Розглянемо послідовний критерій з граничними значеннями

$$C_1' = (1-\beta)/\alpha, \quad C_0' = \beta/(1-\alpha),$$

для яких попередні нерівності перетворюються на рівності. Нехай a, b – імовірності похибок для цього критерію. Тоді з наведених нерівностей для α, β отримуємо $a/(1-b) \le 1/C_1' = \alpha/(1-\beta)$, $b/(1-a) \le C_0' = \beta/(1-\alpha)$, звідки множенням на знаменники та додаванням виводимо нерівність

$$a + b \le \alpha + \beta.$$

Отже, для будь-яких заданих $\alpha, \beta \in (0,1)$ існує послідовний критерій (з граничними значеннями $c_i = \ln C_i'$), що має не більшу за $\alpha + \beta$ суму *ймовірностей похибок першого та другого роду*.

Для наближеного обчислення середньої кількості необхідних спостережень у послідовному критерії використовується **тотожність Вальда**:

$$\mathbf{E}_\theta\ S_\tau = m(\theta)\mathbf{E}_\theta\tau, \quad m(\theta) \equiv \mathbf{E}_\theta S_1 = \mathbf{E}_\theta h(\xi_1).$$

На відміну від подібної схеми з розділу гіллястих процесів, де випадкова кількість доданків τ не залежала від самих доданків, у даній ситуації величина τ явно залежить від сум S_k, адже вона будується за цими сумами. Тим не менше випадкова подія $\{\tau \ge k\}$, так само як і її доповнення $\{\tau < k\} = \{\tau \le k-1\}$, визначається виключно вектором $(S_1, ..., S_{k-1})$ і за теоремою *про векторні перетворення незалежних величин* не залежить від величини ξ_k. Тому

$$\mathbf{E}_\theta\ S_\tau = \sum_{k\geq 1} \mathbf{E}_\theta \left(h(\xi_k)\ 1\!\!\mathrm{I}_{\ \{k\leq\tau\}}\right) = \sum_{k\geq 1} \mathbf{E}_\theta h(\xi_k)\ \mathbf{P}_\theta(k\leq\tau) =$$

$$m(\theta)\sum_{k\geq 1}\sum_{j\geq k} \mathbf{P}_\theta(\tau=j) = m(\theta)\sum_{j\geq 1} j\ \mathbf{P}_\theta(\tau=j) = m(\theta)\mathbf{E}_\theta\tau.$$

За означенням моменту τ сума S_τ або (1) вперше перевищила рівень c_1, або (0) вперше стала меншою за c_0, або ж (n) цього не сталося на інтервалі $[1,n]$ і просто настав фінальний момент n. Як вказано вище, за законом великих чисел імовірність останньої події прямує до нуля при $n\to\infty$. Імовірність $\mathbf{P}_{\theta_0}$ (за нульової гіпотези) першої події дорівнює ймовірності *похибки першого роду* α, а ймовірність другої події збігається з доповненням до одиниці $1-\alpha$. Крім того, при виконанні умови (1) справедлива нерівність $S_\tau\geq c_1$, а за умови (0) маємо $S_\tau\leq c_0$. Припустимо, що останні нерівності є наближеними рівностями, тобто суми S_n не занадто перестрибують відповідні рівні. Тоді $\mathbf{E}_{\theta_0}S_\tau\approx\alpha c_1+(1-\alpha)c_0$. Отже, внаслідок *тотожності Вальда* середня кількість спостережень для послідовного критерію наближено дорівнює

$$\mathbf{E}_{\theta_0}\tau\approx(\alpha c_1+(1-\alpha)c_0)/m(\theta_0).$$

Приклад. Нехай *кратна вибірка* $X=(\xi_1,...,\xi_n)$ утворена нормальними спостереженнями: $\xi_1\simeq N(\mu,\sigma^2)$ з невідомим середнім $\theta=\mu$ та відомою дисперсією σ^2. У наведеному вище прикладі застосування звичайного *критерію відношення вірогідностей* для перевірки гіпотези $H_0:\theta=\theta_0=\mu_0$ проти альтернативи $H_1:\theta=\theta_1=\mu_1$ встановлено, що за даних похибок α,β першого та другого роду найбільш потужний критерій потребує не менше ніж $n(\alpha,\beta)\equiv\sigma^2(x_\alpha+x_\beta)^2\ /\ (\mu_1-\mu_0)^2$ спостережень, де x_α,x_β – відповідні квантилі стандартного нормального розподілу.

Для порівняння обчислимо величини з виразу для $\mathbf{E}_{\theta_0}\tau$:

$$c_1=\ln C_1'=\ln((1-\beta)/\alpha),\quad c_0=\ln C_0'=\ln(\beta/(1-\alpha)),$$

$$m(\theta_0)=\mathbf{E}_{\theta_0}h(\xi_1)=\mathbf{E}_{\theta_0}\ln\frac{f(\xi_1,\theta_1)}{f(\xi_1,\theta_0)}=$$

$$\mathbf{E}_{\theta_0}\left(-(\xi_1-\mu_1)^2/2\sigma^2+(\xi_1-\mu_0)^2/2\sigma^2\right)=-(\mu_1-\mu_0)^2/2\sigma^2.$$

Відношення об'ємів для послідовного та простого критерію дорівнює

$$\frac{\mathbf{E}_{\theta_0}\tau}{n(\alpha,\beta)}\approx 2\ \frac{\alpha\ln(\alpha/(1-\beta))+(1-\alpha)\ln((1-\alpha)/\beta)}{(x_\alpha+x_\beta)^2}.$$

Якщо обрати $\alpha=\beta=0.05$, то $x_\alpha=x_\beta\approx 1.64$ і останнє відношення наближено дорівнює 0.49. Отже, для даних похибок має місце *подвійна економія числа спостережень* у порівнянні з *критерієм відношення вірогідностей*.

Вправи

(1) Кратна вибірка $X = (\xi_1, ..., \xi_n)$ утворена спостереженнями з показниковими розподілами $Exp(\theta)$. Знайти розподіл випадкової величини τ для однобічного критерію з $c_0 = 0$.

(2) Випадкові величини $(\xi_k, k = \overline{1,n})$ незалежні і мають рівномірні розподіли $U(0,\theta)$. Довести, що кількість спостережень послідовного критерію перевірки гіпотези $H_0 : \theta = 1$ проти альтернативи $H_1 : \theta = 2$ дорівнює $\tau = \inf(k \le n : \xi_k > 1)$ при $\inf \varnothing = n$. Довести, що ймовірності похибок першого та другого роду дорівнюють відповідно 0 і 2^{-n}, та $\mathbf{E}(\tau \mid H_0) = n$, $\mathbf{E}(\tau \mid H_1) = 2 - 2^{1-n}$.

(3) N осіб проходять аналіз крові, з імовірністю p фіксується захворювання кожної незалежно від інших..Процедура перевірки зводиться до аналізу сумарної проби кожної підгрупи з k осіб. Якщо результат негативний, то він поширюється на всіх у підгрупі, інакше повторно аналізується кожна особа підгрупи. Довести, що при малих p те значення k, що мінімізує середню кількість тестів, наближено дорівнює $2N\sqrt{p}$.

3.18. Метод найменших квадратів

У регресійному аналізі вивчається проблема кількісного впливу одних змінних (наприклад, відсоткового складу різних домішок у сплаві) на числові значення інших змінних (наприклад, міцності сплаву). Одночасно враховується стохастичний характер величин, що спостерігаються.

3.18.1. Модель лінійної регресії

В теорії лінійної регресії розглядається модель лінійної залежності спостережень від параметрів:

$$\xi_j = \sum\nolimits_{i=1}^{k} \theta_i \, t_{ij} + \varepsilon_j, \; j = \overline{1,n},$$

де $X = (\xi_j, \; j = \overline{1,n})$ – вектор спостережень (*вибірка*),

$\theta = (\theta_i, \; i = \overline{1,k}) \in \mathbb{R}^k$ – невідомий вектор параметрів, що підлягає оцінці, його розмірність $k < n$,

матриця $T = (t_{ij}, \; i = \overline{1,k}, \; j = \overline{1,n})$ вважається відомою,

вектор $\varepsilon = (\varepsilon_j, \; j = \overline{1,n})$ містить випадкові похибки вимірювань.

Природне припущення полягає в тому, що випадкові похибки ε_i є *незалежними у сукупності*, *однаково розподіленими*, мають нульове *середнє:* $\mathbf{E}\varepsilon_i = 0$ (відсутня систематична похибка) та *дисперсію* σ^2.

Усі вектори будемо розглядати як вектори-стовпчики.

У векторній формі модель лінійної регресії має вигляд

$$X = T'\theta + \varepsilon,$$

де символ $'$ визначає транспонування. Вектор похибок задовольняє умови

$$\mathbf{E}_\theta \varepsilon = 0, \quad \mathrm{Cov}_\theta\ \varepsilon = \sigma^2 I$$

з одиничною матрицею I.

Для оцінки невідомого параметра θ К.Гаусс запропонував і використав метод найменших квадратів (МНК).

Означення. Оцінкою методу найменших квадратів *векторного параметра θ називається статистика*

$$\widehat{\theta} = \arg\min\ _{\theta \in \mathbb{R}^k}\ (X - T'\theta)^2.$$

Ясно, що $\widehat{\theta}$ обчислюється за X і T, тобто є *оцінкою*. Існування $\widehat{\theta}$ випливає з неперервності та обмеженості знизу квадратичної функції.

Теорема (про співвідношення між оцінкою МНК та ОМВ). *Нехай похибки ε_j незалежні у сукупності і однаково нормально розподілені:*

$$\varepsilon \simeq N_n(0,\ \sigma^2 I).$$

Тоді оцінка МНК збігається з оцінкою максимальної вірогідності.

Доведення. За умовою вибірка

$$X \simeq T'\theta + N_n(0,\ \sigma^2 I) = N_n(T'\theta,\ \sigma^2 I)$$

є *нормальним вектором*, а логарифмічна *функція вірогідності* дорівнює

$$\ln L(X,\theta) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}(X - T'\theta)^2.$$

Отже, точка максимуму $L(X,\theta)$ за θ збігається з оцінкою *МНК* ⊡

Теорема (про рівняння методу найменших квадратів).

(а) Вектор $\widehat{\theta} = \widehat{\theta}(X) \in \mathbb{R}^k$ збігається з оцінкою методу найменших квадратів тоді й тільки тоді, коли він є розв'язком ***рівняння МНК***

$$V\widehat{\theta} = TX,$$

де V – симетрична матриця розміру $k \times k$, що має вигляд:

$$V \equiv TT'.$$

(б) Якщо rang $T = k$, то $\det V \neq 0$ і єдина оцінка МНК дорівнює

$$\widehat{\theta} = V^{-1}TX.$$

Доведення. *Необхідність.*

(а) Обчислимо головну частину приросту квадратичної функції в околі точки θ (диференціал Гато):

$$(X - T'(\theta + h))^2 - (X - T'\theta)^2 = 2h'T(X - T'\theta) + o(h), \; h \to 0.$$

Якщо θ є точкою екстремуму, то ліва частина не змінює знак при зміні знака h, тому диференціал Гато є нульовим

$$d[(X - T'\theta)^2, h] = 2h'T(X - T'\theta) = 2h'(TX - V\theta) = 0, \; \forall h \in \mathbb{R}^n,$$

що і доводить (а).

(б) З неперервності та невід'ємності квадратичної функції випливає існування хоча б однієї точки локального мінімуму. Тому *рівняння МНК* завжди має принаймні один розв'язок. З (а) робимо висновок, що єдина оцінка *МНК* дорівнює $V^{-1}TX$.

Достатність є очевидним наслідком наступної теореми.

Теорема (про розклад суми квадратів відхилень). *Нехай θ – істинне значення параметра, а $\widehat{\theta}$ – довільний розв'язок рівняння МНК. Тоді* ***повна сума квадратів*** *відхилень дорівнює сумі двох сум квадратів:*

$$SS \equiv (X - T'\theta)^2 = (X - T'\widehat{\theta})^2 + (T'(\widehat{\theta} - \theta))^2,$$

і набуває найменшого значення в точці $\theta = \widehat{\theta}$.

Означення. *Суми в зображенні повної суми квадратів відхилень*

$$SS_v = (X - T'\widehat{\theta})^2, \quad SS_r = (T'(\widehat{\theta} - \theta))^2$$

називаються відповідно сумою квадратів відхилень від регресії *та* залишковою сумою квадратів.

Доведення. Обчислимо

$$(X - T'\theta)^2 = (X - T'\widehat{\theta} + T'(\widehat{\theta} - \theta))^2 =$$
$$(X - T'\widehat{\theta})^2 + (T'(\widehat{\theta} - \theta))^2 + 2(T'(\widehat{\theta} - \theta))'(X - T'\widehat{\theta}) =$$
$$(X - T'\widehat{\theta})^2 + (T'(\widehat{\theta} - \theta))^2 + 2(\widehat{\theta} - \theta)'(TX - V\widehat{\theta}) =$$
$$(X - T'\widehat{\theta})^2 + (T'(\widehat{\theta} - \theta))^2,$$

оскільки $\widehat{\theta}$ є розв'язком *рівняння МНК* $\boxdot$

Теорема (про незміщеність і розподіл оцінки МНК). *Нехай* $\operatorname{rang} T = k$ *і має місце нормальна модель: $\varepsilon \simeq N_n(0, \, \sigma^2 I)$. Тоді:*

(а) оцінка МНК $\widehat{\theta}$ є незміщеною для θ, причому $\widehat{\theta} \simeq N_k(\theta, \, \sigma^2 V^{-1})$,

(б) статистика $\widehat{\sigma}^2_{n-k} = SS_v \; /(n-k)$ є незміщеною оцінкою для σ^2,

(в) статистики $SS_v \,/\, \sigma^2$ та $SS_r \,/\, \sigma^2$ незалежні й мають хі-квадрат розподіли з $n-k$ і k ступенями свободи відповідно,

(г) випадкові вектори $\widehat{\theta}$ та $X - T'\widehat{\theta}$ незалежні.

Зауваження. Властивості незміщеності виконуються і без припущення нормальності моделі.

Доведення. За теоремою *про лінійні перетворення нормальних векторів* статистика $\widehat{\theta} = V^{-1}TX$ має *нормальний розподіл*, тому твердження *(а)* випливає з рівнянь

$$\mathbf{E}_\theta\widehat{\theta} = \mathbf{E}V^{-1}T(T'\theta + \varepsilon) = \theta + V^{-1}T\mathbf{E}\varepsilon = \theta,$$
$$\mathsf{Cov}_\theta(\widehat{\theta}) = V^{-1}T\mathsf{Cov}(X)(V^{-1}T)' = \sigma^2 V^{-1}TT'V^{-1} = \sigma^2 V^{-1},$$

де використано теорему *про коваріаційну матрицю лінійного перетворення.* Останні тотожності доведено без припущення нормальності похибок.

Твердження *(б)* також виконується без такого припущення. Дійсно, з урахуванням теореми *про розклад суми квадратів відхилень*, попередніх тотожностей та симетричності матриці V отримуємо

$$(n-k)\mathbf{E}_\theta\widehat{\sigma}^2_{n-k} = \mathbf{E}_\theta(X - T'\widehat{\theta})^2 = \mathbf{E}_\theta(X - T'\theta)^2 - \mathbf{E}_\theta(T'(\widehat{\theta} - \theta))^2 =$$
$$\mathbf{E}_\theta \textstyle\sum_{j=1}^n \varepsilon_j^2 - \mathbf{E}_\theta(\widehat{\theta} - \theta)'V(\widehat{\theta} - \theta) =$$
$$n\sigma^2 - \mathbf{E}_\theta \textstyle\sum_{i,j=1}^k (\widehat{\theta}_i - \theta)V_{ij}(\widehat{\theta}_j - \theta) = n\sigma^2 - \sum_{i,j=1}^k V_{ij}\ \mathsf{Cov}(\widehat{\theta}_i, \widehat{\theta}_j) =$$
$$n\sigma^2 - \textstyle\sum_{i.j=1}^k V_{ij}\sigma^2(V^{-1})_{ij} = n\sigma^2 - \sigma^2\mathrm{Tr}(VV^{-1}) = (n-k)\sigma^2.$$

Для доведення *(в)* зауважимо, що за теоремою *про рівняння методу найменших квадратів*

$$\widehat{\theta} = V^{-1}TX = V^{-1}T(T'\theta + \varepsilon) = \theta + V^{-1}T\varepsilon.$$

Звідси отримуємо

$$SS_r = (T'(\widehat{\theta} - \theta))^2 = (T'(\theta + V^{-1}T\varepsilon) - T'\theta)^2 = (T'V^{-1}T\varepsilon)^2 \equiv (\Pi\varepsilon)^2.$$

Тут $n \times n$-матриця $\Pi \equiv T'V^{-1}T$ має ранг k, симетрична та ідемпотентна: $\Pi^2 = T'V^{-1}TT'V^{-1}T = T'V^{-1}VV^{-1}T = T'V^{-1}T = \Pi$. Тому існує *ортонормована* матриця U, що приводить її до діагональної матриці:

$$U\Pi U' = I_n(k) = (\delta_{ij}1\!\!1_{i\le k},\ i,j = \overline{1,n}).$$

Аналогічно обчислюємо

$$SS_v = (X - T'\widehat{\theta})^2 = (T'\theta + \varepsilon - T'(\theta + V^{-1}T\varepsilon))^2 =$$
$$((I - T'V^{-1}T)\varepsilon)^2 = ((I - \Pi)\varepsilon)^2.$$

Позначимо $\zeta = U\varepsilon/\sigma$. Тоді $\zeta \simeq N_n(0,\ I)$ – *стандартний нормальний вектор* за теоремою *про лінійні перетворення нормальних векторів*, оскільки $\varepsilon/\sigma \simeq N_n(0, I)$ – стандартний нормальний вектор за умовою.

Тому внаслідок теореми *про розклад суми квадратів відхилень*

$$SS = SS_v + SS_r = ((I-\Pi)\varepsilon)^2 + (\Pi\varepsilon)^2 =$$
$$\varepsilon'(I-\Pi)'(I-\Pi)\varepsilon + \varepsilon'\Pi'\Pi\varepsilon = \varepsilon'(I-\Pi)\varepsilon + \varepsilon'\Pi\varepsilon,$$

де мають місце рівності відповідно перших та других доданків. Отже,

$$SS_v/\sigma^2 + SS_r/\sigma^2 = \varepsilon'(I-\Pi)\varepsilon/\sigma^2 + \varepsilon'\Pi\varepsilon/\sigma^2 =$$
$$\zeta'U(I-\Pi)U'\zeta + \zeta'U\Pi U'\zeta =$$
$$\zeta'(I-I_n(k))\zeta + \zeta'I_n(k)\zeta = \textstyle\sum_{i=k+1}^{n}\zeta_i^2 + \sum_{i=1}^{k}\zeta_i^2 = \chi^2_{n-k} + \chi^2_k,$$

де величини в правій частині незалежні за теоремою *про векторні перетворення незалежних величин* і мають потрібні *хі-квадрат розподіли*

Для доведення *(г)* розглянемо випадкові вектори $T'(\widehat{\theta} - \theta) = \Pi\varepsilon$ та $X - T'\widehat{\theta} = (I-\Pi)\varepsilon$. Вони є лінійними функциями від нормального вектора ε, тому вони є за теоремою *про незалежність лінійних форм нормального вектора* незалежними за умови некорельованості. Остання випливає з рівностей

$$\mathbf{E}_\theta(X - T'\widehat{\theta})(T'(\widehat{\theta}-\theta))' = \mathbf{E}_\theta(I-\Pi)\varepsilon\varepsilon'\Pi = \sigma^2(I-\Pi)\Pi = 0$$

$\boxdot$

Зауваження. Відношення статистик із пункту *(в)* теореми *про незміщеність і розподіл оцінки МНК* має *розподіл Фішера* після нормування на кількості ступенів свободи і використовується для перевірки гіпотези про суттєвість залежності теоретичного середнього спостережень від параметра, що оцінюється. У дисперсійному аналізі (ANOVA) вказані статистики зображують у вигляді таблиці

Джерело	Сума квадратів	Ступені свободи	Середня сума квадратів
Залишок моделі	$SS_r = (T'(\widehat{\theta}-\theta))^2$	k	SS_r/k
Відхилення від регресії	$SS_v = (X - T'\widehat{\theta})^2$	$n-k$	$SS_v/(n-k)$
Загалом сума та відношення	$SS\ = (X - T'\theta)^2$	n	$\widehat{\phi}_{k,n-k} = \frac{SS_r}{k}/\frac{SS_v}{n-k}$

Тут θ – гіпотетичне значення параметра. Зокрема, нульове значення θ відображає відсутність залежності спостережень від точок вимірювання. Великі значення статистики $\widehat{\phi}_{k,n-k}$ з *розподілом Фішера* та $k, n-k$ ступенями свободи вказують на порушення гіпотези щодо $\theta = 0$, оскільки свідчать, що похибка SS_r у оцінюванні θ суттєво перевищує похибку SS_v у поясненні регресії. Одночасно використовується **коефіцієнт детермінації** $R^2 = SS_v/SS$, який відображає частку суми квадратів, що пояснена регресійною моделлю, у повній сумі квадратів відхилень. Можна довести (**вправа**), що статистика R^2 має *бета-розподіл*.

Теорема (теорема Гаусса – Маркова). *Нехай* rang $T = k$, $\widehat{\theta}$ *– оцінка методу найменших квадратів параметра* θ*, а вектор*

$$z = Z\theta \in \mathbb{R}^m, \ m \le k,$$

породжений лінійним перетворенням $Z : \mathbb{R}^k \to \mathbb{R}^m$.

Тоді оцінка $\widehat{z} = Z\widehat{\theta} = ZV^{-1}TX$ *є лінійною незміщеною оцінкою для* z*, і має найменше середнє квадратичне відхилення від* z *(тобто найменші діагональні елементи коваріаційної матриці) у класі лінійних за* X *незміщених оцінок вектора* z.

У припущенні нормальності моделі $\widehat{z} \simeq N(z, \ \sigma^2 ZV^{-1}Z')$.

Доведення. Нехай $\widetilde{z} = CX$ є довільною лінійною незміщеною оцінкою для z. Тоді

$$\mathbf{E}_\theta \widetilde{z} = C \ \mathbf{E}_\theta X = CT'\theta = z = Z\theta, \ \forall \theta \in \mathbb{R}^k,$$

звідки виводимо тотожність

$$Z = CT'.$$

Звідси випливає ортогональність матриць $ZV^{-1}T$ і $C - ZV^{-1}T$:

$$ZV^{-1}T(C - ZV^{-1}T)' = ZV^{-1}(TC' - TT'V^{-1}Z') =$$
$$ZV^{-1}(TC' - Z') = ZV^{-1}(TC' - (CT')') = 0.$$

Тому за теоремою *про коваріаційну матрицю лінійного перетворення*

$$\mathrm{Cov}_\theta(\widetilde{z}) = C\mathrm{Cov}_\theta(X)C' = C\mathrm{Cov}_\theta(\varepsilon)C' = \sigma^2 CC' =$$
$$\sigma^2(ZV^{-1}T + C - ZV^{-1}T)(ZV^{-1}T + C - ZV^{-1}T)' =$$
$$\sigma^2 ZV^{-1}T(ZV^{-1}T)' + ZV^{-1}T(C - ZV^{-1}T)' +$$
$$(C - ZV^{-1}T)(ZV^{-1}T)' + \sigma^2(C - ZV^{-1}T)(C - ZV^{-1}T)' =$$
$$\sigma^2 ZV^{-1}Z' + \sigma^2(C - ZV^{-1}T)(C - ZV^{-1}T)' =$$
$$\sigma^2 ZV^{-1}Z' = \mathrm{Cov}_\theta(Z\widehat{\theta}) = \mathrm{Cov}_\theta(\widehat{z}),$$

де нерівність матриць означає одночасну нерівність всіх відповідних діагональних елементів і випливає з невід'ємної визначеності другого доданку в лівій частині нерівності, що має вигляд $\sigma^2 DD'$. Останнє твердження теореми випливає з теореми *про лінійні перетворення нормальних векторів* та наведеного виразу для $\mathrm{Cov}_\theta(\widehat{z})$ ⊡

Вправи

(1) Довести, що без припущення нормальності моделі оцінка МНК $\widehat{\theta}$ та оцінка $\widehat{\sigma}^2_{n-k}$ є незміщеними для θ та σ^2, і $\mathrm{Cov}_\theta(\widehat{\theta}) = \sigma^2 V^{-1}$.

(2) Довести, що випадкові вектори $X - T'\widehat{\theta}$ і $T'(\widehat{\theta} - \theta)$ (а) ортогональні м.н., (б) некорельовані, (в) у випадку нормальної моделі незалежні.

(3) Нехай $X = (\xi_j,\ j = \overline{1,n})$ – довільна вибірка така, що $\mathbf{E}\xi_j = \mu$, $\mathbf{D}\xi_j = \sigma^2$, $\mathrm{Cov}(\xi_j, \xi_k) = \rho\sigma^2,\ j \neq k$. Довести, що в класі лінійних за спостереженнями ξ_j оцінок найменшу дисперсію має вибіркове середнє $\widehat{\mu}_n$.

(4) Нехай $\xi_k = \theta_{l(k)} + \varepsilon_k, k = \overline{1,n}$, де $l(k) = 1$ при $k \le m$ та $l(k) = 2$ при $k > m$, а випадкові похибки ε_k – незалежні з розподілом $N(0, \sigma^2)$. Знайти оцінку МНК для (θ_1, θ_2).

(5) Нехай у моделі лінійної регресії $X = T'\theta + \varepsilon$ координати вектора похибок ε незалежні, а $\widehat{\theta}$ – ОМВ вектора θ. (а) Якщо ε_j рівномірно розподілені на інтервалі $[-\sigma, \sigma]$, то $\widehat{\theta}$ мінімізує функцію контрасту $\max_{1 \le j \le n} \left|\xi_j - (T'\theta)_j\right|$. (б) Якщо ε_j мають щільності Лапласа $(2\sigma)^{-1} \exp(-|y|/\sigma)$, то $\widehat{\theta}$ мінімізує суму абсолютних відхилень $\sum_{j=1}^n \left|\xi_j - (T'\theta)_j\right|$. Знайти ОМВ для σ.

(6) Випадковий вектор (ξ, η) має сумісну щільність $f(x, y)$, а вагова функція $w(x, y) > 0$ така, що величина $(\xi^2 + \eta^2)w(\xi, \eta)$ інтегровна. Визначимо зважену оцінку МНК $(a^*, b^*) = \arg\ \min_{(a,b)\in\mathbb{R}^2} \mathbf{E}w(\xi, \eta)(\eta - a - b\xi)^2$. Нехай сумісна щільність вектора (ξ^*, η^*) пропорційна функції $f(x, y)w(x, y)$. Довести рівності $b^* = \mathrm{Cov}(\xi^*, \eta^*)/\mathbf{D}\xi^*$, $a^* = \mathbf{E}\eta^* - b^*\mathbf{E}\xi^*$.

(7) Припустимо, що похибки ε_j незалежні однаково розподілені центровані та $\sigma^2 = \mathbf{E}\varepsilon_1^2 < \infty$, причому $n^{-1}V_n \equiv n^{-1}TT' \to \Lambda, n \to \infty$. Вивести з центральної граничної теореми для випадкових векторів асимптотичну нормальність оцінки МНК: $\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \to^W N_k(0, \sigma^2\Lambda^{-1}),\ n \to \infty$.

3.18.2. Проста лінійна регресія

Розглянемо для прикладу випадок $k = 2$. Нехай

$$\theta = (a, b),\ t_{1j} = 1,\ t_{2j} = x_j,\ j = \overline{1,n}.$$

Модель простої лінійної регресії має вигляд

$$\xi_j = a + bx_j + \varepsilon_j, \quad j = \overline{1,n},$$

де b – нахил регресії, a – її зміщення.

Матриці T, V, V^{-1} задаються рівностями

$$T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & \ldots & x_n \end{pmatrix},$$

$$V = TT' = n \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix}, \quad \overline{x} \equiv \tfrac{1}{n}\sum\nolimits_{j=1}^{n} x_j, \ \overline{x^2} = \tfrac{1}{n}\sum\nolimits_{j=1}^{n} x_j^2,$$

$$V^{-1} = \tfrac{1}{n\sigma_x^2} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}, \ \sigma_x^2 \equiv \overline{x^2} - (\overline{x})^2 = \tfrac{1}{n}\sum\nolimits_{j=1}^{n}(x_j - \overline{x})^2.$$

Звідси обчислюємо оцінки МНК

$$V^{-1}T = \tfrac{1}{n\sigma_x^2} \begin{pmatrix} \overline{x^2} - x_1\overline{x} & \overline{x^2} - x_2\overline{x} & \ldots & \overline{x^2} - x_n\overline{x} \\ x_1 - \overline{x} & x_2 - \overline{x} & \ldots & x_n - \overline{x} \end{pmatrix},$$

$$\widehat{b} = \frac{1}{n}\sum\nolimits_{j=1}^{n} \xi_j (x_j - \overline{x})/\sigma_x^2,$$

$$\widehat{a} = \sum\nolimits_{j=1}^{n} \xi_j(\overline{x^2} - x_j\overline{x})/n\sigma_x^2 = \sum\nolimits_{j=1}^{n} \xi_j(\overline{x^2} - (\overline{x})^2 - (x_j - \overline{x})\overline{x})/n\sigma_x^2 =$$
$$\sum\nolimits_{j=1}^{n} \xi_j(\overline{x^2} - (\overline{x})^2)/n\sigma_x^2 - \overline{x}\widehat{b} = \frac{1}{n}\sum\nolimits_{j=1}^{n} \xi_j - \overline{x}\widehat{b} = \widehat{\mu}_n - \overline{x}\widehat{b},$$

Оцінка $\widehat{b}$ називається *вибірковим коефіцієнтом кореляції* векторів (ξ_j), (x_j). За теоремою *про незміщеність і розподіл оцінки МНК* у припущенні нормальності моделі вектор $(\widehat{a}, \widehat{b})$ є *нормальним вектором на площин*і:

$$(\widehat{a}, \widehat{b}) \simeq N_2(a, \ b, \ \sigma^2\overline{x^2}/n\sigma_x^2, \ \sigma^2/n\sigma_x^2, \ -\sigma^2\overline{x}(\overline{x^2})^{-1/2}),$$

(**вправа**) і не залежить від незміщеної оцінки дисперсії похибок:

$$\widehat{\sigma}_{n-2}^2 \equiv \frac{1}{n-2}SS_v = \frac{1}{n-2}\sum\nolimits_{j=1}^{n}(\xi_j - \widehat{a} - \widehat{b}x_j)^2.$$

Остання після ділення на $\sigma^2/(n-2)$ має *хі-квадрат розподіл* з $n-2$ ступенями свободи. Тому, зокрема, статистика

$$\widehat{\tau}_{n-2} = \sqrt{n}(\widehat{b} - b)\sigma_x/\widehat{\sigma}_{n-2}$$

вже не містить σ та має *розподіл Стьюдента* з $n-2$ ступенями свободи (**вправа**). Це твердження застосовується для побудови надійних інтервалів та перевірки гіпотез щодо нахилу b.

Далі, за *теоремою Гаусса - Маркова* для даного $x \in \mathbb{R}$ статистика

$$\widehat{y} = \widehat{a} + \widehat{b}x = \widehat{\mu}_n + \widehat{b}(x - \overline{x}) \simeq N(y, \ \sigma^2(1 + (x - \overline{x})^2/\sigma_x^2)/n),$$

є незміщеною лінійною оцінкою мінімальної дисперсії для прогнозованого значення $y = a + bx$ (**вправа**). Нарешті, статистика

$$\widehat{\tau}_{n-2} = \sqrt{n}(\widehat{y} - y)/\left(\widehat{\sigma}_{n-2}\sqrt{1 + (x - \overline{x})^2/\sigma_x^2}\right)$$

має *розподіл Стьюдента* з $n-2$ ступенями свободи, що дає можливість інтервального оцінювання та перевірки гіпотез для y.

Приклад. Розглянемо вибірку з 10 страхових вимог X та відповідних страхових виплат Y :

x	2.10	2.40	2.50	3.20	3.60	3.80	4.10	4.20	4.50	5.00
y	2.18	2.06	2.54	2.61	3.67	3.25	4.02	3.71	4.38	4.45

Графічне зображення цих даних підтверджує, що тут може бути наявною лінійна залежність.

Обчислення дають такі результати:

$\overline{x} = 3.54$, $\overline{x^2} = 13.376$, $\overline{y} = 3.287$, $\overline{y^2} = 11.520$, $\overline{xy} = 12.381$,

$\sigma_x^2 = 0.844$, $\sigma_y^2 = 0.716$, $\sum_{j=1}^{10} y_j(x_j - \overline{x}) = 7.45$,

$\widehat{b} = 0.88$, $\widehat{a} = 0.16$.

Вправи

(1) Знайти оцінку (a, b) методу найменших квадратів перпендикулярних відстаней від регресії, що мінімізує суму $\sum_{j=1}^{n}(\xi_j - a - bx_j)^2/(1 + b^2)$.

(2) Визначимо для вектора $\xi = (\eta, \zeta)$ змішані моменти $\mu_{ij} = \mathbf{E}\eta^i\zeta^j, i, j \geq 0$, та для кратної векторної вибірки $X = (\xi_1, ..., \xi_n)$ вибіркові змішані моменти $\widehat{\mu}_{ij} = n^{-1}\sum_{k=1}^{n}\eta_k^i\zeta_k^j$. (а) Узагальнити метод моментів на клас змішаних моментів. (б) Нехай має місце модель лінійної регресії $\eta_k = a + b\zeta_k + \varepsilon_k, k = \overline{1, n}$, з некорельованими однаково розподіленими центрованими похибками ε_k. Довести, що оцінка методу моментів параметра $\theta = (a, b)$ має вигляд

$$\widehat{b} = \left(\sum_{k=1}^{n}(\eta_k - \widehat{\eta}_n)(\xi_k - \widehat{\xi}_n)\right)/\sum_{k=1}^{n}(\xi_k - \widehat{\xi}_n)^2, \qquad \widehat{a} = \widehat{\eta}_n - \widehat{b\xi}_n.$$

(3) Нехай $(\xi_1, ..., \xi_n)$ – кратна вибірка з нормального розподілу на площині, тобто $\xi_j = (\zeta_{1j}, \zeta_{2j}) \simeq N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. (а) Довести, що ОМВ для вектора $(\sigma_1^2, \sigma_2^2, \rho)$ при відомих середніх μ_1, μ_2 має вигляд: $\widehat{\sigma}_i^2 = n^{-1}\sum_{j=1}^{n}(\zeta_{ij} - \mu_i)^2$, $i = 1, 2$, $\widehat{\rho} = \left(\sum_{j=1}^{n}(\zeta_{1j} - \mu_1)(\zeta_{2j} - \mu_2)\right)/n\widehat{\sigma}_1\widehat{\sigma}_2$. (б) При невідомих μ_1, μ_2 та $n > 5$ ці оцінки визначаються як у попередній задачі.

3.18.3. Поліноміальна регресія

Попередню задачу лінійної апроксимації спостережень можна узагальнити, якщо розглянути схему **поліноміальної регресії**:

$$\xi_j = \sum\nolimits_{i=0}^{k} \theta_i \, x_j^i + \varepsilon_j, \ j = \overline{1,n},$$

де (x_j) – задані точки спостережень. Незважаючи на нелінійний характер залежності від x, дана модель залишається лінійною за невідомим параметром $\theta = (\theta_0, ..., \theta_k)$. Тому для аналізу поліноміальної регресії можна застосувати всі результати розділу про лінійну регресію, якщо визначити матрицю перетворень $T = (x_j^i, \ i = \overline{0,k}, \ j = \overline{1,n})$.

Припустимо, що точки спостережень (x_j) "рівномірно розподілені" на відрізку $(0,1)$, тобто $\sum_{j=1}^n g(x_j) \approx \int_0^1 g(x)dx$ для $g \in C(0,1)$. Тоді матриця $V = TT'$ дорівнює

$$V = \left(\sum\nolimits_j x_j^i x_j^l, \ 0 \le i,l \le k \right) \approx (1/(i+l+1), \ 0 \le i,l \le k).$$

Можна довести, що визначник цієї матриці дуже близький до нуля навіть при великих k. Отже, обчислення оцінки *МНК*, що пов'язане з обертанням матриці V, стикатиметься з великими труднощами.

Для поліпшення якості оцінювання в поліноміальній схемі доцільно розглянути її в еквівалентному (внаслідок перепараметризації) вигляді

$$\xi_j = \sum\nolimits_{i=0}^{k} \theta_i \, \varphi_i(x_j) + \varepsilon_j, \ j = \overline{1,n},$$

де $\varphi_i(x)$ – деякі поліноми ступеня i від x. Матриця V тут має вигляд

$$V = \left(\sum\nolimits_{j=1}^{n} \varphi_i(x_j) \ \varphi_l(x_j), \ 0 \le i,l \le k \right).$$

Для поліпшення якостей оцінки *МНК* поліноми φ_i обирають так, щоб матриця V була діагональною. З огляду на її визначення розглянемо таку білінійну форму (псевдо-скалярний добуток) від функцій f, g:

$$(f,g) \equiv \sum\nolimits_{j=1}^{n} f(x_j) \ g(x_j).$$

Визначимо рекурентно послідовність **поліномів Чебишева**

$$\varphi_0(x) \equiv 1, \quad \varphi_k(x) \equiv x^k - \sum\nolimits_{i=0}^{k-1} \frac{(x^k, \varphi_i)}{(\varphi_i, \varphi_i)} \varphi_i(x), \ k \ge 1.$$

За індукцією з цього означення виводимо, що для всіх $i < k$

$$(\varphi_k, \varphi_i) = (x^k, \varphi_i) - (\varphi_i, \varphi_i)(x^k, \varphi_i)/(\varphi_i, \varphi_i) = 0.$$

Отже, поліноми Чебишева попарно ортогональні, тому коваріаційна матриця $V = ((\varphi_i, \varphi_i)\delta_{ik})$ є діагональною і оцінка МНК дорівнює

$$\widehat{\theta} = V^{-1}TX = \left(\sum\nolimits_{j=1}^{n} \xi_j \ \varphi_i(x_j) \ / \ (\varphi_i, \varphi_i), \quad i = \overline{0, k}\right).$$

Оскільки *коваріаційна матриця* оцінки $\widehat{\theta}$ за теоремою *про незміщеність і розподіл оцінки МНК* дорівнює діагональній матриці $\sigma^2 V^{-1}$, то координати $\widehat{\theta}_i$ некорельовані, а за нормальної моделі – незалежні.

3.18.4. Дисперсійний аналіз

На відміну від регресійного аналізу, де залежність спостережень від незалежної змінної (кількісного фактора) конкретизується як лінійна через матрицю плану спостережень T, у дисперсійному аналізі можливий будь-який вид залежності спостережень (вимірювань) від значення якісного фактора (номера ділянки, назви фірми-поставника товару, статі обслідуваних і т.п.).

Модель однофакторного **дисперсійного аналізу** (**ANOVA**), що відображає наявність одного визначального фактора, має вигляд

$$\xi_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad j = \overline{1, n_i}, i = \overline{1, k}.$$

Спостереження ξ_{ij} інтерпретується як результат j-го вимірювання при i-му значенні фактора. Всього маємо k значень (рівнів) фактора, та n_i спостережень на i-му рівні.

Сума $\mu + \alpha_i$ є середнім значенням спостережень (**відгуком**) при i-му рівні фактора, а параметр μ – спільним середнім усіх спостережень. Різниця α_i між вказаними середніми називається **ефектом** i-го рівня. Задача полягає у оцінюванні параметрів $\mu + \alpha_i$ за спостереженнями ξ_{ij}. Оскільки кількість параметрів $k + 1$ перевищує кількість досяжних значень $\mu + \alpha_i, i = \overline{1, k}$, то для однозначності оцінювання на доданки α_i необхідно накласти одну додаткову умову. Тому вважають, що *ефекти* α_i задовольняють умову

$$\sum\nolimits_{i=1}^{k} n_i \alpha_i = 0.$$

Доданок ε_{ij} є похибкою вимірювань. Постулюється, що випадкові величини $(\varepsilon_{ij}, j = \overline{1, n_i}, i = \overline{1, k})$ незалежні у сукупності центровані: $\mathbf{E}\varepsilon_{ij} = 0$, і мають нормальний розподіл $\varepsilon_{ij} \simeq N(0, \sigma^2)$ з невідомою дисперсією σ^2.

Якщо розглянути вектор параметрів $\theta = (\mu, \alpha_1, ..., \alpha_k)$, перенумерувати спостереження та відповідно визначити матрицю T з нулів та одиниць, модель *дисперсійного аналізу* можна звести до моделі *лінійної регресії*.

Однак і прямим обчисленням можна отримати *оцінки методу найменших квадратів*

$$(\widehat{\mu}, \widehat{\alpha}_1, ...\widehat{\alpha}_k) \equiv \operatorname{argmin}_{(\mu,\alpha_i)} \sum_{i,j} (\xi_{ij} - \mu - \alpha_i)^2.$$

що мають вигляд

$$\widehat{\mu} = \overline{\xi}_{\bullet\bullet}, \quad \widehat{\alpha}_i = \overline{\xi}_{i\bullet} - \overline{\xi}_{\bullet\bullet}, \ i = \overline{1,k}.$$

Тут використано позначення для усереднених значень

$$\overline{\xi}_{\bullet\bullet} = n^{-1} \sum_{i,j} \xi_{ij}, \quad \overline{\xi}_{i\bullet} = n_i^{-1} \sum_{j=1}^{n_i} \xi_{ij}.$$

Доведення спирається на той факт, що вектори з координатами

$$(\overline{\xi}_{\bullet\bullet} - \mu), \ \ (\overline{\xi}_{i\bullet} - \overline{\xi}_{\bullet\bullet} - \alpha_i), \ \ (\xi_{ij} - \overline{\xi}_{i\bullet}) \text{ при } j = \overline{1, n_i}, i = \overline{1,k}$$

є попарно ортогональними для довільних μ, α_i (**Вправа**).

З цієї ортогональності виводимо розклад суми квадратів відхилень:

$$SS = SS_v + SS_r, \quad \text{де}$$
$$SS \equiv \sum_{i,j} (\xi_{ij} - \widehat{\mu})^2 \text{ – повна сума квадратів,}$$
$$SS_v \equiv \sum_i n_i \widehat{\alpha}_i^2 \text{ – міжгрупова сума квадратів,}$$
$$SS_r \equiv \sum_{i,j} \left(\xi_{ij} - \widehat{\mu} - \widehat{\alpha}_i\right)^2 \text{ – залишкова сума квадратів.}$$

Можна довести, що координати випадкового вектора

$$\zeta \equiv (\overline{\xi}_{\bullet\bullet}, \ (\overline{\xi}_{i\bullet} - \overline{\xi}_{\bullet\bullet}, \ i = \overline{1,k}), \ (\xi_{ij} - \overline{\xi}_{i\bullet}, \ i = \overline{1,k}, \ j = \overline{1, n_i}))$$

попарно некорельовані (**вправа**), тому за теоремою *про лінійні перетворення нормальних векторів* та наслідком *про незалежність лінійних форм нормального вектора* вказані координати незалежні у сукупності. Звідси виводимо, що випадкові величини $\widehat{\mu}$, SS_v та SS_r незалежні.

Перше питання, що виникає у дисперсійному аналізі, полягає у тому, чи існує взагалі залежність *відгуків* від рівня фактору. Тому формулюється така нульова гіпотеза щодо *ефектів*

$$H_0 : \alpha_i = 0, \ \forall i = \overline{1,k}.$$

З нормальності вектора ζ та незалежності його координат виводимо, що за умови H_0 випадкові величини SS/σ, SS_v/σ та SS_r/σ мають *хі-квадрат розподіли* з $n-1$, $k-1$ та $n-k$ ступенями свободи відповідно (**вправа**). Даний факт відображається у такій таблиці ANOVA

Джерело	Сума квадратів	Ступені свободи	Середня сума квадратів
Залишкова сума квадратів	SS_r	$n-k$	$SS_r/(n-k)$
Міжгрупова сума квадратів	SS_v	$k-1$	$SS_v/(k-1)$
Повна сума квадратів	SS	$n-1$	

Тому для перевірки гіпотези H_0 використовується статистика

$$\widehat{\phi}_{k-1,n-k} = (SS_v/(k-1)) \ / \ (SS_r/(n-k)),$$

що має *розподіл Фішера* з $k-1, n-k$ ступенями свободи. Великі значення цієї статистики вказують на порушення гіпотези H_0, тому статистичний критерій $(\widehat{\phi}_{k-1,n-k}, [w_{nk\alpha}, \infty))$ з належним критичним значенням $w_{nk\alpha}$ має рівень α та є консистентним для перевірки гіпотези H_0.

Оцінювання дисперсії σ^2 похибок ε_{ij} можливе через незміщені оцінки

$$\widehat{\sigma}_i^2 \equiv (n_i-1)^{-1} \sum_{j=1}^{n_i} (\xi_{ij} - \widehat{\mu} - \widehat{\alpha}_i)^2,$$

$$\widehat{\sigma}^2 \equiv (n-k)^{-1} \sum_{i=1}^{k} (n_i-1)\widehat{\sigma}_i^2 = (n-k)^{-1} \sum_{i,j} (\xi_{ij} - \widehat{\mu} - \widehat{\alpha}_i)^2.$$

Випадкові величини $\widehat{\sigma}_i^2, \widehat{\sigma}^2$ не залежать від $\widehat{\mu}$ та SS_v і мають після ділення на σ^2 *хі-квадрат розподіли* з $n_i - 1$ та $n-k$ ступенями свободи.

Наступне питання полягає у тому, чи суттєво відрізняються *відгуки* від різних рівнів фактора. Можна розглянути гіпотезу $\alpha_i - \alpha_j = 0$ для деяких $i \neq j$. Більш загальною є гіпотеза $H_0 : \sum_{i=1}^{k} c_i \alpha_i = 0$, де вектор $c = (c_i)$ є **контрастом**, тобто $\sum_{i=1}^{k} c_i = 0$. Позначимо $\sigma_c^2 \equiv \sum_{i=1}^{k} c_i^2/n_i$.

З наведених вище тверджень про нормальність та незалежність робимо висновок (**вправа**), що статистика

$$\widehat{t}_{n-k} = \left(\sum_{i=1}^{k} c_i(\widehat{\alpha}_i - \alpha_i) \right) / \sigma_c \widehat{\sigma}$$

має *розподіл Стьюдента* з $n-k$ ступенями свободи. Великі значення статистики свідчать про порушення гіпотези $H_0 : \sum_{i=1}^{k} c_i \alpha_i = 0$. Це твердження можна використати як для побудови інтервальних оцінок контрасту $\sum_{i=1}^{k} c_i \alpha_i$ так і для перевірки гіпотези про його нульове значення.

Зауважимо, що одночасне використання критерію Стьюдента для різних *контрастів* c за однією вибіркою призводить до похибки у вірогідному рівні критерію, оскільки статистики критеріїв є залежними. Для подолання цієї обставини застосовують S-метод Шеффера. Він випливає з нерівності Коші

$$\left(\sum_{i=1}^{k} c_i(\widehat{\alpha}_i-\alpha_i)\right)^2 \sigma_c^{-2} = \left(\sum_{i=1}^{k} c_i(\overline{\varepsilon}_{i\bullet}-\overline{\varepsilon}_{\bullet\bullet})\right)^2 \sigma_c^{-2} \le \sum_{i=1}^{k} n_i(\overline{\varepsilon}_{i\bullet}-\overline{\varepsilon}_{\bullet\bullet})^2,$$

де права частина не залежить від c та після ділення на σ^2 має хі-квадрат розподіл з $k-1$ ступенем свободи. Звідси виводимо, що для всіх $\varepsilon > 0$

$$\mathbf{P}\left(\sup_{(c_i)}\left(\sum_{i=1}^{k} c_i(\widehat{\alpha}_i-\alpha_i)\right)^2 /(k-1)\sigma_c^2\widehat{\sigma}^2 \ge \varepsilon\right) \le \mathbf{P}(\widehat{\phi}_{k-1,n-k} \ge \varepsilon),$$

де величина $\widehat{\phi}_{k-1,n-k}$ має *розподіл Фішера* з $k-1, n-k$ ступенями свободи. Цим визначається множина нульових контрастів за заданим рівнем.

Вправа. Розглянемо лінійну двохфакторну модель, у якій двовимірна вибірка $X = (\xi_{ij}, i = \overline{1,n}.j = \overline{1,m})$ задовольняє рівняння

$$\xi_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i = \overline{1,n},\ j = \overline{1,m},$$

де похибки $\varepsilon_{ij} \simeq N(0,\sigma^2)$ незалежні у сукупності, а вектор невідомих параметрів $\theta = (\mu,\ (\alpha_i, i = \overline{1,n}),\ (\beta_j, j = \overline{1,m}),\ \sigma^2)$ такий, що виконуються умови центрованості: $\sum_{i=1}^{n} \alpha_i = 0$, $\sum_{j=1}^{m} \beta_j = 0$. Довести, що вектор

$$\widehat{\theta} = (\overline{\xi}_{\bullet\bullet},\ (\overline{\xi}_{i\bullet} - \overline{\xi}_{\bullet\bullet}, i = \overline{1,n}),\ (\overline{\xi}_{\bullet j} - \overline{\xi}_{\bullet\bullet}, j = \overline{1,m}),\ \widehat{\sigma}^2)$$

є оптимальною оцінкою для θ, що складається з чотирьох незалежних груп. Тут групові середні статистики мають вигляд

$$\overline{\xi}_{\bullet\bullet} = \sum_{i=1}^{n}\sum_{j=1}^{m} \xi_{ij}/nm, \quad \overline{\xi}_{i\bullet} = \sum_{j=1}^{m} \xi_{ij}/m, \quad \overline{\xi}_{\bullet j} = \sum_{i=1}^{n} \xi_{ij}/n,$$

$$\widehat{\sigma}^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} (\xi_{ij} - \overline{\xi}_{i\bullet} - \overline{\xi}_{\bullet j} + \overline{\xi}_{\bullet\bullet})^2/(n-1)(m-1).$$

3.19. Квадратично інтегровні стаціонарні випадкові процеси

Розглянемо задачі статистичного прогнозу для стаціонарних випадкових процесів другого порядку. Такі процеси ефективно використовуються в теорії передачі інформації, фінансовій математиці тощо.

Позначимо через

$$L_2(\mathbf{P}) = \{\xi : \Omega \to \mathbb{C},\ \mathbf{E}\,|\xi|^2 < \infty\}$$

гільбертів простір квадратично інтегровних комплекснозначних випадкових величин зі скалярним добутком $(\xi,\eta) = \mathbf{E}\xi\overline{\eta}$, де $\overline{x}$ комплексно спряжене до x, із нормою $\|\xi\|^2 = (\xi,\xi) = \mathbf{E}\,|\xi|^2$ та відповідною середньоквадратичною (с.к.) границею

$$\xi = l.i.m\, \xi_n \iff \mathbf{E}\,|\xi_n - \xi|^2 \to 0,\ n \to \infty.$$

Елементи $L_2(\mathbf{P})$ розрізняються з точністю до рівності *м.н.*

Нехай $G \subset L_2(\mathbf{P})$ – довільна підмножина. Позначимо породжений G **лінійний підпростір**

$$\pounds(G) = \left\{\eta = \sum\nolimits_{k=1}^{n} c_k\xi_k,\ \xi_k \in G,\ c_k \in \mathbb{C}\right\}$$

та його **с.к. замикання**

$$\aleph(G) = \left\{\xi \in L_2(\mathbf{P}),\ \xi = l.i.m\,\eta_n,\ \eta_n \in \pounds(G)\right\}.$$

Назвемо **ізометрією** підпросторів гільбертових просторів таку їх лінійну бієкцію, що зберігає норму (отже, і скалярний добуток).

Теорема (про продовження ізометрії). *Нехай* $H_k, k = 1, 2,$ – *пара гільбертових просторів з нормами* $\|\cdot\|_k$, *кожен з яких містить щільний лінійний підпростір* $H_{0k} \subset H_k$, $H_k = \aleph(H_{0k})$. *Тоді довільна ізометрія підпросторів* $\Im_0 : H_{01} \to H_{02}$ *має єдине продовження до ізометрії* $\Im : H_1 \to H_2$, $\Im\,|_{H_{01}} = \Im_0$.

Доведення. Нехай $h_1 \in H_1,\ h_1 = l.i.m\ h_{1n},\ \ h_{1n} \in H_{01}$.

Покладемо $h_{2n} = \Im_0 h_{1n}$. Зі збіжності, а отже і фундаментальності h_{1n} виводимо фундаментальність h_{2n} :

$$\|h_{2n} - h_{2m}\|_2 = \|\Im_0(h_{1n} - h_{1m})\|_2 = \|h_{1n} - h_{1m}\|_1 \to 0,\ n, m \to \infty,$$

а тому і збіжність цієї послідовності внаслідок повноти H_2. Позначимо

$$\Im h_1 = l.i.m\ h_{2n}.$$

Ця границя визначена коректно і є шуканою ізометрією, оскільки з неперервності норми випливає тотожність

$$\|\Im h_1\|_2 = \lim\|h_{2n}\|_2 = \lim\|\Im_0 h_{1n}\|_2 = \lim\|h_{1n}\|_1 = \|h_1\|_1\,,$$

а також і взаємна однозначність ⊡

3.19.1. Стаціонарні випадкові процеси

Означення. *Нехай параметрична множина* $T = \mathbb{Z}$ *або* $T = \mathbb{R}$. *Квадратично інтегровний комплекснозначний випадковий процес*

$$(\xi_t,\ t \in T)\ \subset\ L_2(\mathbf{P})$$

називається **стаціонарним (стаціонарним процесом другого порядку,** *або ж* **стаціонарним у широкому розумінні)**, *якщо*

$$\mathbf{E}\xi_t = m,\ \ \mathrm{Cov}(\xi_t,\ \overline{\xi_s}) \equiv \mathbf{E}(\xi_t - m)(\overline{\xi_s - m}) = r(t - s),\ \ \forall t, s \in T,$$

для деяких комплексних m, $r(t)$, $t \in T$. Функція $r(t)$ називається коваріаційною функцією *процесу.*

Приклади

(а) Стохастична гармоніка. Нехай ζ_0 *квадратично інтегровна* центрована випадкова величина: $\mathbf{E}\zeta_0 = 0, \mathbf{E}\,|\zeta_0|^2 < \infty$, а λ, θ – дійсні сталі. Тоді процес $\xi_t = \zeta_0 \exp(i\lambda t + i\theta)$ є *стаціонарним* із нульовим *середнім* та *коваріаційною функцією* $r(t) = \mathbf{E}\,|\zeta_0|^2 \exp(i\lambda t)$.

(б) Процес зі скінченним спектром. Нехай $(\zeta_k, k = \overline{1, n})$ набір із центрованих некорельованих випадкових величин: $\mathbf{E}\zeta_k = 0$, $\mathbf{E}\zeta_k\overline{\zeta_l} = 0, k \neq l$, а λ_k, θ_k – дійсні сталі. Тоді процес

$$\xi_t = \sum\nolimits_{k=1}^{n} \zeta_k \exp(i\lambda_k t + i\theta_k)$$

є *стаціонарним* із нульовим середнім та *коваріаційною функцією*

$$\mathrm{Cov}(\xi_t,\ \overline{\xi_s}) = \mathbf{E}\sum\nolimits_{k=1}^{n} \zeta_k \exp(i\lambda_k t + i\theta_k) \sum\nolimits_{j=1}^{n} \overline{\zeta_j} \exp(-i\lambda_j s - i\theta_j) =$$

$$\sum\nolimits_{k=1}^{n} \mathbf{E}\,|\zeta_k|^2 \exp(i\lambda_k(t-s)) = r(t-s).$$

Означення. *Процес $(\xi_t,\ t \in T)$ є* с.к. неперервним, *якщо*

$$\mathbf{E}\,|\xi_t - \xi_s|^2 \equiv \|\xi_t - \xi_s\|^2 \to 0,\ t \to s,\ \forall s.$$

Зауваження (про критерій с.к. неперервності). *С.к. неперервність стаціонарного процесу еквівалентна неперервності його коваріаційної функції в нулі.*

Дійсно, після центрування процесу маємо

$$\|\xi_t - \xi_s\|^2 = \mathbf{E}(\xi_t - \xi_s)(\overline{\xi_t - \xi_s}) = \mathbf{E}(|\xi_t|^2 - \xi_t\overline{\xi_s} - \xi_s\overline{\xi_t} + |\xi_s|^2) =$$

$$2\ \mathrm{Re}\ (r(0) - r(t-s)),$$

$$|r(t) - r(0)|^2 = \left|\mathbf{E}(\xi_t - \xi_0)\overline{\xi_0}\right|^2 \leq \|\xi_t - \xi_0\|^2 r(0)\ \boxdot$$

Теорема (про властивості коваріаційної функції). *Функція $r(t)$ є коваріаційною функцією стаціонарного процесу тоді й тільки тоді, коли вона є* **додатно визначеною,** *тобто для всіх $n \geq 1$, $t_k \in T$, $c_k \in \mathbb{C}$*

$$\sum\nolimits_{j,k=1}^{n} c_k\overline{c_j}\, r(t_k - t_j) \geq 0.$$

Доведення. *Необхідність* отримуємо обчисленням форм через ξ_t:

$$\textstyle\sum_{j,k=1}^{n} c_k\overline{c_j}\, r(t_k - t_j) = \sum_{j,k=1}^{n} c_k\overline{c_j}\, \mathbf{E}\xi_{t_k}\overline{\xi_{t_j}} =$$

$$\mathbf{E}\textstyle\sum_{j,k=1}^{n} c_k\overline{c_j}\ \xi_{t_k}\overline{\xi_{t_j}} = \mathbf{E}\left|\sum_{k=1}^{n} c_k\ \xi_{t_k}\right|^2 \geq 0.$$

Достатність випливає з існування відповідних нормальних випадкових послідовностей, доведення не наводиться $\boxdot$

Теорема (теорема Бохнера про додатно визначені функції).

Функція $r(t)$ є неперервною додатно визначеною тоді й тільки тоді, коли знайдеться скінченна міра Лебега – Стілтьєса F на $\mathfrak{B}(\mathbb{R})$ така, що має місце **спектральне зображення**

$$r(t) = \int_{\mathbb{R}} \exp(i\lambda t)\, dF(\lambda), \quad \forall t \in T.$$

Означення. *Міра F називається* **спектральною мірою коваріаційної функції** *r, а відповідна невласна функція розподілу $F(x) \equiv F((-\infty, x))$ – її* **спектральною функцією.**

Доведення. *Достатність* перевіряється безпосередньо:

$$\sum\nolimits_{j,k=1}^{n} c_k \overline{c_j}\, r(t_k - t_j) = \int_{\mathbb{R}} \sum\nolimits_{j,k=1}^{n} c_k \overline{c_j}\ \exp(i\lambda(t_k - t_j)) dF(\lambda) =$$

$$\int_{\mathbb{R}} \left| \sum\nolimits_{k=1}^{n} c_k \exp(i\lambda t_k) \right|^2 dF(\lambda) \geq 0.$$

Необхідність доведемо для випадку послідовностей: при $T = \mathbb{Z}$. Відповідне твердження називається **теоремою Герглотца**.

Для довільної неперервної додатно визначеної функції $r(n), n \in \mathbb{Z}$, розглянемо таку послідовність функцій аргумента $\lambda \in [-\pi, \pi]$:

$$f_N(\lambda) = \tfrac{1}{2\pi N} \sum\nolimits_{k,l=1}^{N} r(k-l) \exp(-ik\lambda + il\lambda) =$$
$$\tfrac{1}{2\pi} \sum\nolimits_{|n| \leq N} (1 - |n|/N)\, r(n) \exp(-in\lambda),$$

де зроблено заміну змінної $n = k - l$.

Ці функції дійсні та невід'ємні за умовою *додатної визначеності* після вибору $t_k = k$, $c_k = \exp(-ik\lambda)$, та обмежені. Тому інтеграли Лебега

$$F_N(B) = \int_{B\ \cap\ [-\pi,\pi]} f_N(\lambda)\, d\lambda,$$

задають невід'ємні скінченні міри. За ортогональністю гармонік

$$\int_{-\pi}^{\pi} \exp(i\lambda k) \exp(-i\lambda j) d\lambda = 2\pi\ \delta_{kj},$$

з означення f_N та функції $x^+ = \max(x, 0)$ обчислюємо

$$\int_{-\pi}^{\pi} \exp(i\lambda n) dF_N(\lambda) = (1 - |n|/N)^+\, r(n).$$

Міри F_N зосереджені на компактному інтервалі $[-\pi, \pi]$ і рівномірно обмежені: $F_N(\mathbb{R}) = \int_{-\pi}^{\pi} \exp(i\lambda 0) dF_N(\lambda) = r(0) < \infty$. За *теоремою Прохорова про критерій слабкої компактності* нормована послідовність функцій розподілу $\{F_N(-\infty,x)/r(0),\ N\geq 1\}$ слабко компактна, оскільки верхня межа в критерії нульова вже при $c>\pi$. Тому за означенням *слабкої компактності* для деякої скінченної міри F та для деякої підпослідовності $F_{N_k}\to^W F,\ N_k\to\infty$. Використовуючи цю збіжність у попередній рівності при $N=N_k$, дістанемо сформульоване зображення:

$$\int_{-\pi}^{\pi}\exp(i\lambda n)dF(\lambda)=\lim_{k\to\infty}\int_{-\pi}^{\pi}\exp(i\lambda n)dF_{N_k}(\lambda)=$$
$$\lim_{k\to\infty}(1-|n|/N)^{+}r(n)=r(n),\ \forall n\in\mathbb{Z}\ \boxdot$$

Теорема (теорема Хінчина про спектральне зображення коваріаційної функції).

(а) Функція $r(t)$, $t\in T$, є коваріаційною функцією с.к. неперервного стаціонарного процесу другого порядку тоді й тільки тоді, коли виконується спектральне зображення з теореми Бохнера про додатно визначені функції. Міра F (невласна функція розподілу F) у цьому зображенні називається ***спектральною мірою (спектральною функцією)*** *процесу.*

(б) У випадку $T=\mathbb{Z}$ міра F зосереджена на відрізку $[-\pi,\pi]$.

Зауваження. За теоремою *про формулу обертання для характеристичної функції* справедлива тотожність

$$F(t)-F(s)=(2\pi)^{-1}\lim_{\varepsilon\downarrow 0}\int_s^t dx\int_{-\infty}^{\infty}\exp(-iux-\varepsilon^2u^2)\,r(u)du.$$

Доведення (а) є очевидним наслідком теореми *про властивості коваріаційної функції*, *теореми Бохнера про додатно визначені функції* та зауваження про критерій с.к. неперервності.

Твердження (б) випливає з *теореми Герглотца* $\boxdot$

Вправи

(1) Випадкові величини $(\varepsilon_n, n\in\mathbb{Z})$ такі, що $\mathbf{E}\varepsilon_n=0$, $\mathbf{E}\varepsilon_n\overline{\varepsilon_k}=\delta_{nk}$. Тоді для довільних сталих $(c_k, k=\overline{0,m})$ послідовність $\xi_n=\sum_{k=0}^m c_k\varepsilon_{n-k}$ є стаціонарною другого порядку. Знайти її коваріаційну функцію.

(2) Для квадратично сумованої коваріаційної функції: $\sum_{t\in\mathbb{Z}}|r(t)|^2<\infty$ спектральна функція F має щільність $f(\lambda)=(2\pi)^{-1}\sum_{t\in\mathbb{Z}}\exp(-it\lambda)r(t)$.

(3) Величини $(\eta_n, n\geq 0)$ – незалежні однаково розподілені, а $(\nu(t), t\geq 0)$ – незалежний від них процес Пуассона. Довести, що випадковий процес $\xi_t=\exp(i\eta_{\nu(t)})$ є стаціонарним другого порядку, і знайти його коваріаційну функцію.

(4) Довести, що при $\alpha>0$ процес $\xi_t=\exp(-\alpha t)W(\exp(2\alpha t)), t\in\mathbb{R}_+$, є стаціонарним другого порядку (процесом Орнштейна-Уленбека), та знайти його коваріаційну функцію. Тут $W(s)$ – вінерів процес.

3.19.2. Міри з ортогональними значеннями, стохастичні інтеграли

Означення. *Сім'я випадкових величин* $(\mu(B),\ B \in \mathfrak{B}(\mathbb{R})) \subset L_2(\mathbf{P})$ *називається* випадковою мірою з ортогональними значеннями, *якщо для деякої скінченної міри* m *на* $\mathfrak{B}(\mathbb{R})$ (структурної міри) *виконуються умови*

(а) $\mathbf{E}\mu(A)\,\overline{\mu(B)} = m(A \cap B),\ \forall A, B \in \mathfrak{B}(\mathbb{R}),$

(б) $\mu(A \cup B) = \mu(A) + \mu(B)$ *м.н. для всіх несумісних* $A, B \in \mathfrak{B}(\mathbb{R}),$

(в) $\mathbf{E}\mu(A) = 0.$

Приклад. Нехай $(\zeta_k,\ k \in \mathbb{Z})$ – попарно *некорельовані* випадкові величини з нульовим середнім. Тоді $\mu(A) = \sum_{k\in A} \zeta_k$ є мірою з ортогональними значеннями та зі структурною мірою $m(A) = \sum_{k\in A} \mathbf{E}\,|\zeta_k|^2$:

$$\mathbf{E}\mu(A)\,\overline{\mu(B)} = \sum\nolimits_{k\in A} \sum\nolimits_{j\in B} \mathbf{E}\zeta_k\overline{\zeta_j} = \sum\nolimits_{k=j\in A\cap B} \mathbf{E}\,|\zeta_k|^2 = m(A \cap B),$$

адитивність та центрованість μ очевидні. Зауважимо, що наведений вище стаціонарний процес зі скінченим спектром $\xi_t = \sum_{k=1}^{n} \zeta_k \exp(ikt)$ можна зобразити у вигляді $\zeta_t = \int \exp(i\lambda t)\mu(d\lambda)$, де інтеграл слід розуміти як інтеграл за точковою мірою.

Означення. *Нехай* $(\mu(B),\ B \in \mathfrak{B}(\mathbb{R}))$ *– випадкова міра з ортогональними значеннями та зі структурною мірою* m. *Позначимо через:*

$H_2(\mu) = \aleph(\mu(B),\ B \in \mathfrak{B}(\mathbb{R}))$ *– замкнений лінійний підпростір* $L_2(\mathbf{P})$, *що породжений її значеннями, та через*

$L_2(m)$ *– гільбертів простір комплекснозначних борелєвих функцій, що інтегровні з квадратом за мірою* m : $\|f\|_m^2 \equiv \int |f|^2\, dm < \infty$, *зі скалярним добутком* $(f, g)_m = \int f\overline{g}dm$.

Теорема (про ізометрію просторів H2 і L2). *Нехай випадкова міра з ортогональними значеннями* $(\mu(B),\ B \in \mathfrak{B}(\mathbb{R}))$ *має структурну міру* m. *Існує єдина лінійна ізометрія* $\Im : L_2(m) \to H_2(\mu)$ *така, що*

$$\Im(\mathbb{1}_B) = \mu(B),\ \forall B \in \mathfrak{B}(\mathbb{R}).$$

Доведення. Розглянемо лінійні підпростори

$$H_{01} = \pounds(\mathbb{1}_B, B \in \mathfrak{B}(\mathbb{R})) \subset L_2(m), \quad H_{02} = \pounds(\mu(B), B \in \mathfrak{B}(\mathbb{R})) \subset H_2(\mu).$$

Вказане в теоремі відображення породжує ізометрію H_{01} на H_{02}, оскільки за умовою

$$\|\mathbb{1}_B\|_m^2 = m(B) = \mathbf{E}\mu(B)\overline{\mu(B)} = \|\mu(B)\|^2 = \|\Im(\mathbb{1}_B)\|^2,$$

і внаслідок ортогональності значень для попарно несумісних B_k

$$\left\|\sum c_k 1\!\!\mathrm{I}_{B_k}\right\|_m^2 = \int \left|\sum c_k 1\!\!\mathrm{I}_{B_k}(x)\right|^2 m(dx) =$$
$$\int \sum\nolimits_k c_k 1\!\!\mathrm{I}_{B_k}(x) \sum\nolimits_j \overline{c_j} 1\!\!\mathrm{I}_{B_j}(x)\ m(dx) =$$
$$\int \sum\nolimits_{k,j} c_k\overline{c_j} 1\!\!\mathrm{I}_{B_k\cap B_j}(x) m(dx) = \sum |c_k|^2 m(B_k) =$$
$$\sum\nolimits_{k,j} c_k\overline{c_j}\ \mathbf{E}\mu(B_k)\overline{\mu(B_j)} = \mathbf{E}\left|\sum c_k\mu(B_k)\right|^2 = \left\|\sum c_k\mu(B_k)\right\|^2.$$

За означенням *інтегралу Лебега – Стілтьєса* за мірою m вкладення підпростору $H_{01} \subset L_2(m)$ є щільним, оскільки цей інтеграл збігається з границею інтегралів від відповідних *простих* функцій. За означенням простору $H_2(\mu)$ він є замиканням у $L_2(\mathbf{P})$ лінійної оболонки $H_{02} = \pounds(\mu(B),\ B \in \mathfrak{B}(\mathbb{R}))$, а остання внаслідок адитивності (б) має вигляд

$$H_{02} = \left\{\sum c_k\mu(B_k),\ c_k \in \mathbb{C}, B_k \in \mathfrak{B}(\mathbb{R}), B_k \cap B_j = \varnothing, k \neq j\right\}.$$

Отже, включення $H_{02} \subset H_2(\mu)$ також щільне. Тому твердження теореми випливає з теореми *про продовження ізометрії* ⊡

Означення. *Нехай* $f \in L_2(m)$. Стохастичним інтегралом *Вінера – Дуба від функції* f *за випадковою мірою* μ *з ортогональними значеннями називається ізометрія* $\Im(f) \in H_2(\mu)$, *що існує та єдина за теоремою про ізометрію просторів H2 і L2 і позначається як*

$$\int f\ d\mu \equiv \Im(f).$$

Функція f *називається* спектральною характеристикою *випадкової величини* $\int f\ d\mu$.

За цим означенням існує аналогія з визначенням інтегралу Лебега:

$$\int f\ d\mu = l.i.m_{n\to\infty}\ \sum\nolimits_k c_{nk}\mu(B_{nk}),$$

де сталі c_{nk} та множини B_{nk} входять до визначення простих функцій $f_n(x) = \sum_k c_{nk} 1\!\!\mathrm{I}_{B_{nk}}(x)$, що апроксимують у нормі $L_2(m)$ функцію f.

Наслідок. *Стохастичний інтеграл* є лінійним, неперервним за f у метриці простору $L_2(m)$ та задовольняє умову

$$\mathbf{E}\left(\int f\ d\mu\right)\left(\overline{\int g\ d\mu}\right) = \int f\overline{g}\ dm.$$

Доведення. Лінійність та неперервність є властивостями лінійної *ізометрії*. Остання рівність полягає у збіжності скалярних добутків ізометричних елементів, яка випливає з ізометричної збіжності норм відповідних елементів та з тотожності

$$4(f,g) = \|f+g\|^2 - \|f-g\|^2 + i\,\|f+ig\|^2 - i\,\|f-ig\|^2 \ \boxdot$$

Теорема (про продовження міри з ортогональними значеннями). *Нехай алгебра множин $\mathfrak{A}(\mathbb{R})$ породжує сигма-алгебру $\mathfrak{B}(\mathbb{R})$. Якщо існує система випадкових величин $(\mu(B),\ B \in \mathfrak{A}(\mathbb{R}))$, яка задовольняє при $B \in \mathfrak{A}(\mathbb{R})$ умови (а),(б),(в) в означенні випадкової міри з ортогональними значеннями, то існує така міра з ортогональними значеннями $(\mu^*(B),\ B \in \mathfrak{B}(\mathbb{R}))$, що має ту ж структурну міру m та $\mu^*(B) = \mu(B)$ м.н., для всіх $B \in \mathfrak{A}(\mathbb{R})$.*

Доведення. Позначимо

$$\mathfrak{B}_0 = \{B \in \mathfrak{B}(\mathbb{R}) : \inf(m(A\Delta B),\ A \in \mathfrak{A}(\mathbb{R})) = 0\}.$$

Клас $\mathfrak{B}_0$ містить алгебру $\mathfrak{A}(\mathbb{R})$. Клас $\mathfrak{B}_0$ є замкненим відносно доповнень, скінченних об'єднань та зліченних об'єднань множин, що попарно не перетинаються. Отже, $\mathfrak{B}_0$ є *сигма-алгеброю* і містить $\mathfrak{A}(\mathbb{R})$, тому $\mathfrak{B}_0 = \mathfrak{B}(\mathbb{R})$.

Якщо $B \in \mathfrak{B}(\mathbb{R})$, а $A_n \in \mathfrak{A}(\mathbb{R})$ такі, що $m(A_n\Delta B) \to 0$, то

$$\|\mu(A_n) - \mu(A_m)\|^2 = \mathbf{E}(\mu(A_n) - \mu(A_m))(\overline{\mu(A_n) - \mu(A_m)}) =$$

$$m(A_n) + m(A_m) - 2m(A_n \cap A_m) = m(A_n\Delta A_m) \to 0,\ n.m \to \infty.$$

Отже, послідовність $\mu(A_n)$ фундаментальна в $L_2(\mathbf{P})$, тому існує границя $\mu^*(B) = l.i.m\ \mu(A_n)$. Вона визначена коректно і задовольняє умови (а),(б),(в), оскільки ці умови зберігаються при переході до *с.к.* границі $\boxdot$

3.19.3. Процеси з ортогональними приростами

Означення. *Квадратично інтегровний процес $(\zeta(t),\ t \in T) \subset L_2(\mathbf{P})$ називається процесом з* ортогональними приростами, *якщо:*

(а) $\mathbf{E}\zeta[s,\ t)\overline{\zeta[u,\ v)} = 0$ *при всіх* $s < t \le u < v$,
де приріст процесу $\zeta[a,b) \equiv \zeta(b) - \zeta(a)$,

(б) $\sup_{s\,<\,t}\ \mathbf{E}\,|\zeta[s,t)|^2 < \infty$,

(в) $\mathbf{E}\zeta[s,\ t) = 0$.

Теорема (про спектральну функцію процесу з ортогональними приростами). *Нехай $(\zeta(t),\ t \in T)$ процес із ортогональними приростами. Існує неспадна обмежена функція F така, що*

$$\mathbf{E}\,|\zeta[s,t)|^2 = F(t) - F(s), \qquad \mathbf{E}\zeta(I)\overline{\zeta(J)} = F(I \cap J),$$

для всіх інтервалів I, J, де значення функції від інтервалу слід розуміти як відповідний приріст.

Означення. *Функція F називається* спектральною функцією *процесу* $(\zeta(t),\ t \in T)$.

Доведення. Функція $\mathbf{E}\,|\zeta(I)|^2$ є адитивною на алгебрі об'єднань інтервалів, що не перетинаються: $I \cap J = \varnothing$, оскільки

$$\mathbf{E}\,|\zeta(I \cup J)|^2 = \mathbf{E}(\zeta(I) + \zeta(J))(\overline{\zeta(I) + \zeta(J)}) =$$
$$\mathbf{E}\,|\zeta(I)|^2 + \mathbf{E}\,|\zeta(J)|^2 + 2\,\mathrm{Re}\,\mathbf{E}\zeta(I)\overline{\zeta(J)} = \mathbf{E}\,|\zeta(I)|^2 + \mathbf{E}\,|\zeta(J)|^2\,,$$

отже, є монотонно неспадною функцією інтервалу $I = [s, t)$. Тому з (б) випливає існування монотонної обмеженої функції

$$\lim\nolimits_{s \to -\infty} \mathbf{E}\,|\zeta[s,t)|^2 = F(t),$$

звідки виводиться перше твердження теореми:

$$F(t) + \mathbf{E}\,|\zeta[t,u)|^2 =$$
$$\lim\nolimits_{s \to -\infty} \mathbf{E}\,|\zeta[s,t)|^2 + \mathbf{E}\,|\zeta[t,u)|^2 = \lim\nolimits_{s \to -\infty} \mathbf{E}\,|\zeta[s,u)|^2 = F(u), \quad t < u.$$

Друге твердження після підстановки $I = (I \cap J) \cup (I \backslash J)$ випливає із білінійності *коваріації* та з *ортогональності приростів*:

$$\mathbf{E}\zeta(I)\overline{\zeta(J)} = \mathbf{E}\zeta(I \cap J)\overline{\zeta(J)} + \mathbf{E}\zeta(I \backslash J)\overline{\zeta(J)} = \mathbf{E}\zeta(I \cap J)\overline{\zeta(J)} =$$
$$\mathbf{E}\zeta(I \cap J)\overline{\zeta(J \cap I)} + \mathbf{E}\zeta(I \cap J)\overline{\zeta(J \backslash I)} = \mathbf{E}\zeta(I \cap J)\overline{\zeta(J \cap I)} = F(I \cap J)\ \boxdot$$

Приклад. Якщо $(\eta_k,\ k \in \mathbb{Z})$ – послідовність попарно некорельованих випадкових величин із нульовим *середнім*, то процес $\zeta(t) = \sum_{k\,<\,t} \eta_k$ є процесом із *ортогональними приростами*, причому $F(t) = \sum_{k\,<\,t} \mathbf{E}\,|\eta_k|^2$.

Теорема (про випадкову міру, породжену випадковим процесом із ортогональними приростами). *Для процесу з ортогональними приростами $(\zeta(t),\ t \in T)$ існує єдина випадкова міра з ортогональними значеннями μ на $\mathfrak{B}(\mathbb{R})$ така, що*

$$\mu([s,t)) = \zeta[s,t),\ \forall s < t.$$

Відповідна структурна міра m є мірою Лебега – Стілтьєса, що породжується спектральною мірою процесу:

$$m([s,t)) = \mathbf{E}\,|\mu([s,t))|^2 = \mathbf{E}\,|\zeta(t) - \zeta(s)|^2 = F(t) - F(s).$$

Доведення випливає з теореми *про продовження міри з ортогональними значеннями*. Міра μ для $A = \cup_{k=1}^n [a_k, b_k) \in \mathfrak{A}(\mathbb{R})$ визначається як

$$\mu(A) = \sum\nolimits_{k=1}^n \zeta[a_k, b_k).$$

Лінійність та центрованість μ очевидні. *Структурна міра* на алгебрі $\mathfrak{A}(\mathbb{R})$ обчислюється внаслідок теореми *про спектральну функцію процесу з ортогональними приростами*:

$$\mathbf{E}\mu(A)\,\overline{\mu(B)} = \mathbf{E}\sum\nolimits_{I\subset A}\sum\nolimits_{J\subset B}\zeta(I)\overline{\zeta(J)} =$$

$$\sum\nolimits_{I\subset A}\sum\nolimits_{J\subset B}F(I\cap J) = \sum\nolimits_{I=J\subset A\cap B}F(I) = F(A\cap B)\ \boxdot$$

Означення. *Нехай* $f \in L_2(F)$. Стохастичним інтегралом *від функції* f *за процесом* ζ *із ортогональними приростами називається інтеграл за відповідною випадковою мірою з ортогональними значеннями* μ :

$$\int f\,d\zeta \equiv \int f\,d\mu.$$

Зауваження. За властивостями інтегралу за випадковою мірою інтеграл за процесом є лінійним та неперервним за f і задовольняє умову

$$\mathbf{E}\left(\int f\,d\zeta\right)\left(\overline{\int g\,d\zeta}\right) = \int f\,\overline{g}\,dF.$$

3.19.4. Спектральне зображення стаціонарного процесу

Теорема (теорема Карунена про зображення стаціонарного процесу). *Нехай* $(\xi_t,\ t\in T)$ *– стаціонарний процес із нульовим середнім:* $\mathbf{E}\xi_t = 0$, *і коваріаційною функцією* $r(t)$, *що має спектральну міру* F :

$$r(t) = \int_{\mathbb{R}} \exp(i\lambda t)\,dF(\lambda).$$

Розглянемо гільбертів підпростір, що породжений процесом ξ :

$$H_2(\xi) = \aleph(\xi_t,\ t\in T) \subset\ L_2(\mathbf{P}).$$

Існує процес із ортогональними приростами $(\zeta(\lambda),\ \lambda\in\mathbb{R})$ *зі значеннями в* $H_2(\xi)$, *із тією ж спектральною мірою* F :

$$\mathbf{E}\,|\zeta[s,t)|^2 = F(t) - F(s),$$

та такий, що має місце спектральне зображення

$$\xi_t = \int_{\mathbb{R}} \exp(i\lambda t)\,d\zeta(\lambda),\ \forall t\in T.$$

Означення. *Міра* F *називається* спектральною мірою *процесу* (ξ_t). *Якщо ця міра абсолютно неперервна відносно міри Лебега, то її щільність* f *називається* спектральною щільністю *процесу* (ξ_t). *Випадковий* *процес* $(\zeta(\lambda), \lambda \in \mathbb{R})$ *називається* **стохастичним спектральним процесом** *для стаціонарного процесу* (ξ_t).

Зауваження. У випадку $T = \mathbb{Z}$ *спектральна міра* F та процес $\zeta(\lambda)$ зосереджені на відрізку $[-\pi, \pi]$.

Доведення. Позначимо через $L_2(F)$ гільбертів простір комплекснозначних борелєвих функцій, що інтегровні з квадратом за мірою F :

$$L_2(F) = \{f : \|f\|_F^2 \equiv \int |f|^2 \, dF < \infty\}.$$

Розглянемо його лінійний підпростір

$$H_{01} = \pounds(\exp(it\lambda), \ t \in T) \subset L_2(F).$$

За теоремою Вейєрштраса підпростір H_{01} щільний у $L_2(F)$.

За означенням $H_2(\xi)$ також щільно вкладеним є підпростір

$$H_{02} = \pounds(\xi_t, \ t \in T) \subset H_2(\xi).$$

Розглянемо відображення H_{01} на H_{02}, що задається формулою

$$\Im(\exp(it\cdot)) = \xi_t, \ t \in T,$$

для твірних елементів $\exp(it\lambda)$ та продовжується за лінійністю до ізоморфізму на інші елементи H_{01}.

Оскільки за означенням *коваріаційної функції* та *спектральної міри*

$$\|\sum c_k \exp(i\lambda t_k)\|_F^2 = \int |\sum_k c_k \exp(i\lambda t_k)|^2 \, dF(\lambda) =$$
$$\int \sum_k c_k \exp(i\lambda t_k) \sum_j \overline{c_j} \exp(-i\lambda t_j) \ dF(\lambda) =$$
$$\int \sum_{k,j} c_k \overline{c_j} \exp(i\lambda(t_k - t_j)) \ dF(\lambda) = \sum_{k,j} c_k \overline{c_j} \ r(t_k - t_j) =$$
$$\sum_{k,j} c_k \overline{c_j} \ \mathbf{E}\xi(t_k)\overline{\xi(t_j)} = \mathbf{E} \sum_k c_k \xi(t_k) \sum_j \overline{c_j \xi(t_j)} =$$
$$\mathbf{E} \left|\sum c_k \xi(t_k)\right|^2 = \|\sum c_k \xi(t_k)\|^2 = \|\Im\left(\sum c_k \exp(i\lambda t_k)\right)\|^2,$$

то відображення $\Im$ є *ізометрією* H_{01} на H_{02}.

За теоремою *про продовження ізометрії* існує єдине продовження до ізометрії $\Im$ з $H_1 = L_2(F)$ на $H_2 = H_2(\xi)$.

З обмеженості F виводимо, що $\mathbb{1}_{(-\infty, \, t)} \in L_2(F)$. Позначимо

$$\zeta(t) \equiv \Im(\mathbb{1}_{(-\infty, \, t)}) \in H_2(\xi).$$

Тоді за лінійністю $\Im(\mathbb{1}_{[s,t)}) = \Im(\mathbb{1}_{[-\infty,t)} - \mathbb{1}_{(-\infty,s)}) = \zeta(t) - \zeta(s) = \zeta[s,t)$ і внаслідок ізометрії отримуємо

$$\mathbf{E}\zeta[s, \ t)\overline{\zeta[u, \ v)} = \ (\mathbb{1}_{[s,t)}, \mathbb{1}_{[u,v)})_F = 0, \quad \forall s < t \le u < v,$$
$$\mathbf{E} \left|\zeta[s,t)\right|^2 = \left\|\mathbb{1}_{[s,t)}\right\|_F^2 = F(t) - F(s).$$

Оскільки $\mathbf{E}\xi(t) = 0$ і за *нерівністю Коші* математичне сподівання є неперервною функцією на $L_2(\mathbf{P})$, то $\mathbf{E}\eta = 0$ для всіх $\eta \in H_2(\xi)$. Зокрема, $\mathbf{E}\zeta(t) = 0$.

Отже, $\zeta(t)$ – випадковий процес із *ортогональними приростами* та зі *спектральною мірою* F.

Розглянемо лінійний підпростір

$$L_2^0(F) \equiv \left\{ g \in L_2(F) : \Im(g) = \int g(\lambda) d\zeta(\lambda) \right\}.$$

Цей простір замкнений в нормі $L_2(F)$ через неперервність *ізометрії* $\Im$ та неперервність *стохастичного інтегралу.*

Для індикаторної функції інтервалу $g(\lambda) = 1\!\!\mathrm{I}_{[s,t)}(\lambda)$ з лінійності ізометрії $\Im$ отримуємо тотожність

$$\Im(g) = \zeta[s,t) = \int 1\!\!\mathrm{I}_{[s,t)}(\lambda) d\zeta(\lambda) = \int g(\lambda) d\zeta(\lambda).$$

тобто $g \in L_2^0(F)$. Тому за лінійністю простору $L_2^0(F)$ для кожної множини $A \in \mathfrak{A}[\mathbb{R}]$ має місце включення $1\!\!\mathrm{I}_A \in L_2^0(F)$. Далі, для довільної множини $B \in \mathfrak{B}[\mathbb{R}]$ знайдеться послідовність $A_n \in \mathfrak{A}[\mathbb{R}]$ така, що $F(A_n \Delta B) \to 0$, звідки $1\!\!\mathrm{I}_{A_n} \to 1\!\!\mathrm{I}_B, n \to \infty$. Отже, функція $1\!\!\mathrm{I}_B$ та всі *прості* функції належать простору $L_2^0(F)$.

Оскільки за означенням *інтеграла Лебега* клас простих функцій щільний у $L_2(F)$, а простір $L_2^0(F)$ замкнений, то $L_2^0(F) = L_2(F)$.

Підстановкою у отриману тотожність $\Im(g) = \int g(\lambda) d\zeta(\lambda)$ гармоніки $g(\lambda) = \exp(i\lambda t)$ дістанемо *спектральне зображення* для процесу ξ_t :

$$\xi_t = \Im(\exp(it\lambda)) = \int_{\mathbb{R}} \exp(i\lambda t)\, d\zeta(\lambda),\ \forall t \in T.$$

Справедливість зауваження випливає з *теореми Хінчина про спектральне зображення коваріаційної функції*, (б) $\boxdot$

Вправи

(1) Довести, що вінерів процес $(w(t), t \in [0,a])$ є процесом з ортогональними приростами. Знайти його спектральну функцію.

(2) Строго стаціонарна послідовність $(\xi_n, n \in \mathbb{Z})$ центрована: $\mathbf{E}\xi_n = 0.$та має стохастичну спектральну міру $(\zeta(d\lambda))$. Довести середньоквадратичну збіжність $n^{-1} \sum_{k=0}^{n-1} \xi_n \to^{L_2} \zeta(\{0\}), n \to \infty$.

(3) Строго стаціонарна послідовність $(\xi_n, n \in \mathbb{Z})$ така, що виконуються умови $\mathbf{E}\xi_n = 0$, та $\mathbf{E}\left|\sum_{k=0}^{n-1} \xi_k\right|^2 = O(n^{2-\delta}), n \to \infty$, при деякому $\delta > 0$. Довести, що $n^{-1} \sum_{k=0}^{n-1} \xi_k \to^{P1} 0, n \to \infty$.

(4) Нехай $(\zeta_n, n \geq 1)$ – незалежні у сукупності стандартні нормальні величини. (а) Якщо $(\varphi_n(t), n \geq 1)$ – ортонормований базис у $L^2([0,1])$, то випадковий процес $w(t) = \sum_{n\geq 1} \zeta_n \int_0^t \varphi_n(s)ds$ є вінерівським процесом. (б) Сума збіжного у середньоквадратичному на $[0, 2\pi)$ ряду $w(t) = \sum_{n\in\mathbb{Z}} \zeta_n(1 - \exp(int))/in$ є вінерівським процесом на $[0, 2\pi)$ (при $n = 0$ коефіцієнт дорівнює t).

(5) Розклад Карунена – Лоева – Пугачова. Нехай $\xi(t), t \in [0,1]$ с.к. неперервний дійсний випадковий процес з нульовим середнім та коваріаційною функцією $r(t,s) = \mathbf{E}\xi(t)\xi(s)$. Ця функція неперервна та інтегровна з квадратом на $[0,1]^2$. Тому інтегральне рівняння Фредгольма у просторі $L_2[0,1]$:

$$\varphi(t) = \lambda \int_0^1 r(t,s)\varphi(s)ds$$

має для деяких $\lambda = \lambda_n > 0$ розв'язки $\varphi_n \in L_2[0,1], n \geq 1$, що утворюють ортонормований базис. Зокрема, справедливе зображення

$$r(t,s) = \sum\nolimits_{n\geq 1} \lambda_n^{-1}\varphi_n(t)\varphi_n(s).$$

Суму справа можна розглядати як спектральний розклад за лічильною мірою m, що має одиничні стрибки на $\mathbb{N}$. Застосувавши до цього розкладу теорему Карунена, побудуйте випадковий процес $\zeta(t)$ на $[0,1]$ з ортогональними приростами та зі спектральною функцією m. Позначимо $\zeta_n = \zeta(n+0) - \zeta(n-0)$ величини ненульових стрибків цього процесу. Тоді з теореми Карунена отримуємо зображення з попарно ортогональними та нормованими $\zeta_n \in L_2(\mathbf{P})$:

$$\xi(t) = \sum\nolimits_{n\geq 1} \lambda_n^{-1/2}\varphi_n(t)\ \zeta_n$$

Зокрема, для вінерівського процесу $w(t)$ маємо

$$r(t,s) = \min(t,s),\ \lambda_n = \pi^2(n-1/2)^2, \varphi_n(t) = \sqrt{2}\sin(n-1/2)\pi t,$$

$$w(t) = \sqrt{2}\sum\nolimits_{n\geq 1} \zeta_n\ \pi^{-1}(n-1/2)^{-1}\sin(n-1/2)\pi t.$$

3.20. Оптимальний лінійний прогноз стаціонарних послідовностей

За допомогою *спектрального зображення* можна розв'язати задачу знаходження лінійного прогнозу з найменшою дисперсією.

3.20.1. Лінійний прогноз у гільбертовому просторі

Теорема (про лінійний прогноз у гільбертовому просторі). *Нехай H – деякий гільбертів простір, послідовність $(\xi_k, k \in \mathbb{Z}) \subset H$ породжує замкнений лінійний підпростір*

$$H_0 = \aleph(\xi_k,\ k \in \mathbb{Z}),$$

а вектор $\xi \in H$. Тоді

(а) існує єдина **проекція** *ξ на H_0 :*

$$\Pi_0\xi \equiv \arg\min{}_{\eta \in H_0}\ \|\xi - \eta\|^2,$$

(б) проекція є ортогональною: $(\xi - \Pi_0\xi,\ \eta) = 0,\ \forall\eta \in H_0$,

(в) елемент $\zeta \in H_0$ дорівнює $\Pi_0\xi$ тоді й тільки тоді, коли

$$(\xi - \zeta,\ \xi_k) = 0,\ \forall k \in \mathbb{Z},$$

(г) оператор Π_0 є лінійним і ортогональним: $\Pi_0^2 = \Pi_0$.

Доведення

(а) Позначимо лінійні підпростори

$$L_n = \pounds(\xi_k,\ |k| \le n),\ n \ge 1.\ L_\infty = \cup L_n.$$

За означенням $H_0 = \aleph(L_\infty)$, тому елемент $\zeta \in H_0$ дорівнює $\Pi_0\xi$ тоді й тільки тоді, коли

$$\|\xi - \zeta\|^2 \le \|\xi - \eta\|^2,\ \forall\eta \in L_\infty,$$

оскільки за неперервністю норми дана нерівність подовжується з L_∞ на весь підпростір H_0.

Оскільки простір L_n скінченновимірний, то існує єдиний вектор

$$\eta_n = \arg\min{}_{\eta \in L_n}\ \|\xi - \eta\|^2.$$

Дійсно, за процедурою Грама – Шмідта у просторі L_n існує ортонормований базис $\{e_1, ..., e_d\}$. Тому вектор

$$\eta_n = \sum\nolimits_{k=1}^{d} (\xi, e_k) e_k$$

є шуканим, причому $\xi - \eta_n \perp L_n$

Розглянемо такі сталі

$$a = \|\xi\|^2,\ \varepsilon_n = \|\xi - \eta_n\|^2.$$

Оскільки $L_n \subset L_m$ при $m > n$, то

$$\varepsilon_n \downarrow \varepsilon \equiv \lim \varepsilon_n,\quad \xi - \eta_m \perp \eta_n.$$

З $\xi = \xi - \eta_n + \eta_n$ отримуємо також рівності

$$(\xi,\ \eta_n) = \|\eta_n\|^2 = \|\xi\|^2 - \|\xi - \eta_n\|^2 = a - \varepsilon_n.$$

Тому

$$\|\eta_n - \eta_m\|^2 = \|(\eta_n - \xi) - (\eta_m - \xi)\|^2 = \|\eta_n - \xi\|^2 +$$
$$\|\eta_m - \xi\|^2 - 2(\eta_n - \xi, \eta_m - \xi) = \varepsilon_n + \varepsilon_m + 2(\xi, \eta_m - \xi) =$$
$$\varepsilon_n + \varepsilon_m + 2(a - \varepsilon_m) - 2a = \varepsilon_n - \varepsilon_m \to 0.$$

Отже, послідовність η_n фундаментальна й має границю $\zeta \equiv \lim \eta_n$, причому $\|\xi - \zeta\|^2 = \lim \|\xi - \eta_n\|^2 = \varepsilon$.

Для довільного $\eta \in L_\infty$ існує n таке, що $\eta \in L_n$. Тому

$$\|\xi - \eta\|^2 \ge \varepsilon_n \ge \varepsilon = \|\xi - \zeta\|^2, \quad \forall \eta \in L_\infty.$$

Звідси внаслідок твердження, що наведене на початку доведення, випливає *існування* проекції $\Pi_0 \xi = \zeta$.

Для доведення *єдиності* зауважимо, що при наявності двох точок мінімуму $\zeta_i, i = 1, 2$, квадратична форма

$$\|\xi - t\zeta_1 - (1-t)\zeta_2\|^2 = \|\xi - \zeta_2\|^2 - 2t\ (\xi - \zeta_2,\ \zeta_1 - \zeta_2) + t^2 \|\zeta_1 - \zeta_2\|^2$$

набувала б одного й того ж самого найменшого значення для різних значень аргументу $t = 0, 1$, звідки $t^2 \|\zeta_1 - \zeta_2\|^2 = 0$.

Твердження (б) випливає з ортогональності $\xi - \eta_n \perp L_k\ \forall n \ge k$. Оскільки для довільного $\eta \in L_\infty$ знайдеться k таке, що $\eta \in L_k$, то $(\xi - \eta_n, \eta) = 0$ починаючи з деякого n, звідки дістанемо $(\xi - \Pi_0\xi, \eta) = 0$ при $n \to \infty$. Нарешті, за неперервністю скалярного добутку цю рівність подовжуємо на $H_0 = \aleph(L_\infty)$.

Необхідність (в) є очевидним наслідком (б).

Для доведення *достатності* у (в) припустимо, що $(\xi - \zeta,\ \xi_k) = 0$, для всіх $k \in \mathbb{Z}$. Тоді за лінійністю та неперервністю $(\xi - \zeta,\ \eta) = 0,\ \forall \eta \in H_0$. Отже, для довільного $\eta \in H_0$

$$\|\xi - \eta\|^2 = \|\xi - \zeta\|^2 + \|\zeta - \eta\|^2 + 2(\xi - \zeta,\ \zeta - \eta) \ge \|\xi - \zeta\|^2,$$

де рівність $(\xi - \zeta,\ \zeta - \eta) = 0$ випливає з включення $\zeta - \eta \in H_0$.

Лінійність (г) є наслідком (в): якщо $\zeta = \Pi_0\xi,\ \eta = \Pi_0\varkappa$, то

$$(\xi + \varkappa - \zeta - \eta,\ \xi_k) = (\xi - \zeta,\ \xi_k) + (\varkappa - \eta,\ \xi_k) = 0,\ \forall k \in \mathbb{Z},$$

звідки $\Pi_0\xi + \Pi_0\varkappa = \zeta + \eta = \Pi_0(\xi + \varkappa)$.

Ортогональність (д) виводиться з тотожності $\Pi_0\eta = \eta,\ \eta \in H_0$, за якою $\Pi_0^2\eta = \Pi_0(\Pi_0\eta) = \Pi_0\eta = \eta$ $\boxdot$

3.20.2. Спектральний розв'язок задачі лінійного прогнозу

Нехай $(\xi_n,\ n \in \mathbb{Z})$ – *стаціонарна* послідовність із нульовим середнім, F – її *спектральна міра* на $[-\pi, \pi]$. Розглянемо гільбертові підпростори

$$H_2^-(\xi) \equiv \aleph(\xi_n, n \le 0) \subset H_2(\xi) \equiv \aleph(\xi_n, n \in \mathbb{Z}) \subset L_2(\mathbf{P}).$$

Позначимо через

$$\widehat{\xi}_1 = \Pi_0 \xi_1 \equiv \arg\min\ _{\eta \in H_2^-(\xi)} \|\xi_1 - \eta\|^2 \in H_2^-(\xi)$$

оптимальний лінійний прогноз значення ξ_1 за спостереженнями з $H_2^-(\xi)$. Він існує та коректно визначений за теоремою *про лінійний прогноз у гільбертовому просторі*.

Визначимо $\mathbb{Z}_\pm = \{n \in \mathbb{Z} : n \gtreqless 0\}$ та уведемо підпростори

$$L_2^\pm(F) = \aleph_F(\exp(in\lambda), n \in \mathbb{Z}_\pm),$$

де $\aleph_F$ – оператор замикання лінійної оболонки у просторі $L_2(F)$.

Як і вище, символом $\Im$ позначатимемо *ізометрію* $\Im : L_2(F) \to H_2(\xi)$, що визначена в *теоремі Карунена про зображення стаціонарного процесу* як $\Im(\exp(it\cdot)) = \xi_t$ та породжена стохастичним інтегралом за відповідним процесом із *ортогональними приростами* ζ :

$$\Im(g) = \int_{-\pi}^{\pi} g(\lambda) d\zeta(\lambda).$$

Нагадаємо, що функція g називається *спектральною характеристикою* випадкової величини $\Im(g)$.

Визначимо також підпростіри

$$H_2^\pm(\xi) \equiv \aleph(\xi_n, n \in \mathbb{Z}_\pm) \subset H_2(\xi) \subset L_2(\mathbf{P}).$$

Лема (про звуження ізометрії). *Кожне зі звужень ізометрії* $\Im : L_2^\pm(F) \to H_2^\pm(\xi)$ *є ізометрією.*

Доведення очевидне, оскільки за означенням $\Im(\exp(in\cdot)) = \xi_n$, а послідовності $(\exp(in\lambda), n \in \mathbb{Z}_\pm)$ є базисами в $L_2^\pm(F)$ $\boxdot$

Теорема (про критерій оптимальності спектрального зображення). *Функція* $\widehat{g} \in L_2(F)$ *є спектральною характеристикою оптимальному прогнозу:* $\Im(\widehat{g}) = \widehat{\xi}_1$, *тоді й тільки тоді, коли* $\widehat{g} \in L_2^-(F)$ *і*

$$\int_{-\pi}^{\pi} (\exp(i\lambda) - \widehat{g}(\lambda))\ \overline{\exp(in\lambda)}\ dF(\lambda) = 0, \quad \forall n \le 0.$$

Доведення. З леми *про звуження ізометрії* виводимо, що включення $\widehat{g} \in L_2^-(F)$ еквівалентне включенню $\widehat{\xi}_1 \in H_2^-(\xi)$.

З теореми *про лінійний прогноз у гільбертовому просторі* випливає, що величина $\zeta \in H_2^-(\xi)$ є оптимальним лінійним прогнозом: $\zeta = \widehat{\xi}_1$, тоді й тільки тоді, коли

$$\mathbf{E}(\xi_1 - \zeta)\overline{\xi_n} \equiv (\xi_1 - \zeta,\ \xi_n) = 0,\ \forall n \le 0.$$

Оскільки має місце *ізометрія* $\Im(\exp(i\lambda n)) = \xi_n$, і $\Im(\widehat{g}) = \widehat{\xi}_1$, то відповідні скалярні добутки у $H_2(\xi)$ та $L_2(F)$ збігаються:

$$(\xi_1 - \zeta,\ \xi_n) = \int_{-\pi}^{\pi} (\exp(i\lambda) - \widehat{g}(\lambda))\ \overline{\exp(in\lambda)}\ dF(\lambda).$$

Тому наведений критерій еквівалентний тотожності теореми ⊡

Позначимо через L_2, $L_2^{\pm}$ аналогічні до $L_2(F)$, $L_2^{\pm}(F)$ простори функцій для випадку, коли F – *міра Лебега* на $[-\pi, \pi]$. Як відомо,

$$L_2 = \left\{ g(\lambda) = \sum\nolimits_{n \in \mathbb{Z}} c_n \exp(in\lambda) : \sum |c_n|^2 < \infty \right\},$$

$$L_2^{\pm} = \left\{ g(\lambda) = \sum\nolimits_{n \in \mathbb{Z}_{\pm}} c_n \exp(in\lambda) : \sum\nolimits_{n \in \mathbb{Z}_{\pm}} |c_n|^2 < \infty \right\}.$$

Теорема (про ізоморфізм спектральних просторів). *Нехай виконується умова:*

(L) спектральна міра F має щільність f таку, що

$$0 < \inf f,\ \sup f < \infty.$$

Тоді $$L_2(F) = L_2,\quad L_2^{\pm}(F) = L_2^{\pm}.$$

Доведення. Базисні елементи у всіх трьох пар просторів однакові: це послідовності функцій $\exp(in\lambda)$, де $n \in \mathbb{Z}$ або $\mathbb{Z}_{\pm}$. Тому тотожність просторів випливає з еквівалентності норм, за якими обчислюється замикання:

$$\|g\|_F^2 = \int |g|^2 f(\lambda) d\lambda \le \sup f\ \int |g|^2 d\lambda = \sup f\ \|g\|^2,$$

$$\|g\|_F^2 = \int |g|^2 f(\lambda) d\lambda \ge \inf f\ \int |g|^2 d\lambda = \inf f\ \|g\|^2 \ \boxdot$$

Теорема (про побудову оптимального прогнозу). *Нехай виконується умова (L). Функція $\widehat{g} \in L_2^-$ є спектральною характеристикою оптимального прогнозу: $\Im(\widehat{g}) = \widehat{\xi}_1$ тоді й тільки тоді, коли знайдуться функції $f_{\pm} \in L_2^{\pm}$ такі, що для спектральної щільності f має місце факторизація*

$$f(\lambda) = \frac{f_+(\lambda)}{f_-(\lambda)},\ \lambda \in [-\pi, \pi],$$

причому знаменник має вигляд $f_-(\lambda) = 1 - \exp(-i\lambda)\widehat{g}(\lambda)$.

Зауваження. З включення $\widehat{g} \in L_2^-$ випливає, що функція $\exp(-i\lambda)\widehat{g}(\lambda)$ має ряд Фур'є зі строго від'ємними номерами гармонік, тому останнє зображення для функції f_- означає, що її нульовий коефіцієнт Фур'є має дорівнювати 1. Ця умова еквівалентна включенню $\exp(i\lambda)(1 - f_-(\lambda)) \equiv \widehat{g}(\lambda) \in L_2^-$.

Доведення. *Необхідність.* Нехай $\widehat{\xi}_1$ – оптимальний прогноз із відповідною *спектральною характеристикою* $\widehat{g}$. З теореми *про критерій оптимальності спектрального зображення* виводимо, що $\widehat{g} \in L_2^-(F) = L_2^-$ і

$$\int_{-\pi}^{\pi} (\exp(i\lambda) - \widehat{g}(\lambda)) \exp(-in\lambda)\ f(\lambda) d\lambda =$$
$$\int_{-\pi}^{\pi} \exp(-i(n-1)\lambda)\ f_-(\lambda) f(\lambda) d\lambda = 0,\ \forall n \le 0.$$

Очевидно, що $f_-(\lambda) \equiv 1 - \exp(-i\lambda)\widehat{g}(\lambda) \in L_2^-$. Визначимо функцію $f_+(\lambda) \equiv f_-(\lambda)\ f(\lambda)$. Тоді $f_+ \in L_2$, тому що $f_- \in L_2$, а f обмежена за умовою (L). Отже, $f_+(\lambda) = \sum_{n\in\mathbb{Z}} f_n \exp(in\lambda)$ з $\sum_{n\in\mathbb{Z}} |f_n|^2 < \infty$. Оскільки послідовність функцій $\exp(in\lambda)$ утворює базис у L_2, то з отриманої вище тотожності для добутку $f_-(\lambda)f(\lambda)$ випливає, що

$$f_+(\lambda) = \sum\nolimits_{n\ge 0} f_n \exp(in\lambda)\ \in L_2^+.$$

Достатність. Якщо для *спектральної характеристики* $\widehat{g} \in L_2^-(F)$ та функції $f_\pm \in L_2^\pm$ справджуються умови теореми, то за *теоремою Карунена про зображення стаціонарного процесу* величина $\widehat{\xi}_1 \equiv \Im(\widehat{g})$ задовольняє умову (б) теореми *про лінійний прогноз у гільбертовому просторі* $\boxdot$

Теорема (про обчислення оптимального прогнозу). *Нехай виконується умова (L), а $f_\pm$ – чисельник і знаменник у факторизації спектральної щільності f з теореми* про побудову оптимального прогнозу.

Оптимальний прогноз для ξ_1 має вигляд

$$\widehat{\xi}_1 = \sum\nolimits_{n<0} b_n \xi_{n+1},$$

де коефіцієнти b_n визначаються з розкладу

$$f_-(\lambda) = 1 - \exp(-i\lambda)\widehat{g}(\lambda) = 1 - \sum\nolimits_{n<0} b_n \exp(in\lambda).$$

Похибка прогнозу дорівнює

$$\mathbf{E}\left|\xi_1 - \widehat{\xi}_1\right|^2 = \int_{-\pi}^{\pi} f_+(\lambda) d\lambda.$$

Доведення. За теоремою *про побудову оптимального прогнозу*

$$\widehat{g}(\lambda) = \exp(i\lambda)(1 - f_-(\lambda)) = \sum\nolimits_{n<0} b_n \exp(i\lambda + in\lambda).$$

За визначенням *ізометрії* з *теореми Карунена про зображення стаціонарного процесу*

$$\widehat{\xi}_1 = \int_{-\pi}^{\pi} \widehat{g}(\lambda)\, d\zeta(\lambda) =$$

$$\int_{-\pi}^{\pi} \sum\nolimits_{n<0} b_n \exp(i(n+1)\lambda)\, d\zeta(\lambda) = \sum\nolimits_{n<0} b_n \xi_{n+1}.$$

Для обчислення похибки скористаємося означенням *ізометрії* та теоремою *про побудову оптимального прогнозу:*

$$\mathbf{E}\left|\xi_1 - \widehat{\xi}_1\right|^2 = \mathbf{E}\left(\xi_1 - \widehat{\xi}_1\right)\left(\overline{\xi_1 - \widehat{\xi}_1}\right) = \mathbf{E}\left(\xi_1 - \widehat{\xi}_1\right)\overline{\xi_1} =$$

$$\int_{-\pi}^{\pi} (\exp(i\lambda) - \widehat{g}(\lambda))\overline{\exp(i\lambda)} f(\lambda) d\lambda = \int_{-\pi}^{\pi} f_-(\lambda) f(\lambda) d\lambda = \int_{-\pi}^{\pi} f_+(\lambda) d\lambda \boxdot$$

Теорема (про натуральну факторизацію спектральної щільності). *Нехай виконується умова:*

(R) логарифм спектральної щільності $f(\lambda)$ *розкладається в абсолютно збіжний ряд Фур'є:*

$$\ln f(\lambda) = \sum\nolimits_{k\in\mathbb{Z}} a_k \exp(ik\lambda),\ A \equiv \sum\nolimits_{k\in\mathbb{Z}} |a_k| < \infty.$$

Тоді функції

$$f_+(\lambda) \equiv \exp\left(\sum\nolimits_{k\geq 0} a_k \exp(ik\lambda)\right),$$

$$f_-(\lambda) \equiv \exp\left(-\sum\nolimits_{k<\ 0} a_k \exp(ik\lambda)\right)$$

задовольняють умови теореми про побудову оптимального прогнозу, а вказані у ній елементи факторизації $f_\pm(\lambda)$ *визначені однозначно.*

Зокрема, спектральна характеристика оптимального прогнозу $\widehat{\xi}_1$ *дорівнює* $\widehat{g}(\lambda) = \exp(i\lambda)(1 - f_-(\lambda))$, *та обчислюється з використанням формул обертання* $a_k = (2\pi)^{-1} \int_{[-\pi,\pi]} \exp(-ik\lambda) \ln f(\lambda) d\lambda$. *Похибка цього прогнозу має вигляд (теорема Сеге – Колмогорова):*

$$\mathbf{E}\left|\xi_1 - \widehat{\xi}_1\right|^2 = 2\pi \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln f(\lambda) d\lambda\right).$$

Доведення. Зауважимо, що з (R) випливає умова (L):

$$\exp(-A) = \exp\left(-\sum |a_k|\right) \leq f(\lambda) \leq \exp\left(\sum |a_k|\right) = \exp(A),$$

а також обмеженість сум

$$f_\pm(\lambda) = \sum_{m\ge 0}\left(\pm\sum_{k\ (\ge)(<)\ 0} a_k \exp(ik\lambda)\right)^m/m! \le$$
$$\sum_{m\ge 0} A^m/m! = \exp(A) < \infty,$$

звідки $f_\pm \in L_2^\pm$, причому $f_-(\lambda)$ має вигляд $1-\sum_{n<0} b_n \exp(in\lambda)$.

Єдиність визначення компонент факторизації випливає з обмеженості знизу елементів канонічної факторизації. Дійсно, якщо функції $h_\pm$ також задовольняють вказане означення, то $h_+(\lambda)/f_+(\lambda) = h_-(\lambda)/f_-(\lambda)$ для всіх $\lambda \in \mathbb{R}$. З обмеженості: $\inf f_\pm > 0$ випливає включення $h_\pm/f_\pm \in L_2$, а з означення функцій $f_\pm, h_\pm$ виводимо, що $h_\pm/f_\pm \in L_2^\pm$. Оскільки $L_2^+ \cap L_2^- = \{c, c \in \mathbb{R}\}$, то з наведеної тотожності випливає пропорційність $h_\pm = cf_\pm$ для деякої сталої c. З умови на коефіцієнт при нульовій гармоніці для h_-, f_- обчислюємо $c = 1$.

Формула для обчислення коефіцієнтів a_k є наслідком попарної ортогональності гармонічних функцій на $L_2[-\pi,\pi]$, вираз для $\widehat{g}$ наведений у теоремі *про обчислення оптимального прогнозу*. З тієї ж теореми:

$$\mathbf{E}\left|\xi_1 - \widehat{\xi}_1\right|^2 = \int_{-\pi}^{\pi} f_+(\lambda)d\lambda = \int_{-\pi}^{\pi} \exp\left(\sum_{k\ge 0} a_k \exp(ik\lambda)\right) d\lambda =$$
$$\int_{-\pi}^{\pi} \exp(a_0)\left(1+\sum_{m\ge 1}\left(\sum_{k>0} a_k \exp(ik\lambda)\right)^m / \ m!\right) d\lambda = 2\pi \exp(a_0) \boxdot$$

3.20.3. Приклади обчислення оптимального прогнозу

1. **Процес рухомого середнього** (MA: Moving Average). Нехай *спектральна щільність стаціонарної* послідовності $(\xi_n, n \in \mathbb{Z})$ має вигляд

$$f(\lambda) = (2\pi)^{-1}\left|P(\exp(-i\lambda))\right|^2,\ P(z) = \sum_{k=0}^{m} p_k z^k,$$

де дійсний поліном P не має коренів на одиничному колі і $p_0 = 1$.

Позначимо через ζ – *стохастичний спектральний процес* для (ξ_n), що має *ортогональні прирости*. Тоді випадкові величини

$$\varepsilon_n = \int_{-\pi}^{\pi} \frac{\exp(in\lambda)}{P(\exp(-i\lambda))}\, d\zeta(\lambda),\ n \in \mathbb{Z},$$

ортонормовані:

$$\mathbf{E}\varepsilon_n\overline{\varepsilon_k} = \int_{-\pi}^{\pi} \frac{\exp(i(n-k)\lambda)}{\left|P(\exp(-i\lambda))\right|^2}\, f(\lambda)d\lambda = \delta_{nk}.$$

Обчислимо

$$\sum_{k=0}^{m} p_k \varepsilon_{n-k} = \int_{-\pi}^{\pi} \sum_{k=0}^{m} p_k \frac{\exp(i(n-k)\lambda)}{P(\exp(-i\lambda))} d\zeta(\lambda) =$$

$$\int_{-\pi}^{\pi} \exp(in\lambda) d\zeta(\lambda) = \xi_n.$$

Отже, відповідна *стаціонарна* послідовність $(\xi_n,\ n \in \mathbb{Z})$ є послідовністю **рухомого середнього:**

$$\xi_n = \sum\nolimits_{k=-m}^{0} p_{-k} \varepsilon_{n+k}.$$

Вправа. Нехай величини з послідовності $(\varepsilon_n,\ n \in \mathbb{Z})$ незалежні однаково розподілені. Довести, що для $(\xi_n,\ n \geq 1)$ виконується закон великих чисел.

Розглянемо приклад, у якому $P(z) = 1 - pz,\ |p| < 1$. Відповідна *коваріаційна функція* обчислюється зі свого спектрального зображення

$$r(0) = 1 + p^2,\ r(\pm 1) = -p.$$

Послідовність ξ_n має зображення

$$\xi_n = \varepsilon_n - p\ \varepsilon_{n-1}.$$

Елементи факторизації *спектральної щільності* дорівнюють

$$f_+(\lambda)\ = P(\exp(i\lambda))/2\pi,$$

$$f_-(\lambda)\ = 1/P(\exp(-i\lambda)) = 1 + \sum\nolimits_{n>0}\ p^n \exp(-in\lambda),$$

$$\widehat{g}(\lambda) = -\sum\nolimits_{n>0} p^n \exp(i\lambda(-n+1)).$$

Отже, оптимальний прогноз має вигляд

$$\widehat{\xi}_1 = -\sum\nolimits_{n>0} p^n \xi_{-n+1}\ \boxdot$$

2. Процес авторегресії (AR:AutoRegression). Нехай *спектральна щільність стаціонарної* послідовності $(\xi_n, n \in \mathbb{Z})$ дорівнює

$$f(\lambda) = \frac{1}{2\pi\,|Q(\exp(-i\lambda))|^2},\ Q(z) = \sum\nolimits_{k=0}^{m} q_k z^k,$$

де дійсний поліном Q не має коренів на одиничному колі.

Тоді випадкові величини

$$\varepsilon_n = \int_{-\pi}^{\pi} \exp(in\lambda) Q(\exp(-i\lambda))\ d\zeta(\lambda),\ n \in \mathbb{Z},$$

ортонормовані, де ζ – відповідний *стохастичний спектральний процес* зі спектральною щільністю f. Обчислимо

$$\sum\nolimits_{k=0}^{m} q_k \xi_{n-k} = \int_{-\pi}^{\pi} \sum\nolimits_{k=0}^{m} q_k \exp(i(n-k)\lambda) d\zeta(\lambda) = \varepsilon_n.$$

Отже, послідовність $(\xi_n,\ n \in \mathbb{Z})$ є розв'язком **рівняння авторегресії**:

$$\sum\nolimits_{k=-m}^{0} q_{-k} \xi_{n+k} = \varepsilon_n.$$

Елементи факторизації *спектральної щільності* дорівнюють

$$\begin{aligned} f_+(\lambda) &= 1/2\pi Q(\exp(i\lambda)),\\ f_-(\lambda) &= Q(\exp(-i\lambda)) = \textstyle\sum_{k=-m}^{0} q_{-k}\exp(ik\lambda),\\ \widehat{g}(\lambda) &= -\textstyle\sum_{k=-m}^{-1} q_{-k}\exp(i\lambda(k+1)). \end{aligned}$$

Отже, оптимальний прогноз має вигляд

$$\widehat{\xi}_1 = -\sum_{k=-m}^{-1} q_{-k}\xi_{k+1} \ \boxdot$$

3. **Процес із дробово-раціональною спектральною щільністю (ARMA: AutoRegression and Moving Average)**. Нехай *спектральна щільність стаціонарного* процесу $(\xi_n, n \in \mathbb{Z})$ є дробово-раціональною:

$$f(\lambda) = \frac{P(\exp(i\lambda))}{Q(\exp(i\lambda))},$$

де P, Q – комплексні поліноми, що не мають коренів на одиничному колі. Якщо поліном P або Q має корінь α, $|\alpha| < 1$, то в розкладі f є множник $1 - \alpha\exp(-i\lambda)$. Оскільки функція f – дійсна, то $f(\lambda) = \overline{f(\lambda)}$. З цього рівняння робимо висновок, що у розкладі f знайдеться також і множник $1 - \overline{\alpha}\exp(i\lambda)$. За наявності кореня з $|\beta| > 1$ відповідні нормовані множники також утворюють для $\alpha = 1/\overline{\beta}$ внесок вигляду

$$(\beta - \exp(i\lambda))(\overline{\beta} - \exp(-i\lambda)) = |\beta|^2 (1 - \alpha\exp(-i\lambda) - \alpha)(1 - \overline{\alpha}\exp(i\lambda)).$$

Групуючи такі множники попарно, приведемо щільність до вигляду

$$f(\lambda) = c\frac{\prod_{k=1}^{l}(1-\alpha_k\exp(-i\lambda))(1-\overline{\alpha_k}\exp(i\lambda))}{\prod_{j=1}^{m}(1-\beta_j\exp(-i\lambda))(1-\overline{\beta_j}\exp(i\lambda))} =$$

$$c\frac{\prod_{k=1}^{l}|1-\alpha_k\exp(-i\lambda)|^2}{\prod_{j=1}^{m}\left|1-\beta_j\exp(-i\lambda))\right|^2},$$

де $|\alpha_k| < 1$, $\left|\beta_j\right| < 1$, а стала c повинна бути дійсною. Звідси виводимо, що процес ARMA є розв'язком різницевого рівняння

$$\prod_{j=1}^{m}(1-\beta_j\Delta)\xi_n = \prod_{k=1}^{l}(1-\alpha_k\Delta)\varepsilon_n$$

з оператором Δ зворотного часового зсуву (тобто $\Delta\xi_n = \xi_{n-1}$, $\Delta\varepsilon_n = \varepsilon_{n-1}$) та з ортонормованими ε_n, що визначаються як

$$\varepsilon_n = \int_{-\pi}^{\pi}\exp(in\lambda)f_-(\lambda)d\zeta(\lambda).$$

Тут $f_{-}(\lambda)$ – відповідний знаменник факторизації:

$$f_{-}(\lambda) = \frac{\prod_{j=1}^{m}(1-\beta_j \exp(-i\lambda))}{\prod_{k=1}^{l}(1-\alpha_k \exp(-i\lambda))} =$$

$$\prod\nolimits_{j=1}^{m}(1-\beta_j \exp(-i\lambda)) \prod\nolimits_{k=1}^{l} \sum\nolimits_{n\geq 0} \alpha_k^n \exp(-in\lambda) \in L_2^{-}.$$

Оптимальний лінійний прогноз задається спектральною функцією $\widehat{g}(\lambda) = (1 - f_{-}(\lambda))\exp(i\lambda)$ $\boxdot$

Вправи

(1) Нехай (ξ_n) – стаціонарна послідовність, випадкова величина $\eta \in H_2(\xi)$ має спектральну характеристику $g(\lambda)$, а θ – оператор зсуву. Довести, що спектральна характеристика $\theta\eta$ дорівнює $\exp(i\lambda)g(\lambda)$.

(2) Знайти оптимальний прогноз для $\widehat{\xi}_1$, якщо коваріаційна функція має вигляд $r(t) = (1+|t|)\exp(-|t|)$.

(3) Знайти оптимальний прогноз для $\widehat{\xi}_1$, якщо спектральна щільність дорівнює (а) $f(\lambda) = (25+24\lambda)^{-1}$, (б) $f(\lambda) = 2+2\cos\lambda$.

3.20.4. Регулярні стаціонарні послідовності

Розглянемо певне узагальнення послідовностей *рухомого середнього*. Нехай $(\xi_n,\ n\in\mathbb{Z})$ – *стаціонарна* послідовність, що є *квадратично інтегровною* з нульовим середнім і *коваріаційною функцією* $r(n)$.

Позначимо лінійні підпростори

$$L_n = \pounds(\xi_k,\ k\leq n),\ H_n = \aleph(L_n)\subset H_{n+1},$$

$$L_\infty = \cup L_n,\quad H_\infty = \cup H_n,\ H_{-\infty} = \cap H_n.$$

Нехай лінійне відображення $\theta : L_\infty \to L_\infty$ визначається рівністю

$$\theta\xi_n \equiv \xi_{n+1},\ \forall n\in\mathbb{Z},$$

на твірних елементах L_∞ та продовжується за лінійністю на всі скінченні лінійні форми з L_∞ :

$$\theta \sum\nolimits_{k\leq n} c_k\xi_k \equiv \sum\nolimits_{k\leq n} c_k\xi_{k+1}.$$

Оскільки послідовність $(\xi_n,\ n\in\mathbb{Z})$ – *стаціонарна*, то відображення θ є *ізометрією* простору L_n на L_{n+1} :

$$\left\|\theta\sum\nolimits_{k\leq n} c_k\xi_k\right\|^2 = \left\|\sum\nolimits_{k\leq n} c_k\xi_{k+1}\right\|^2 = \mathbf{E}\left|\sum\nolimits_{k\leq n} c_k\xi_{k+1}\right|^2 =$$

$$\sum\nolimits_{k\leq n, j\leq n} c_k\overline{c}_j \mathbf{E}\xi_{k+1}\overline{\xi_{j+1}} = \sum\nolimits_{k\leq n, j\leq n} c_k\overline{c}_j\ r(k-j) = \sum\nolimits_{k\leq n, j\leq n} c_k\overline{c}_j \mathbf{E}\xi_k\overline{\xi_j} =$$

$$\mathbf{E}\left|\sum\nolimits_{k\leq n} c_k\xi_k\right|^2 = \left\|\sum\nolimits_{k\leq n} c_k\xi_k\right\|^2 = \left\|\sum\nolimits_{k\leq n} c_k\xi_k\right\|^2.$$

За неперервністю внаслідок теореми *про продовження ізометрії* відображення θ продовжується до ізометрії H_n на H_{n+1}.

Означення. Ортогональним проектором H_∞ *на* H_n *називається лінійний оператор, що діє на елементи* $\xi \in H_\infty$ *за формулою*

$$\Pi_n \xi \equiv \arg\min{}_{\eta \in H_n} \|\xi - \eta\|^2 .$$

Цей оператор існує та коректно визначений за теоремою *про лінійний прогноз у гільбертовому просторі.*

Означення. *Стаціонарна послідовність* $(\xi_n,\ n \in \mathbb{Z})$ *з нульовим середнім* $\mathbf{E}\xi_n = 0$ *називається* регулярною, *якщо* $\cap_{\ n\in\mathbb{Z}} H_n = \{0\}$.

Теорема (теорема Вольда про структуру регулярної стаціонарної послідовності). *Нехай стаціонарна послідовність* $(\xi_n,\ n \in \mathbb{Z})$ *є регулярною. Тоді для всіх* $n \in \mathbb{Z}$*:*

(а) Включення $H_n \subset H_{n+1}$ *є строгим (не щільним).*

(б) Величини $\varepsilon_{n+1} \equiv \xi_{n+1} - \Pi_n \xi_{n+1} \in H_{n+1}$ *невироджені, попарно ортогональні,* $\varepsilon_{n+1} \perp H_n$.

(в) Має місце рівність $\varepsilon_{n+1} = \theta \varepsilon_n$.

(г) Послідовність $(\varepsilon_k,\ k \le n)$ *є ортогональним базисом у* H_n.

(д) Існують комплексні $(c_k, k \le 0)$, *з* $c_0 = 1$, *такі, що справедливе зображення* ***рухомого середнього***

$$\xi_n = \sum{}_{k \le 0} c_k\, \varepsilon_{k+n} = \sum\nolimits_{k \le n} c_{k-n}\, \varepsilon_k,$$

де ряд збігається у нормі $L_2(\mathbf{P})$ *– у середньо-квадратичному.*

(е) Спектральна міра стаціонарної послідовності $(\xi_n,\ n \in \mathbb{Z})$ *має таку спектральну щільність на* $[-\pi, \pi]$:

$$f(\lambda) = (2\pi)^{-1} \mathbf{E}\, |\varepsilon_0|^2 \left| \sum\nolimits_{k \le 0} c_k \exp(ik\lambda) \right|^2 .$$

Доведення

(а) Припустимо, що $H_n = H_{n+1}$ для деякого n.

Тоді існують лінійні форми $\eta_n^m = \sum_{k \le n} c_k^m \xi_k \in L_n$ такі, що

$$\left\| \xi_{n+1} - \eta_n^m \right\| \to 0,\ m \to \infty.$$

Оскільки θ – *ізометрія*, то для величин $\eta_{n-1}^m \equiv \theta^{-1} \eta_n^m \in L_{n-1}$ справедливі співвідношення

$$\left\| \xi_n - \eta_{n-1}^m \right\| = \left\| \theta(\xi_n - \eta_{n-1}^m) \right\| = \left\| \xi_{n+1} - \eta_n^m \right\| \to 0,\ m \to 0.$$

Отже, $\xi_n \in H_{n-1}$ та $H_{n-1} = H_n$, звідки за індукцією. $H_{-\infty} = H_n$, що суперечить умові *регулярності*.

(б) Невиродженість є наслідком (а), оскільки з $\varepsilon_{n+1} = 0$ випливало б, що $\xi_{n+1} = \Pi_n \xi_{n+1} \in H_n$, що суперечить регулярності. Ортогональність $\varepsilon_{n+1} \perp H_n$ є наслідком властивості (б) ортогонального проектора з теореми *про лінійний прогноз у гільбертовому просторі*. Нарешті, зі включення $\varepsilon_n \in H_n$ виводимо попарну ортогональність $\varepsilon_k \perp \varepsilon_n,\ k < n$.

(в) Оскільки θ – *ізометрія* L_{n-1} на L_n, то для довільного $\eta_{n-1} \in L_{n-1}$

$$\left\|\xi_n - \eta_{n-1}\right\| = \left\|\theta(\xi_n - \eta_{n-1})\right\| = \left\|\xi_{n+1} - \theta\eta_{n-1}\right\|.$$

Отже, нижні границі за η_{n-1} в обох частинах цієї рівності є однаковими, а відповідні аргументи збігаються до векторів, пов'язаних відображенням θ. Тому за означенням *ортогонального проектора* $\Pi_n \xi_{n+1} = \theta \Pi_{n-1} \xi_n$ і

$$\varepsilon_{n+1} = \xi_{n+1} - \Pi_n \xi_{n+1} = \theta(\xi_n - \Pi_{n-1}\xi_n) = \theta\varepsilon_n.$$

(г) Нехай для фіксованого n величина $\zeta \in H_n$ така, що $\zeta \perp \varepsilon_k,\ \forall k \le n$. За означенням H_n для деякої послідовності скінченних лінійних форм

$$\eta_n^m = \sum\nolimits_{k \le n} c_k^m \xi_k \in L_n$$

має місце збіжність $\eta_n^m \to \zeta,\ m \to \infty$. Оскільки $\varepsilon_n \perp \xi_k \in H_k$ при $k < n$ за властивістю (б), то

$$(\xi_n, \varepsilon_n) = (\xi_n - \Pi_{n-1}\xi_n, \varepsilon_n) = (\varepsilon_n, \varepsilon_n) = \left\|\varepsilon_n\right\|^2,$$

$$c_n^m \left\|\varepsilon_n\right\|^2 = c_n^m (\xi_n, \varepsilon_n) = (\eta_n^m,\ \varepsilon_n) \to (\zeta, \varepsilon_n) = 0,\ m \to \infty.$$

Звідси отримуємо: $\eta_n^m - \Pi_{n-1}\eta_n^m = c_n^m \varepsilon_n \to 0,\ m \to \infty$. Отже,

$$\zeta = \lim\nolimits_{m\to\infty} \eta_n^m = \lim\nolimits_{m\to\infty} \Pi_{n-1}\eta_n^m \in H_{n-1}.$$

За індукцією $\zeta \in H_{-\infty} = \{0\}$, тобто $\zeta = 0$. Отже, послідовність $(\varepsilon_k,\ k \le n)$ є базисом у H_n. Ортогональність цього базису доведено у твердженні (б).

(д) З попереднього пункту та означення базису випливає існування сталих $c_k \in \mathbb{C}$ таких, що

$$\xi_0 = \sum\nolimits_{k \le 0} c_k\, \varepsilon_k,$$

де $c_0 = 1$, оскільки за означенням (б) та за ортогональністю базису ε_n

$$\xi_0 - \varepsilon_0 = \Pi_{-1}\xi_0 = \sum\nolimits_{k \le -1} c_k\, \varepsilon_k = \xi_0 - c_0\varepsilon_0.$$

Звідси внаслідок (в) та за означенням неперервної *ізометрії* θ дістанемо зображення (д) для довільного n.

(е) Вираз для *спектральної щільності* виводиться з (д) та з означення *спектральної міри*, оскільки при $n \geq 0$

$$\int_{-\pi}^{\pi} \exp(in\lambda) f(\lambda) d\lambda = r(n) = \mathbf{E}\xi_n \overline{\xi_0} = \mathbf{E} \sum\nolimits_{k \leq 0} c_k \varepsilon_{k+n} \overline{\sum\nolimits_{j \leq 0} c_j \varepsilon_j} =$$

$$\sum\nolimits_{k \leq 0} \sum\nolimits_{j \leq 0} c_k \overline{c_j}\ \mathbf{E}\varepsilon_{k+n} \overline{\varepsilon_j} = \mathbf{E} \left|\varepsilon_0\right|^2 \sum\nolimits_{k \leq 0} c_k\ \overline{c_{k+n}}\ \mathbb{I}_{k \leq -n} =$$

$$\mathbf{E} \left|\varepsilon_0\right|^2 (2\pi)^{-1} \int_{-\pi}^{\pi} \exp(in\lambda) \left| \sum\nolimits_{k \leq 0} c_k \exp(ik\lambda) \right|^2 d\lambda, \ \forall n \in \mathbb{Z} \ \boxdot$$

Теорема (про оптимальний прогноз для регулярної послідовності). *Позначимо через* $\widehat{\xi}_1 = \Pi_0 \xi_1$ *найкращий у середньоквадратичному лінійний прогноз значення* ξ_1 *за спостереженнями* $(\xi_k, \ k \leq 0)$. *Для регулярної послідовності мають місце тотожності*

$$\widehat{\xi}_1 = \sum\nolimits_{k<0} c_k\ \varepsilon_{k+1} = \xi_1 - \varepsilon_1, \quad \mathbf{E} \left| \xi_1 - \widehat{\xi}_1 \right|^2 = \mathbf{E} \left|\varepsilon_1\right|^2.$$

Доведення очевидне. Зауважимо, що побудова прогнозу складається з 2-х етапів: обчислення коефіцієнтів c_k, які визначаються *коваріаційною функцією*, або ж її *спектральною щільністю*, та відшукання базису (ε_k) з рівнянь зображення *рухомого середнього* $\boxdot$

Приклад. Нехай послідовність $(\xi_n, n \in \mathbb{Z})$ задовольняє зображення *рухомого середнього* з коефіцієнтами $c_n = \rho^{-n}, n \leq 0$ з $\rho \in (-1, 1)$, та з ортонормованими похибками (ε_n), тобто має *спектральну щільність*

$$f(\lambda) = (2\pi)^{-1} \left| \sum\nolimits_{k \leq 0} \rho^{-k} \exp(ik\lambda) \right|^2 = (2\pi)^{-1} \left|1 - \rho \exp(-i\lambda)\right|^{-2}.$$

Тоді зі вказаного зображення внаслідок мультиплікативності функції ρ^n виводимо, що дана послідовність є розв'язком системи рівнянь

$$\xi_n = \varepsilon_n + \sum\nolimits_{k \leq n-1} \rho^{n-k} \varepsilon_k = \varepsilon_n + \rho \xi_{n-1}, \ n \in \mathbb{Z}.$$

Отже, оптимальна оцінка ξ_1 за спостереженнями $\xi_k, k \leq 0$, має вигляд $\widehat{\xi}_1 = \xi_1 - \varepsilon_1 = \rho \xi_0$.

Додаток

Список рекомендованої літератури

Основна література

1. *Дороговцев А.Я.* Элементы общей теории меры и интеграла – Киев, Вища школа, 1989.

2. *Дороговцев А.Я.* Математичний аналіз, у двох частинах – Київ, Вища школа, 1993.

3. *Ядренко М.Й.* Дискретна математика – Київ, Експрес, 2003.

4. *Ширяев А. Н.* Вероятность, в 2-х кн. – Москва, МЦНМО, 2004.

5. *Гихман И.И., Скороход А.В., Ядренко М.Й.* Теория вероятностей и математическая статистика – Киев, Вища школа, 1988.

6. *Скороход А.В.* Елементи теорії ймовірностей та випадкові процеси – Київ, Вища школа, 1975.

7. *Скороход А.В.* Лекції з теорії випадкових процесів – Київ, Либідь, 1990.

8. *Ивченко Г.И., Медведев Ю.И.* Математическая статистика – Москва, Высшая школа, 1984.

9. *Березанский Ю.М., Ус Г.Ф., Шефтель З.Г.* Функциональный анализ – Киев, Вища школа, 1990.

10. *Карташов М.В.* Теорія ймовірностей та математична статистика – Київ, ТВіМС, 2004.

11. *Дороговцев А.Я.,Сільвестров Д.С.,Скороход А.В.,Ядренко М.Й.* Теория вероятностей. Сборник задач – Киев, Вища школа, 1976.

12. Теория вероятностей. Методические указания к ведению практических занятий, Киев, КГУ, 1983.

13. Завдання до лабораторних занять з курсу "Математична статистика", Київ, механіко-математичний факультет КНУ, 2003.

Додаткова література

1. *Гнеденко Б.В.* Курс теории вероятностей – Москва, Наука, 1987.

2. *Скороход А.В.* Вероятность – Москва, ВИНИТИ, 1989.

3. *Боровков А.А.* Теория вероятностей – Москва, Наука, 1987.

4. *Севастьянов Б.А.* Курс теории вероятностей и математической статистики – Москва, Наука, 1982.

5. *Коваленко И.Н., Гнеденко Б.В.* Теория вероятностей – Киев, Вища школа, 1990.

6. *Розанов А.А.* Теория вероятностей, случайные процессы и математическая статистика. – Москва, Наука, 1985.

7. *Лоев М.* Теория вероятностей – Москва, ИЛ, 1962.

8. *Неве Ж.* Математические основы теории вероятностей – Москва, ИЛ, 1969.

9. *Ламперти Дж.* Вероятность – Москва, Наука, 1973.

10. *Феллер В.* Введение в теорию вероятностей и ее приложения – В двух томах, Москва, Мир, 1984.

11. *Уиттл П.* Вероятность. – М.: Наука, 1982.

12. *Секей Г.* Парадоксы в теории вероятностей и математической статистике. – Москва, Мир, 1990.

13. *Крамер Г.* Математические методы статистики – Москва, Мир, 1975.

14. *Закс Ш.* Теория статистических выводов – Москва, Мир, 1975.

15. *Уилкс С.* Математическая статистика – Москва, Наука, 1967.

16. *Леман Э.* Проверка статистических гипотез – Москва, Наука, 1979.

17. *Леман Э.* Теория точечного оценивания – Москва, Наука, 1991.

18. *Питмен Э.* Основы теории статистических выводов – Москва, Мир, 1986.

19. *Гихман И.И., Скороход А.В.* Введение в теорию случайных процессов, Москва, Наука, 1977.

20. *Карлин С.* Основы теории случайных процессов, Москва, Мир, 1971.

21. *Мешалкин Л.Д.* Сборник задач по теории вероятностей. – Москва, МГУ, 1963.

22. *Зубков А.М.,Севастьянов Б.А.,Чистяков В.П. Сборник задач по теории вероятностей – Москва, Наука, 1989.*

23. *Климов Г.П., Кузьмин А.Д.* Вероятность, процессы, статистика. Задачи с решениями – Москва, МГУ, 1986.

24. *Чибисов Д.М., Пагурова В.И.* Задачи по математической статистике – Москва, МГУ, 1990.

25. *Ивченко Г.И., Медведев Ю.И., Чистяков А.В.* Сборник задач по математической статистике – Москва, Высшая школа, 1989.

26. Справочник по прикладной статистике, под ред. *Э.Ллойд, У.Ледерман*, т.1,2 – Москва, Финансы и статистика, 1989.

27. *Айвазян С.А.,Енюков И.С.,Мешалкин Л.Д.* Прикладная статистика, Исследование зависимостей – Москва, Финансы и статистика, 1985.

28. *Турчин В.М.* Теорія ймовірностей, Основні поняття, приклади, задачі – Київ, А.С.К., 2004.

29. *Gray R.M.* Probability, Random Processes and Ergodic Properties – Springer-Verlag, 1987.

30. *Durrett R.* Probability, Theory and Examples – Wadsworth Publishing, 1996.

31. *Kallenberg O.* Foundation of Modern Probability – Springer-Verlag, 1997.

Грецький алфавіт

A, α	альфа	I, ι	йота	P, ρ	ро
B, β	бета	$K, \varkappa$	каппа	Σ, σ	сигма
Γ, γ	гама	Λ, λ	ламбда	T, τ	тау
Δ, δ	дельта	M, μ	мю	Υ, υ	іпсилон
E, ε	епсилон	N, ν	ню	Φ, φ	фі
Z, ζ	дзета	Ξ, ξ	ксі	X, χ	хі
H, η	ета	O, o	омікрон	Ψ, ψ	псі
Θ, θ	тета	Π, π	пі	Ω, ω	омега

Список позначень

$\equiv$ – рівність за означенням.

$a_n \sim b_n$ – *асимптотична еквівалентність* послідовностей: $a_n/b_n \to 1$ при $n \to \infty$.

$n! = n(n-1)..2 \cdot 1$ – факторіал числа n.

$\mathbb{N}, \mathbb{Z}, \mathbb{Z}_+, \mathbb{R}, \mathbb{R}_+, \mathbb{C}$ – множини натуральних, цілих, дійсних, невід'ємних цілих, невід'ємних дійсних та комплексних чисел.

$C(I), C_b(I)$ – простори неперервних та обмежених неперервних функцій на множині I.

$\operatorname{Re} z$, $\operatorname{Im} z$ – дійсна та уявна частини числа z.

rang A – ранг матриці A.

diam $U = \sup_{x,y \in U} \ |x-y|$ – діаметр множини $U \subset \mathbb{R}^n$.

Ω – *простір елементарних подій*, *універсальна* множина.

$\varnothing$ – *неможлива* подія, порожня множина.

2^Ω – клас всіх підмножин множини Ω.

$\mathbf{P}(A \mid B)$ – *умовна ймовірність* події A за умови B.

$\varphi(x), \Phi(x)$ – щільність, функція стандартного *нормального розподілу*.

$\mathfrak{A}(\mathbb{R})$, $\mathfrak{A}(\mathbb{R}^n)$ – *алгебри* множин в $\mathbb{R}$ та $\mathbb{R}^n$, що породжуються напівінтервалами та паралелепіпедами.

$\mathfrak{B}(\mathbb{R})$, $\mathfrak{B}(\mathbb{R}^n)$ – *борелєві* сигма-алгебри в $\mathbb{R}$ та $\mathbb{R}^n$.

$\sigma[\mathfrak{A}]$ – сигма-алгебра, що *породжена* класом множин $\mathfrak{A}$.

$\mathbb{I}_B$ – індикаторна функція висловлювання B.

$\mathbb{I}_{\{A\}}(\omega)$ – *індикаторна величина* випадкової події A.

$\delta_{kj} = \mathbb{I}_{k=j}$ – символ Кронекера.

$I = (\delta_{kj})$ – одинична матриця.

$\xi \simeq \eta$ – збіжність *функцій розподілу* випадкових величин ξ, η.

$B(n,p)$ – випадкова величина, що має *біноміальний розподіл* із числом випробувань n та ймовірністю успіху p.

$G(p)$ – випадкова величина, що має *геометричний розподіл* з імовірністю успіху p.

$\Pi(\lambda)$ – випадкова величина з *розподілом Пуассона* і параметром λ.

$B_1 \times B_2 = \{(b_1, b_2), b_i \in B_i\}$ – декартів добуток множин.

$x^+ = \max(0, x)$ – додатна частина числа x.

$x^- = \max(0, -x)$ – від'ємна частина числа x.

ξ^+, ξ^- – *додатна та від'ємна частини* випадкової величини ξ.

м.н. – майже напевне.

$\mathbf{E}\xi, \mathbf{M}\xi$ – *математичне сподівання* випадкової величини ξ.

$\mathbf{D}\xi$ – *дисперсія* випадкової величини ξ.

$\operatorname{argmin} F$, $\operatorname{argmax} F$ – значення аргументу, для якого досягається найменше чи найбільше значення функції F (у припущенні його існування).

$f^{(-1)}(B) = \{x : f(x) \in B\}$ – прообраз множини B при відображенні f.

$L_2(\mathbf{P})$ – простір *квадратично інтегровних* випадкових величин.

$N(\mu, \sigma^2)$ – випадкова величина з *нормальним розподілом*, середнім μ та дисперсією σ^2.

$N_n(m, V)$ – *нормальний вектор* у $\mathbb{R}^n$ із середнім m та коваріаційною матрицею V.

$\mathfrak{L}(\mathbb{R}^n, \mathbb{R}^m)$ – клас лінійних перетворень із $\mathbb{R}^n$ в $\mathbb{R}^m$.

$\mathfrak{L}_n(\mathbb{R}^n)$ – клас невироджених лінійних перетворень простору $\mathbb{R}^n$.

$U(a,b)$ – випадкова величина, *рівномірно розподілена* на (a,b).

$Exp(\lambda)$ – випадкова величина, що має *показниковий розподіл* із параметром λ.

$\Gamma(\alpha)$ – повна гама-функція в точці α.

$\Gamma(\lambda, \alpha)$ – випадкова величина з *гама-розподілом* і параметрами λ, α.

χ_n^2 – величина з *хі-квадрат розподілом* і n ступенями свободи.

τ_n – величина з *розподілом Стьюдента* і n ступенями свободи.

$f_{n,m}$ – величина, що має *розподіл Фішера* з n, m ступенями свободи.

$\operatorname{Cov}(\xi, \eta)$ – *коваріація* випадкових величин ξ, η.

$\operatorname{Cov}(\xi)$ – *коваріаційна матриця* випадкового вектора ξ.

Π_a – *кут* із правою верхньою вершиною a.

$\xi_n \to^P \xi$ – збіжність випадкових величин *за ймовірністю.*

$\xi_n \to^{P1} \xi$ – збіжність випадкових величин *з імовірністю 1.*

$\xi_n \to^L \xi$ – збіжність випадкових величин *у середньому.*

$F_n \to^O F$ – збіжність функцій розподілу *в основному.*

$\xi_n \to^W \xi$ – *слабка збіжність* випадкових величин.

$Int(B), Cl(B)$ – внутрішність та замикання множини B.

$\overline{\lim}_{n\to\infty}\mathfrak{F}_n$ – *залишкова сигма-алгебра*, що породжена послідовністю сигма-алгебр $\mathfrak{F}_n$.

Θ – параметрична множина довільної природи.

$\mathbf{P}_\theta(\cdot)$ – імовірність на класі подій при значенні параметра θ.

$\mathbf{E}_\theta\xi$ – *математичне сподівання* величини ξ при значенні параметра θ (інтеграл за мірою $\mathbf{P}_\theta$).

$\mathbf{D}_\theta\xi$ – *дисперсія* величини ξ при значенні параметра θ.

X – статистична *вибірка* зі значеннями у просторі S.

$(S,\ \Sigma,\ \lambda)$ – *вибірковий простір* із сигма-алгеброю Σ та мірою λ.

$F_1 \ll F_0$ – міра F_1 *абсолютно неперервна відносно міри* F_0.

$X = (\xi_1, ..., \xi_n)$ – *кратна вибірка* об'єму n.

$\widehat{\theta}$ – *оцінка* параметра θ.

$L(X, \theta)$ – *вибіркова функція вірогідності*.

$f(\xi, \theta)$ – *функція вірогідності спостереження*.

$\widehat{F}_n(x)$ – *емпірична функція розподілу*.

$\xi_{(k)}$ – k-та *порядкова статистика*.

$\widehat{\mu}_n,\ \widehat{\sigma}_n^2$ – *вибіркове середнє* та *вибіркова дисперсія*.

$\widehat{s}_n^2$ – *нормована вибіркова дисперсія*.

$\mu_k,\ \mu_k^0$ – нецентральний *теоретичний момент* та *центральний теоретичний момент* k-го порядку.

$\widehat{\mu}_{kn},\ \widehat{\mu}_{kn}^0$ – нецентральний *вибірковий момент* та *центральний вибірковий момент* k-го порядку.

$U(X, \theta)$ – вибіркова *функція впливу*.

$u(\xi, \theta)$ – *функція впливу спостереження*.

ОМВ – *оцінка максимальної вірогідності*.

$H_0 : \theta \in \Theta_0$ – *нульова гіпотеза* щодо значення параметра θ.

$H_1 : \theta \in \Theta_1$ – *альтернативна* гіпотеза щодо значення θ.

$(\widehat{\varkappa}(X),\ D_1)$ – *статистичний критерій* з *критичною областю* D_1.

$(\widehat{\varkappa}(X),\ \pi(t))$ – *рандомізований критерій* перевірки гіпотези.

$(\widehat{\varkappa}(X),\ \pi(t),\ t_0)$ – *простий рандомізований критерій*.

МНК – *метод найменших квадратів*.

$\pounds(G)$ – лінійна оболонка множини G.

$\aleph(G)$ – замикання лінійної оболонки множини G.

н.с.д. D – найбільший спільний дільник множини D.

Предметний покажчик- Драфт

Зміст