М.В.Карташов

Теорія ймовірностей та математична статистика

Річний курс для математиків та статистиків

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

Розширене електронне видання

Київ
ТВіМС
2004

УДК 519.2

Затверджено Міністерством освіти і науки України,
лист N 1/11-1854 від 28.04.2004 року

Рецензенти:
доктор фіз-мат наук, професор В.В.Булдигін
(НТУУ "Київський політехнічний інститут")

доктор фіз-мат наук, професор М.І.Портенко
(Інститут математики НАНУ)

ПІДРУЧНИК

Карташов Микола Валентинович

Теорія ймовірностей та математична статистика

Видавництво МП "ТВіМС"

Карташов М.В. Теорія ймовірностей та математична статистика. Річний курс для математиків та статистиків. – Київ, Видавництво ТВіМС, 2004, 307 с.

Підручник містить матеріал курсів ”Теорія ймовірностей”, ”Математична статистика” і призначений для студентів університетів математичних та статистичних спеціальностей. Виклад грунтується на понятті інтегралу Лебега, однак всі необхідні властивості останнього визначаються та виводяться, що сприяє доступності курсу.

Основні поняття та теореми ілюстровано на прикладах. Для повнішого засвоєння курсів наведені вправи для самостійного розв'язання. У додатку вміщено алфавітний перелік термінів та назв усіх теорем курсу.

Для студентів університетів.

ISBN 966-95255-5-1

Передмова

Підручник містить матеріал двосеместрових лекцій з теорії ймовірностей та математичної статистики, які викладалися для студентів спеціальностей математика, статистика III курсу механіко-математичного факультету Київського національного університету імені Тараса Шевченка протягом 1988-2003 рр. Об'єм матеріалу розрахований на 105 академічних лекційних годин, по 3 години на тиждень.

На відміну від попередньої традиції [5], виклад суттєво спирається на поняття інтегралу Лебега, як це зроблено в [4,6,7]. Основні необхідні властивості інтегралу доводяться для подальшого використання. Без доведення використовуються початкові результати загальної теорії міри – теорема Каратеодорі про продовження міри, теорема Фубіні про кратний та повторний інтеграли, теорема про заміну змінної та теорема Радона – Нікодима про абсолютно неперервні міри. Відповідний курс викладається на факультеті одночасно з курсом теорії ймовірностей [1], однак для досить 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. Для повнішого засвоєння курсу та при розв'язанні задач корисним є володіння основами дискретної математики [3].

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 рукопису – професорам М.І. Портенку та В.В. Булдигіну, за цінні зауваження і поради, які дозволили поліпшити виклад.

Підручник видано в рамках реалізації програми Європейської комісії Tempus-Tacis NP 22012-2001 ”Improvement of education in economical-statistical area in Ukraine”. Користуючись нагодою, дякую за підтримку координатору проекту професору Д.С. Сільвестрову та адміністратору проекту Е.Д.Сільвестровій.

Розділ 1

Теорія ймовірностей

Вступ

Створення теорії ймовірностей пов'язане з іменами Паскаля, Ферма, Гюйгенса, Якоба Бернуллі, Муавра (17-18 ст). Ці вчені вивчали випадкові явища, що спостерігаються в азартних іграх.

Імовірнісні проблеми теорії похибок вимірювань, похибок стрільби, статистики народонаселення розв'язувалися в працях Лапласа, Пуассона, Гаусса в 18-19 століттях.

У 19-20 ст. значні досягнення в теорії ймовірностей та статистиці були отримані Чебишевим, Марковим, Ляпуновим, Буняковським, зокрема, у вигляді граничних теорем.

Сучасна теорія ймовірностей була створена в працях Р. Мізеса, С.Н. Бернштейна, Е. Бореля, А.М. Колмогорова, О.Я. Хінчина, Б. В. Гнеденка, П. Леві, Ю.В. Прохорова, та інших вчених 20 ст.

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

З погляду ймовірнісників, всі явища класифікуються 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. На практиці детермінованих явищ немає.

(нн) *Недетерміновані неповторювані.* Наприклад: наявність життя на Марсі. Висловлювання щодо її ймовірності 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 всим 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$=0.5080$
К.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.

Як же обґрунтувати аксіоми теорії ймовірностей ? Виявляється, досить вказати на такі властивості відносної *частоти* як відношення кількості сприятливих випробувань до кількості всіх випробувань:

(*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 подій* (висловлювань щодо результату експерименту).

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, результати яких не можна передбачити однозначно (наприклад, підкидання монети чи грального кубика), або ж можна передбачити.

Певний фіксований результат стохастичного експерименту, який не можна виразити через сукупність інших результатів, називається елементарною подією. Наприклад, випадання шістки при підкиданні кості – на відміну від результату, що полягає у випаданні парної кількості очок.

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.*

Так як 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 – як ω .

1.1. Приклади стохастичних експериментів та відповідних просторів елементарних подій

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

2. *Підкидання двох монет.* Простір елементарних подій містить пари можливих варіантів підкидань однієї монети:

$$\Omega = \{\text{АА,АР,РА,РР}\}.$$

Фіксація результату підкидання двох монет потребує задання впорядкованої пари з двох елементів – результатів підкидання першої та другої монети.

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, отже

$$\Omega = \{\text{У, НУ, ННУ,...,НН...НУ,...}\}.$$

З тим же результатом можна ототожнити простір елементарних подій із множиною натуральних чисел $\mathbb{N} = \{n,\ n \geq 1\}$.

За припущення, що є протилежним до скінченності кількості підкидань, треба постулювати, що простір елементарних подій містить також точку $\infty = \{\text{ НН...Н...}\}$.

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

1.2. 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, для яких справджуються певні висловлювання про результат стохастичного експерименту.*

Зауважимо, що не всі підмножини простору елементарних подій повинні бути випадковими подіями. Підмножин може бути більше, ніж можна сформулювати висловлювань. Наприклад, у випадку, коли $\Omega = [0, 1]$, маємо не менше континуума підмножин, але не більш ніж зліченну кількість висловлювань (якщо кількість букв у висловлюванні скінченна).

1.3. Властивості класу випадкових подій

Властивості висловлювань та логічні операції над ними призводять до наступних властивостей класу випадкових подій – підмножин Ω.

1. Множина $\Omega \subset \Omega$ ("щось та відбудеться") є випадковою подією, яка називається універсальною.

2. Множина $\varnothing$, що не містить жодної елементарної події ("нічого не відбудеться"), є випадковою подією, і називається неможливою подією.

3. Якщо елементарна подія ω належить випадковій події A, то кажуть, що ω сприяє A (позначення $\omega \in A$).

4. Якщо $A \subset B$, то кажуть, що подія A спричиняє подію B (або ж міститься в ній).

5. Для події A її доповнення $\overline{A}$ є випадковою подією, яка полягає в тому, що не відбудеться A.

6. Для двох подій A, B їх об'єднання $A \cup B$ є випадковою подією, яка полягає в тому, що відбудеться A або B.

7. Для двох подій A, B їх переріз (або ж перетин) $A \cap B$ є випадковою подією, яка полягає в тому, що відбудуться A і B одночасно.

8. Якщо $A \cap B = \varnothing$, то події A, B називаються несумісними (або ж такими, що не перетинаються). Події послідовності $(A_n, n \geq 1)$ називаються попарно несумісними, якщо $A_i \cap A_j = \varnothing$ для всіх $i \neq j$.

9. Для двох подій A, B їх різниця $A \setminus B$ 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, $A \Delta B = (A \backslash B) \cup (B \backslash A)$.

11. Для послідовності подій $(A_n, n \geq 1)$ їх об'єднання $\cup A_n$ є випадковою подією, яка полягає в тому, що відбудеться хоча б одна подія із цієї послідовності. Аналогічно, переріз $\cap A_n$ є випадковою подією, яка полягає в тому, що відбудуться всі події із даної послідовності.

12. 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.

13. 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 (позначення $A_n \downarrow A \equiv \lim A_n$) є випадкова подія, яка полягає в тому, що відбудуться одночасно всі події послідовності, тобто $A = \cap A_n$.У такому випадку кажуть, що події A_n монотонно збігаються до A.

14. 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** (або ж відбудуться всі події з деякої нескінченної підпослідовності A_{n_k}). Цю подію можна записати у вигляді

$$\overline{\lim} A_n = \cap_{m\geq 1} \cup_{n\geq m} A_n.$$

Дійсно, елементарна подія ω належить правій частині тоді і тільки тоді, коли для довільного $m \geq 1$ знайдеться номер $n \geq m$ такий, що $\omega \in A_n$, тобто коли відбудеться нескінченна кількість серед подій A_n.

15. 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**. Цю подію можна записати у вигляді

$$\underline{\lim} A_n = \cup_{m\geq 1} \cap_{n\geq m} A_n.$$

Дійсно, елементарна подія ω належить правій частині тоді і 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.

Вправа. Обчислити $\underline{\lim} A_n, \overline{\lim} A_n$ у випадку, коли $\Omega = \mathbb{R}$, а події $A_n = (-\infty, a_n)$.

2. Аксіоми теорії ймовірностей та властивості ймовірності

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

2.1. Аксіоми класу випадкових подій

Нехай $\mathcal{F}$ – клас усіх випадкових подій у деякому стохастичному експерименті із простором елементарних подій Ω . Тоді:

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

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

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

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

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

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

Очевидно, ця властивість еквівалентна умові

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

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

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

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

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

$$\mathcal{F} = \bigcap_{S \supset \mathfrak{A},\ S\ \ \sigma-\text{алгебра}} S.$$

Вправа. Це визначення коректне, оскільки хоча б одна така сигма-алгебра існує (клас $S = 2^{\Omega}$ усіх підмножин Ω), причому переріз будь-якої кількості сигма-алгебр є сигма-алгеброю.

Теорема (про монотонний клас). *Якщо клас множин $\mathcal{F}$ 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.

Першу групу аксіом теорії ймовірностей можна сформулювати так:
"Клас усіх випадкових подій є сигма-алгеброю".

Вправа. Довести, що з наведених аксіом випливають всі вказані вище властивості класу випадкових подій. Вказівка. 1. Аксіома F1. 2. $\varnothing = \overline{\Omega}$. 5. Аксіома F2. 6. Аксіома F3. 7. $A \cap B = \overline{\overline{A} \cup \overline{B}}$.9. $A \setminus B = A \cap \overline{B}$.11. $\cap A_n = \overline{\cup \overline{A_n}}$. 12. Аксіома F3. 13. Властивість 11. 14. Аксіома F3 і властивість 11. 15. Аксіома F3 і властивість 11.

2.2. Приклади випадкових подій та їх класів

1. *Підкидання двох монет*, $\Omega = \{$АА,АР,РА,РР$\}$. Події

{Перший аверс}={АА,АР}, {Хоча б один аверс}={АА,АР,РА}.

2. *Підкидання монети до першого успіху*, $\Omega = \{$У, НУ,..,Н...НУ,...$\}$. Подія {Парна кількість підкидань до першого успіху} $=$ {НУ, НННУ,...НН...НУ,...}.

3. *Вибір випадкової точки на одиничному відрізку*, $\Omega = [0, 1]$. Клас усіх випадкових подій має містити всі інтервали $[0, x)$ та їх перетворення за допомогою операцій доповнення, зліченного об'єднання та інших операцій, що зберігають належність до сигма-алгебри.

Вправа.

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

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

2.3. Аксіоми ймовірності

У теорії ймовірностей із кожною випадковою подією пов'язують числову міру її вірогідності – ймовірність. Оскільки ймовірність розглядається як границя відносних частот, то на основі наведених вище властивостей частот (невід'ємність, нормованість та адитивність) природно постулювати такі аксіоми ймовірності:

Означення. Ймовірність *P є числовою функцією на класі $\mathcal{F}$ всіх випадкових подій з такими властивостями:*

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

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

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

$$P(\bigcup_{n\geq 1} A_n) = \sum_{n\geq 1} P(A_n).$$

Функція P називається адитивною ймовірністю, якщо замість аксіоми (P3) виконується слабша умова

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

$$P(A\cup B) = P(A) + P(B).$$

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

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

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

Другу групу аксіом теорії ймовірностей можна сформулювати так:

"Ймовірність є сигма-адитивною невід'ємною нормованою функцією на класі всіх випадкових подій".

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

Ω – простір елементарних подій (абстрактна множина),

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

P – ймовірність на сигма-алгебрі подій $\mathcal{F}$ із властивостями (P1 – P3).

2.4. Властивості ймовірності

З аксіом теорії ймовірностей випливають такі властивості ймовірності.

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

1. Ймовірність доповнення події $A \in F$

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

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

2. Ймовірність неможливої події**:** $P(\varnothing) = 0$, оскільки $\varnothing = \overline{\Omega}$. Зауважимо, що з $P(A) = 0$ не випливає, що A – неможлива подія.

3. Ймовірність вкладеної різниці: для $A, B \in F$, $A \subset B$

$$P(B \setminus A) = P(B) - P(A),$$

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

4. Монотонність ймовірності: для $A, B \in F$, $A \subset B$

$$P(A) \leq P(B),$$

виводиться з невід'ємності та попереднього пункту:

$$P(B) - P(A) = P(B \setminus A) \geq 0.$$

5. Множина значень ймовірності

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

оскільки $\varnothing \subset A \subset \Omega$ і має місце монотонність ймовірності

6. Ймовірність об'єднання двох подій $A, B \in \mathcal{F}$

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

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

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

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

$$P(\bigcup\nolimits_{1 \le k \le n} A_k) = \sum\nolimits_{1 \le\ i \le\ n} P(A_i) - \sum\nolimits_{1 \le i < j \le n} P(A_i \cap A_j) +$$

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

При $n = 2$ ця формула збігається із властивістю 6, а для $n > 2$ вона доводиться за індукцією із використанням властивості ймовірності об'єднання. Дійсно, нехай формула справедлива для заданого n. Позначаючи $B = \cup_{k=1}^{n} A_k$, маємо

$$P(\cup_{k=1}^{n+1} A_k) = P(B \cup A_{n+1}) = P(B) + P(A_{n+1}) - P(B \cap A_{n+1}) =$$
$$\sum\nolimits_{1 \le k \le n} P(A_k) - \sum\nolimits_{1 \le k < j \le n} P(A_k \cap A_j) + .. + P(A_{n+1}) - P(\cup_{k=1}^{n} (A_k \cap A_{n+1}))$$
$$= \sum\nolimits_{1 \le k \le n+1} P(A_k) - \sum\nolimits_{1 \le k < j \le n} P(A_k \cap A_j) + .. - \sum\nolimits_{1 \le k \le n} P(A_k \cap A_{n+1}) + ..,$$

звідки за індукцією дістанемо шукане твердження.

8. Напівадитивність ймовірності:

$$P\left(\bigcup{}_{1\le k\le n} A_k\right) \le \sum{}_{1\le k\le n} P(A_k)$$

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

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

9. Неперервність ймовірності**.**
(а) Якщо $A_n \uparrow A$, то $P(A_n) \uparrow P(A)$.
(б) Якщо $A_n \downarrow A$, то $P(A_n) \downarrow P(A)$.

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

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

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

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

10. Зліченна напівадитивність ймовірності

$$P\left(\bigcup A_k\right) \le \sum P(A_k)$$

випливає із напівадитивності та неперервності ймовірності:

$$P\left(\bigcup A_k\right) = \lim{}_{n\to\infty}\ P\left(\bigcup\nolimits_{1\le k\le n} A_k\right) \le \lim{}_{n\to\infty} \sum\nolimits_{1\le k\le\ n} P(A_k).$$

11. Еквівалентність неперервності в нулі та сигма-адитивності для адитивної ймовірності.

Будемо говорити, що ймовірність P на класі випадкових подій $\mathcal{F}$ неперервна в нулі, якщо для будь-якої послідовності подій $A_n \in \mathcal{F}$ такої, що $A_n \downarrow \varnothing$, має місце збіжність $P(A_n) \downarrow 0$.

Для того, щоб адитивна ймовірність P на сигма-алгебрі подій $\mathcal{F}$ була сигма-адитивною, необхідно і достатньо, щоб вона була неперервною в нулі на $\mathcal{F}$.

Доведення. Якщо P сигма-адитивна, то вона неперервна в нулі за властивістю неперервності ймовірності.

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

12. Продовження неперервної ймовірності на сигма-алгебру.

Для того, щоб адитивна ймовірність P на алгебрі подій $\mathfrak{A}$ мала продовження до сигма-адитивної ймовірності на породженій сигма-алгебрі $\mathcal{F} = \sigma[\mathfrak{A}]$, необхідно і достатньо, щоб вона була неперервною в нулі на алгебрі $\mathfrak{A}$ – тобто для будь-якої послідовності подій $A_n \in \mathfrak{A}$ таких, що $A_n \downarrow \varnothing$, повинна мати місце збіжність $P(A_n) \downarrow 0$.

Доведення. Якщо ймовірність P сигма-адитивна, то вона неперервна в нулі за властивістю неперервності.

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

Вправа.

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

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

(3). Для довільних подій A, B справедлива нерівність
$P(A \cup B)P(A \cap B) \leq P(A)P(B)$.

3. Приклади ймовірнісних просторів

Для застосувань ймовірнісних методів на практиці необхідно будувати відповідні ймовірнісні простори. Тому розглянемо конкретні приклади таких просторів.

3.1. Класичне означення ймовірності

Нехай простір $\Omega = \{\omega_1, ..., \omega_n\}$ – скінченний, а клас подій $\mathcal{F}$ 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, $P(\{\omega_k\}) = 1/n$, $k = \overline{1, n}$. 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* $(\Omega, \mathcal{F}, P)$ *зі скінченним простором елементарних подій* Ω *прийняте* класичне означення ймовірності, *якщо для будь-якої події* $A \subset \Omega$ *її ймовірність дорівнює відношенню кількості елементарних подій, що сприяють* A, *до кількості усіх елементарних подій:*

$$P(A) = |A| \ / \ |\Omega| \,,$$

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

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.* Ймовірність парної кількості очок дорівнює $|\{2,4,6\}|\,/6=3/6=1/2$.

2. *Підкидання навмання двох симетричних монет.* Ймовірність хоча б одного аверса $|\{\text{AP,PA,AA}\}|\,/4=3/4$.

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 слід вважати рівноможливими, ймовірність кожної дорівнює 2^{-n}.

Припустимо, що частинка в момент k змінює свою координату (стрибає) на величину ε_k. Позначимо $S_n=\varepsilon_1+...+\ \varepsilon_n$ результуюче положення частинки. Тоді ймовірність наявності k додатних та $n-k$ від'ємних стрибків дорівнює

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

Зокрема, при $n = 2m$ та $k = m$ імовірність повернення в початок має вигляд

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

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$\}$. Тоді за формулою включення–виключення

$$P(A) = P(\cup_{k=1}^n A_k) =$$

$$\sum\nolimits_{1 \le i \le n} P(A_i) - \sum\nolimits_{1 \le i < j \le n} P(A_i \cap A_j) + .. + (-1)^{n-1}\ P(A_1 \cap \ldots \cap A_n) =$$

$$C_n^1 \frac{(n-1)!}{n!} - C_n^2 \frac{(n-2)!}{n!} + \ldots + (-1)^{n-1} C_n^n \frac{0!}{n!} =$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n-1} \frac{1}{n!} \to 1 - \frac{1}{e},\ n \to \infty.$$

Вправа. 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 ?

3.2. Дискретний імовірнісний простір

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

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

Вправа. Клас $\digamma = 2^{\Omega}$ тоді й тільки тоді, коли вимірні всі одноточкові множини: $\{\omega_n\} \in \digamma$ для всіх n.

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

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

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

Отже, для кожної ймовірності P на $\digamma$ визначений дискретний розподіл на одноточкових множинах.

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

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

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

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

Доведення. Адитивність є наслідком адитивності суми.

Можна вважати, що $\Omega = \mathbb{N}$. Нехай $A_n \downarrow \varnothing$. Тоді $\quad a_n = \inf A_n \uparrow \infty$,

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

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 $p_n = P(\{\omega_n\}) = 2^{-n}, n \geq 1$. Ця послідовність дійсно є дискретним розподілом.

Тому ймовірність парної кількості підкидань до успіху дорівнює $\sum_{k=1}^{\infty} 2^{-2k} = 1/3$.

2. Припустимо, що при підкиданні монети, окрім випадіння аверсу (А) та реверсу (Р), можливе також 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 $P(\{\omega_n\}) = 3^{-n}$. Ймовірність 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 $P(A) = p$, $P(B) = q$, $P(A \cap B) = r$?

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

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

то шукана необхідна і достатня умова має вигляд:

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

Вправа. Нехай $(\Omega,\ \mathcal{F},\ P)$ – довільний імовірнісний простір. Випадкова подія A називається атомом, якщо $P(A) > 0$ та для будь-якої події з $B \subset A$ випливає або $P(B) = 0$, або $P(A \backslash B) = 0$. Атоми A, B вважаються однаковими, якщо $P(A \Delta B) = 0$. Довести такі твердження:. (а) Якщо A, B – атоми, то або $P(A \cap B) = 0$, або $P(A \Delta B) = 0$; (б) Множина всіх атомів на даному просторі не більш ніж зліченна; (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 числа $p \in [0, P(A_0)]$ знайдеться подія A така, що $P(A) = p$.

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 ймовірностей окремих точок до ймовірностей інтервалів.

Нехай $\Omega = [a, b)$, сигма-алгебра подій $\mathcal{F}$ містить всі інтервали $[x, y)$, і на $\mathcal{F}$ задана деяка ймовірність 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

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

Дійсно, ймовірності $P([a + (b - a)\frac{k}{n},\ a + (b - a)\frac{k+1}{n}))$ при $0 \leq k < n$ мають не залежати від k та в сумі становити 1, отже кожна з них дорівнює $1/n$. Звідси $P([a, a+(b-a)\frac{k}{n})) = \frac{k}{n}$. Обираючи підпослідовність $k/n \uparrow (y-a)/(b-a)$, із неперервності ймовірності виводимо, що

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

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 $\digamma$ містить усі підмножини Ω із визначеним d-вимірним об'ємом. Будемо говорити, що у просторі $(\Omega, \ \digamma, \ P)$ прийнято* геометричне означення ймовірностей, *якщо для всіх $A \in \digamma$*

$$P(A) = |A| \ / \ |\Omega| \, ,$$

де $|\bullet|$ – d-вимірний об'єм множини, 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 є випадковими. Прийшовши до місця, кожен чекає на іншого протягом $1/4$ 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 інтерпретувати на користь прийняття геометричного означення ймовірностей. Подія, що відповідає зустрічі, має вигляд

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

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

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*. Внаслідок інваріантності довжини відносно поворотів одну вершину можна зафіксувати. Тому $\Omega = [0, 2\pi]$, множина елементарних подій, які сприяють A, містить дві дуги довжини $\pi/3$, а ймовірність A дорівнює

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

(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

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

(в) Хорда визначається *випадковою точкою перетину з перпендикулярним до неї радіусом* всередині кола. Простір $\Omega = [0, 1]$, а елементарні події з A утворюють кільце внутрішнього радіуса $\sqrt{3}/2$. Тому

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

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.

Припустимо, що довжина голки не перевищує відстані між паралельними прямими на площині: $2l \leq 2a$. Тому перетин голки із найближчою прямою можливий тоді й тільки тоді, коли $x \leq l \sin \varphi$. Обчислюючи за геометричним означенням площу відповідної множини елементарних подій, знаходимо

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

Відзначимо, що цей результат дає можливість статистичного оцінювання числа π через емпіричну частоту перетинів $\nu_n(A)$ з формули $\pi \approx 2l \ n \ / \ a\nu_n(A)$.

Вправа. 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 ?

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 $\Omega = \{\omega_1, \ldots, \omega_n\}$ прийнято класичне означення ймовірностей, $A, B \subset \Omega$, $n_A = |A|$, $n_B = |B|$, $n_{AB} = |A \cap B|$. За означенням, $P(A) = n_A/n$. Припустимо, стає відомим, що подія B вже відбулася. Тоді кількість усіх можливих у наступному 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 слід обрати рівною

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

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

$$P(A \;/\; B) = P(A \cap B) \;/\; P(B).$$

Приклади.

1. *Зміна шансів на хоча б одну шістку при двох підкиданнях кості, якщо відома сума очок.* Ймовірність 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

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

2. *Вибір двох тузів із колоди без повернення.* Ймовірність 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 $B \in \mathcal{F}$ і $P(B) > 0$. Тоді*

(а) $P(\,\cdot\,/\,B) \geq 0$,

(б) $P(\Omega\,/\,B) = P(B\,/\,B) = 1$,

(в) $P(\cup A_n\,/\,B) = \sum P(A_n\,/\,B)$ *для попарно несумісних* A_n.

Доведення.

(а) За означенням відношення невід'ємних чисел,

(б) випливає з рівності $P(B \cap B) = P(B)$,

(в) з означення умовної ймовірності та сигма-адитивності ймовірності:

$$P(\cup A_n\,/\,B) = P((\cup A_n) \cap B)/P(B) = P(\cup(A_n \cap B))/P(B) = \sum P(A_n \cap B)/P(B) = \sum P(A_n\,/\,B),$$

оскільки події $A_n \cap B$ попарно несумісні $\boxdot$

Наслідок. Умовна ймовірність $P(A\,/\,B)$ 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).

4.1. Ймовірність 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 ймовірності 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* $A, B \in \mathcal{F}$ *і* $P(B) > 0$. *Тоді*

$$P(A \cap B) = P(B) \cdot P(A \,/\, B).$$

Доведення очевидне $\boxdot$

Зауваження. У випадку, коли $P(B) = 0$, наведена формула також справедлива, оскільки за монотонністю ймовірності $P(A \cap B) \leq P(B) = 0$, а права частина завжди нульова.

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

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

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

5. Формули повної ймовірності та Байєса

Розглянемо деякий двохетапний стохастичний експеримент (наприклад, послідовний вибір 2 тузів 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 висловлювання полягає в тому, що* $(H_k,\ k \geq 1)$ *утворюють* розбиття *простору* Ω.

5.1. Формула повної ймовірності

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

$$P(A) = \sum\nolimits_{k\geq 1} P(A \mathbin{/} H_k)\ P(H_k).$$

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

$$P(A) = \sum\nolimits_{k\geq 1} P(A \cap H_k) = \sum\nolimits_{k\geq 1} P(A \mathbin{/} H_k)\ P(H_k).$$

Зауваження. Дана формула справедлива також у випадку, коли $P(H_k) = 0$ для деяких k, оскільки за означенням добуток 0 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 ймовірностей $P(H_k) = 1/n$ та $P(A \ / \ H_k) = k/n$. Отже, $P(A) = \sum_{k=1}^{n}(k/n)/n = (n+1)/2n$.

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 $k \le m$ щасливих білетів. Позначимо через $A_n, \ n \ge 1$, 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 ймовірностей $P(A_1) = k/m$. Тому за формулою повної ймовірності

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

За індукцією доводимо, що $P(A_j) = k/m$ при $j \le m$.

5.2. 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* $(H_k, \ k \ge 1)$ – *повна група подій, і* $B \in \mathcal{F}$ *із* $P(B) > 0$. *Тоді*

$$P(H_k \ / \ B) = \frac{P(B \ / \ H_k) \ P(H_k)}{\sum_{j \ge 1} P(B \ / \ H_j) \ P(H_j)}.$$

Доведення. За означенням умовної ймовірності та за формулою повної ймовірності

$$P(H_k\ /\ B) = \frac{P(H_k \cap B)}{P(B)} = \frac{P(B\ /\ H_k)\ P(H_k)}{\sum P(B\ /\ H_j)\ P(H_j)}.$$

Приклади.

1. *Оцінка невідомого складу урни за кольором випадково 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, вона виявилася білою. Як треба змінити ймовірності гіпотез щодо складу урни ?

Нехай A – подія, що полягає у виборі білої кулі. За умовою рівноможливості гіпотез $P(H_k) = 1/(n+1)$, $P(H_k/A) = \frac{(k/n)(1/(n+1))}{\sum_{k=0}^{n}(k/n)(1/(n+1))} = \frac{2k}{n(n+1)}$.

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 2 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

$P(H_1) = 10000/(10000 + 80000) = 1/9$,

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

6. Незалежні випадкові події

У теорії ймовірностей використовується специфічне поняття незалежності, яке можна назвати статистичною незалежністю. Події вважають незалежними, якщо наявна інформація про одну з них не змінює шансів для іншої.

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

Використовуючи означення умовної ймовірності $P(A \,/\, B) = P(A \cap B) \,/\, P(B)$, отримуємо еквівалентне, але більш загальне та симетричне означення попарної незалежності.

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

$$P(A \cap B) = P(A)\, P(B).$$

Зауваження. При формулюванні складних стохастичних експериментів часто вказують ймовірності 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 $P(A_1 \cap A_2) = 1/36$. Еквівалентним підходом є прийняття рівноможливості $P(A_k) = 1/6$ 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 $P(A) = |A| / |\Omega|$. Якщо ж постулювати випадковість вибору кожної координати точки та незалежність координат, маємо еквівалентний вираз $P(A) = (|A_1| / 1) \times (|A_2| / 1) = |A| / 1$.

Вправа. Скільки існує пар незалежних подій у ймовірнісному просторі з простою кількістю елементарних подій при класичному означенні ймовірностей ?

6.1. Перетворення незалежних подій

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 $B_1 \cap B_2 = \varnothing$, то A і $B_1 \cup B_2$ також незалежні.

(г) 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.

Доведення.

(а) $P(A \cap \Omega) = P(A) = P(A)\ P(\Omega)$.

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

(в) $P(A \cap (B_1 \cup B_2)) = P((A \cap B_1) \cup (A \cap B_2)) = P(A \cap B_1) + P(A \cap B_2) = P(A)(P(B_1) + P(B_2)) = P(A)\ P(B_1 \cup B_2)$.

(г) $P(A \cap (B_2 \setminus B_1)) = P(A \cap B_2 \setminus A \cap B_1) = P(A \cap B_2) - P(A \cap B_1) = P(A)(P(B_2) - P(B_1)) = P(A)\ P(B_2 \setminus B_1)$ $\boxdot$

Вправа. Довести, що подія A не залежить від себе самої тоді й тільки тоді, коли $P(A) \in \{0, 1\}$.

Більш загальне поняття незалежності стосується ймовірностей одночасної появи декількох 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* $1 \le k_1 < k_2 < \ldots < k_m \le n$

$$P(A_{k_1} \cap A_{k_2} \cap \ldots \cap A_{k_m}) = \prod_{1 \le\ i\ \le m} P(A_{k_i}).$$

Вправа. Обчислити ймовірність об'єднання незалежних у сукупності подій через імовірності окремих подій.

6.2. Приклад Бернштейна залежних у сукупності величин

Очевидно, що з незалежності в сукупності випливає попарна незалежність подій з послідовності $(A_k,\ k = \overline{1,n})$ – тобто незалежність кожної пари подій $(m = 2)$. Приклад С.Н.Бернштейна свідчить, що обернене твердження хибне – *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

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

$P(\text{С}\cap\text{Ж}) = P(\text{Ж}\cap\text{Ч}) = P(\text{С}\cap\text{Ч}) = 1/4 = P(\text{С}\cap\text{Ж}\cap\text{Ч}) \neq 1/8.$
Отже, події С,Ж,Ч незалежні попарно, але залежні в сукупності.

7. Дискретні випадкові величини

Поняття випадкової події дозволяє вивчати якісні (дихотомічні) наслідки стохастичних експериментів. На практиці важливо вивчати також кількісні результати (наприклад: кількість аверсів при 2 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* $\xi : \Omega \to \mathbb{R}$, *що 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 *прообраз* $\{\omega : \xi(\omega) = x_n\}$ *є випадковою подією, тобто*

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

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

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

7.1. Розподіл дискретної випадкової величини

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

$$(p_n \equiv P(\xi = x_n),\ n \geq 1).$$

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 вираз вигляду $P(\xi = x)$ треба розуміти як $P(\{\omega : \xi(\omega) = x\})$.

Розподіл дискретних випадкових величин часто зображують у вигляді таблиці

$$\begin{bmatrix} \xi & x_1 & x_2 & \dots & x_n & \dots \\ P & p_1 & p_2 & \dots & p_n & \dots \end{bmatrix}$$

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

$$P(g(\xi) \in B) = \sum_{n:g(x_n)\in B} p_n.$$

Доведення є наслідком зображення

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

7.2. Індикаторна випадкова величина

Означення. Нехай A – випадкова подія. Її Індикаторною величиною 1_A називається двозначна випадкова величина

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

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

7.3. Проста випадкова величина

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

Очевидно, що сума, добуток та функції від простих величин є простими.

Якщо ξ – проста випадкова величина з множиною різних значень $\{x_1, ..., x_n\}$, то $C_k = \{\xi = x_k\} \in \mathcal{F}$ утворюють повну групу подій і має місце зображення

$$\xi = \sum_{k=1}^{n} x_k 1_{C_k}.$$

7.4. Схема стохастичних випробувань Бернуллі

Значна кількість дискретних розподілів пов'язана із такою схемою випробувань.

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) ймовірність успіху не залежить від номера випробування.

Для побудови ймовірнісного простору, який відповідає схемі Бернуллі, позначимо через n кількість випробувань, p – ймовірність успіху в одному випробуванні, $q = 1 - p$ – ймовірність 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: $\mathcal{F} = 2^\Omega$.

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

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

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

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

Звідси

$$P(\{\omega : \omega = \varepsilon\}) = \prod_{k=1}^n P(\{\omega : \omega_k = \varepsilon_k\}) =$$
$$\prod_{k=1}^n p^{\varepsilon_k} q^{1-\varepsilon_k} = p^{\nu_n(\varepsilon)} q^{n-\nu_n(\varepsilon)},$$

де функція
$$\nu_n(\omega) = \sum_{k=1}^{n} \omega_k = |\{k : \omega_k = 1\}|$$
збігається з загальною кількістю успіхів в елементарній події $\omega = (\omega_1, ..., \omega_n)$. Як в усякому дискретному ймовірнісному просторі, ймовірність будь-якої події дорівнює сумі ймовірностей сприятливих елементарних подій.

7.5. 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 ξ зауважимо, що за означенням $\xi(\omega) = \nu_n(\omega)$, де функцію ν_n визначено вище при описанні схеми Бернуллі. Тому

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

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

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

Послідовність $(p_n(k),\ k = \overline{0, n})$ називається біноміальним розподілом імовірностей.

Вправа. (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.

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 шуканий висновок $\boxdot$

Вправа. У якому випадку найбільш імовірне значення біноміального розподілу досягається у двох точках ?

7.6. Геометричний розподіл

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

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

$$P_N(\xi = k) = \sum_{\omega_{k+1},..\omega_N \;\in\{0,1\}} P_N(\{(0...01\omega_{k+1}, ..., \omega_N)\}) =$$

$$P_N(\bigcup_{\omega_{k+1},..\omega_N \;\in\{0,1\}} \{(0...01\omega_{k+1}, ..., \omega_N)\}) = P_N(\{(0..01)\}) = q^{k-1}p.$$

Отже, геометричний розподіл імовірностей має вигляд

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

7.7. Негативний геометричний розподіл

Означення. Дискретна випадкова величина ξ має *негативний геометричний 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

$$P(\xi = k) = C_{k-1}^{r-1}\ (1-p)^{k-r}p^{r},\ \ k = 1, 2, \ldots$$

Дійсно, 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 $n-1$ місці.

7.8. 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.

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

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

Зауваження. Якщо проінтерпретувати N як загальну кількість виробів, що виготовлені на підприємстві, n – як кількість бракованих серед них, m – як об'єм 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

Вправа. Найбільш імовірне значення k дорівнює $(n+1)(m+1)\ /\ (N+1)-1 \simeq n\ m\ /\ N$. Вказівка: відношення послідовних імовірностей має вигляд $(k+1)(N-n-m+k+1)/(n-k)(m-k)$.

7.9. Розподіл Пуассона

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

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

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.

7.10. 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 із ймовірністю успіху p_n. Позначимо ν_n – кількість успіхів у n-ій серії.*

Припустимо, що одночасно $n \to \infty$ та $p_n \to 0$ так, що $np_n \to \lambda > 0$. Тоді

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

Доведення.

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

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

$$(np_n)^k \to \lambda^k, \quad (1 - \tfrac{np_n}{n})^n \to e^{-\lambda} \boxdot$$

Приклад. За 1 хвилину сервер виконав 2000 з'єднань у мережі, з імовірністю 0.001 кожне з них було перерване, незалежно від інших. Тоді ймовірність того, що перервалися принаймні 2 з'єднання, наближено дорівнює $P(\nu_{1000} \ge 2) \approx 1 - e^{-2} - \frac{2^1}{1!} e^{-2} \approx 0.594$.

Вправа. Довести, що при $np = \lambda$ сума модулів різниць між імовірностями $P(\nu_n = k)$ та їх граничними значеннями не перевищує $2\lambda^2/n$.

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

Теорема (**локальна гранична теорема Муавра-Лапласа**). *Розглянемо схему серій випробувань Бернуллі, в якій n-та серія містить n випробувань із ймовірністю успіху p, $q = 1-p$. Позначимо 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 1):

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

де функція

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

є так званою стандартною нормальною щільністю розподілу.

Доведення теореми спирається на формулу Стірлінга

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

Зауважимо, що в даній граничній схемі одночасно $n \to \infty$, $k \to \infty$, $n - k \to \infty$, 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

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

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

Теорема (інтегральна гранична теорема Муавра-Лапласа). *За припущень локальної теореми для всіх $a < b$ має місце при $n\to\infty$ збіжність*

$$P(\ a < \frac{\nu_n - np}{\sqrt{npq}}\ < b) \to \Phi(b) - \Phi(a),$$

де $\Phi(x) = \int_{-\infty}^{x}\varphi(y)dy$ – *стандартна нормальна функція розподілу.*

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

$$\sum\nolimits_{k:\ (k-np)/\sqrt{npq}\ \in\ (a,b)} p_n(k) = \sum\nolimits_{k=np+a\sqrt{npq}}^{np+b\sqrt{npq}} (\sqrt{npq}\ p_n(k))\frac{1}{\sqrt{npq}},$$

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

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

8. Загальне означення випадкової величини та вектора

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

8.1. Борелєва сигма-алгебра, борелєві множини

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

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

Скінченні об'єднання несумісних напівінтервалів утворюють алгебру множин

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

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 ж 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.

Вправа.

(1) Борелєвими є всі інтервали та відрізки, одноточкові множини, всі відкриті множини (як зліченні об'єднання відкритих інтервалів) та замкнені множини (як доповнення відкритих).

(2) Сигма-алгебра, що породжена відкритими підмножинами $\mathbb{R}$, збігається з борелєвою сигма-алгеброю.

(3) Виходячи з поняття міри Лебега, побудувати приклад неборелєвої множини.

Аналогічно, система n-вимірних кутів

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

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 $B_1 \times ... \times B_n \in \mathfrak{B}(\mathbb{R}^n)$, якщо $B_k \in \mathfrak{B}(\mathbb{R})$. 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 та є сигма-алгеброю, отже, збігається з $\mathfrak{B}(\mathbb{R})$. Тому

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

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

8.2. 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 функція* $\xi : \Omega \to \mathbb{R}$, *така, що для довільної борелевої множини* B *її прообраз* $\{\omega : \xi(\omega) \in B\}$ *є випадковою подією, тобто:*

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

Функції ξ, які задовольняють наведену умову, називаються вимірними відносно сигма-алгебри $\mathcal{F}$. 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 (невласною) випадковою величиною 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 множини, яка не є випадковою подією $\xi = 1_C$, $C \notin \mathcal{F}$.

8.3. Критерій вимірності функції

Теорема (**про критерій вимірності скалярної функції**). *Нехай ξ – довільна функція, $\xi : \Omega \to \mathbb{R}$. Вона є випадковою величиною тоді й тільки тоді, коли прообрази напівінтервалів $\{\xi < x\}$ є випадковими подіями: $\{\xi < x\} \in \mathcal{F}$, $\forall x \in \mathbb{R}$.*

Доведення. З властивостей операції обчислення прообразу випливає, що клас

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

є сигма-алгеброю, оскільки прообраз доповнення та об'єднання множин збігається відповідно з доповненням та 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 шукане твердження є наслідком означення випадкової величини $\boxdot$

Як наслідок отримуємо таку теорему.

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

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

8.4. Випадковий вектор

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

Вправа. Як і для випадкових величин, довести, що функція $\xi : \Omega \to \mathbb{R}^n$ є випадковим вектором тоді й тільки тоді, коли прообраз кута є випадковою подією:

$$\{\xi < x\} \equiv \{\xi_1 < x_1, ..., \xi_n < x_n\} \in \mathcal{F}$$

для всіх $x = (x_1, ..., x_n) \in \mathbb{R}^n$, де ξ_k – k-та координата вектора ξ.

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

Доведення цілком аналогічне доведенню теореми для скалярних величин.

Зокрема, кожна координата випадкового вектора є випадковою величиною, оскільки координатне відображення $\pi_k : \mathbb{R}^n \to \mathbb{R}$ неперервне і є борелєвим.

Виконується й обернене твердження: вектор, утворений із n випадкових величин, є випадковим вектором у $\mathbb{R}^n$. Дійсно, прообраз n-вимірного кута є перерізом прообразів кутів для координатних відображень:

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

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.*

Старт
Початок
Зміст
Сторінка 60 з 509
Назад
Екран
Закрити
Выхід

Доведення теореми випливає з борелєвості (що є наслідком неперервності) відповідних функцій $|x|$, $x+y$, $x-y$, $x\cdot y$, x/y, $\max(x,y)$ від випадкового вектора (ξ,η) $\boxdot$

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

8.5. Вимірність границі випадкових величин

Теорема (про вимірність границі випадкових величин). *Нехай $(\xi_n,\ n\ge 1)$ – випадкові величини і границя $\xi(\omega)=\lim\xi_n(\omega)$ існує при кожній ω. Тоді ξ – випадкова величина.*

Доведення випливає з тотожності

$$\{\lim\xi_n<x\}=\bigcup\nolimits_{k\ge1}\bigcup\nolimits_{m\ge1}\bigcap\nolimits_{n\ge m}\{\xi_n<x-1/k\}.$$

Її справедливість обгрунтовується 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 і тільки тоді, коли для деякого $k<\infty$ всі члени послідовності менші за $x-1/k$, починаючи з деякого номера $\boxdot$

Наслідок. Якщо $(\xi_n,\ n\ge1)$ – випадкові величини, то $\overline{\lim}\xi_n$, $\underline{\lim}\xi_n$ – також випадкові величини (можливо, невласні).

Доведення грунтується на двох попередніх теоремах і випливає з того, що вказані граничні значення є монотонними границями випадкових величин. Зокрема,

$$\overline{\lim}\xi_n=\lim\nolimits_{N\to\infty}\lim\nolimits_{m\to\infty}\ \max\nolimits_{N\le\ n\ \le N+m}\ \xi_n\ \boxdot$$

Вправа.Якщо $(\xi_n,\ n\ge1)$ випадкові величини, то $\inf\xi_n$, $\sup\xi_n$ – також випадкові величини, можливо, невласні.

8.6. Апроксимація простими величинами

У наступній теоремі та в подальшому символом $x_n \uparrow x$ 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) $\lim x_n = x$. Аналогічний зміст матиме запис $x_n \downarrow x$.

Теорема (про апроксимацію випадкових величин простими).

(а) Нехай ξ – невід'ємна випадкова величина. Тоді існує послідовність простих випадкових величин ξ_n таких, що $0 \leq \xi_n \uparrow \xi$ при всіх ω.

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

Доведення.

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

$$\varphi_n(x) = \sum\nolimits_{k=0}^{n2^n-1} k2^{-n}\, 1_{k2^{-n} \leq x < (k+1)2^{-n}} + n\, 1_{n \leq x}.$$

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 не спадає.

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: $0 \le \xi_n = \varphi_n(\xi) \uparrow \xi$ для всіх $\omega \in \Omega$.

(б) Визначимо

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

додатну та від'ємну частини величини ξ. Це невід'ємні випадкові величини за теоремою про вимірність суперпозиції.

Нехай $\xi_n^\pm$ прості величини, що внаслідок (а) апроксимують величини ξ^+ та ξ^- відповідно. Тоді $\xi_n = \xi_n^+ - \xi_n^-$ є шуканою простою величиною, оскільки

$$\xi = \xi^+ - \xi^- = \lim \xi_n^+ - \lim \xi_n^-, \ \ |\xi_n| = \xi_n^+ + \xi_n^- \le \xi^+ + \xi^- = |\xi| \ \boxdot$$

9. Функція розподілу випадкової величини та її властивості

Сукупність ймовірностей подій вигляду $\{\xi = x\}$ не завжди повністю описує величину ξ . Наприклад, для випадкової точки на відрізку всі ці ймовірності дорівнюють нулю. 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*

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

9.1. Властивості функції розподілу

Теорема (про основні властивості функції розподілу). *Нехай ξ – випадкова величина, а $F \equiv F_\xi$ її функція розподілу. 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 границя

$$F(x-0) = \lim\,_{y\,\uparrow\,x,\ y\,<\,x} F(y), \text{ відповідно}$$
$$F(-\infty) = \lim\,_{y\,\downarrow\,-\infty} F(y),\ F(+\infty) = \lim\,_{y\,\uparrow\,+\infty} F(y).$$

Існування та єдиність границь у (2) та (3) випливає з монотонності (1) функції розподілу).

Доведення.

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

(2) Відомо, що границя монотонної функції не залежить від вибору апроксимуючої послідовності аргументів. Тому аргумент y можна вважати цілозначним. 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

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

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

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

(3) Клас відкритих напівінтервалів $(-\infty, x), x \in \mathbb{R}$, є неперервним зліва по відношенню до 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

$$F(x-0) = \lim_{n \uparrow \infty} F(x - \tfrac{1}{n}) = \lim_{n \uparrow \infty} P(\xi < x - \tfrac{1}{n}) =$$
$$P(\xi < x) = F(x) \boxdot$$

Теорема (про властивості функції розподілу). *Справедливі такі рівності:*

(a) $P(a \leq \xi < b) = F_\xi(b) - F_\xi(a)$,

(б) $P(\xi = x) = F_\xi(x+0) - F_\xi(x)$,
(в) $P(\xi \geq x) = 1 - F_\xi(x)$,
(г) $P(\xi \leq x) = F_\xi(x+0)$,
(д) $P(\xi > x) = 1 - F_\xi(x+0)$.

Доведення.

(а) З формули для ймовірності вкладеної різниці подій отримуємо

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

(б) Зі співвідношення $\{\xi = x\} = \cap_{n>1}\{x \leq \xi < x + \frac{1}{n}\}$, неперервності ймовірності та твердження (а) дістанемо

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

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

(г,д) Виводяться з означення функції розподілу та (б).

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

Вправа. Довести, що функція розподілу має не більш ніж зліченну множину точок розриву.

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*

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

значення якої задається рівністю

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

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.*

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 є наслідком невід'ємності та адитивності $\boxdot$

9.2. Неперервність у нулі міри Лебега – Стілтьєса

Теорема (про неперервність у нулі міри Лебега – Стілтьєса). *Адитивна міра 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 $F(\overline{[-c,c)}) \equiv F(-c) + 1 - F(c) < \varepsilon$. Це можливо за умови нормованості функції розподілу.

2. Позначимо $B_n = A_n \cap [-c,c) \in \mathfrak{A}(\mathbb{R})$. Тоді за монотонністю ймовірності $F(A_n \backslash B_n) \le F(\overline{[-c,c)}) < \varepsilon$.

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

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

Це можливо, оскільки 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

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

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

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

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 $n \geq m$, $k \leq n$ з урахуванням того, що $B_n \subset B_k$, а міра F напівадитивна:

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

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

6. Залишилося зазначити, що

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

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

Теорема про еквівалентність неперервності в нулі та сигма-адитивності для адитивних мір наведена вище $\boxdot$

Теорема (про існування міри Лебега – Стілтьєса). *Нехай F – функція розподілу. Тоді 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 $\mathfrak{B}(\mathbb{R})$ $\boxdot$

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

9.3. Обчислення ймовірностей

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*

$$P(\xi \in B) = F_\xi(B),$$

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

Доведення. За означенням шукана 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 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, для скорочення розподілом випадкової величини ξ будемо називати відповідну міру Лебега – Стілтьєса F_ξ на $\mathfrak{B}(\mathbb{R})$.

10. Абсолютно неперервні випадкові величини

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

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

Наступний клас містить виключно неперервні функції розподілу.

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

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

де інтеграл слід розуміти

(а) як інтеграл Рімана – Стілтьєса для кусково-неперервної функції $f_\xi(x)$,

(б) або ж як інтеграл Лебега – Стілтьєса для вимірної функції $f_\xi(x)$.

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

Зауваження. Поняття інтегралу 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), (2) та (3) означення функції розподілу.

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

$$f_\xi(x) = F'_\xi(x) = \frac{dF_\xi(x)}{dx}$$

10.1. Обчислення ймовірностей

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

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

Доведення. Шукана рівність виконується для напівінтервалів $B = (-\infty, x)$ 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 всій породженій σ-алгебрі $\mathfrak{B}(\mathbb{R})$ $\boxdot$

Вправа. Точка $x \in \mathbb{R}$ називається точкою зростання функції розподілу F, якщо $F(x+\varepsilon) > F(x-\varepsilon)$ для всіх $\varepsilon > 0$. Побудуйте приклад дискретної функції розподілу, для якої множина точок зростання збігається з $\mathbb{R}$.

10.2. Класифікація функцій розподілу

Одночасно з класами абсолютно неперервних та дискретних функцій розподілу існує також екзотичний клас сингулярних функцій розподілу – вони є неперервними, але множина їх точок зростання має нульову довжину (міру Лебега).

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

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.

Приклади абсолютно неперервних величин.

10.3. 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.*

Це означає, що ймовірність попадання величини в якусь 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 випливає, що щільність розподілу рівномірної на $[a, b]$ величини дорівнює

$$f_\xi(x) = \frac{1}{b-a} 1_{x \in [a,b]},$$

а функція розподілу має вигляд

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

10.4. Нормальний (Гауссів) розподіл

(a) **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. 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*.*

Для обчислення щільності $f_\xi(x)$ застосовується (**Вправа**) наведений нижче наслідок про розподіл лінійного перетворення випадкової величини.

10.5. Показниковий (експоненційний) розподіл

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

$$f(x) = \lim \tfrac{1}{h}\, P(x \le \xi < x+h) = \lim \tfrac{1}{h} \int_x^{x+h} f(y)dy$$

при $h \to 0$ – тобто як "щільність ймовірності".

На практиці одночасно із безумовною ймовірністю $P(x \le \xi < x + h)$ важливе значення відіграє умовна ймовірність $P(x \le \xi < x + h \ / \ \xi \ge x)$. 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 із самого початку випробувань.

Тому відповідним аналогом щільності є функція, що дорівнює границі для умовних ймовірностей

$$\lambda(x) = \lim{}_{h \downarrow 0} \tfrac{1}{h} P(x \le \xi < x + h \ / \ \xi \ge x).$$

За означенням умовної ймовірності

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

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

$$P(x \le \xi < x + h \ / \ \xi \ge x) = h\lambda(x) + o(h), \ \ h \to 0.$$

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

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

де використано заміну змінної $u = F(y)$ з урахуванням того, що щільність $f(x)$ є похідною для функції розподілу $F(x)$. Розв'язком цього рівняння є шукана 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

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

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

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 показникової величини 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*

$$P(\xi \geq t + s \;/\; \xi \geq t) = P(\xi \geq s).$$

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

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

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

Вправа. Довести, що в класі абсолютно неперервних величин властивість відсутності післядії мають лише показникові величини. Вказівка: відсутність післядії еквівалентна мультиплікативності $Q(t + s) = Q(t)\; Q(s)$ для ймовірності виживання $Q(t) = P(\xi \geq t)$, а єдиним неперервним розв'язком рівняння $Q(t + s) = Q(t)\; Q(s)$ є експонента.

10.6. Розподіл Вейбула

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 нарешті змінюється етапом неухильного зростання (ефект старіння).

10.7. 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.

10.8. Розподіл Парето

Функції розподілу всіх зазначених вище розподілів експоненційно швидко прямують до 1 при $x \to \infty$. Для наступного розподілу така швидкість є степеневою.

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

$$P(\xi < x) = \begin{cases} 1 - (\lambda x)^{-\alpha},\ x \geq 1/\lambda, \\ 0,\ x < 1/\lambda \end{cases}.$$

10.9. 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

$$P(\xi < x) = P(\varphi < arctg\ x) = \tfrac{1}{2} + \tfrac{1}{\pi} arctg\ x,$$

і щільність ξ має вигляд

$$f_\xi(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

10.10. Функції від випадкової величини

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

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

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

Доведення очевидне, оскільки за означенням прообразу $\{g(\xi) \in B\} = \{\xi \ \in g^{(-1)}(B)\}$.

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 (тобто породжена міра Лебега – Стілтьєса) величин ξ_k, то

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

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*

$$P(\zeta < x) = P(\xi < (x - a)/b) = F_\xi((x - a)/b),$$

а у випадку існування щільності розподілу f_ξ існує щільність лінійного перетворення

$$f_\zeta(x) = b^{-1} f_\xi((x - a)/b).$$

Доведення. Перша формула очевидна.

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

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

Вправа.

(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 $-\ln \xi$ має показниковий розподіл.

11. 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

$$\mu_n = \frac{1}{n}\sum_{k=1}^{n}\zeta_k = \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{m} x_i 1_{\{\zeta_k = x_i\}} =$$
$$\frac{1}{n}\sum_{i=1}^{m} x_i \sum_{k=1}^{n} 1_{\{\zeta_k = x_i\}} = \sum_{i=1}^{m} x_i \frac{\nu_n(x_i)}{n},$$

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, $\frac{\nu_n(x_i)}{n} \to P(\zeta = x_i) = p_i$. Тому $\mu_n \to \sum x_i\, p_i$. 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 $\{x_n,\ n \geq 1\}$ та розподілом $(p_n = P(\xi = x_n), n \geq 1)$.* Математичним сподіванням дискретної величини *ξ називається сума ряду*

$$M\xi = \sum_{n\geq 1} x_n\, p_n,$$

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

Зауваження. Враховуючи асимптотичну властивість, що викладена у прикладі, а також з метою 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.

11.1. Обчислення математичного сподівання функції від дискретної випадкової 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 – довільна числова функція. Тоді*

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

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

Доведення.

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

$$Mg(\xi) = \sum\nolimits_n z_n\ P(g(\xi) = z_n) =$$

$$\sum\nolimits_n z_n\ P(\cup_{\ j:\ g(x_j)\ =\ z_n}\{\xi = x_j\}) = \sum\nolimits_n z_n\ \sum\nolimits_{j:\ g(x_j)\ =\ z_n} P(\xi = x_j) =$$

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

Вправа. Довести, що для цілозначної невід'ємної величини $M\xi = \sum_{n\geq 1} P(\xi \geq n) = \sum_{n\geq 0} P(\xi > n)$.

11.2. 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 $M\xi \geq 0$.

(б –однорідність) $M(c\xi) = cM\xi$.

(в –адитивність) $M(\xi + \eta) = M\xi + M\eta$.

(г –монотонність) Якщо $\xi \geq \eta$, то $M\xi \geq M\eta$.

(д –неперервність знизу) Якщо ξ_n, η прості величини і $0 \leq \xi_n \uparrow \xi \equiv \lim \xi_n \geq \eta$ при всіх ω, то границя $\lim\ M\xi_n \geq M\eta$.

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 ж 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 $g((\xi, \eta)) = \xi + \eta$:

$$M(\xi + \eta) = \sum\nolimits_{i,j} (x_i + y_j)\ P(\xi = x_i, \eta = y_j) =$$
$$\sum\nolimits_{i,j} x_i\ P(\xi = x_i, \eta = y_j) + \sum\nolimits_{i,j} y_j\ P(\xi = x_i, \eta = y_j) =$$
$$\sum\nolimits_i x_i \sum\nolimits_j P(\xi = x_i, \eta = y_j) + \sum\nolimits_j y_j \sum\nolimits_i P(\xi = x_i, \eta = y_j) =$$
$$\sum\nolimits_i x_i\ P(\xi = x_i) + \sum\nolimits_j y_j\ P(\eta = y_j) = M\xi + M\eta.$$

Рівність кратних та повторних сум випливає з абсолютної збіжності результуючих рядів.

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

$$0 \le M(\xi - \eta) = M\xi - M\eta.$$

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

Розглянемо для довільного фіксованого $\varepsilon > 0$ випадкові події $A_n = \{\xi_n \ge \eta - \varepsilon\}$. За умовою $A_n \uparrow \Omega$, звідки за неперервністю ймовірності $P(A_n) \to 1$. Оскільки величина η проста, то її найбільше значення $m = \max \eta < \infty$ скінченне і не залежить від елементарної події ω.

Для всіх ω за означенням A_n справедлива нерівність

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

отже, за монотонністю математичного сподівання

$$\lim M\xi_n \ge \lim M(\eta - m\, 1_{\overline{A_n}} - \varepsilon) = M\eta - m \lim P(\overline{A_n}) - \varepsilon = M\eta - \varepsilon,$$

оскільки $P(\overline{A_n}) \to 0$. Переходячи в останній нерівності до границі $\varepsilon \to 0$, доводимо шукану нерівність $\boxdot$

11.3. Приклади обчислення математичного сподівання

11.3.1. Індикаторна величина

$$M 1_A = 1 \cdot P(A) + 0 \cdot P(\overline{A}) = P(A).$$

11.3.2. Біноміальний розподіл

Нехай $\xi \simeq B(n,p)$. Тоді $M\xi = np$.
Наведемо три способи обчислення.

(а) $M\xi = \sum_{k=0}^{n} k\, C_n^k\, p^k q^{n-k} = np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!}\, p^{k-1} q^{n-k} = np.$

(б) Розглянемо генератрису

$$\varphi(z) = Mz^{\xi} = \sum_{k=0}^{n} z^k C_n^k\, p^k q^{n-k} = (pz+q)^n.$$

Оскільки $M\xi = \varphi'(z)\mid_{z=1}$, то $M\xi = np(pz+q)^{n-1}\mid_{z=1} = np$.

(в) Нехай $\chi_k = 1_{Y_k}$ – індикаторна величина успіху в k-му випробуванні Бернуллі. Тоді $M\chi_k = p$ і $\xi = \sum_{k=1}^{n} \chi_k$, отже $M\xi = \sum_{k=1}^{n} M\chi_k = np$.

11.3.3. Геометричний розподіл

Для $\xi \simeq G(p)$ за методом генератрис

$M\xi = \sum_{k\ge1} k\, q^{k-1} p = p(\sum_{k\ge1} q^k)'_q = p/(1-q)^2 = 1/p.$

11.3.4. Розподіл Пуассона

Для $\xi \simeq \Pi(\lambda)$ за означенням

$M\xi = \sum_{k\ge0} k\, \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k\ge1} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda.$

Вправа.

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

(2). Нехай $P(B) > 0$, а ξ – дискретна випадкова величина зі значеннями $\{x_n, n \ge 1\}$. Визначимо умовне математичне сподівання $M(\xi/B) \equiv M(\xi 1_B)/P(B)$.

Довести, що: (а) для такого математичного сподівання виконуються всі властивості математичного сподівання дискретної величини, (б) $M(\xi/B) = \sum_{n\geq 1} x_n P(\xi = x_n\ /B)$

12. Загальне означення математичного сподівання

12.1. Невід'ємні випадкові величини

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

$$M\xi = \lim_{n \to \infty} M\xi_n \leq \infty,$$

де ξ_n – *прості величини такі, що* $0 \leq \xi_n \uparrow \xi$.

Зауваження. Існування вказаної послідовності ξ_n випливає з теореми про апроксимацію випадкових величин простими, а існування скінченої або нескінченної границі $M\xi_n \uparrow M\xi$ є наслідком монотонності цієї послідовності.

12.2. Коректність означення математичного сподівання

Теорема (про коректність визначення математичного сподівання). *Границя* $\lim M\xi_n$ *в означенні математичного сподівання існує (скінченна або нескінченна) і не залежить від вибору простих величин* ξ_n *таких, що* $0 \leq \xi_n \uparrow \xi$, *отже, математичне сподівання визначене коректно.*

Доведення. Існування границі обгрунтовано в зауваженні. Для доведення її єдиності припустимо, що одночасно з ξ_n прості величини ξ'_n такі, що $0 \leq \xi'_n \uparrow \xi$. Тоді $0 \leq \xi_n \uparrow \xi \geq \xi'_k$ для кожного k. Отже, за властивістю (д) неперервності знизу математичного сподівання дискретної величини існують монотонні границі $\lim M\xi_n \geq \lim M\xi'_k$. Переставляючи ξ'_n і ξ_n місцями, дістанемо рівність $\lim M\xi_n = \lim M\xi'_n$, не зважаючи на скінченність чи нескінченність цих границь $\boxdot$

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

$$M\xi = \sup(M\eta : \ \eta \text{ проста}, \ 0 \leq \eta \leq \xi).$$

Доведення.

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

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

12.3. Інтегровні невід'ємні випадкові величини

Означення. *Невід'ємна випадкова величина ξ називається* інтегровною, *якщо $M\xi < \infty$.*

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

$$M(\xi + \zeta) = M\xi + M\zeta.$$

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 $0 \le \xi_n \uparrow \xi$, $0 \le \zeta_n \uparrow \zeta$, 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 теореми про коректність визначення математичного сподівання

$$M(\xi + \zeta) = \lim M(\xi_n + \zeta_n) = \lim M\xi_n + \lim M\zeta_n = M\xi + M\zeta < \infty$$

за властивістю границі суми збіжних числових послідовностей $\boxdot$

12.4. Математичне сподівання знакозмінних 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*

$$M\xi = M\xi^+ - M\xi^-.$$

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

Вправа.Випадкова величина ξ інтегровна тоді й тільки тоді, коли при деякому $h > 0$ абсолютно збігається ряд $I(\xi, h) = \sum_{n\in\mathbb{Z}} nh\, P(nh \le \xi < nh+h)$. Ця збіжність не залежить від h, причому $M\xi = \lim_{h\to 0} I(\xi, h)$.

12.5. 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) $Mc = c$.
(2 –центрованість) Якщо $\xi = 0$ *м.н., то* $M\xi = 0$.
(3 –невід'ємність) Якщо $\xi \ge 0$ *м.н., то* $M\xi \ge 0$.
(4 –додатність) Якщо $\xi \ge 0$*, і* $M\xi = 0$*, то* $\xi = 0$ *м.н.*
(5 –адитивність) $M(\xi+\zeta) = M\xi + M\zeta$*, якщо* ξ, ζ *– інтегровні.*
(6 –однорідність) $M(c\xi) = cM(\xi)$.
(7 –монотонність) Якщо $\xi \le \zeta$ *м.н., то* $M\xi \le M\zeta$.
(8 –інваріантність) Якщо $\xi = \zeta$ *м.н., то* $M\xi = M\zeta$.
(9 –опуклість) $|M\xi| \le M\,|\xi|$.

Зауважимо, що властивість лінійності математичного сподівання полягає в однорідності та адитивності.

Доведення.

Твердження (1) очевидне, оскільки стала є простою випадковою величиною.

При доведенні (2) можна вважати, що $\xi \ge 0$, 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 $M\xi = M\xi^{+} - M\xi^{-} = 0$. Далі, з $0 \le \zeta \le \xi$, де ζ – проста, а $\xi = 0$ м.н., виводимо, що $\zeta = 0$ м.н. Тоді $\zeta = \sum_{k=1}^{n} c_k 1_{A_k}$, де

$c_1 = 0, P(A_1) = 1, P(A_k) = 0, \forall k > 1$, і $M\zeta = 0 \cdot 1 + 0 = 0$. Тому за теоремою про інваріантне зображення математичного сподівання $M\xi = 0$.

В умовах (3) маємо $\xi^- = 0$ м.н. і внаслідок (2) $M\xi^- = 0$. Оскільки $\xi^+ \geq 0$, з невід'ємності математичного сподівання дискретної величини та теореми про інваріантне зображення математичного сподівання дістанемо $M\xi^+ \geq 0$. Тому $M\xi = M\xi^+ - M\xi^- \geq 0$.

Для доведення (4) припустимо, що $\xi \neq 0$ майже напевне. Тоді $P(\xi \geq \varepsilon) > 0$ для деякого $\varepsilon > 0$, тому що

$$\{\xi \geq 0\} \cap \{\xi \neq 0\} = \{\xi > 0\} = \cup_{k>0}\{\xi > 1/k\}.$$

Оскільки $\xi \geq \varepsilon \, 1_{\{\xi \geq \varepsilon\}}$, то за теоремою про інваріантне зображення математичного сподівання

$$M\xi \geq M\varepsilon \, 1_{\{\xi \geq \varepsilon\}} = \varepsilon \, P(\xi \geq \varepsilon) > 0,$$

що суперечить умові.

Для доведення (5) зауважимо, що у випадку невід'ємних $\zeta, \xi \geq 0$ адитивність доведена вище. Для знакозмінних величин запишемо тотожність

$$\xi^+ + \zeta^+ - (\xi + \zeta)^+ = \xi^- + \zeta^- - (\xi + \zeta)^- = \eta.$$

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 $M(\xi^\pm + \zeta^\pm - \eta) = M(\xi^\pm + \zeta^\pm) - M\eta$.

Отже,

$$
\begin{gathered}
M(\xi+\zeta)=M(\xi+\zeta)^{+}-M(\xi+\zeta)^{-}= \\
M(\xi^{+}+\zeta^{+}-\eta)-M(\xi^{-}+\zeta^{-}-\eta)= \\
M(\xi^{+}+\zeta^{+})-M\eta-M(\xi^{-}+\zeta^{-})+M\eta= \\
M\xi^{+}+M\zeta^{+}-M\xi^{-}-M\zeta^{-}=M\xi+M\zeta.
\end{gathered}
$$

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:

$$
M(c\xi)=\lim M(c\xi_n)=\lim cM(\xi_n)=cM\xi.
$$

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) $M\zeta-M\xi=M(\zeta-\xi)\geq 0$.

Доведення (8) випливає з твердження (7), якщо його застосувати до нерівностей $\xi \leq \zeta,\ -\xi \leq -\zeta$.

Властивість (9) отримуємо з (7) та нерівностей $-|\xi| \leq \xi \leq |\xi|$ $\boxdot$

12.6. Перехід до границі під знаком математичного сподівання

В даному розділі збіжність випадкових величин будемо розуміти як поточкову, тобто збіжність при всіх $\omega \in \Omega$.

Теорема (теорема Лебега про монотонну збіжність). *Нехай* $0 \leq \xi_n \uparrow \xi$. *Тоді* $M\xi_n \uparrow M\xi \leq \infty$.

Доведення.

За теоремою про апроксимацію випадкових величин простими оберемо для кожного n прості $0 \le \zeta_{nk} \uparrow \xi_n$ при $k \to \infty$. Позначимо

$$\zeta_k = \max_{n \,\le\, k} \zeta_{nk}.$$

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 $\lim M\zeta_k = M\zeta = M\xi$. Отже, з тих же нерівностей за монотонністю математичного сподівання

$$M\xi = \lim M\zeta_k \le \lim M\xi_k \le M\xi,$$

тобто $\lim M\xi_k = M\xi$, незалежно від того, скінченне $M\xi$ чи ні $\boxdot$

Вправа. Доведіть справедливість формули для математичного сподівання дискретної (не простої) випадкової величини.

Теорема (теорема Лебега про мажоровану збіжність).
Нехай $|\xi_n| \le \eta$, причому $M\eta < \infty$.

(а –Нерівності Фату). Тоді

$$M \ \underline{\lim}_{\ n\to\infty} \xi_n \le \underline{\lim}_{\ n\to\infty} M\xi_n \le \overline{\lim}_{\ n\to\infty} M\xi_n \le M \ \overline{\lim}_{\ n\to\infty} \xi_n.$$

(б –Мажорована збіжність) Якщо додатково $\xi_n \to \xi, n \to \infty$, то

$$M\xi_n \to M\xi, \quad M\,|\xi_n - \xi| \to 0,\ n \to \infty.$$

Зауваження. Для справедливості лівої нерівності Фату достатньо, щоб $\xi_n \geq 0$.

Доведення.

(а) Доведемо ліву нерівність. Позначимо $\zeta_n = \inf_{\,k \geq n} \xi_k$.

За означенням нижньої границі $\zeta_n \uparrow \underline{\lim}\xi_n$ і $\zeta_n \geq \inf_{\,k \geq n}(-\eta) = -\eta$. Отже, $0 \leq \zeta_n + \eta \uparrow \underline{\lim}\ \xi_n + \eta$. За адитивністю математичного сподівання та за теоремою Лебега про монотонну збіжність

$$M\underline{\lim}\xi_n + M\eta = M(\underline{\lim}\xi_n + \eta) = M \lim(\zeta_n + \eta) = \lim M(\zeta_n + \eta) =$$
$$\lim M\zeta_n + M\eta = \underline{\lim}M\zeta_n + M\eta \leq \underline{\lim}M\xi_n + M\eta,$$

оскільки $\zeta_n \leq \xi_n$.

Звідси випливає ліва нерівність Фату. Очевидно, що вказане доведення справедливе для довільних невід'ємних величин ξ_n, адже в даному доведенні можна обрати $\eta \equiv 0$.

Права нерівність Фату виводиться з лівої після підстановки $-\xi_n$ замість ξ_n з урахуванням тотожностей $\underline{\lim}(-\xi_n) = -\overline{\lim}\xi_n$, $\overline{\lim}(-\xi_n) = -\underline{\lim}\xi_n$.

Перше твердження (б) випливає з (а), якщо врахувати, що $\underline{\lim}\xi_n = \overline{\lim}\xi_n = \xi$ за означенням границі числової послідовності, звідки

$$M\xi = M\ \underline{\lim}\xi_n \leq \underline{\lim}\ M\xi_n \leq \overline{\lim}\ M\xi_n \leq M\ \overline{\lim}\xi_n = M\xi,$$

тобто всі нерівності тут є рівностями.

Для доведення другого твердження пункту (б) теореми застосуємо перше до мажоровано збіжної послідовності

$$\xi_n' = |\xi_n - \xi| \to 0, \;\; 0 \le \xi_n' \le 2\eta \; \boxdot$$

Вправа. Довести, що

$$P(\underline{\lim} A_n) \le \underline{\lim}\; P(A_n) \le \overline{\lim}\; P(A_n) \le P(\overline{\lim} A_n).$$

12.7. Абстрактний інтеграл Лебега

Означення. Інтегралом Лебега за мірою μ *від вимірної функції* $\xi(\omega)$ *на вимірному просторі* $(\Omega, \mathcal{F}, \mu)$ *називається число*

$$\int_\Omega \xi(\omega)\mu(d\omega) \equiv \mu(\Omega) \int_\Omega \xi(\omega) P(d\omega) = \mu(\Omega)\; M\xi\;,$$

у припущенні невід'ємності або інтегровності ξ, *де ймовірність* P *на* $\mathcal{F}$ *визначається як* $P(A) = \mu(A)/\mu(\Omega)$. *Інтегровна величина* ξ *називається* інтегровною за мірою μ.

Інтегралом від ξ на множині $A \in \mathcal{F}$ називається інтеграл на Ω від функції $\xi(\omega)1_A(\omega)$.

Зауваження. Інтеграл Лебега має всі властивості, що викладені в теоремах про коректність визначення 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 і інтегралів Лебега – Стілтьєса та за мірою Лебега, що розглянуті нижче.

12.8. Інтеграл Лебега – Стілтьєса

Означення. *Нехай* F *– нормована міра Лебега – Стілтьєса на* $\mathfrak{B}(\mathbb{R})$, *а* $g(x)$ *– борелєва функція.* Інтеграл Лебега – Стілтьєса *від* g *за мірою* F

визначається як математичне сподівання випадкової величини $g : \Omega \to \mathbb{R}$ *на ймовірнісному просторі* $(\Omega, \mathcal{F}, P) \equiv (\mathbb{R}, \mathfrak{B}(\mathbb{R}), F)$, *тобто як монотонна границя інтегралів від апроксимуючих простих борелевих функцій із лінійним продовженням на знакозмінні борелеві функції.*

Обидва позначення для інтегралу Лебега – Стілтьєса

$$\int_{-\infty}^{\infty} g(x)dF(x) = \int_{\mathbb{R}} g(x)F(dx)$$

мають однаковий зміст.

Інтегралом Лебега – Стілтьєса від 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 ірою 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

$$\lim \sum_{k=0}^{n} g_k(F(a_{k+1}) - F(a_k)) \equiv (LS) \int_{[a,b]} g(x)F(dx),$$

де $a = a_0 < a_1 < .. < a_{n+1} = b$ – розбиття відрізку $[a, b]$, а g_k – відповідно верхні або нижні межі функції g на відрізках $[a_k, a_{k+1}]$

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.

12.9. Інтеграл за мірою Лебега

Означення. *Нехай $g(x)$ – борелева функція.* Інтеграл за мірою Лебега *на $[a, b]$ від g визначається як нормований інтеграл Лебега – Стілтьєса за* мірою Лебега $(b - a)F_{ab}(\cdot)$:

$$\int_a^b g(x)dx \equiv (b - a)\int_{-\infty}^{\infty} g(x)dF_{ab}(x),$$

що породжена функцією рівномірного розподілу

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

Зауваження. Можна довести, що для кусково-неперервних функцій g інтеграл за мірою Лебега збігається з інтегралом Рімана

12.10. Теореми Фубіні та Радона – Нікодима.

Вимірним простором будемо називати пару, що складається з абстрактної множини Ω з виділеною сигма-алгеброю $\mathcal{F}$ її підмножин, або ж трійку, з додатковим третім елементом – мірою μ на $\mathcal{F}$.

Теорема Фубіні часто застосовується для обчислення кратних інтегралів.

Означення. *Вимірний простір* $(\Omega, \mathcal{F}, \mu)$ *називається* прямим добутком *вимірних просторів* $(\Omega_k, \mathcal{F}_k, \mu_k)$, $k = 1, 2$, *якщо:*

(a) $\Omega = \Omega_1 \times \Omega_2 \equiv \{\omega = (\omega_1, \omega_2),\ \omega_k \in \Omega_k\}$.

(б) $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2 \equiv \sigma[\{A_1 \times A_2, \quad A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}]$.

(в) $P(A_1 \times A_2) = P_1(A_1)P_2(A_2)$, $\forall A_k \in \mathcal{F}_k$.

Теорема (теорема Фубіні про кратний та повторні інтеграли). *Нехай простір* $(\Omega, \mathcal{F}, \mu) = (\Omega_1, \mathcal{F}_1, \mu_1) \otimes (\Omega_2, \mathcal{F}_2, \mu_2)$ *є прямим добутком, а* $\xi(\omega) = \xi(\omega_1, \omega_2)$ *випадкова величина на* $(\Omega, \mathcal{F}, \mu)$ *Величина* ξ *інтегровна за мірою* μ *тоді й тільки тоді, коли функція* $\xi(\cdot, \omega_2)$ *для* μ_2*-майже всіх* ω_2 *інтегровна за мірою* μ_1 *і вимірна величина* $\int_{\Omega_1} \xi(\omega_1, \omega_2)\mu_1(d\omega_1)$ *є інтегровною за мірою* μ_2. *У цьому випадку справедлива рівність*

$$\int_{\Omega} \xi(\omega)\mu(d\omega) = \int_{\Omega_2} \left[\int_{\Omega_1} \xi(\omega_1, \omega_2)\mu_1(d\omega_1) \right] \mu_2(d\omega_2).$$

Доведення не наводиться.

Означення. Числова функція $\mu : \mathcal{F} \to \mathbb{R}$ на вимірному просторі $(\Omega, \mathcal{F})$ називається сигма-скінченною мірою, якщо існує послідовність множин $C_n \in \mathcal{F}$ таких, що кожне зі звужень $\mu(\cdot \cap C_n)$ є мірою на $\mathcal{F}$. Інтеграл Лебега за такою мірою μ визначається як границя інтегралів за відповідними звуженнями.

Прикладом таких мір є точкова міра та міра Лебега.

Означення. *Нехай* ν *– міра, а* μ *– сигма-скінченна міра на просторі* $(\Omega, \mathcal{F})$. *Міра* ν абсолютно неперервна *відносно* μ, *позначення* $\nu \ll \mu$, *якщо для кожної множини* $A \in \mathcal{F}$ *із* $\mu(A) = 0$ *випливає* $\nu(A) = 0$.

Теорема (теорема Радона – Нікодима). *Нехай* $(\Omega, \mathcal{F}, \mu)$ *– вимірний простір зі сигма-скінченною мірою* μ, *а міра* ν *задана на* $\mathcal{F}$ *і абсолютно неперервна:* $\nu \ll \mu$. *Тоді знайдеться* $\mathcal{F}$*-вимірна інтегровна за мірою* μ *функція* f *на* Ω *така, що*

$$\nu(A) = \int_A f(\omega)\mu(d\omega), \ \forall A \in \mathcal{F}.$$

Функція f *називається щільністю міри* ν *відносно* μ *та позначається як* $f(x) = \frac{d\nu}{d\mu}(x)$.

Доведення не наводиться.

13. Обчислення математичного сподівання

13.1. Обчислення математичного сподівання функції від випадкової величини через її функцію розподілу

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 борелева функція. Випадкова величина $g(\xi)$ інтегровна тоді й тільки тоді, коли функція $g(x)$ інтегровна за мірою $F_\xi(\cdot)$, і за цим припущенням має місце рівність*

$$Mg(\xi) = \int_{-\infty}^{\infty} g(x)\, dF_\xi(x).$$

Зокрема, дане твердження справедливе при $g(x) \equiv x$.

Зауваження. Сформульоване твердження можна розглядати як теорему про заміну змінної ω на x за допомогою відображення ξ в інтегралі Лебега

$$\int_\Omega g(\xi(\omega)) P(d\omega) = \int_{\mathbb{R}} g(x)\ F_\xi(dx).$$

Доведення.

Нехай $g(x) = \sum c_k 1_{x \in B_k}$ проста функція. Тоді тотожність (завжди визначених) інтегралів випливає з лінійності математичного сподівання та інтегралу Лебега – Стілтьєса, а також із теореми про обчислення ймовірностей,пов'язаних із випадковою величиною ξ :

$$Mg(\xi) = \sum g(c_k) P(\xi \in B_k) = \sum g(c_k) F_\xi(B_k) = \int_{-\infty}^{\infty} g(x)\, dF_\xi(x).$$

Нехай $g(x)$ невід'ємна борелєва функція. Розглянемо прості 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 інтегралу Лебега – Стілтьєса. Зауважимо, що рівність є справедливою як для інтегровних, так і для неінтегровних g. Тому функції $g(\xi)$ та $g(x)$ є інтегровними одночасно.

Для знакозмінних g використаємо лінійність математичного сподівання та інтегралу Лебега – Стілтьєса:

$$Mg(\xi) = Mg^+(\xi) - Mg^-(\xi) = \int_{-\infty}^{\infty} g^+(x)\, dF_\xi(x) - \int_{-\infty}^{\infty} g^-(x)\, dF_\xi(x) \boxdot$$

Вправа. Довести, що (а) для невід'ємної випадкової величини
$M\xi = \int_0^\infty (1 - F_\xi(x))dx$, (б) для інтегровної величини
$M\xi = \int_0^\infty (F_\xi(-x) + 1 - F_\xi(x))dx$.

Теорема про обчислення математичного сподівання функції від дискретної величини наведена в розділі про математичне сподівання дискретної величини.

13.2. Обчислення математичного сподівання функції від випадкової величини через її щільність

Теорема (про обчислення математичного сподівання функції від неперервної величини). *Нехай випадкова величина ξ має щільність $f_\xi(x)$, а $g(x)$ – деяка 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

$$Mg(\xi) = \int_{-\infty}^{\infty} g(x)\ f_\xi(x)\,dx.$$

Зауваження. У випадку, коли функції $g(x),\ f_\xi(x)$ кусково неперервні, інтеграл Лебега в правій частині можна обчислювати як інтеграл Рімана.

Доведення. Розглянемо клас тих функцій $g(x)$, 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 випадковою величиною, через її щільність, цей клас містить всі індикаторні функції $g(x) = 1_B(x)$ для борелєвих B. За лінійністю він містить всі прості функції $g(x)$. За теоремами Лебега про монотонну збіжність та про мажоровану збіжність вказаний клас замкнений відносно монотонної та мажорованої збіжності. Тому цей клас містить всі інтегровні функції.

Отже, остаточно дана теорема випливає з теореми про обчислення математичного сподівання функції від випадкової величини $\boxdot$

13.3. Приклади обчислення математичного сподівання

1. Рівномірний розподіл: для $\xi \simeq U(a,b)$:
$M\xi = \int_a^b x\frac{1}{b-a}dx = (a+b)/2.$

2. Показниковий розподіл: для $\xi \simeq Exp(\lambda)$:
$M\xi = \int_0^\infty x\ \lambda\exp(-\lambda x)dx = 1/\lambda.$

3. Нормальний розподіл: для $\xi \simeq N(\mu, \sigma^2)$:

$M\xi = M(\mu + \sigma\ \zeta) = \mu + \sigma\ M\zeta = \mu$, де ζ – стандартна нормальна величина, $M\zeta = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx = 0.$

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.

14. 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. *Нехай випадкова величина* ξ квадратично інтегровна, *тобто* $M\xi^2 < \infty$. Дисперсією ξ *називається число*

$$D\xi = M(\xi - M\xi)^2.$$

Корінь квадратний $\sigma_\xi = \sqrt{D\xi}$ *називається* середньоквадратичним відхиленням ξ.

Вправа. Дисперсія скінченна тоді й тільки тоді, коли ξ квадратично інтегровна.Вказівка: за нерівністю $2|\xi| \le 1 + \xi^2$ з квадратичної інтегровності випливає інтегровність ξ.

14.1. Властивості дисперсії

Теорема (про властивості дисперсії). *За умови квадратичної інтегровності* ξ :

(а) $D\xi \ge 0$,
(б) $D\xi = 0$ *тоді й тільки тоді, коли* $\xi = M\xi$ *майже напевне,*
(в) $D\xi = M\xi^2 - (M\xi)^2$,
(г) $D(a\xi + b) = a^2 D\xi$,
(д) $D\xi = \min_{c\in \mathbb{R}} M(\xi - c)^2$,

$M\xi = arg\ \min_{c\in\ \mathbb{R}}\ M(\xi - c)^2.$

Доведення.

(а) випливає з позитивності математичного сподівання, оскільки $(\xi - M\xi)^2 \geq 0$.

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

(г) За означенням дисперсії та за лінійністю математичного сподівання $D(a\xi + b) = M(a\xi + b - M(a\xi + b))^2 = M(a\xi - aM\xi)^2 = a^2 D\xi$.

Для доведення (д) обчислимо

$$M(\xi - c)^2 = M(\xi - M\xi + M\xi - c)^2 =$$
$$M(\xi - M\xi)^2 + (M\xi - c)^2 + 2M(\xi - M\xi)(M\xi - c) =$$
$$M(\xi - M\xi)^2 + (M\xi - c)^2 \geq M(\xi - M\xi)^2 = D\xi,$$

причому рівність має місце тоді й тільки тоді, коли $c = M\xi$ $\boxdot$

Теорема (про обчислення дисперсії). *Нехай величина ξ має функцію розподілу $F_\xi(x)$. Тоді*

$$D\xi = \int_{-\infty}^{\infty} (x - M\xi)^2\, dF_\xi(x) = \int_{-\infty}^{\infty} x^2\, dF_\xi(x) - (M\xi)^2.$$

Якщо існує щільність розподілу $f_\xi(x)$, то

$$D\xi = \int_{-\infty}^{\infty} (x - M\xi)^2\, f_\xi(x)dx = \int_{-\infty}^{\infty} x^2\, f_\xi(x)dx - (M\xi)^2.$$

Доведення є очевидним наслідком теореми про обчислення математичного сподівання функції від випадкової величини, а саме для функції $g(x) = (x - M\xi)^2$ $\boxdot$

14.2. Приклади обчислення дисперсії

1.*Індикаторна величина*

$$D1_A = 1 \cdot P(A) + 0 \cdot P(\overline{A}) - (P(A))^2 = P(A)(1 - P(A)).$$

2. *Біноміальний розподіл*: $\xi \simeq B(n,p)$, $D\xi = npq$.
Розглянемо генератрису

$$\varphi(z) = Mz^\xi = \sum\nolimits_{k=0}^{n} z^k C_n^k \, p^k q^{n-k} = (pz+q)^n.$$

Оскільки за формулою про обчислення математичного сподівання функції від дискретної величини

$$M\xi(\xi-1) = \sum\nolimits_{k=0}^{n} k(k-1) C_n^k \, p^k q^{n-k} = \varphi''(z) \mid_{\, z=1},$$

то

$$D\xi = M\xi(\xi-1) + M\xi - (M\xi)^2 =$$
$$n(n-1)p^2(pz+q)^{n-2} \,|_{z=1} + np - (np)^2 = npq.$$

3. *Розподіл Пуассона*: $\xi \simeq \Pi(\lambda)$, $D\xi = \lambda$.
За означенням

$$D\xi = M\xi(\xi-1) + M\xi - (M\xi)^2 =$$
$$\sum\nolimits_{k\geq 0} k(k-1)\, \frac{\lambda^k}{k!} e^{-\lambda} + \lambda - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

4. *Рівномірний розподіл:* $\xi \simeq U(a,b)$, $D\xi = (b-a)^2/12$.

$$D\xi = \int_a^b x^2 \frac{1}{b-a} dx - (a+b)^2/4 = (b-a)^2/12.$$

5. *Показниковий розподіл:* $\xi \simeq Exp(\lambda)$, $D\xi = 1/\lambda^2$.

$$D\xi = \int_0^\infty x^2 \; \lambda \exp(-\lambda x) dx - 1/\lambda^2 = 1/\lambda^2.$$

6. *Нормальний розподіл:* $\xi \simeq N(\mu, \sigma^2)$, $D\xi = \sigma^2$.

За формулою (г) про дисперсію лінійного перетворення

$$D\xi = M(\mu + \sigma\,\zeta) = \sigma^2 D\zeta = \sigma^2,$$

де ζ – стандартна нормальна величина, для якої

$$D\zeta = M\zeta^2 = \int_{-\infty}^\infty x^2 \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx =$$

$$\int_0^\infty 2y \frac{1}{\sqrt{2\pi}} \exp(-y) \frac{dy}{\sqrt{2y}} = \Gamma(\tfrac{3}{2}) \sqrt{\tfrac{2}{\pi}} = 1.$$

Наслідок: параметр σ^2 нормального розподілу збігається з дисперсією відповідної нормальної величини.

15. Імовірнісні нерівності

Теорема (загальна нерівність Чебишева). *Нехай функція* $g : \mathbb{R}_+ \to \mathbb{R}_+$ *не спадає, а* $\xi \geq 0$. *Тоді для довільної сталої* $c > 0$

$$P(\xi \geq c) \leq \frac{Mg(\xi)}{g(c)}.$$

Доведення. З невід'ємності та монотонності g виводимо нерівність $g(\xi) \geq g(c) 1_{\{\xi\,\geq\, c\}}$, $\forall \omega \in \Omega$. Звідси за монотонністю математичного сподівання дістанемо шукану нерівність:

$$Mg(\xi) \geq Mg(c) 1_{\{\xi\,\geq\, c\}} = g(c) P(\xi\ \geq c) \boxdot$$

Вправа.

(1) У припущенні скінченності правої частини доведіть нерівність $P(\xi \geq x) \leq \exp(-ax) M \exp(a\xi)$.

(2) Довести при $x > 0$ нерівність $\int_x^\infty \exp(-y^2/2) dy \leq x^{-1} \exp(-x^2/2)$.

Теорема (нерівність Чебишева для дисперсій). *Якщо існує* $D\xi$, *то для довільного* $\varepsilon > 0$

$$P(|\xi - M\xi| \geq \varepsilon) \leq D\xi \ / \ \varepsilon^2.$$

Доведення. Із загальної нерівності Чебишева для невід'ємної випадкової величини $(\xi - M\xi)^2$ та функції $g(x) \equiv x$ отримуємо

$$P(|\xi - M\xi| \geq \varepsilon) = P((\xi - M\xi)^2 \geq \varepsilon^2) \leq M(\xi - M\xi)^2 / \varepsilon^2 = D\xi \ / \ \varepsilon^2 \boxdot$$

Вправа. Для якої величини ξ нерівність Чебишева перетворюється на рівність ?

Правило 3σ. 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, 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 3 сигма.

Вправа. Довести при $t > 0$ нерівність $P(|\xi - M\xi| \geq t\ M\,|\xi|) \leq 1/t$.

Теорема (нерівність Йенсена). *Нехай функція $g(x)$ опукла донизу (як x^2), а випадкові величини ξ та $g(\xi)$ інтегровні. Тоді*

$$g(M\xi) \leq Mg(\xi).$$

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 $x = \xi$, $x_0 = M\xi$ та скористаємося монотонністю математичного сподівання:

$$Mg(\xi) \geq M(g(M\xi) + k(M\xi)(\xi - M\xi)) =$$
$$g(M\xi) + k(M\xi)M(\xi - M\xi) = g(M\xi)\ \boxdot$$

Теорема (нерівність Ляпунова). *Якщо $1 \leq s \leq t$, то*

$$(M\,|\xi|^s)^{1/s} \leq (M\,|\xi|^t)^{1/t}.$$

Доведення. Оскільки функція $g(x) = x^{t/s}$, $x \geq 0$, опукла донизу, то шукана нерівність випливає з нерівності Йенсена

$$(M|\xi|^s)^{t/s} = g(M|\xi|^s) \le Mg(|\xi|^s) = M|\xi|^t \ \boxdot$$

Теорема (**нерівність Гельдера**) *Якщо числа* $p, q > 0$ *спряжені:* $1/p + 1/q = 1$, *то*

$$M|\xi\eta| \le (M|\xi|^p)^{1/p}\,(M|\eta|^q)^{1/q}.$$

Доведення. Підставимо в елементарну нерівність $xy \le x^p/p + y^q/q$ вирази $x = |\xi|/(M|\xi|^p)^{1/p}$ та $y = |\eta|/(M|\eta|^q)^{1/q}$ та використаємо монотонність математичного сподівання:

$$M|\xi\eta|/(M|\xi|^p)^{1/p}(M|\eta|^q)^{1/q} \le$$
$$M|\xi|^p/pM|\xi|^p + M|\eta|^q/qM|\eta|^q = 1/p + 1/q = 1 \ \boxdot$$

Теорема (**нерівність Коші**)

$$(M\xi\eta)^2 \le M\xi^2\, M\eta^2.$$

Доведення. Випливає з нерівності Гельдера при $p = q = 2$ $\boxdot$

Зауважимо, що рівність у нерівності Коші має місце тоді й тільки тоді, коли $\eta = c\xi$ майже напевне для деякої сталої c, оскільки при виборі $c = \sqrt{M\eta^2/M\xi^2}$ справедлива тотожність

$$M(\eta - c\xi)^2 = 2c(\sqrt{M\xi^2\, M\eta^2} - M\xi\eta)$$

Теорема (**нерівність Мінковського**) *Якщо число* $p \ge 1$, *то*

$$(M|\xi+\eta|^p)^{1/p} \le (M|\xi|^p)^{1/p} + (M|\eta|^p)^{1/p}.$$

Зауваження. Величину $\|\xi\| \equiv (M\,|\xi|^p)^{1/p}$ можна розглядати як норму у лінійному просторі інтегровних у степені p випадкових величин: $L_p(P) = \{\xi : M\,|\xi|^p < \infty\}$. 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

$$M\,|\xi+\eta|^p \le M\,|\xi|\,|\xi+\eta|^{p-1} + \ M\,|\eta|\,|\xi+\eta|^{p-1} \le$$
$$(M\,|\xi|^p)^{1/p}(M\,|\xi+\eta|^{pq-q})^{1/q} + (M\,|\eta|^p)^{1/p}(M\,|\xi+\eta|^{pq-q})^{1/q} =$$
$$((M\,|\xi|^p)^{1/p} + \ (M\,|\eta|^p)^{1/p})(M\,|\xi+\eta|^p)^{1/q} \ \boxdot$$

16. Сумісна функція розподілу вектора та її властивості

Означення. *Нехай* $\xi = (\xi_1, ..., \xi_n)$ *випадковий вектор*. Сумісною функцією розподілу *(або ж* сукупною функцією розподілу*) вектора* ξ *називається дійсна функція* $F_\xi(x) : \mathbb{R}^n \to \mathbb{R}$, *яка в точці* $x = (x_1, ..., x_n)$ *дорівнює*

$$F_\xi(x) \equiv F_\xi(x_1...x_n) = P(\xi < x) \equiv P(\xi_1 < x_1, ..., \xi_n < x_n).$$

В 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.

16.1. Загальні властивості сумісної функції розподілу

Теорема (про властивості сумісної функції розподілу). *Нехай* ξ *– випадковий вектор, а* $F = F_\xi$ *його сумісна функція розподілу. Тоді ця функція:*

(1) неспадна: $F(x) \le F(y)$ *при всіх* $x \le y$,

(2) нормована: $F(x) \to 0$ *при* $\min_k x_k \downarrow -\infty$, *та* $F(x) \to 1$ *при* $\min_k x_k \uparrow \infty$,

(3) неперервна зліва: $F(x - 0) = F(x)$ *для всіх* x, *тобто*

$$\lim_{y \uparrow x,\ y<x} F(y) = F(x).$$

Зауваження. Існування границь 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 $y \uparrow x$ $\boxdot$

16.2. 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 та з наведеного зображення отримуємо

$$P(\xi \in [a, b)) = P(\xi \in \Pi_b \setminus \Pi_{b'}) - P(\xi \in \Pi_{a'} \setminus \Pi_a) =$$
$$P(\xi \in \Pi_b) - P(\xi \in \Pi_{b'}) - P(\xi \in \Pi_{a'}) + P(\xi \in \Pi_a) = \Delta_{[a,b)}F,$$

де за означенням $\Delta_{[a,b)}F = F(b) - F(b') - F(a') + F(a)$.

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 $[a, b)$ *– результат послідовних часткових приростів*

$$\Delta_{[a,b)}F = \Delta^n_{[a_n,b_n)}\Delta^{n-1}_{[a_{n-1},b_{n-1})}...\Delta^1_{[a_1,b_1)}F.$$

Теорема (про ймовірність значення всередині паралелепіпеда). *Якщо випадковий вектор* ξ *має сумісну функцію розподілу* F, *то*

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

Доведення проводиться за індукцією. Індуктивне припущення має вигляд

$$P(\{\xi_1 \in [a_1, b_1)..., \xi_k \in [a_k, b_k)\ \} \cap A) =$$
$$\Delta^k_{[a_k,b_k)}...\Delta^1_{[a_1,b_1)}P(\{\xi_1 < x_1, ..., \xi_k < x_k\} \cap A),$$

де A – довільна подія. Справедливість індуктивного переходу випливає з очевидної тотожності

$$P(\{\xi_k \in [a_k, b_k)\ \} \cap A \cap B) = \Delta^k_{[a_k,b_k)}P(\{\xi_k < x_k\} \cap A \cap B)$$

для довільних подій A, B.

Теорема (**про прирости сумісної функції розподілу**). *Нехай ξ – випадковий вектор, а $F = F_\xi$ його сумісна функція розподілу. Тоді виконується умова:*

(4) для довільного паралелепіпеда $[a, b)$ приріст

$$\Delta_{[a,b)} F = P(\xi \in [a, b)) \ge 0.$$

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), (2), (3) і (4), називається* сумісною функцією розподілу.

Вправа.

(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)..Функція $F(x_1, x_2) = 1/2 + 2\pi^{-1} \arctan \max(x_1, x_2)$ задовольняє умови (1),(3), однак не є сумісною функцією розподілу.

(3). Функція $F(x_1, x_2) = \min(x_1+x_2, 1) 1_{x_1 \ge 0,\ x_2 \ge 0}$ задовольняє умови (1),(2),(3), однак не є сумісною функцією розподілу.

16.3. 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*

$$\mathfrak{A}(\mathbb{R}^n) = \{A = \bigcup_{k=1}^m [a(k), b(k)), \quad [a(i), b(i)) \cap [a(j), b(j)) = \varnothing, i \ne j\},$$

значення якої задаються рівністю

$$F(A) = \sum\nolimits_{k=1}^{m} \Delta_{[a(k),\, b(k))} F.$$

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 очевидне $\boxdot$

Теорема (про неперервність у нулі міри Лебега – Стілтьєса). *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 аналогічної теореми в одновимірному випадку $\boxdot$

Теорема (про міру Лебега – Стілтьєса в евклідовому просторі). *Нехай F – сумісна функція розподілу в $\mathbb{R}^n$. Тоді на борелєвій σ-алгебрі $\mathfrak{B}(\mathbb{R}^n)$ існує єдина невід'ємна, нормована та 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 про продовження міри з алгебри на σ-алгебру $\mathfrak{B}(\mathbb{R})$ $\boxdot$

Означення. *Побудована за теоремою про міру Лебега – Стілтьєса в евклідовому просторі міра F називається* мірою Лебега – Стілтьєса, що породжена сумісною функцією розподілу F.

16.4. Обчислення ймовірностей через сумісну функцію розподілу

Теорема (про обчислення ймовірностей, пов'язаних із випадковим вектором). *Нехай випадковий вектор ξ має сумісну функцію розподілу F_ξ. Тоді для довільної множини $B \in \mathfrak{B}(\mathbb{R}^n)$ виконується рівність*

$$P(\xi \in B) = F_\xi(B),$$

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

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 $\mathfrak{B}(\mathbb{R}^n)$ $\boxdot$

Міру Лебега – Стілтьєса F_ξ будемо називати розподілом випадкового вектора ξ.

16.5. 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 *величини ξ та функції розподілу F_ξ.*

З нормованості та монотонності функції розподілу випливають такі властивості щільності:

(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 інтегралу (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.

Вправа.

(1). Нехай сумісна функція розподілу F має сумісну щільність f. Довести, що приріст на паралелепіпеді $[a,b)$ дорівнює інтегралу.

$$\Delta_{[a,b)}F = \int_{a_1}^{b_1} ... \int_{a_n}^{b_n} f(x_1, ..., x_n)dx_1...dx_n.$$

(2) Якщо випадковий вектор $\xi = (\xi_1, ..., \xi_n)$ має сумісну щільність f, то всі його координати мають щільності,

$$f_{\xi_k}(x) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, ...x_{k-1}, x, x_{k+1}, .., x_n)dx_1...dx_{k-1}dx_{k+1}, ..dx_n.$$

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

$$f_\xi(x_1, ..., x_n) = \frac{\partial}{\partial x_1} ... \frac{\partial}{\partial x_n} F_\xi(x_1, ..., x_n).$$

16.6. Обчислення ймовірностей через сумісну щільність

Теорема (**про обчислення ймовірностей, пов'язаних із випадковим вектором, через його сумісну щільність**). *Якщо випадковий вектор ξ має сумісну щільність $f_\xi(x)$, то*

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

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 всій 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 інтегралу Лебега отримаємо $L(B^\pm) = 0$ $\boxdot$

16.7. Функції від випадкового вектора

Теорема (**про обчислення ймовірностей та математичного сподівання функції від випадкового вектора**). *Якщо випадковий вектор*

Старт
Початок
Зміст
Сторінка 120 з 509
Назад
Екран
Закрити
Выхід

ξ має сумісну функцію розподілу $F_\xi(x)$, а $g : \mathbb{R}^n \to \mathbb{R}^m$ борелева функція, то

$$P(g(\xi) \in B) = \int_{g^{(-1)}(B)} dF_\xi(x),\ \forall B \in \mathfrak{B}(\mathbb{R}^m),$$

$$Mg(\xi) = \int_{\mathbb{R}^m} g(x)\, dF_\xi(x).$$

У випадку існування сумісної щільності ці інтеграли Лебега – Стілтьєса можна замінити на відповідні інтеграли за мірою Лебега із 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 $\{\xi \cdot \eta < 1/2\}$? З умови випадковості робимо висновок, що вектор (ξ, η) рівномірно розподілений на квадраті $[0,1]^2$, тобто має сумісну щільність $f(x,y) = 1_{(x,y)\in[0,1]^2}$.

Позначимо $B = \{(x,y) \in [0,1]^2 : x \cdot y < 1/2\}$. Тоді

$P(\xi \cdot \eta < 1/2) = P((\xi,\eta) \in B) = \int_B f(x,y)dxdy =$

$1/2 + \int_{1/2}^1 dx/2x = (1 + \ln 2)/2.$

17. Числові характеристики випадкових векторів

Означення. Математичним сподіванням випадкового вектора $\xi = (\xi_1, ..., \xi_n)$ *називається вектор, складений з математичних сподівань координат* ξ :

$$M\xi = (M\xi_1, ..., M\xi_n),$$

за умови, що останні визначені коректно.

Зауваження. Математичне сподівання випадкового вектора має всі основні властивості математичного сподівання – позитивність, монотонність, лінійність тощо.

17.1. Коваріація та кореляція випадкових величин

Означення. *Нехай* ξ, η – *квадратично інтегровні величини. Їх* коваріацією *називається число*

$$\mathrm{Cov}(\xi, \eta) \equiv M(\xi - M\xi)(\eta - M\eta) = M\xi\eta - M\xi M\eta,$$

що є скінченним за нерівністю Коші:

$$|\mathrm{Cov}(\xi, \eta)| \le \sqrt{M(\xi - M\xi)^2 M(\eta - M\eta)^2} = \sqrt{D\xi \; D\eta} = \sigma_\xi \sigma_\eta.$$

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* $\mathrm{Cov}(\xi,\eta) = 0$.

Вправа. Виведіть із нерівності Коші, що рівність $\rho(\xi,\eta) = \pm 1$ має місце тоді й тільки тоді, коли $\eta = a \pm b\xi$ для деяких сталих $a, b > 0$.

Якщо простір квадратично інтегровних центрованих величин

$$L_2^0(P) \equiv \{\xi : \ M\xi^2 < \infty, \ M\xi = 0\}$$

розглядати як гільбертів простір зі скалярним добутком $(\xi,\eta) = M\xi\eta$, то $\|\xi\|^2 = D\xi$ і $\rho(\xi,\eta) = \cos(\xi \wedge \eta)$. Некорельованість величин інтерпретується як їх ортогональність у даному просторі.

17.2. Коваріаційна матриця випадкового вектора

Означення. *Нехай $\xi = (\xi_1, ..., \xi_n)$ випадковий вектор. Його* коваріаційною матрицею *називається матриця*

$$\mathrm{Cov}(\xi) \equiv (\mathrm{Cov}(\xi_i,\xi_j), \ i,j = \overline{1,n}).$$

Теорема (про властивості коваріаційної матриці). *Коваріаційна матриця випадкового вектора симетрична та невід'ємно визначена.*

Доведення. Симетричність очевидна. Невід'ємна визначеність (тобто невід'ємність квадратичних форм $c'\mathrm{Cov}(\xi)c \geq 0, \ \forall c \in \mathbb{R}^n$) випливає з наступної теореми $\boxdot$

Теорема (про дисперсію лінійної форми від випадкового вектора). *Нехай $\xi = (\xi_1, ..., \xi_n)$ випадковий вектор із коваріаційною матрицею $Cov(\xi)$, стала $c \in \mathbb{R}^n$, а $\eta = c'\xi \equiv \sum_{k=1}^n c_k\xi_k$ лінійна форма від ξ. Тоді*

$$D\eta = c'\mathrm{Cov}(\xi)c = \sum\nolimits_{i,\ j} c_i \mathrm{Cov}(\xi_i,\xi_j)c_j.$$

Доведення. Піднесемо до квадрату та скористаємося лінійністю математичного сподівання:

$$D\eta = M(\eta - M\eta)^2 = M(\sum c_k(\xi_k - M\xi_k))^2 =$$
$$M\sum_{i,j} c_i(\xi_i - M\xi_i)c_j(\xi_j - M\xi_j) = \sum_{i,j} c_i \mathrm{Cov}(\xi_i, \xi_j)c_j \boxdot$$

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*

$$\eta = \ C\xi = \ (\sum_{k=1}^{n} c_{ik}\xi_k, \ i = 1, m)$$

відповідне лінійне перетворення ξ. *Тоді*

$$\mathrm{Cov}(\eta) = C\ \mathrm{Cov}(\xi)\ C'.$$

Доведення. Обчислимо за означенням коваріації nf pf лінійністю математичного сподівання

$$\mathrm{Cov}(\eta)_{i,j} = M(\eta_i - M\eta_i)(\eta_j - M\eta_j) =$$
$$M\left(\sum{}_k C_{ik}(\xi_k - M\xi_k)\sum{}_l C_{jl}(\xi_l - M\xi_l)\right) =$$
$$\sum{}_{k,l} C_{ik}\mathrm{Cov}(\xi_k, \xi_l)(C')_{lj} = (C\ \mathrm{Cov}(\xi)\ C')_{ij} \boxdot$$

18. 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* $B_k \in \mathfrak{B}(\mathbb{R})$

$$P(\xi_1 \in B_1, ..., \xi_n \in B_n) = \prod\nolimits_{k=1}^{n} P(\xi_k \in B_k).$$

18.1. 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*

$$F_{(\xi_1,...,\xi_n)}(x_1, ..., x_n) \equiv P(\xi_1 < x_1, ..., \xi_n < x_n) =$$

$$\prod\nolimits_{k=1}^{n} P(\xi_k < x_k) \equiv \prod\nolimits_{k=1}^{n} F_{\xi_k}(x_k).$$

Доведення. Позначимо $(*)$ рівняння в означенні незалежності. Розглянемо клас множин

$$\mathfrak{B}_1^* = \{B_1 \in \mathfrak{B}(\mathbb{R}) : (*) \ \forall \ B_k = (-\infty, x_k), k = \overline{2, n}\}.$$

З теореми про перетворення незалежних подій випливає, що клас $\mathfrak{B}_1^*$ містить множину $\mathbb{R}$ та замкнений відносно вкладеної різниці та об'єднання множин, які не перетинаються. Крім того, за умовою $(-\infty, x_1) \in \mathfrak{B}_1^*$ для всіх x_1. Тому $\mathfrak{B}_1^*$ містить алгебру $\mathfrak{A}(\mathbb{R})$, яка породжується напівінтервалами за допомогою вказаних операцій.

З властивості неперервності ймовірності випливає, що клас $\mathfrak{B}_1^*$ замкнений відносно монотонної збіжності. Дійсно. з $B_1(m) \in \mathfrak{B}_1^*$ та $B_1(m) \uparrow B_1$ виводимо, що

$$P(\xi \in B_1(m) \times (-\infty, x_2).. \times (-\infty, x_n)) \to P(\xi \in B_1 \times (-\infty, x_2).. \times (-\infty, x_n)),$$

$$P(\xi_1 \in B_1(m)) \prod_{k=2}^{n} P(\xi_k < x_k) \to P(\xi_1 \in B_1) \prod_{k=2}^{n} P(\xi_k < x_k),$$

звідки $B_1 \in \mathfrak{B}_1^*$.

Отже, за теоремою про монотонний клас $\mathfrak{B}_1^* = \mathfrak{B}(\mathbb{R})$.

Далі, позначимо

$$\mathfrak{B}_2^* {=} \{B_2 \in \mathfrak{B}(\mathbb{R}) : (*)\ \forall B_1 \in \mathfrak{B}(\mathbb{R}), \forall B_k = (-\infty, x_k),\ \ k = 3, n\}.$$

Клас $\mathfrak{B}_2^*$ містить всі інтервали $(-\infty, x_2)$ за доведеним вище. Звідси аналогічно доводимо, що $\mathfrak{B}_2^* = \mathfrak{B}(\mathbb{R})$.Повторюючи це міркування, за індукцією дістаємо шукане твердження $\boxdot$

Вправа. Дискретні випадкові величини $(\xi_1, ..., \xi_n)$ незалежні в сукупності тоді й тільки тоді, коли для всіх $x_1, ...x_n$

$$P(\xi_1 = x_1, ..., \xi_n = x_n) = \prod_{k=1}^{n} P(\xi_k = x_k)$$

Теорема (про критерій незалежності абсолютно неперервних величин). *Нехай випадковий вектор $\xi = (\xi_1, ..., \xi_n)$ має сумісну щільність $f_\xi(x_1, .x_n)$. Величини $(\xi_1, ..., \xi_n)$ незалежні в сукупності тоді й тільки тоді, коли*

$$f_\xi(x_1, ..., x_n) = \prod_{k=1}^{n} f_{\xi_k}(x_k).$$

Доведення.

За умови абсолютної неперервності умова теореми про критерій незалежності випадкових величин еквівалентна рівності

$$\int_{-\infty}^{x_1} .. \int_{-\infty}^{x_n} f_\xi(y_1, ..., y_n) dy_1 ... dy_n = \prod_{k=1}^{n} \int_{-\infty}^{x_k} f_{\xi_k}(y_k) dy_k =$$

$$= \int_{-\infty}^{x_1} .. \int_{-\infty}^{x_n} \prod_{k=1}^{n} f_{\xi_k}(y_k) dy_1 ... dy_n \quad \forall x_k \in \mathbb{R}.$$

Звідси за означенням сумісної щільності та з наслідку про єдиність сумісної щільності виводимо сформульоване твердження $\boxdot$

Вправа. Якщо випадкові величини $(\xi_1, ..., \xi_n)$ незалежні в сукупності та мають щільності, то утворений ними вектор абсолютно неперервний, а його сумісна щільність збігається з добутком щільностей координат.

18.2. Перетворення незалежних величин

Теорема (про перетворення незалежних величин). *Нехай випадкові величини $(\xi_1, ..., \xi_n)$ незалежні в сукупності, а $g_1, ... g_n$ – борелеві функції. Тоді випадкові величини*

$$(g_1(\xi_1), ..., g_n(\xi_n))$$

незалежні в сукупності.

Доведення.

Обчислимо при $B_k \in \mathfrak{B}(\mathbb{R})$

$$P(g_1(\xi_1) \in B_1, ..., g_n(\xi_n) \in B_n) =$$

$$P(\xi_1 \in g_1^{(-1)}(B_1), ..., \xi_n \in g_n^{(-1)}(B_n)) =$$

$$\prod_{k=1}^{n} P(\xi_k \in g_k^{(-1)}(B_k)) = \prod_{k=1}^{n} P(g_k(\xi_k) \in B_k) \boxdot$$

Теорема (**про векторні перетворення незалежних величин**). *Нехай випадкові величини* $(\xi_1, ..., \xi_n)$ *незалежні в сукупності, а* $f : \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^{n-k} \to \mathbb{R}$ – *борелеві функції. Тоді випадкові величини*

$$f(\xi_1, ..., \xi_k),\ g(\xi_{k+1}, ..., \xi_n)$$

незалежні.

Доведення. За означенням незалежності досить довести, що

$$P((\xi_1, .\xi_k) \in B,\ (\xi_{k+1}, .\xi_n) \in C) = P((\xi_1, .\xi_k) \in B)\ P((\xi_{k+1}, .\xi_n) \in C)$$

для всіх $B \in \mathfrak{B}(\mathbb{R}^k)$, $C \in \mathfrak{B}(\mathbb{R}^{n-k})$. Ця рівність для прямокутників $B = \prod_{i=1}^k B_i$, $C = \prod_{i=k+1}^n C_i$ випливає з означення незалежності. Для множин B, C із відповідних алгебр $\mathfrak{A}(\mathbb{R}^k)$, $\mathfrak{A}(\mathbb{R}^{n-k})$, що є об'єднаннями прямокутників, виводимо її з адитивності ймовірності. Нарешті, для борелєвих B, C шукана рівність справедлива, оскільки відповідні класи множин B, C замкнені відносно монотонної збіжності за властивістю неперервності ймовірностей.

Вправа.

(1) Знайти функцію розподілу максимуму та мінімуму незалежних однаково розподілених випадкових величин через їх спільну функцію розподілу.

(2) 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. Знайти відповідні сумісні розподіли. Узагальнити ці твердження на випадок n незалежних однаково розподілених показникових величин.

(3) Випадкові вектори $(\xi_1, ...\xi_n)$ з $\mathbb{R}^d$ називаються незалежними в сукупності, якщо для довільних $B_k \in \mathfrak{B}(\mathbb{R}^d), k = \overline{1, n}$, випадкові події $\{\xi_k \in B_k\}$ є

незалежними у сукупності. Довести теореми про критерій незалежності випадкових векторів, критерій незалежності абсолютно неперервних векторів, про перетворення незалежних векторів.

19. Математичне сподівання добутку незалежних величин

19.1. Математичне сподівання добутку

Теорема (про математичне сподівання добутку незалежних величин). *Якщо величини ξ, η незалежні та інтегровні, то добуток $\xi\eta$ є інтегровним і*

$$M\xi\eta = M\xi \; M\eta.$$

Доведення.
(а) Нехай $\xi = 1_A$, $\eta = 1_B$. Тоді події A, B незалежні і

$$M\xi\eta = M \; 1_A 1_B = M 1_{A\cap B} = P(A \cap B) = P(A)P(B) = M\xi \; M\eta.$$

(б) Нехай величини ξ, η незалежні і прості, $\xi = \sum x_i 1_{A_i}$, $\eta = \sum y_j 1_{B_j}$. Тоді за означенням незалежності випадкових величин події $\{\xi = x_i\} = A_i$ і $\{\eta = y_j\} = B_j$ незалежні, а випадкова величина $\xi\eta$ дорівнює $x_i y_j$ на множині елементарних подій $A_i \cap B_j$.За теоремою про обчислення математичного сподівання функції від дискретної величини для функції $g((\xi, \eta)) = \xi\eta$

$$M(\xi\eta) = \sum_{i,j} x_i y_j \; P(\xi = x_i, \eta = y_j) = \sum_{i,j} x_i y_j P(A_i \cap B_j) =$$
$$\sum_{i,j} x_i y_j P(A_i)P(B_j) = \sum_i x_i P(A_i) \sum_j y_j P(B_j) = M\xi \; M\eta.$$

(в) Якщо $\xi, \eta \geq 0$, за теоремою про апроксимацію випадкових величин простими побудуємо прості апроксимуючі величини $0 \leq \xi_n \uparrow \xi$, $0 \leq \eta_n \uparrow \eta$. Тоді $0 \leq \xi_n \eta_n \uparrow \xi\eta$, причому ξ_n, η_n незалежні за визначенням $\xi_n = \varphi_n(\xi)$, $\eta_n = \varphi_n(\eta)$ та за теоремою про перетворення незалежних величин.

З твердження (б) та з теореми Лебега про монотонну збіжність із рівняння $M\xi_n\eta_n = M\xi_n\ M\eta_n$ граничним переходом дістанемо шукану рівність.

(г) Для знакозмінних інтегровних ξ, η за теоремою про перетворення незалежних величин величини $\xi^{\pm}, \eta^{\pm}$ незалежні. Тому за лінійністю

$$M\xi\eta = M(\xi^+\eta^+ - \xi^-\eta^+ - \xi^+\eta^- + \xi^-\eta^-) =$$
$$M\xi^+M\eta^+ - M\xi^-M\eta^+ - M\xi^+M\eta^- + M\xi^-M\eta^- = M\xi M\eta \boxdot$$

19.2. Дисперсія суми незалежних величин

Теорема (про дисперсію суми незалежних величин). *Нехай випадкові величини* $(\xi_1, ..., \xi_n)$ *попарно незалежні. Тоді*

$$D\left(\sum\nolimits_{k=1}^n \xi_k\right) = \sum\nolimits_{k=1}^n D(\xi_k).$$

Доведення. Обчислимо за означенням дисперсії та за лінійністю математичного сподівання

$$D\left(\sum\nolimits_{k=1}^n \xi_k\right) = M\left(\sum\nolimits_{k=1}^n (\xi_k - M\xi_k)\right)^2 = \sum\nolimits_{i,j} M(\xi_i - M\xi_i)(\xi_j - M\xi_j) =$$
$$\sum\nolimits_{i,j}(M(\xi_i\xi_j) - M\xi_i M\xi_j) = \sum\nolimits_{i=j}(M(\xi_i\xi_j) - M\xi_i M\xi_j) =$$
$$\sum\nolimits_i (M\xi_i^2 - (M\xi_i)^2) = \sum\nolimits_{k=1}^n D\xi_k.$$

де при $i \neq j$ використано теорему про властивості дисперсії та про математичне сподівання добутку незалежних величин $\boxdot$

Приклад. Обчислення дисперсії біноміального розподілу. Нехай $\chi_k = 1_{Y_k}$ – індикаторна величина успіху в k-му випробуванні Бернуллі. За умовою ці величини незалежні. Тоді загальна кількість успіхів дорівнює $\xi =$

$\sum_{k=1}^{n}\chi_k$, отже за формулою про дисперсію індикаторної величини $D\xi = \sum_{k=1}^{n} D\chi_k = np(1-p)$.

Вправа.

(1). Для незалежних ξ, η обчислити $D(\xi\eta)$.

(2). Випадкові величини $\xi_1, ..., \xi_n$ незалежні та однаково розподілені, причому $M(\xi_1 - M\xi_1)^3 = 0$. Довести, що випадкові величини $\mu = \frac{1}{n}\sum_{k=1}^{n}\xi_k$ та $\sum_{k=1}^{n}(\xi_k - \mu)^2$ некорельовані.

20. Розподіл суми незалежних величин. Гама-розподіл

20.1. Розподіл суми незалежних величин

Для обчислення функції розподілу суми незалежних величин доведемо лему.

Лема. Нехай випадкові величини ξ, η незалежні та мають функції розподілу F_ξ, F_η. Тоді для довільної множини $B \in \mathfrak{B}(\mathbb{R}^2)$

$$P((\xi, \eta) \in B) = \int_{-\infty}^{\infty} F_\xi(B_y)\, dF_\eta(y),$$

де $B_y = \{x : (x, y) \in B\}$ – переріз множини B на рівні y.

Доведення.

(а) Для прямокутників $B = C \times D$ шукана рівність випливає з незалежності ξ, η та з тотожності $(C \times D)_y = C$ при $y \in D$ та $(C \times D)_y = \varnothing$ при $y \notin D$, звідки $F_\xi(B_y) = F_\xi(C) 1_{y \,\in\, D}$ і

$$P((\xi, \eta) \in B) = \int_{-\infty}^{\infty} F_\xi(C) 1_{y \,\in\, D}\, dF_\eta(y) = F_\xi(C) F_\eta(D) = P(\xi \in C, \eta \in D).$$

(б) Оскільки відображення $B \to B_y$ є адитивним відносно об'єднання попарно несумісних множин, а обидві частини рівності леми – лінійні за B, то ця рівність залишається справедливою для множин з алгебри $\mathfrak{A}(\mathbb{R}^2)$.

(в) Нарешті, з теореми Лебега про монотонну збіжність для правої частини та з властивості неперервності ймовірності для лівої виводимо, що клас борелєвих множин, для яких рівність леми має місце, замкнений відносно монотонної збіжності. Тому за теоремою про монотонний клас дана рівність виконується для всіх $B \in \mathfrak{B}(\mathbb{R}^2)$ $\boxdot$

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.

Нехай $B^a = \{(x, y) : x + y < a\}$. Тоді $P(\xi + \eta < a) = P((\xi, \eta) \in B^a)$ і переріз $B^a_y = \{x : x < a - y\}$. Після підстановки B^a у попередню лему отримуємо

$$P(\xi + \eta < a) = P((\xi, \eta) \in B^a) = \int_{-\infty}^{\infty} F_\xi(B^a_y)\, dF_\eta(y) =$$
$$\int_{-\infty}^{\infty} F_\xi(a - y)\, dF_\eta(y) = F_\xi * F_\eta(a) \boxdot$$

Наслідок (про властивості згортки).

(а) Згортка функцій розподілу є функцією розподілу.

(б) Згортка є комутативною та асоціативною операцією.

Доведення очевидне $\boxdot$

Вправа.

(1) Для незалежних дискретних випадкових величин ξ, η

$$P(\xi + \eta = a) = \sum_x P(\xi = x) P(\eta = a - x),$$

де сума поширюється на множину значень величини ξ і називається згорткою дискретних розподілів величин ξ та η.

(2) Якщо ξ, η – незалежні величини з розподілами Пуассона з параметрами λ, μ відповідно. Тоді сума $\xi + \eta$ має розподіл Пуассона з параметром $\lambda + \mu$.

Означення. Згорткою щільностей розподілу f, g *називається функція*

$$f * g(a) = \int_{-\infty}^{\infty} f(a-y)\, g(y) dy.$$

Теорема (про щільність розподілу суми незалежних величин). *Нехай випадкові 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 $a \in \mathbb{R}$ справедливі тотожності

$$F_\xi * F_\eta(a) = \int_{-\infty}^{\infty} F_\xi(a-y)\, dF_\eta(y) = \int_{-\infty}^{a} F_\xi(a-y)\, f_\eta(y) dy =$$

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{a-y} f_\xi(x) f_\eta(y) dx \right) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_\xi(x) f_\eta(y) 1_{x+y\,<\,a}\, dx\, dy =$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_\xi(u-y) f_\eta(y) 1_{u\,<\,a}\, du\, dy = \int_{-\infty}^{a} f_\xi * f_\eta(u)\, du,$$

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 $\boxdot$

Вправа.

Старт

Початок

Зміст

Сторінка 135 з 509

Назад

Екран

Закрити

Выхід

(1) Випадкові величини $\xi_1, ..., \xi_n, \eta$ незалежні у сукупності, $P(\xi_k = 0) = P(\xi_k = 1) = 1/2$, а η рівномірно розподілена на $[0, 1]$. Довести, що $\xi = \sum_{k=1}^{n} 2^{-k}\xi_k + 2^{-n}\eta$ також рівномірно розподілена на $[0, 1]$.

(2, Роббінс) Випадкові величини $\xi_1, ..., \xi_n$ незалежні у сукупності, а $\zeta_k = \xi_1 + .. + \xi_k$ – їх частинні суми. Довести, що

$$P(\zeta_1 \le x_1, ..., \zeta_n \le x_n) \ge \prod_{k=1}^{n} P(\zeta_k \le x_k),\ \forall x_1, ..., x_n$$

(3) Випадкові величини $\xi_1, ..., \xi_n$ незалежні у сукупності та рівномірно розподілені на $[0, 1]$. Тоді щільність суми $\xi_1 + ... + \xi_n$ дорівнює

$$u_n(x) = \frac{1}{(n-1)!} \sum_{k=0}^{n} (-1)^k C_n^k (\max(x - k, 0))^{n-1}.$$

(4) Випадкові величини ξ_1, ξ_2 незалежні та мають показниковий розподіл $Exp(\lambda)$. Тоді різниця $\zeta = \xi_1 - \xi_2$ має розподіл Лапласа зі щільністю $f_\zeta(x) = \frac{1}{2}\lambda \exp(-\lambda |x|)$.

20.2. Розподіли Ерланга, Гама та хі-квадрат

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 щільністю:*

$$f_n(x) = \lambda(\lambda x)^{n-1} \exp(-\lambda x)/(n - 1)!,\ x \ge 0.$$

Доведення.
За умовою $f_1(x) = \lambda \exp(-\lambda x), x \ge 0$.
Тут та нижче при $x < 0$ всі щільності $f_n(x) = 0$.

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 рекурентне 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 $n \geq 2$ рекурентні рівняння

$$g_n(x) = \lambda \int_0^x g_{n-1}(y)dy, \quad g_1(x) = \lambda,\ \forall x \geq 0,$$

звідки за індукцією знаходимо

$$g_n(x) = \lambda(\lambda x)^{n-1}/(n-1)!, \quad f_n(x) = \lambda(\lambda x)^{n-1}\exp(-\lambda x)/(n-1)!\ \boxdot$$

Якщо замінити в означенні розподілу Ерланга факторіал 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* $\lambda, \alpha > 0$, *позначення* $\zeta \simeq \Gamma(\lambda, \alpha)$, *якщо її щільність дорівнює*

$$f(x) = \lambda(\lambda x)^{\alpha-1}\exp(-\lambda x)\ /\ \Gamma(\alpha),\ x \geq 0.$$

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 має гама-розподіл*

$$\zeta_1 + \zeta_2 \simeq \Gamma(\lambda, \alpha+\beta).$$

Доведення. За теоремою про щільність розподілу суми незалежних величин

$$f_{\zeta_1+\zeta_2}(x) = f_{\zeta_1} * f_{\zeta_2}(x) =$$

$$\int_0^x \lambda(\lambda y)^{\alpha-1}\exp(-\lambda y)\ \lambda(\lambda x-\lambda y)^{\beta-1}\exp(-\lambda x+\lambda y)dy/\Gamma(\alpha)\Gamma(\beta) =$$

$$= \lambda^{\alpha+\beta}\ x^{\alpha+\beta-1}\ \exp(-\lambda x)\ \int_0^1 (1-t)^{\alpha-1}t^{\beta-1}dt\ /\ \Gamma(\alpha)\Gamma(\beta)$$

$$= \lambda(\lambda x)^{\alpha+\beta-1}\exp(-\lambda x)\ /\ \Gamma(\alpha+\beta),\ x \ge 0,$$

де зроблено заміну $t = y/x$ та використані означення та властивість повної бета-функції $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ ⊡

Частковим випадком гама-розподілу є розподіл хі-квадрат, який широко застосовується в статистиці.

Теорема (**про хі-квадрат розподіл**). *Нехай* $(\zeta_n,\ n \ge 1)$ *незалежні в сукупності стандартні нормальні величини:* $\zeta_n \simeq N(0,1)$. *Сума квадратів*

$$\chi_n^2 = \zeta_1^2 + \ldots + \zeta_n^2$$

має гама-розподіл $\Gamma(1/2,\ n/2)$, *який називається хі-квадрат розподілом із* n *ступенями свободи.*

Доведення. Перевіримо спочатку, що $\zeta_1^2 \simeq \Gamma(1/2, 1/2)$. Ця рівність є наслідком теореми про обчислення розподілу функції від випадкової величини :

Старт
Початок
Зміст
◀◀ ▶▶
◀ ▶
Сторінка 138 з 509
Назад
Екран
Закрити
Вихід

$$P(\zeta_1^2 < x) = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)dy = 2\int_0^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)dy =$$
$$\int_0^x \frac{1}{\sqrt{2\pi}} u^{-1/2}\ \exp(-u/2)du = \frac{1}{\Gamma(1/2)} \int_0^x (u/2)^{-1/2}\ \exp(-u/2)du/2.$$

Отже, при $n = 1$ теорема доведена. Оскільки величини ζ_k^2 незалежні в сукупності, то доданки в сумі $\chi_{n+1}^2 = \chi_n^2 + \zeta_n^2$ 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 $\chi_{n+1}^2 \simeq \Gamma(1/2, n/2) + \Gamma(1/2, 1/2) \simeq \Gamma(1/2, n/2 + 1/2)$, що доводить теорему за індукцією $\boxdot$

Вправа.

(1) 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.

(2) Довести формулу Ліувілля перетворення кратних інтегралів

$$\int_0^\infty ... \int_0^\infty f(x_1 + ... + x_n) x_1^{\alpha_1 - 1} ... x_n^{\alpha_n - 1} dx_1 ... dx_n =$$
$$\frac{\Gamma(\alpha_1)...\Gamma(\alpha_n)}{\Gamma(\alpha_1 + ... + \alpha_n)} \int_0^\infty f(y) y^{\alpha_1 + ... + \alpha_n - 1} dy.$$

21. Нормальні випадкові вектори

У даному розділі, як і в інших, всі вектори розглядаються як вектори-стовпчики. Символ $'$ означає транспонування векторів та матриць.

Означення. *Випадковий вектор* $\zeta = (\zeta_1, ..., \zeta_n)'$ *називається* n*-вимірним* стандартним нормальним вектором, *позначення* $\zeta \simeq N_n(0, I)$, *якщо випадкові величини* $\zeta_1, ..., \zeta_n$ *є незалежними в сукупності стандартними нормальними величинами.*

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

$$f_\zeta(x) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}} \exp(-x_k^2/2) =$$

$$= (2\pi)^{-n/2} \exp(-\frac{1}{2} \sum_{k=1}^{n} x_k^2) = (2\pi)^{-n/2} \exp(-\frac{1}{2} x\prime x).$$

Означення. *Випадковий вектор* $\xi = (\xi_1, ..., \xi_n)'$ *називається* n*-вимірним* 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 є 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.

21.1. 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*

$$f_\xi(x) = (2\pi)^{-n/2} |\det V|^{-1/2} \exp(-\frac{1}{2}(x-m)'V^{-1}(x-m)), \ x \in \mathbb{R}^n,$$

де $m \in \mathbb{R}^n$, а матриця $V \in \mathfrak{L}_n(\mathbb{R}^n)$ є симетричною додатно визначеною. Якщо $\xi = m + A\,\zeta$, $\zeta \simeq N_n(0,I)$, 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. Останнє позначення для випадку, коли $m = 0$ і $V = I$ – одинична матриця, відповідає за означенням стандартному нормальному вектору.

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

$$\Pi_x = (-\infty, x) = \prod_{k=1}^n (-\infty, x_k).$$

За теоремою про обчислення ймовірностей, пов'язаних із випадковою величиною, через її щільність

$$F_\xi(x) = P(\xi < x) = P(m + A\,\zeta \in \Pi_x) = P(\zeta \in A^{-1}(\Pi_x - m)) =$$

$$\int_{y \in A^{-1}(\Pi_x - m)} f_\zeta(y)dy = \int_{y:\ m+Ay\in\Pi_x} f_\zeta(y)dy =$$

$$\int_{\Pi_x} f_\zeta(A^{-1}(u-m))\left|\det A^{-1}\right| du,$$

де використана заміна змінної $u = m + Ay$ з якобіаном $\left|\det A^{-1}\right|$. Звідси за означенням сумісної щільності

$$f_\xi(u) = f_\zeta(A^{-1}(u-m))\left|\det A^{-1}\right| =$$

$$|\det A|^{-1}\ (2\pi)^{-n/2}\ \exp(-\frac{1}{2}(A^{-1}(u-m))'A^{-1}(u-m)) =$$

$$(2\pi)^{-n/2}\ |\det V|^{-1/2}\ \exp(-\frac{1}{2}(u-m)'V^{-1}(u-m)),$$

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 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 і $\xi = m + A\,\zeta$ за означенням ζ
$\boxdot$

21.2. Параметри розподілу нормального вектора

Теорема (про інтерпретацію параметрів нормального розподілу). *Нехай нормальний вектор* $\xi \simeq N_n(m, V)$. *Тоді*

$$M\xi = m, \ \mathrm{Cov}(\xi) = V.$$

Доведення. Очевидно, що за лінійністю

$$M\xi = M(m + A\,\zeta) = m + A\,M\zeta = m + A0 = m.$$

Далі, за теоремою про коваріаційну матрицю лінійного перетворення випадкового вектора

$$\mathrm{Cov}(\xi) = \mathrm{Cov}(\xi - m) = \mathrm{Cov}(A\zeta) = A\mathrm{Cov}(\zeta)A' = AIA' = V \ \boxdot$$

21.3. 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*

$$f_\xi(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times$$

$$\exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x_1-\mu_1}{\sigma_1}\frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right),$$

причому $M\xi_k = \mu_k, \ D\xi_k = \sigma_k^2, \ k = 1, 2,$ *а коефіцієнт кореляції між* ξ_1 *і* ξ_2 *дорівнює*

$$\rho(\xi_1,\xi_2) = \mathrm{Cov}(\xi_1,\xi_2)/\sigma_1\sigma_2 = M(\xi_1-\mu_1)(\xi_2-\mu_2)/\sigma_1\sigma_2 = \rho.$$

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 формули обертання матриці

$$V^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2} \end{pmatrix},$$

та з теореми про сумісну щільність нормального вектора знаходимо шукану формулу для щільності. Подальше випливає з теореми про інтерпретацію параметрів нормального розподілу та з означення коефіцієнта кореляції ⊡

Вправа. Нехай $(\xi_1,\xi_2) \simeq N_2$ з $M\xi_k = 0, D\xi_k = 1, \rho = \mathrm{Cov}(\xi_1,\xi_2)$. Довести, що: (а) $P(\xi_1\cdot\xi_2 < 0) = (\arccos\rho)\ /\ \pi$, (б) щільність відношення ξ_1/ξ_2 дорівнює $\sqrt{1-\rho^2}/\ \pi(1-2\rho x+x^2)$.

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 $Cov(\xi_1,\xi_2) = 0$.*

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 шукане твердження випливає з теореми про критерій незалежності абсолютно неперервних величин ⊡

Вправа. Нехай (ξ_1,ξ_2) нормальний вектор на площині з нульовим середнім, причому $\mathrm{Cov}(\xi_1,\xi_2) < 0$ Довести, що

$$P(\xi_1 \ge x_1,\ \xi_2 \ge x_2) \le P(\xi_1 \ge x_1)\ P(\xi_2 \ge x_2), \quad \forall x_k \ge 0.$$

21.4. Лінійні перетворення нормальних векторів

Теорема (про лінійні перетворення нормальних векторів).

(а) Якщо $\xi \simeq N_n(m, V)$, стала $a \in \mathbb{R}^n$, і перетворення $T \in \mathfrak{L}_n(\mathbb{R}^n)$ невироджене, то $\eta \equiv a + T\xi \simeq N_n(a + Tm,\ TVT')$.

(б) Якщо $\zeta \simeq N_n(0, I)$ стандартний нормальний вектор, а матриця U – ортонормована, 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.

Доведення.

(а) Нехай $\xi = m + A\,\zeta$, де вектор ζ – стандартний нормальний. Тоді

$\eta = a + T\xi = a + Tm + TA\,\zeta$.

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 $MU\zeta = UM\zeta = 0$, $\mathrm{Cov}(U\zeta) = UIU' = I$

(в) Нехай $\xi_k = m_k + A_k\zeta_k$, де ζ_k – стандартні нормальні вектори. Тоді вектори $\zeta_k = A_k^{-1}(\xi_k - m_k)$ знову 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) $\boxdot$

Вправа. Нехай $f_k(x_1, x_2)$ щільність двовимірного нормального розподілу 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, однак його координати нормально розподілені.

Теорема (про збереження нормальності розподілу). *Нехай $T \in \mathfrak{L}(\mathbb{R}^n, \mathbb{R}^k)$ для деякого $k < n$, і $rang(T) = k$.*

Якщо $\xi \simeq N_n(m, V)$, то $T\xi \simeq N_k(Tm,\ TVT')$.

Доведення.

Нехай $\xi = m + A\ \zeta$, де вектор ζ – стандартний нормальний. Тоді $\eta = T\xi = Tm + T\zeta = Tm + B\zeta \in \mathbb{R}^k$, де $B = TA$, $\mathrm{rang}(B) = k$.

Оскільки розмірність ядра $Ker(B) \equiv \{x : Bx = 0\}$ 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. Зауважимо, що

$$BU = (Be_1, ..., Be_n) = (Be_1, ..., Be_k, 0, ..., 0)$$

за означенням U. 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:

$$\mathrm{Cov}(T\xi) = T\mathrm{Cov}(\xi)T' = TVT' \boxdot$$

Вправа. Якщо ζ_k – незалежні нормальні стандартні 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.

21.5. Незалежність і некорельованість нормальних величин

Наслідок (про незалежність лінійних форм нормального вектора) *Нехай $\xi \simeq N_n(m, V)$, і $c, d \in \mathbb{R}^n$, $rang(c, d) = 2$. Для того, щоб лінійні форми $c'\xi$ і $d'\xi$ були незалежні, необхідно й достатньо, щоб вони були некорельовані*

$$\mathrm{Cov}(c'\xi,\ d'\xi) \equiv c'Vd = 0.$$

Доведення. За теоремою про збереження нормальності розподілу вектор $(c'\xi,\ d'\xi)$ є нормальним вектором на площині. Тому твердження теореми випливає з теореми про незалежність координат нормального вектора на площині.

Наслідок (про нормальність суми незалежних нормальних векторів) *Нехай $\xi_k \simeq N_n(m_k, V_k)$ – незалежні нормальні випадкові вектори. Тоді сума $\xi_1 + \xi_2 \simeq N_n(m_1 + m_2, V_1 + V_2)$ теж має нормальний розподіл.*

Доведення. За теоремою про лінійні перетворення нормальних векторів, (в), складений вектор (ξ_1, ξ_2) є нормальним. Тому його лінійна форма $\xi_1 + \xi_2$ є нормальним вектором за теоремою про збереження нормальності розподілу. Оскільки координати ξ_1 незалежні і некорельовані з координатами ξ_2, то

$$M(\xi_1 + \xi_2) = m_1 + m_2,$$

$$\mathrm{Cov}(\xi_1+\xi_2)=\mathrm{Cov}(\xi_1)+\mathrm{Cov}(\xi_2)=V_1+V_2,$$

отже, розподіл нормального вектора $\xi_1+\xi_2 \simeq N_n(m_1+m_2, V_1+V_2)$ $\boxdot$

Вправа. Узагальнений нормальний вектор визначається як лінійне перетворення $\xi = m + A\,\zeta$ стандартного нормального з довільною матрицею $A \in \mathfrak{L}(\mathbb{R}^n, \mathbb{R}^n)$. Довести, що: (а) розподіл ξ однозначно визначається вектором m та матрицею $V = AA'$, (б) теорему про інтерпретацію параметрів, (в) теорему про лінійні перетворення, (г) теорему про збереження нормальності розподілу для класу узагальнених 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 $M\xi_\varepsilon = m$, $\mathrm{Cov}(\xi_\varepsilon) = (A+\varepsilon T)(A+\varepsilon T)' = V+\varepsilon^2 J$, то розподіл $\xi_\varepsilon \simeq N_n(m, V+\varepsilon^2 J)$ повністю визначається параметрами ε та m, V. Нарешті, границя $P(\xi \in \Pi_x) = \lim_{\varepsilon\to 0} P(\xi_\varepsilon \in \Pi_x)$ залежить лише від m, V та x.

22. Збіжність за ймовірністю та її властивості.

Означення. *Послідовність випадкових величин* $(\xi_n,\ n \ge 1)$ збігається за ймовірністю *до величини* ξ *(позначення* $\xi_n \to^P \xi$*), якщо для кожного* $\varepsilon > 0$

$$P(\ |\xi_n - \xi| \ge \varepsilon) \to 0 \text{ при } n \to \infty.$$

Зауваження. Якщо одночасно $\xi_n \to^P \eta$, то $P(|\eta - \xi| \ge \varepsilon) = 0$ для всіх $\varepsilon > 0$, звідки $\eta - \xi = 0$ м.н., тобто границя за ймовірністю визначена однозначно з точністю до рівності майже напевне.

Вправа. Доведіть, що клас усіх випадкових величин із метрикою Леві

$$\rho(\xi, \eta) = \inf(\varepsilon > 0 : P(|\xi - \eta| \ge \varepsilon) \le \varepsilon)$$

є повним метричним простором, збіжність в якому є збіжністю за ймовірністю. Вказівка. Якщо $P(|\xi - \eta| \ge \varepsilon) \le \delta$, то $\rho(\xi, \eta) \le \max(\varepsilon, \delta)$. Звідси

$$\rho(\xi, \eta) = \inf(\max(\varepsilon, \delta) > 0 : P(|\xi - \eta| \ge \delta) \le \varepsilon).$$

Тому $\rho(\xi_n, \xi) \to 0$ тоді й тільки тоді, коли $\lim P(|\xi_n - \xi| \ge \delta) = 0,\ \forall \delta > 0$. З нерівності $P(|\xi - \eta| \ge \varepsilon + \delta) \le P(|\xi - \zeta| \ge \varepsilon) + P(|\zeta - \eta| \ge \delta)$ випливає нерівність трикутника. Доведення повноти спирається на можливість виділення з послідовності, що є фундаментальною за ймовірністю, підпослідовності, яка фундаментальна майже напевне.

22.1. Властивості збіжності за ймовірністю

Теорема (про властивості збіжності за ймовірністю). *Нехай* $\xi_n \to^P \xi$.

(а) Якщо $f \in C(\mathbb{R})$, *то* $f(\xi_n) \to^P f(\xi)$.

(б) Якщо $|\xi_n| \le c$, *то* $M\xi_n \to M\xi$.

(в) Якщо $V(x)$ *додатна неперервна функція,* $V(x)/x \to \infty$ *при* $x \to \infty$, *та* $\sup_n M\ V(|\xi_n|) < \infty$, *то* $M\xi_n \to M\xi$.

Доведення.

(а) Для довільного $\gamma > 0$ оберемо $c > 0$ так, щоб $P(|\xi| > c) \le \gamma$. Позначимо

$$w_{c,\delta}(f) = \sup_{|x| \le c,\ |x-y| \le \delta} |f(x) - f(y)|$$

модуль неперервності функції f у δ-околі відрізка $[-c, c]$. 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 з припущень $|\xi_n - \xi| < \delta$ і $|\xi| \le c$ виводимо, що $|f(\xi_n) - f(\xi)| < \varepsilon$. Отже, за вказаним вибором $\delta > 0$ має місце включення

$$\{|f(\xi_n) - f(\xi)| \ge \varepsilon\} \subset \{|\xi| > c\} \cup \{|\xi_n - \xi| \ge \delta\},$$

і за напівадитивністю ймовірності

$$P(|f(\xi_n) - f(\xi)| \ge \varepsilon) \le P(|\xi| > c) + P(|\xi_n - \xi| \ge \delta) \le$$

$$\gamma + P(|\xi_n - \xi| \ge \delta) \le 2\gamma$$

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

$$|\xi_n - \xi| \le \varepsilon + 2c\, 1_{\{|\xi_n - \xi| \ge \varepsilon\}}$$

і з монотонності математичного сподівання отримуємо

$$\overline{\lim}_{n\to\infty} M\, |\xi_n - \xi| \le \varepsilon + 2c\, \overline{\lim}_{n\to\infty} P(|\xi_n - \xi| \ge \varepsilon) = \varepsilon.$$

Зважаючи на довільність $\varepsilon > 0$, звідси виводимо шукане твердження.

(в) 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) $|\tau_c(x)| \le c, \quad \tau_c \in C_b.$

Отже, згідно з доведеними властивостями (а,б)

$$\tau_c(\xi_n) \to^P \tau_c(\xi),\ M\tau_c(\xi_n) \to M\tau_c(\xi),\ n \to \infty.$$

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 $\varepsilon(c)$

$$|x - \tau_c(x)|\ \le |x|\ 1_{|x|\,>\,c} \le \varepsilon(c) V(|x|),\ \forall x \in \mathbb{R}.$$

Далі, з неперервності V випливає збіжність $V(|\xi_n|) \to^P V(|\xi|)$, звідки за лівою нерівністю Фату (що виконується для довільних невід'ємних величин $V(|\xi_n|)$)

$$MV(|\xi|) = M\underline{\lim}_{n\to\infty} V(|\xi_n|) \le \underline{\lim}_{n\to\infty} MV(|\xi_n|) < \infty.$$

Тому з означення τ_c та попередніх оцінок дістанемо нерівність

$$\overline{\lim}_{n\to\infty}\ |M\xi_n - M\xi|\ \le\ \overline{\lim}_{n\to\infty}\ M\,|\xi_n - \tau_c(\xi_n)| +$$
$$\overline{\lim}_{n\to\infty}\ |M\tau_c(\xi_n) - M\tau_c(\xi)|\ + M\ \ |\tau_c(\xi) - \xi|\ \le$$
$$\varepsilon(c)\sup_{n\ge1} MV(|\xi_n|) + 0 + \varepsilon(c)MV(|\xi|).$$

Вибором c праву частину можна зробити як завгодно малою $\boxdot$

Вправа.

(1) Послідовність випадкових величин $(\xi_n,\ n\ge1)$ обмежена за ймовірністю, тобто $\lim_{c\to\infty}\sup_{n\ge1} P(|\xi_n|\ge c) = 0$, а $\eta_n \to^P 0,\ n\to\infty$. Довести, що $\xi_n\eta_n \to^P 0$.

(2) Нехай $\xi_n \to^P \xi,\ \eta_n \to^P \eta$, а $f\in C(\mathbb{R}^2)$. Довести, що $f(\xi_n,\eta_n) \to^P f(\xi,\eta)$.

22.2. Інші види збіжності випадкових величин

Означення. *Послідовність випадкових величин* $(\xi_n, n\ge1)$ збігається з імовірністю 1 *до величини* ξ *або ж* збігається майже напевне *(позначення* $\xi_n \to^{P1} \xi$*), якщо*

$$P(\{\omega : \exists\ \lim_{n\to\infty}\ \xi_n(\omega) = \xi(\omega)\}) = 1.$$

Означення. *Послідовність випадкових величин* $(\xi_n\ , n\ge1)$ *збігається до величини* ξ у середньому *порядку* q *(позначення* $\xi_n \to^{Lq} \xi$*), якщо* $M\,|\xi_n - \xi|^q \to 0$ *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*

$$P1 \Longrightarrow P \Longleftarrow L_1 \Longleftarrow L_q, q>1.$$

Вправа. Наведіть приклади, для яких інші імплікації є хибними. Вказівка. Нехай $(\Omega, \digamma, P) = ([0,1),\ \mathfrak{B}[0,1),\ L)$. Тоді
$(P \not\Rightarrow P1)$ $\xi_n = 1_{\omega \in [\sqrt{n}, \sqrt{n+1})\ mod\ 1}$, $(P \not\Rightarrow L_1)$ $\xi_n = n^2 1_{\omega \in [1/(n+1),\ 1/n)}$.

Доведення.

За означенням збіжність $\xi_n \to^{P1} \xi$ з імовірністю 1 еквівалентна такому твердженню

$$\exists O \in \digamma : P(O) = 1,\ \forall \omega \in O, \forall \varepsilon > 0$$
$$\exists N = N(\varepsilon, \omega) < \infty : |\xi_n(\omega) - \xi(\omega)| < \varepsilon.$$

Отже, випадковий номер $N_\varepsilon(\omega) = \sup(n : |\xi_n - \xi| \ \geq \varepsilon)$ – скінченний майже напевне.

Тому імплікація $P1 \Longrightarrow P$ випливає з $P(\overline{O}) = 0$ та включення

$$\{|\xi_n - \xi| \geq \varepsilon\} \subset \overline{O} \cup (O \cap \{n \leq N_\varepsilon(\omega)\}) \downarrow \overline{O} \cup \varnothing,\ n \to \infty.$$

Друга імплікація $P \Longleftarrow L_1$ є наслідком загальної нерівності Чебишева з функцією $g(x) = x$ та випадковою величиною $|\xi_n - \xi|$:

$$P(|\xi_n - \xi| \geq \varepsilon) \leq M\,|\xi_n - \xi| \ / \ \varepsilon \to 0,\ n \to \infty.$$

Нарешті, імплікація $L_1 \Longleftarrow L_q$ є наслідком нерівності Ляпунова

$$M\,|\xi_n - \xi| \leq (M\,|\xi_n - \xi|^q)^{1/q} \boxdot$$

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:

$$P1 \Longrightarrow L_1,\ P \Longrightarrow L_1.$$

22.3. Критерій збіжності майже напевне

За означенням послідовність випадкових величин $(\xi_n,\ n \geq 1)$ збігається майже напевне до величини ξ, якщо подія

$$A = \{\ \exists\ \lim \xi_n = \xi\} = \cap_{k\geq 1} \cup_{m\geq 1} \cap_{n\geq m} \{|\xi_n - \xi| \leq 1/k\}$$

має одиничну ймовірність: $P(A) = 1$.

Теорема (про критерій збіжності майже напевне). *Послідовність $(\xi_n,\ n \geq 1)$ збігається майже напевне до величини ξ тоді й тільки тоді, коли*

$$\sup_{k\geq n}\ |\xi_n - \xi| \to^P 0.$$

Доведення. Позначимо $A_{km} = \cap_{n\geq m}\{|\xi_n - \xi| \leq 1/k\}$.

Оскільки ці події монотонно не зростають за k, то $\cup_{m\geq 1} A_{km} \downarrow A$ при $k \to \infty$ і за властивістю неперервності ймовірності $P(A) = 1$ тоді й тільки тоді, коли $P(\cup_{m\geq 1} A_{km}) = 1$ для кожного k. Так як A_{km} монотонно не спадають за m, то остання рівність еквівалентна $\lim_{m\to\infty} P(A_{km}) = 1$ для кожного k. За означенням збіжності за ймовірністю це твердження еквівалентне умові теореми, оскільки

$$P(A_{km}) = P(\sup_{n\geq m} |\xi_n - \xi| \leq 1/k) = 1 - P(\sup_{n\geq m} |\xi_n - \xi| > 1/k)\ \boxdot$$

Вправа.

(1) 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 майже напевне.

(3) У дискретному ймовірнісному просторі збіжність майже напевне еквівалентна збіжності за ймовірністю.

(4, теорема Єгорова) Якщо $\xi_n \to^{P1} \xi, n \to \infty$, то для кожного $\varepsilon > 0$ знайдеться подія A_ε така, що $P(A_\varepsilon) \geq 1 - \varepsilon$ та $\xi_n(\omega) \to \xi(\omega),\ n \to \infty$, рівномірно за $\omega \in A_\varepsilon$.

(5) Випадкові величини $(\xi_n,\ n \geq 1)$ незалежні, однаково рівномірно розподілені на $[0, 1]$. Довести, що множина $\{\xi_n,\ n \geq 1\}$ щільна в $[0, 1]$ майже напевне.

23. Закон великих чисел

Означення. *Для послідовності випадкових величин* $(\xi_n,\ n \ge 1)$ *виконується* закон великих чисел, *якщо частинні суми* $S_n = \xi_1 + ... +\ \xi_n$ *задовольняють граничне співвідношення*

$$(S_n - MS_n)\ /\ n \to^{P} 0, \text{ при } n \to \infty.$$

У випадку однаково розподілених випадкових величин $MS_n = nM\xi_1$, тому твердження закону великих чисел еквівалентне збіжності за ймовірністю середніх арифметичних

$$S_n\ /\ n \to^{P} M\xi_1, \text{ при } n \to \infty.$$

23.1. Теорема Чебишева про закон великих чисел

Теорема (теорема Чебишева про закон великих чисел). *Нехай випадкові величини* $(\xi_n,\ n \ge 1)$ *попарно незалежні і*

$$D(\xi_n) = o(n),\ n \to \infty.$$

Тоді для послідовності $(\xi_n,\ n \ge 1)$ *справджується закон великих чисел.*

Доведення. З умови теореми випливає більш сильне твердження – збіжність центрованих та нормованих сум S_n до нуля в середньому квадратичному. Дійсно, за теоремою про дисперсію суми незалежних величин та за теоремою Штольця про границю відношення послідовностей

$$\lim M((S_n - MS_n)\ /\ n)^2 = \lim DS_n\ /\ n^2 =$$

$$\lim \sum\nolimits_{k=1}^{n} D\xi_k\ /\ n^2 = \lim D\xi_n/(n^2 - (n-1)^2) = \lim D\xi_n/(2n-1) = 0.$$

Завершує доведення теореми застосування імплікації $L_2 \Longrightarrow P$, що базується на нерівності Чебишева для дисперсій

$$P(|S_n - MS_n| \geq n\varepsilon) \leq DS_n \ / \ n^2\varepsilon^2 \to 0 \ \boxdot$$

Зауваження. Остання умова теореми на дисперсії доданків виконується, зокрема, якщо дисперсії $D\xi_n$ обмежені. Наприклад, для однаково розподілених величин дисперсії не залежать від n, отже, закон великих чисел виконується.

Вправа.

(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) Величини $(\xi_n, \ n \geq 1)$ незалежні та $P(\xi_n = \pm n^\alpha) = 1/2$. Для яких α має місце закон великих чисел ?

(3 – теорема Бернштейна про закон великих чисел). Нехай $(\xi_n, \ n \geq 1)$ – така послідовність випадкових величин, що $\sup D\xi_n < \infty$ та для коефіцієнтів кореляції $\lim_{|n|\to\infty} \sup_k \rho(\xi_{n+k}, \xi_k) = 0$. Тоді для неї виконується закон великих чисел.

Розглянемо деякі застосування ЗВЧ.

23.2. 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 у кожному. Тоді*

$$\nu_n \ / \ n \to^P p.$$

Доведення виводимо з подання $\nu_n = \sum_{k=1}^n \chi_k$, де доданки χ_k– індикатори успіхів – незалежні і однаково розподілені випадкові величини,

$$M\chi_k = p, \quad D(\nu_n/n) = p(1-p)/n \to 0 \boxdot$$

23.3. 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 функції $f \in C[0,1]$ називається поліном*

$$B_n(f,x) = \sum\nolimits_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k\, x^k (1-x)^{n-k}, \ x \in [0,1].$$

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 $\nu_n(x)$ – 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

$$B_n(f,x) = Mf(\nu_n(x)\ /\ n).$$

Покладемо

$$w_\delta(f) = \sup{}_{|x-y|\,\le\,\delta}\ |f(x) - f(y)|\,, \quad L(f) = \sup|f(x)|\,.$$

Використаємо ідеї доведень пунктів (а) і (б) теореми про властивості збіжності за ймовірністю:

$$\left| f(\frac{\nu_n(x)}{n}) - f(x)\right| \le w_\delta(f) + 2L(f)\ 1_{\{\left|\frac{\nu_n(x)}{n} - x\right| > \delta\}}\ ,$$

звідки

$$|B_n(f,x) - f(x)| \le w_\delta(f) + 2L(f)\ P\left(\left|\frac{\nu_n(x)}{n} - x\right| > \delta\right) \le$$

$$w_\delta(f) + 2L(f)\ D(\frac{\nu_n(x)}{n})/\delta^2 = w_\delta(f) + 2L(f)\ \frac{x(1-x)}{n\delta^2} \le$$
$$w_\delta(f) + 2L(f)\ /\ 4n\delta^2 \le 2\varepsilon,$$

рівномірно за x, починаючи з деякого n, якщо обрати сталу δ так, щоб $w_\delta(f) \le \varepsilon$ для наперед заданого довільного $\varepsilon > 0$ $\boxdot$

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

$$|B_n(f,x) - f(x)| \le D\delta + E\ /\ n\delta^2 = O(1/\sqrt[3]{n})$$

рівномірно за x.

Вправа.

(1) Узагальнити теорему Бернштейна на функції багатьох змінних з класу $C([0,1]^n)$.

(2) Довести, що для $f \in C(\mathbb{R}_+)$ маж місце збіжність

$$\exp(-nx)\sum_{k=0}^{\infty} f\left(\frac{k}{n}\right)(nx)^k/\ k! \to f(x),\ n \to \infty,$$

рівномірно на кожному обмеженому інтервалі. Вказівка: ліва частина дорівнює $Mf((\xi_1 + ... + \xi_n)/n)$, де ξ_k незалежні та мають розподіл Пуассона $\Pi(x)$.

(3) Перетворенням Лапласа неперервної обмеженої функції $f \in C_b(\mathbb{R}_+)$ називається функція $\varphi(p) \equiv \int_0^\infty \exp(-pt) f(t) dt$ від $p \in \mathbb{R}_+$. Довести формулу обернення Уідлера:

$$f(s) = \lim_{n\to\infty} \frac{(-1)^{n-1}}{(n-1)!} \left(\frac{n}{s}\right)^n \varphi^{(n-1)}(\tfrac{n}{s}),$$

де $\varphi^{(k)}(p)$ – k-та похідна φ. Вказівка: звести величину під знаком границі до вигляду $Mf(s(\xi_1 + ... + \xi_n)/n)$, де ξ_k незалежні та мають показниковий розподіл $Exp(1)$.

23.4. 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:

$$I \equiv \int_0^1 f(x)dx = Mf(\alpha),$$

де α – рівномірно розподілена на $[0, 1]$ випадкова величина.

Теорема (про наближення інтегралу методом Монте-Карло). *Нехай функція $f \in L_2[0, 1]$ інтегровна в квадраті, $(\alpha_k,\ k \geq 1)$ – послідовність незалежних рівномірно розподілених на $[0, 1]$ випадкових величин, а $I_k = f(\alpha_k)$. Тоді має місце збіжність*

$$(I_1 + ... + \ I_n)\ /\ n \to^P I,\ n \to \infty.$$

Доведення. Випадкові величини $(I_k,\ k=\overline{1,n})$ є незалежними в сукупності, однаково розподіленими за теоремою про перетворення незалежних величин, та мають скінченні математичні сподівання $I = Mf(\alpha_k)$ 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 $M(I - \frac{I_1+\ldots+I_n}{n})^2 = s^2/n = O(1/n)$ $\boxdot$

На відміну від схем наближеного обчислення інтегралів із регулярно розміщеними вузлами метод Монте-Карло є збіжним також і для розривних функцій.

24. Збіжність з імовірністю 1 та її властивості

Послідовність випадкових величин $(\xi_n \,, n \ge 1)$ збігається з імовірністю 1 до величини ξ, якщо подія

$$A = \{\exists \lim \xi_n = \xi\} = \cap_{\ k\ge 1}\ \underline{\lim}_{n\to\infty}\{|\xi_n - \xi| \le 1/k\} = \overline{\cup_{\ k\ge 1}\overline{\lim}_{\ n\to\infty}\{|\xi_n - \xi| > 1/k\}}$$

має одиничну ймовірність: $P(A) = 1$. 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*

$$\{\overline{\lim}_{n\to\infty}\xi_n > \varepsilon\} \subset \overline{\lim}_{n\to\infty}\{\xi_n > \varepsilon\},$$
$$\overline{\lim}_{n\to\infty}\{\xi_n \ge \varepsilon\} \subset \{\overline{\lim}_{n\to\infty}\xi_n \ge \varepsilon\}.$$

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.

24.1. Лема Бореля – Кантеллі

Теорема (лема Бореля – Кантеллі). *Нехай $(A_n,\ n \ge 1)$ довільна послідовність випадкових подій.*

(a) Якщо ряд $\sum_{n\ge 1} P(A_n)$ збігається, то $P(\overline{\lim} A_n) = 0$.

(б) Якщо події $(A_n,\ n \geq 1)$ *незалежні в сукупності і ряд* $\sum_{n\geq 1} P(A_n)$ *розбігається, то* $P(\overline{\lim} A_n) = 1$.

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 $P(\overline{\lim} A_n) = \lim P(B_m)$.

(а) За властивістю напівадитивності ймовірності

$P(B_m) \ \leq \ \sum_{n\geq m} P(A_n) \to 0, m \to \infty$,

як залишок збіжного ряду, що і доводить цей пункт.

(б) Обчислимо з властивості неперервності ймовірності, враховуючи означення незалежності та теорему про перетворення незалежних подій,

$$P(B_m) = \lim_{N\to\infty} P(\cup_{n=m}^{N} A_n) = 1 - \lim_{N\to\infty} P(\cap_{n=m}^{N} \overline{A_n}) =$$

$$1 - \lim_{N\to\infty} \prod_{n=m}^{N} P(\overline{A_n}) = 1 - \lim_{N\to\infty} \prod_{n=m}^{N} (1 - P(A_n)) \geq$$

$$1 - \lim_{N\to\infty} \exp(-\sum_{n=m}^{N} P(A_n)) = 1 - \exp(-\sum_{n=m}^{\infty} P(A_n)) = 1,$$

оскільки будь-який залишок розбіжного ряду розбігається, і справедлива елементарна нерівність $1 - x \leq \exp(-x)$ $\boxdot$

Вправа.

(1) Якщо для деякої події A ряд $\sum_{n\geq 1} P(A_n \cap A)$ розбігається, то $P(\overline{\lim} A_n) \leq 1 - P(A)$.

(2) Випадкові величини $(\xi_n \ ,n \ \geq \ 1)$ незалежні однаково розподілені і $M\xi_1^+ \ = \ M\xi_1^- \ \leq \ \infty$. Довести, що $P(\xi_n \ = \ o(n),\ \ n \ \to \ \infty) \ = \ 1$ тоді й тільки тоді, коли $M\,|\xi_1| < \infty$.

(3) Випадкові величини $(\xi_n \ ,n \ \geq \ 1)$ незалежні однаково розподілені і $M\xi_1 \ = \ 0$. (а) Якщо $M\,|\xi_1|^\alpha \ < \ \infty$, то $P(|\xi_n| \ = \ o(n^{1/\alpha}),\ \ n \ \to \ \infty) \ = \ 1$. (б) Якщо $M\exp(\alpha\xi_1) < \infty$, то $P(\xi_n = o(\ln n),\ n \to \infty) = 1$.

24.2. 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* $(\xi_k\ , k = \overline{1, n})$ *послідовність незалежних у сукупності випадкових величин така, що* $M\xi_k = 0$, $D\xi_k < \infty$, *а* $S_k = \xi_1 + ... + \xi_k$. *Тоді для кожного* $\varepsilon > 0$

$$P(\max_{\ 1\le\ k\ \le n}\ |S_k| \ge \varepsilon) \le DS_n\ /\ \varepsilon^2.$$

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 $1_A = \sum_{k=1}^n 1_{A_k}$. Відзначимо, що для всіх елементарних подій справедлива нерівність $|S_k|\, 1_{A_k} \ge \varepsilon 1_{A_k}$. Тому за монотонністю математичного сподівання

$$DS_n = MS_n^2 \ge MS_n^2 1_A = \sum_{k=1}^n MS_n^2 1_{A_k} =$$

$$\sum_{k=1}^n M(S_k + S_n - S_k)^2\ 1_{A_k} = \sum_{k=1}^n MS_k^2 1_{A_k} +$$

$$\sum_{k=1}^n M(S_n - S_k)^2\ 1_{A_k} + 2\sum_{k=1}^n MS_k 1_{A_k}\ (S_n - S_k) \ge$$

$$\sum_{k=1}^n M\varepsilon^2 1_{A_k} + 0 + 2\sum_{k=1}^n MS_k 1_{A_k}\ M(S_n - S_k) =$$

$$\varepsilon^2 \sum_{k=1}^n P(A_k) = \varepsilon^2 P(A),$$

де використані невід'ємність другого доданку, та незалежність множників $S_k 1_{A_k}$ 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 з множників є центрованим за умовою: $M(S_n - S_k) = MS_n - MS_k = 0$, отже, математичне сподівання добутку, що дорівнює добуткові за теоремою про математичне сподівання добутку незалежних величин, є нульовим $\boxdot$

Вправа. Нехай g – неперервна неспадна опукла донизу функція, а послідовність (ξ_k) задовольняє умови теореми про нерівність Колмогорова. Тоді ліва частина цієї нерівності не перевищує $Mg(|S_n|)/g(\varepsilon)$

25. Посилений закон великих чисел

Означення. *Для послідовності випадкових величин* $(\xi_n,\ n \geq 1)$ *виконується* посилений закон великих чисел, *якщо їх частинні суми* $S_n = \xi_1 + ... + \xi_n$ *задовольняють граничне співвідношення*

$$(S_n - MS_n)\ /\ n \to^{P1} 0,\ \text{при}\ n \to \infty.$$

Як і для закону великих чисел, у випадку однаково розподілених випадкових величин це співвідношення еквівалентне збіжності середніх арифметичних

$$S_n\ /\ n \to^{P1} M\xi_1,\ \text{при}\ n \to \infty.$$

25.1. Загальна 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* $D\xi_n < \infty$ *та*

$$\sum_{n \geq 1} D\xi_n\ /\ n^2 < \infty.$$

Тоді для послідовності $(\xi_n\ , n \geq 1)$ *має місце посилений закон великих чисел.*

Доведення. Не обмежуючи загальності, можна вважати, що $M\xi_n = 0$. Дійсно, в загальному випадку досить розглянути послідовність незалежних величин $\xi_n' = \xi_n - M\xi_n$ із тими ж дисперсіями. Очевидно, що твердження посиленого закону великих чисел для ξ_n і ξ_n' збігаються.

Позначимо $\zeta_n = \max_{1\le k \le 2^n} |S_k|$. Для кожного k знайдеться єдиний номер n такий, що $2^{n-1} \le k < 2^n$, причому $k \to \infty$ тоді й тільки тоді, коли $n \to \infty$. При такому виборі для всіх k має місце нерівність

$$|S_k / k| \le \zeta_n / 2^{n-1}.$$

Тому для збіжності $S_k / k \to^{P1} 0$ достатньо, щоб

$$\zeta_n / 2^n \to^{P1} 0, \; n \to \infty.$$

Для заданого $\varepsilon > 0$ розглянемо випадкові події

$$A_n(\varepsilon) = \{\zeta_n / 2^n \ge \varepsilon\}.$$

За нерівністю Колмогорова

$$\sum_{n\ge1} P(A_n(\varepsilon)) = \sum_{n\ge1} P(\max_{1\le k \le 2^n} |S_k| \ge \varepsilon 2^n) \le \sum_{n\ge1} DS_{2^n} / 2^{2n}\varepsilon^2 =$$

$$\sum_{n\ge1} \sum_{k \le 2^n} D\xi_k / 2^{2n}\varepsilon^2 = \varepsilon^{-2} \sum_{k\ge1} D\xi_k \sum_{n:\, 2^n \ge k} 2^{-2n} \le$$

$$2\varepsilon^{-2} \sum_{k\ge1} D\xi_k / k^2 < \infty.$$

Отже, за лемою Бореля – Кантеллі (а), $P(\overline{\lim} A_n(\varepsilon)) = 0$ для кожного $\varepsilon > 0$. Згідно з зауваженням про верхню границю величин та подій $\{\overline{\lim}\,(\zeta_n / 2^n) \ge \varepsilon\} \subset \overline{\lim} A_n(\varepsilon)$, тому $P(\{\overline{\lim}\,(\zeta_n / 2^n) \ge \varepsilon\}) = 0 \;\forall \varepsilon > 0$, звідки $\overline{\lim}_{n\to\infty} (\zeta_n / 2^n) < \varepsilon$ майже напевне для кожного $\varepsilon > 0$.

Тому остання верхня границя нульова, $\zeta_n / 2^n \to^{P1} 0$ і виконується посилений закон великих чисел, як показано вище $\boxdot$

25.2. Теорема Бореля

Теорема (про асимптотику м.н. відносної частоти успіху). *Нехай ν_n – кількість успіхів у перших n випробуваннях нескінченної послідовності випробувань Бернуллі з імовірністю успіху p у кожному. Тоді*

$$\nu_n \,/\, n \to^{P1} p.$$

Доведення випливає з зображення $\nu_n = \sum_{k=1}^{n} \chi_k$, де доданки – індикатори успіхів – незалежні в сукупності і однаково розподілені, $M\chi_k = p$ $\boxdot$

25.3. Критерій Колмогорова для однаково розподілених величин

Теорема (критерій Колмогорова посиленого закону великих чисел). *Нехай $(\xi_n,\ n \geq 1)$ послідовність незалежних в сукупності однаково розподілених випадкових величин, $S_n = \xi_1 + ... + \ \xi_n$.*

Для того, щоб існувала стала a така, що $S_n \,/\, n \to^{P1} a$, необхідно і достатньо, щоб $M\,|\xi_1| < \infty$. У цьому випадку $a = M\xi_1$.

Доведення.

Достатність. (1) Позначимо $\xi_n' = \xi_n 1_{\{|\xi_n| \,\leq\, n\}}$, $\xi_n'' = \xi_n - \xi_n'$. Нехай S_n', S_n'' означають суми відповідних величин. Доведемо, що три доданки в правій частині тотожності

$$(S_n - MS_n)/n = (S_n' - MS_n')/n + S_n'' \,/\, n - MS_n'' \,/\, n \equiv a_n + b_n - c_n$$

збігаються майже напевне до нуля.

(1a). Випадкові величини $(\xi_n'\ , n \geq 1)$ 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 $(\xi_n,\ n \ge 1)$

$$\sum_{n\ge1} D\xi_n'/n^2 \le \sum_{n\ge1} M(\xi_n')^2/n^2 = \sum_{n\ge1} M\xi_n^2\ 1_{\{|\xi_n|\,\le\, n\}}\ /n^2 =$$

$$\sum_{n\ge1} \tfrac{1}{n^2} M\xi_1^2\ 1_{\{|\xi_1|\,\le\, n\}} = \sum_{n\ge1} \tfrac{1}{n^2} \sum_{k=1}^{n} M\xi_1^2\ 1_{\{k-1<\,|\xi_1|\,\le k\}} =$$

$$\sum_{k\ge1} M\xi_1^2\ 1_{\{k-1<\,|\xi_1|\,\le k\}} \sum_{n\ge k} \tfrac{1}{n^2} \le \sum_{k\ge1} M\xi_1^2\ 1_{\{k-1<\,|\xi_1|\,\le k\}}\ \tfrac{2}{k} \le$$

$$2\sum_{k\ge1} M\,|\xi_1|\ 1_{\{k-1<\,|\xi_1|\,\le k\}} \le 2\ M\,|\xi_1| < \infty,$$

де використані нерівності $\xi_1^2\ /\ k \le\ |\xi_1|$ на випадкових подіях $\{k-1<|\xi_1|\le k\}$, що є попарно несумісними. Отже, за теоремою Колмогорова $(S_n' - MS_n')/n \equiv a_n \to 0$ майже напевне.

(1b). Далі, знайдемо

$$\sum_{n\ge1} P(\xi_n'' \ne 0) = \sum_{n\ge1} P(|\xi_n| > n) = \sum_{n\ge1} P(|\xi_1| > n) =$$
$$\sum_{n\ge1} \sum_{k\ge n} M1_{k<|\xi_1|\le k+1} =$$
$$\sum_{k\ge1} M\ k\ 1_{\{k<|\xi_1|\le k+1\}} \le \sum_{k\ge1} M\,|\xi_1|\ 1_{\{k<|\xi_1|\le k+1\}}\ \le M\,|\xi_1| < \infty.$$

За лемою Бореля – Кантеллі (a), $P(\overline{\lim}\{\xi_n'' \ne 0\}) = 0$, тобто $P(\underline{\lim}\{\xi_n'' = 0\}) = 1$, звідси $S_n''\ /\ n \equiv b_n \to 0$ майже напевне.

(1c). Нарешті, за теоремою Штольця та за умовою однакової розподіленості

$$\lim MS_n''\ /\ n = \lim M\xi_n''/1 = \lim M\xi_n 1_{|\xi_n|>n} =$$
$$\lim M\xi_1 1_{\{|\xi_1|>n\}} = M \lim \xi_1 1_{\{|\xi_1|>n\}} = M0 = 0$$

внаслідок теореми Лебега про мажоровану збіжність та за припущенням $M\,|\xi_1| < \infty$. Отже, $c_n \to 0$. Достатність доведена.

Необхідність. (2) Зауважимо, що

$$\lim \tfrac{\xi_n}{n} = \lim(\tfrac{S_n}{n}) - \lim(\tfrac{n-1}{n}\tfrac{S_{n-1}}{n-1}) = a - a = 0$$ майже напевне.

Розглянемо незалежні події $A_n = \{|\xi_n| > n\}$. Оскільки $\overline{\lim} A_n \subset \{\overline{\lim}(\xi_n/n) \neq 0\}$, то $P(\overline{\lim} A_n) = 0$ і за лемою Бореля – Кантеллі,(б), наступний ряд має збігатися

$$\infty > \sum_{n\geq 1} P(A_n) = \sum_{n\geq 1} P(|\xi_n| > n) = \sum_{n\geq 1} P(|\xi_1| > n) =$$
$$\sum_{n\geq 1} \sum_{k\geq n} P(k < |\xi_1| \leq k+1) =$$
$$\sum_{k\geq 1} M\, k\, 1_{\{k<|\xi_1|\leq k+1\}} \geq \sum_{k\geq 1} M(|\xi_1| - 1)\, 1_{\{k<|\xi_1|\leq k+1\}} \geq M\,|\xi_1| - 2,$$

де використані однакова розподіленість випадкових величин ξ_n, та очевидна нерівність $k\ 1_{\{k<|\xi_1|\leq k+1\}} \geq (|\xi_1| - 1)\ 1_{\{k<|\xi_1|\leq k+1\}}$, що виконується для всіх елементарних подій. Отже, $M\,|\xi_1| < \infty$ і внаслідок (а) отримуємо $a = M\xi_1$ $\boxdot$

Вправа. (Хаусдорф) Нехай $(\xi_n,\ n \geq 1)$ послідовність незалежних однаково розподілених випадкових величин, $M\xi_1 = \mu,\ M\xi_1^2 < \infty,\ S_n = \xi_1 + \ldots + \xi_n$. Тоді $S_n - n\mu = o(n^{1/2+\delta}), n \to \infty$, майже напевне, для кожного $\delta > 0$.

26. 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.

Означення. *Нехай G, $(G_n,\ n \ge 1) \subset \mathfrak{M}_{01}$* узагальнені функції розподілу*. Будемо говорити, що має місце збіжність* в основному *$G_n \to^O G$, 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 $G \in M_{01}$ 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 послідовності $x'_m \uparrow x$, $x''_m \downarrow x$, $x'_m, x''_m \in D$. Тоді за монотонністю

$$G(x'_m) = \lim_{n\to\infty} G_n(x'_m) \le \underline{\lim}_{n\to\infty} G_n(x) \le$$
$$\overline{\lim}_{n\to\infty} G_n(x) \le \lim_{n\to\infty} G_n(x''_m) = G(x''_m).$$

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 $y < x < z,\quad y, z \in D$:

$$H(y) \le G(x) \le \sup(H(z),\ y < x,\ y \in D) = H(z).$$

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

$$G(x'_m - \tfrac{1}{m}) \le H(x'_m) = \lim_{n\to\infty} G_n(x'_m) \le \underline{\lim}_{n\to\infty} G_n(x) \le$$
$$\overline{\lim}_{n\to\infty} G_n(x) \le \lim_{n\to\infty} G_n(x''_m) = H(x''_m) \le G(x''_m + \tfrac{1}{m}).$$

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 $\lim G_n(x) = G(x)$ $\boxdot$

Вправа.

(1) Навести приклад послідовності функцій розподілу, які збігаються в основному і не збігаються поточково.

(2) Клас $\mathfrak{M}_{01}$ замкнений відносно збіжності в основному.

(3) Довести, що зі збіжності функцій розподілу в основному $F_n \to^O F$ до неперервної функції розподілу F випливає рівномірна за x збіжність.

(4) Нехай F_n, F – функції розподілу і $F_n \to^O F$ Довести, що існує ймовірнісний простір $(\Omega,\ \mathcal{F},\ P)$ та випадкові величини ξ_n, ξ на ньому з функціями розподілу F_n, F відповідно такі, що $\xi_n \to^{P1} \xi,\ \ n \to \infty$.

27. Слабка збіжність функцій розподілу і випадкових величин

Означення. *Нехай* ξ, $(\xi_n,\ n \geq 1)$ *– послідовність випадкових величин із функціями розподілу* F, $(F_n,\ n \geq 1)$ *відповідно. Будемо говорити, що має місце* слабка збіжність

$$\xi_n \to^W \xi, \quad F_n \to^W F,$$

якщо для кожної неперервної обмеженої функції g *(позначення* $\forall g \in C_b(\mathbb{R})$*) збігаються математичні сподівання*

$$Mg(\xi_n) \to Mg(\xi),\ n \to \infty,$$

або ж 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.

Позначимо $Cl(B)$ –замикання, $Int(B)$ –внутрішність множини $B \subset \mathbb{R}$.

27.1. Однозначність слабкої границі

Теорема (про однозначність слабкої границі).
(a) 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.

(a) 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

$$g_m(x, A) = \exp(-m\, \rho(x, A)).$$

неперервна і обмежена за x. За означенням замикання $\rho(x, A) = 0$ тоді й тільки тоді, коли $x \in Cl(A)$. Звідси випливає, що

$$g_m(x, A) \downarrow 1_{Cl(A)}(x),\ m \to \infty.$$

Підставляючи $g = g_m$ у рівність умови (а), з теореми Лебега про мажоровану збіжність при $m \to \infty$ дістанемо

$$F(Cl(A)) = \int_{-\infty}^{\infty} 1_{Cl(A)}(x) dF(x) = \lim\ _{m\to\infty} \int_{-\infty}^{\infty} g_m(x) dF(x) =$$
$$\lim\ _{m\to\infty} \int_{-\infty}^{\infty} g_m(x) dG(x) = \int_{-\infty}^{\infty} 1_{Cl(A)}(x) dG(x) = G(Cl(A)).$$

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 $F = G$ $\boxdot$

Вправа. 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.

27.2. 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* $\xi_n \to^W \xi$.

Доведення теореми є очевидним наслідком пунктів (а),(б) теореми про властивості збіжності за ймовірністю

$$g(\xi_n) \to^P g(\xi), \quad Mg(\xi_n) \to Mg(\xi),$$

де перше співвідношення є наслідком неперервності g та (а), а друге – обмеженості g (а отже і $g(\xi_n)$) та (б) $\boxdot$

Вправа.

(1) Нехай $\xi_n \to^W \xi$, а функція $g \in C(\mathbb{R})$ неперервна. Довести, що $g(\xi_n) \to^W g(\xi)$.

(2) Нехай $g \in C[0,1]$. Обчислити границю $\lim_{n\to\infty} \int_0^1 .. \int_0^1 g\left(\sqrt[n]{x_1..x_n}\right) dx_1..dx_n$.

Теорема (про властивості слабкої збіжності). *Нехай* $F_n \to^W F$.

(а) Для довільної множини $B \in \mathfrak{B}(\mathbb{R})$

$$F(Int(B)) \le \underline{\lim}_{n\to\infty} F_n(B) \le \overline{\lim}_{n\to\infty} F_n(B) \le F(Cl(B)).$$

(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. Використаємо функцію g_m із теореми про однозначність слабкої границі та теорему Лебега про монотонну збіжність:

$$\overline{\lim}\ _{n\to\infty} F_n(B) = \overline{\lim}\ _{n\to\infty} \int_{-\infty}^{\infty} 1_{x\ \in\ B}\ dF_n(x) \le$$

$$\overline{\lim}_{n\to\infty}\int_{-\infty}^{\infty} g_m(x,B)dF_n(x) = \int_{-\infty}^{\infty} g_m(x,B)dF(x) \downarrow F(Cl(B)),\ m\to\infty,$$

де використані співвідношення $g_m(x,B)\downarrow 1_{x\,\in\,Cl(B)} \geq 1_{x\,\in\,B}$.

Для доведення лівої нерівності перейдемо до доповнень:

$$\lim F_n(B) = 1 - \overline{\lim}\ F_n(\overline{B}) \geq 1 - F(Cl(\overline{B})) = F(\overline{Cl(\overline{B})}) = F(Int(B)).$$

(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 $F(B)$ $\boxdot$

27.3. Еквівалентність збіжностей слабкої та в основному

Теорема (про еквівалентність слабкої збіжності та в основному). *Нехай $(F_n,\ n\geq 1)$ послідовність функцій розподілу. Для того, щоб $F_n \to^W F$, необхідно і достатньо, щоб $F_n \to^O F$ і F була функцією розподілу.*

Зауваження. З теореми про збіжність в основному та на щільній множині випливає, що границя в основному функцій розподілу завжди належить класу $\mathfrak{M}_{01}$.(**Вправа**) Однак ця границя не обов'язково є функцією розподілу. Наприклад: $F_n(x) = 1_{n\,<\,x} \to 0, n\to\infty, \forall x$, де $0\in\mathfrak{M}_{01}$.

Доведення.

(а) *Необхідність*. Нехай $F_n \to^W F$. Для довільної точки x неперервності функції $F(x)$ та для множини $B = (-\infty, x)$ маємо $\partial B = \{x\}$ та

$F(\partial B) = F(x+0) - F(x) = 0$. Тому за теоремою про властивості слабкої збіжності $F(x) = F(B) = \lim F_n(B) = \lim F_n(x)$, що і доводить збіжність в основному.

(б) *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

$$g_\delta(x) = \sum\nolimits_{k=0}^{m-1} g(a_k)\, 1_{a_k \le x < a_{k+1}}, \quad L(g) = \sup|g(x)|\,.$$

З неперервності функції g випливає, що вона рівномірно неперервна на відрізку $[-c, c]$, тому модуль неперервності

$$w_{g,c}(\delta) \equiv \sup\nolimits_{|x| \le c} |g(x) - g_\delta(x)| \to 0$$

при $\delta \to 0$. Оцінимо

$$\overline{\lim}_{n\to\infty} \left| \int_{-\infty}^{\infty} g(x) dF_n(x) - \int_{-\infty}^{\infty} g(x) dF(x) \right| \le$$

$$\overline{\lim}_{n\to\infty} \left| \int_{|x|>c} g(x) dF_n(x) \right| + \overline{\lim}_{n\to\infty} \left| \int_{|x|>c} g(x) dF(x) \right| +$$

$$\overline{\lim}_{n\to\infty} \int_{|x|\le c} |g(x) - g_\delta(x)|\, dF_n(x) + \overline{\lim}_{n\to\infty} \int_{|x|\le c} |g(x) - g_\delta(x)|\, dF(x) +$$

$$\overline{\lim}_{n\to\infty} \left| \int_{|x|\le c} g_\delta(x) dF_n(x) - \int_{|x|\le c} g_\delta(x) dF(x) \right| \le$$

$$L(g)\overline{\lim}_{n\to\infty}(F_n(-c)+1-F_n(c)+F(-c)+1-F(c))+2w_{g,c}(\delta)+$$

$$\overline{\lim}_{n\to\infty}\left|\sum\nolimits_{k=0}^{m-1} g(a_k)(F_n(a_{k+1})-F_n(a_k)-F(a_{k+1})+F(a_k))\right| \leq$$

$$2L(g)(F(-c)+1-F(c))+2w_{g,c}(\delta)+$$

$$\sum\nolimits_{k=0}^{m-1}\ |g(a_k)|\ \ \overline{\lim}_{n\to\infty}\left(|F_n(a_{k+1})-F(a_{k+1})|+|F_n(a_k)-F(a_k)|\right) \leq$$

$$2L(g)\varepsilon+2w_{g,c}(\delta)+0,$$

оскільки за умови збіжності в основному $F_n(a_k)\to F(a_k)$ у кожній точці $a_k\in D$. Праву частину останньої нерівності можна зробити вибором ε і δ як завгодно малою.

Отже, $\int gdF_n \to \int gdF$ при $g\in C_b(\mathbb{R})$ та $F_n \to^W F$ $\boxdot$

28. Слабка компактність

Означення. *Нехай* $\mathfrak{M} \subset \mathfrak{M}_{01}$ *деякий клас узагальнених функцій розподілу. Клас* $\mathfrak{M}$ *називається* компактним в основному, *якщо будь-яка послідовність* $(G_n,\ n \geq 1) \subset \mathfrak{M}$ *містить збіжну* *в основному* *підпослідовність:*

$$\exists(G_{n_k},\ k \geq 1) \subset (G_n,\ n \geq 1), \quad G_{n_k} \to^O G,\ n_k \to \infty,$$

для деякої *узагальненої функції розподілу* $G \in \mathfrak{M}_{01}$.

28.1. 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 $K_0 = \mathbb{N}$ множину всіх натуральних чисел, $D = \{x_1, ..., x_n, ...\}$ зліченну щільну множину на числовій осі.

Розглянемо числову послідовність $\{G_n(x_1),\ n \in K_0\} \subset [0, 1]$. Так як $[0, 1]$ – 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 діагональним 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 $i \geq j$ $\min K_i \in K_i \subset K_j$.

Так як 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 $G_n \to^O G$ $\boxdot$

28.2. Теорема Прохорова про слабку компактність

Означення. *Клас $\mathfrak{F}$ функцій розподілу називається* слабко компактним, *якщо будь-яка послідовність $(F_n,\ n \geq 1) \subset \mathfrak{F}$ містить* *слабко збіжну* *підпослідовність:*

$$\exists (F_{n_k},\ k \geq 1) \subset (F_n,\ n \geq 1), \quad F_{n_k} \to^W F,$$

де F – деяка функція розподілу.

Зауважимо, що замкненість множини $\mathfrak{F}$ не передбачається, тобто включення $F \in \mathfrak{F}$ 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* $\mathfrak{F}$ *функцій розподілу є слабко компактним тоді й тільки тоді, коли*

$$\lim_{c\to\infty} \sup_{F \in \mathfrak{F}} F(\overline{[-c,c)}) = 0, \text{ або } \lim_{c\to\infty} \inf_{F \in \mathfrak{F}} F([-c,c)) = 1.$$

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 $F(\overline{[-c,c)}) = 1 - F([-c,c))$.

(a) *Достатність*.

Нехай $\mathfrak{M}_{01}$ – визначений вище клас узагальнених функцій розподілу. Оскільки $\mathfrak{F} \subset \mathfrak{M}_{01}$, то за теоремою Хеллі про компактність в основному для кожної послідовності $(F_n,\ n \ge 1) \subset \mathfrak{F}$ знайдемо узагальнену функцію розподілу $G \in \mathfrak{M}_{01}$ та підпослідовність $(F_{n_k},\ k \ge 1) \subset (F_n,\ n \ge 1)$, 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

$$G(-c) + 1 - G(c) = \lim_{k\to\infty}(F_{n_k}(-c) + 1 - F_{n_k}(c)) =$$

$$\lim_{k\to\infty} F_{n_k}(\overline{[-c,c)}) \le \sup_{F \in \mathfrak{F}} F(\overline{[-c,c)}) \to 0,\ c \to \infty.$$

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 $F_{n_k} \to^W G$. Отже, клас $\mathfrak{F}$ є слабко компактним.

(б) *Необхідність.*

Припустимо, що умова Прохорова не виконана, тобто відповідна границя $\varepsilon > 0$. Оскільки функція $\varepsilon(c) \equiv \sup_{F \in \mathfrak{F}} F(\overline{[-c, c)})$ 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 $F_n \in \mathfrak{F}$, така, що $F_n(\overline{[-n, n)}) \geq \varepsilon/2$. За означенням слабкої компактності $\mathfrak{F}$ 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 деякого $\delta > 0$ 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 $\varepsilon_k(m) \geq \delta/2$, що суперечить умові (а). Тоді для кожного фіксованого c, 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) $\boxdot$

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 більша за 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 застосування теореми Прохорова доводить слабку компактність послідовності $\boxdot$

29. 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 $t \in \mathbb{R}$:*

$$\varphi_\xi(t) \equiv M \exp(it\xi) = \int_{-\infty}^{\infty} \exp(itx)\, dF(x) \equiv \varphi_F(t).$$

Рівність посередині випливає з теореми про обчислення математичного сподівання функції від випадкової величини.

В останньому означенні та надалі математичне сподівання (інтеграл Лебега) від комплекснозначних величин визначається за лінійністю:

$$M(\xi_1 + i\xi_2) = M\xi_1 + iM\xi_2.$$

Вправа. Довести, що для комплекснозначних величин
$|M\xi| \le M\,|\xi|\,,\ M\overline{\xi} = \overline{M\xi}$.

29.1. Однозначність відповідності

Теорема (про однозначну відповідність між функціями розподілу та характеристичними функціями). *Відповідність між функціями розподілу та характеристичними функціями є взаємно однозначною.*

Зауваження. Існують формули обертання для характеристичних функцій, з яких можна явно обчислити функцію розподілу за її характеристичною функцією. Наприклад, для всіх $a < b$, що є точками неперервності функції розподілу F, справедлива тотожність

$$F(b) - F(a) = \frac{1}{2\pi} \lim_{\varepsilon \downarrow 0} \int_a^b dx \int_{-\infty}^{\infty} \exp(-itx - \varepsilon^2 t^2)\, \varphi_F(t)\, dt,$$

де абсолютна збіжність кратного інтегралу при $\varepsilon > 0$ обумовлена обмеженістю характеристичної функції. Зауважимо, що така збіжність при $\varepsilon = 0$ може порушуватись.

Доведення теореми.

Припустимо, що функції розподілу F, G мають однакові характеристичні функції: $\varphi_F(t) = \varphi_G(t)$ при всіх $t \in \mathbb{R}$. Позначимо клас функцій

$$\mathfrak{G} = \left\{ g \in C_b(\mathbb{R}) : \int_{-\infty}^{\infty} g(x) dF(x) = \int_{-\infty}^{\infty} g(x) dG(x) \right\}.$$

За припущенням, кожна гармоніка $\exp(itx) \in \mathfrak{G}$ при $t \in \mathbb{R}$. Клас $\mathfrak{G}$ є лінійним, тому він містить всі тригонометричні поліноми $g(x) = \sum c_k \exp(it_k x)$. 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 $g \in C_b(\mathbb{R})$ та $g(x) = 0$ при $|x| > c - 1$. Оберемо тригонометричний поліном g_ε такий, що $|g(x) - g_\varepsilon(x)| \le \varepsilon$ при $|x| \le c$, що має період $2c$. Після періодичного продовження його на $\mathbb{R}$ отримуємо функцію $g_\varepsilon \in C_b(\mathbb{R})$. За рахунок вибору c різницю

$$\left| \int gdF - \int gdG \right| = \left| \int_{|x| \le c} gdF - \int_{|x| \le c} gdG \right| \le$$

$$\left|\int_{|x|\le c}(g-g_\varepsilon)dF\right|+\left|\int_{|x|\le c}(g-g_\varepsilon)dG\right|+$$

$$\left|\int g_\varepsilon dF-\int g_\varepsilon dG\right|+\left|\int_{|x|>c} g_\varepsilon dF-\int_{|x|>c} g_\varepsilon dG\right|\le$$

$$2\varepsilon+0+(F(\overline{[-c,c]})+G(\overline{[-c,c]})(\sup g+\varepsilon)$$

можна зробити як завгодно малою. Тому $g\in\mathfrak{G}$.

Оскільки кожна неперервна обмежена функція є обмеженою поточковою границею фінітних неперервних обмежених функцій, то $\mathfrak{G}=C_b(\mathbb{R})$. Звідси та з теореми про однозначність слабкої границі, пункт (а), виводимо, що $F=G$ $\boxdot$

Вправа.

(1) Якщо ξ – цілозначна випадкова величина, то її характеристична функція $\varphi(t)$ має період 2π і $P(\xi=n)=(2\pi)^{-1}\int_{-\pi}^{\pi}\exp(-itn)\varphi(t)dt$.

(2) Нехай випадкова величина ξ інтегровна в будь-якій степені і $M\xi^n=(n+k)!/k!$ для деякого k та всіх $n\ge 0$.Довести, що розподіл ξ визначається однозначно та збігається з розподілом Ерланга.

29.2. Властивості характеристичної функції

Теорема (про основні властивості характеристичної функції). *Нехай φ – характеристична функція. Тоді виконуються такі властивості:*

(а –нормованість) $\varphi(0)=1$ і $|\varphi(t)|\le 1$, $\forall t\in\mathbb{R}$,

(б –антисиметрія) $\varphi(-t)=\overline{\varphi(t)}$,

(в –неперервність) $\varphi(t)$ неперервна в нулі та 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, $|\varphi(t)| = \left|\int_{-\infty}^{\infty} \exp(itx)dF(x)\right| \leq \int_{-\infty}^{\infty} |\exp(itx)|\, dF(x) = 1.$

(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) $|\varphi(t) - \varphi(s)| \leq \int_{-\infty}^{\infty} |\exp(itx) - \exp(isx)|\, dF(x) =$
$\int_{-\infty}^{\infty} |\exp(i(t-s)x) - 1|\, dF(x) \to 0,\ t - s \to 0,$
за теоремою Лебега про мажоровану збіжність, оскільки $|\exp(ihx) - 1| \leq 2$ та $|\exp(ihx) - 1| \to 0$ при $h \to 0$ для кожного x.

(г) Сума з умови дорівнює

$$\int_{-\infty}^{\infty} \sum\nolimits_{k,\ j\ =1}^{n} c_k\overline{c_j} \exp(i(t_k - t_j)x)dF(x) =$$
$$\int_{-\infty}^{\infty} \left|\sum\nolimits_{k=1}^{n} c_k \exp(it_kx)\right|^2 dF(x) \geq 0 \boxdot$$

Вправа. Якщо випадкова величина ξ має щільність та інтегровна, то $\left|\varphi_\xi(t)\right| < 1$ для всіх $t \neq 0$.

Теорема (про властивості характеристичної функції).

(а) $\varphi_{a+b\xi}(t) = \exp(ita)\ \varphi_\xi(bt).$

(б) Якщо ξ, η незалежні, то $\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\ \varphi_\eta(t)$.

(в) Якщо ξ інтегровна, то $\varphi_\xi(t) = 1 + it\ M\xi + o(t)$, $t \to 0$.

(г) Якщо ξ квадратично інтегровна, то

$$\varphi_\xi(t) = 1 + it\ M\xi - \tfrac{1}{2}t^2\ M\xi^2 + o(t^2),\ t \to 0.$$

(д) За умови інтегровності або квадратичної інтегровності відповідно

$$M\xi = -i\varphi'_\xi(0), \quad M\xi^2 = -\varphi''_\xi(0).$$

(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

$$\varphi_\xi(t) = \exp(\lambda(\exp(it) - 1)).$$

Доведення.

(а) За мультиплікативністю експоненти та однорідністю математичного сподівання

$$\varphi_{a+b\xi}(t) \ = M\exp(it(a+b\xi)) \ =$$
$$M\exp(ita)\exp(itb\xi) \ = \exp(ita)M\exp(itb\xi) \ = \exp(itb)\varphi_\xi(at).$$

(б) За теоремою про перетворення незалежних величин випадкові величини $\exp(it\xi)$ і $\exp(it\eta)$ також незалежні. Тому

$$\varphi_{\xi+\eta}(t) = M\exp(it(\xi+\eta)) = M\exp(it\xi)\exp(it\eta) =$$
$$M\exp(it\xi)\ M\exp(it\eta) = \varphi_\xi(t)\varphi_\eta(t).$$

внаслідок теореми про математичне сподівання добутку незалежних величин.

(в) З елементарної нерівності $|\exp(itx) - 1 - itx| \le 2|tx|$, $\forall tx \in \mathbb{R}$, випливає, що до інтегралу

$$\frac{\varphi_\xi(t) - 1 - it\ M\xi}{t} = \int_{-\infty}^{\infty} \frac{\exp(itx) - 1 - itx}{t}\ dF(x)$$

можна застосувати теорему Лебега про мажоровану збіжність при $t \to 0$, тому що $\int_{-\infty}^{\infty} 2|x|\, dF(x) = 2M|\xi| < \infty$. Оскільки підінтегральна функція в правій частині поточково прямує до нуля, то права частина прямує до нуля при $t \to 0$, звідки дістанемо шукане співвідношення.

(г) Аналогічно застосовуємо нерівність
$|\exp(itx) - 1 - itx + t^2x^2/2| \le |tx|^2$ до інтегралу

$$\int_{-\infty}^{\infty} \frac{\exp(itx) - 1 - itx + t^2x^2/2}{t^2} dF(x) =$$

$$(\varphi_\xi(t) - 1 - it\ M\xi + t^2 M\xi^2/2)\ /\ t^2.$$

(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

$$\varphi_\zeta(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{x^2}{2} + itx)dx =$$

$$\exp(-\frac{t^2}{2})\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-(x-it)^2/2)dx = \exp(-\frac{t^2}{2}).$$

У загальному випадку досить застосувати (а) до означення $\xi = \mu + \sigma\zeta$.

(ж) За формулою про обчислення математичного сподівання функції від дискретної величини

$$\varphi_\xi(t) = \sum\nolimits_{n\geq 0}\exp(itn)\lambda^n\exp(-\lambda)/n! \ = \exp(\lambda(\exp(it)-1))\boxdot$$

Вправа

(1) Випадкова величина ξ з розподілом $P(\xi = \pm 2n) = C/(n^2\ln n)$ 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) Довести, що (а) характеристична функція випадкової величини зі щільністю Коші $(\pi(1+x^2))^{-1}$ дорівнює $\exp(-|t|)$, (б) для незалежних однаково розподілених випадкових величин $\xi_1, .\xi_n$ зі щільністю Коші нормована сума $(\xi_1 + .. + \xi_n)/n$ має таку ж саму щільність, (в) для вказаної послідовності не справджується закон великих чисел.

(4) Характеристична функція рівномірного розподілу на $[-a, a]$ дорівнює $\frac{\sin at}{at}$.

(5) Якщо величина ξ має гама-розподіл $\Gamma(\lambda, \alpha)$, то $\varphi_\xi(t) = (1 - it/\lambda)^{-\alpha}$.

(6) За означенням випадкова величина ξ симетрична, якщо ξ та $-\xi$ мають однакові функції розподілу. Довести, що симетричність еквівалентна дійснозначності характеристичної функції.

(7, 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.

29.3. Теорема Леві

Теорема (теорема Леві про критерій слабкої збіжності). *Нехай $(F_n,\ n \ge 1)$ послідовність функцій розподілу з характеристичними функціями φ_n.*

Для того, щоб мала місце слабка збіжність $F_n \to^W F,\ n \to \infty$, до деякої функції розподілу F, необхідно і достатньо, щоб характеристичні функції $\varphi_n(t)$ збігались при кожному $t \in \mathbb{R}$ 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 функціями при кожному $t \in \mathbb{R}$.

(б) Достатність. Нехай $\varphi_n(t) \to \varphi(t),\ n \to \infty, \forall t$, 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

$$\frac{1}{2a}\int_{-a}^{a}(1-\varphi_n(t))dt = \frac{1}{2a}\int_{-a}^{a}\left(\int_{-\infty}^{\infty}(1-\exp(itx))dF_n(x)\right)\ dt =$$

$$\frac{1}{2a}\int_{-\infty}^{\infty}\left(\int_{-a}^{a}(1-\exp(itx))dt\right)dF_n(x) =$$

$$\int_{-\infty}^{\infty}\left(1-\frac{\sin ax}{ax}\right)\ dF_n(x) \ge \int_{|x|\,\ge\, b}\left(1-\frac{\sin ax}{ax}\right)\ dF_n(x) \ge$$

$$\int_{|x|\,\geq\, b} \frac{1}{2}\ dF_n(x) \geq \frac{1}{2}\ F_n(\overline{[-b,b)}).$$

Справедливість цих нерівностей випливає, по-перше, з невід'ємності

$$1 - \frac{\sin ax}{ax} \geq 0,$$

оскільки $|\sin y| \leq |y|$, а по-друге, з оцінки

$$\left|\frac{\sin ax}{ax}\right| \leq \frac{1}{|ax|} \leq \frac{1}{2},\ \forall\, |x| \geq b = 2/a.$$

З отриманої нерівності та з теореми Лебега про мажоровану збіжність знаходимо

$$\overline{\lim}_{n\to\infty} F_n(\overline{[-b,b)}) \leq \frac{2}{2a}\overline{\lim}_{n\to\infty}\int_{-a}^{a}(1-\varphi_n(t))dt =$$

$$\frac{1}{a}\int_{-a}^{a}(1-\varphi(t))dt \to 0,\ a\to 0,\ b\to\infty,$$

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 $(F_n,\ n\geq 1)$ є слабко компактною.

За означенням слабкої компактності знайдемо підпослідовність $(F_{n_k},\ k\geq 1) \subset (F_n,\ n\geq 1)$, та функцію розподілу F такі, що $F_{n_k} \to^W F$. 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 була б бути не меншою за ε.

Отже, від супротивного $F_n \to^W F$ $\boxdot$

Вправа.

(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) Випадкові величини $(\xi_n, \ n \geq 1)$ та ξ такі, що $|\xi_n| \leq c, |\xi| \leq c$ для деякої сталої c, та $M\xi_n^k \to M\xi^k, n \to \infty$, при всіх $k \geq 1$. Довести, що $\xi_n \to^W \xi$.

(3) Випадкові величини $(\xi_n, \ n \geq 1)$ незалежні однаково розподілені на відрізку $[0, 1)$ і $|M \exp(2\pi i k \xi_1)| < 1$ для всіх $k \in \mathbb{N}$. Доведіть, що $S_n = (\xi_1 + .. + \xi_n)(mod \ 1) \to^W \chi$, де величина χ рівномірно розподілена на $[0, 1)$.

(4) Нехай ξ – довільна випадкова величина з характеристичною функцією φ_ξ, а величини ζ_T не залежать від ξ і мають щільності $(1-\cos Tx)/\ (\pi T x^2)$, $x\in\mathbb{R}$. (а) Обчислити характеристичні функції $\varphi_{\zeta_T}(t) = (1-|t|/T)1_{|t|<T}$. (б) Вивести звідси слабку збіжність $\xi+\zeta_T \to^W \xi$, $T\to\infty$. (в) Довести формулу обертання для точок неперервності функції розподілу ξ :

$$F_\xi(b)-F_\xi(a)=\frac{1}{2\pi}\lim_{T\to\infty}\int_a^b dx\int_{-T}^{T}\exp(-itx)(1-|t|/T)\,\varphi_\xi(t)\,dt$$

Теорема (**про добуток слабко збіжної послідовності зі збіжною**). *Нехай* $(\xi_n,\ n\ge 1)$ *послідовність випадкових величин така, що* $\xi_n \to^W \xi$, $n\to\infty$, *а числова послідовність* $a_n \to a$. *Тоді*

$$a_n\xi_n \to^W a\xi.$$

Доведення. Позначимо φ_n – характеристичну функцію ξ_n. Оскільки послідовність функцій розподілу F_n величин ξ_n слабко збігається, то вона слабко компактна. Тому за теоремою Прохорова про критерій слабкої компактності $\sup_n F_n(\overline{[-c,c)})\to 0$, $c\to\infty$. Звідси

$$\begin{gathered}|\varphi_n(t)-\varphi_n(s)| \ \le \\ \int_{|x|>c}|e^{itx}-e^{isx}|\,dF_n(x)+\int_{|x|\le c}|e^{itx}-e^{isx}|\,dF_n(x)\le \\ 2\sup\nolimits_n\int_{|x|>c}dF_n(x)+\sup\nolimits_{|x|\le c}|e^{itx}-e^{isx}|\int_{|x|\le c}dF_n(x)\le \\ 2\sup\nolimits_n F_n(\overline{[-c,c)})+\sup\nolimits_{|x|\le c}\ \left|e^{i(t-s)x}-1\right|\to 0,\ |t-s|\to 0.\end{gathered}$$

Отже, функції $\varphi_n(t)$ є рівномірно за n і t неперервними.
Тому зі збіжності $\varphi_n(t)\to\varphi(t)$ та з оцінки

$$|\varphi_n(a_n t) - \varphi(at)| \le |\varphi_n(a_n t) - \varphi_n(at)| + |\varphi_n(at) - \varphi(at)|$$

випливає збіжність $\varphi_n(a_n t) \to \varphi(at)$, що за теоремою Леві про критерій слабкої збіжності доводить шукане твердження $\boxdot$

Вправа.

(1) Доведіть, що $a_n + \xi_n \to^W a + \xi$.

(2) Доведіть, що в попередніх твердженні та теоремі послідовність сталих $a_n \to a$ можна замінити на послідовність випадкових величин $\eta_n \to^P a$.

29.4. Сумісна характеристична функція та слабка збіжність векторів

Означення. Сумісною характеристичною функцією *випадкового вектора $\xi = (\xi_1, ..., \xi_d)$ називаеться така функція векторного аргументу $t = (t_1, ..., t_d) \in \mathbb{R}^d$:*

$$\varphi_\xi(t) = M\exp(it'\xi) = M\exp\left(\sum\nolimits_{k=1}^{d} it_k\xi_k\right).$$

Сумісна характеристична функція має такі ж властивості, що і звичайна характеристична функція. Зокрема, відповідність між сумісними функціями розподілу та сумісними характеристичними функціями є взаємно однозначною.

Означення. *Випадкові d-вимірні вектори ξ_n слабко збігаються: $\xi_n \to^W \xi$, якщо $Mg(\xi_n) \to Mg(\xi)$ 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.

Звідси випливає таке твердження.

Теорема (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 $t'\xi_n \to^W t'\xi$.*

Доведення.

Нехай φ_n, φ – сумісні характеристичні функції векторів ξ_n, ξ. За теоремою Леві про критерій слабкої збіжності при кожному $t \in \mathbb{R}^d$ збіжність $t'\xi_n \to^W t'\xi, n \to \infty$, еквівалентна збіжності характеристичних функцій $M\exp(is\ t'\xi_n) \to M\exp(is\ t'\xi)$ 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 $\{st,\ s \in \mathbb{R}, t \in \mathbb{R}^d\} = \mathbb{R}^d$ $\boxdot$

Вправа.

(1) Нехай $\xi_n \to^W \xi \in \mathbb{R}^d$, а функція $g \in C(\mathbb{R}^d)$ неперервна. Довести, що $g(\xi_n) \to^W g(\xi)$.

(2) Довести, що сумісна характеристична функція нормального випадкового вектора $\xi \simeq N_n(m, V)$ дорівнює $\varphi_\xi(t) = \exp(it'm - t'Vt\ /\ 2)$.

30. 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.

Теорема (класична центральна гранична теорема). *Нехай* $(\xi_n,\ n \geq 1)$ *послідовність незалежних у сукупності однаково розподілених випадкових величин зі скінченними середніми та дисперсіями:* $\mu = M\xi_1$, $\sigma^2 = D\xi_1$, *а* $S_n = \xi_1 + \ldots + \ \xi_n$. *Тоді*

$$(S_n - MS_n)/\sqrt{DS_n} \equiv (S_n - n\mu)/\sigma\sqrt{n} \to^W \zeta \simeq N(0,1),\ n \to \infty.$$

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

$$P(\frac{S_n - n\mu}{\sigma\sqrt{n}} < x) \to \Phi(x) \equiv \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}\ y^2)dy,\ n \to \infty.$$

Зокрема, звідси при довільному розподілі окремих доданків за правилом 3 сигма для досить великих n випливає наближена рівність

$$P(|S_n - n\mu| < 3\sigma\sqrt{n}) \approx 0.997.$$

Доведення. Рівності $MS_n = n\mu$, $DS_n = \sigma^2 n$ є очевидними 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):

$$\varphi_n(t) \equiv M \exp(it((S_n - n\mu)/\sigma\sqrt{n})) =$$
$$\exp(-it\ n\ \mu/\sigma\sqrt{n})\ \varphi_{S_n}(t/\sigma\sqrt{n}) = \exp(-it\ n\ \mu/\sigma\sqrt{n})\ \varphi^n(t/\sigma\sqrt{n}).$$

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

$$\ln\varphi_n(t) = -it\ n\ \mu\ /\ \sigma\sqrt{n} +$$

$$n\left(it\ \mu/\sigma\sqrt{n} - \sigma^2 t^2\ /\ 2\sigma^2 n + o(1/n)\right) = -t^2\ /\ 2\ + o(1),\ n \to \infty.$$

Отже, $\varphi_n(t) \to \exp(-t^2\ /\ 2) = \varphi_\zeta(t)$. Застосування теореми Леві про критерій слабкої збіжності довершує доведення $\boxdot$

Вправа.

(1) Випадкові величини $(\xi_n,\ n \geq 1)$ незалежні, однаково розподілені, $P(\xi_n = 0) = P(\xi_n = 1) = 1/2$. Довести, що сума ряду $\sum_{n\geq 1} 2^{-n}\xi_n$ рівномірно розподілена на $[0,1]$

(2) Випадкові величини $(\xi_n,\ n \geq 1)$ незалежні, однаково розподілені зі щільністю Коші $(\pi(1 + x^2))^{-1}$. Знайти граничний розподіл для нормованих максимумів $n^{-1}\max_{1\leq k\leq n}\xi_k$.

(3) Випадкові величини $(\xi_n,\ n \geq 1)$ 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. (б) Обчислити граничний при $n \to \infty$ розподіл величин $n \min_{1 \le k \le n} \xi_k$.

(4) Вивести з класичної центральної граничної теореми для пуассонівських величин тотожність $\lim_{n\to\infty} \exp(-n) \sum_{k=0}^{n} n^k/k! = 1/2$.

31. Загальна гранична теорема для стандартних серій

31.1. Стандартні послідовності серій

Означення. *Подвійна послідовність випадкових величин* $(\zeta_{nk},\ k = \overline{1, k_n},\ n \geq 1)$ *називається* стандартною послідовністю серій, *якщо виконуються умови:*

(1 –незалежність) випадкові величини в кожній серії $(\zeta_{nk},\ k = \overline{1, k_n})$ *незалежні в сукупності,*

(2 –центрованість) $M\zeta_{nk} = 0,\ \forall k,\ \forall n,$

(3 –нормованість) $\exists D\zeta_{nk} = \sigma_{nk}^2$ *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 $(\xi_n\ , n \geq 1)$ 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

$$\zeta_n = \sum_{k=1}^{k_n} \zeta_{nk},\ \varphi_n(t) = M\exp(it\zeta_n),$$

$$F_{nk}(x) = P(\zeta_{nk} < x),\ \varphi_{nk}(t) = M\exp(it\zeta_{nk}).$$

За адитивністю математичного сподівання, теоремою про дисперсію суми незалежних величин та теоремою про властивості характеристичної функції, пункт (б),

$$M\zeta_n = 0,\ D\zeta_n = 1,\ \varphi_n(t) = \prod\nolimits_{k=1}^{k_n} \varphi_{nk}(t).$$

31.2. Допоміжні леми

Доведення загальної граничної теореми базується на таких лемах.

Лема 1. Для стандартної послідовності серій виконується співвідношення

$$\ln \varphi_n(t) - \sum\nolimits_{k=1}^{k_n} (\varphi_{nk}(t) - 1) \to 0,\ n \to \infty.$$

Доведення. З елементарної нерівності $|\exp(ix) - 1 - ix| \le x^2/2$ та з умови центрованості $M\zeta_{nk} = 0$ означення стандартних серій дістанемо

$$|\varphi_{nk}(t) - 1| = |M(\exp(it\zeta_{nk}) - 1 - it\zeta_{nk})| \le t^2\sigma_{nk}^2 \le t^2 \max_k \sigma_{nk}^2 \to 0.$$

Зокрема, $|\varphi_{nk}(t)| \ge 1/2$, 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 $|\ln(1+z) - z| \le |z|^2$ та попередньої нерівності оцінимо

$$\left|\ln \varphi_n(t) - \sum\nolimits_{k=1}^{k_n} (\varphi_{nk}(t) - 1)\right| =$$

$$\left|\sum\nolimits_{k=1}^{k_n} (\ln \varphi_{nk}(t) - \varphi_{nk}(t) + 1)\right| \le \sum\nolimits_{k=1}^{k_n} |\varphi_{nk}(t) - 1|^2 \le$$

$$\sum\nolimits_{k=1}^{k_n} t^4\sigma_{nk}^4 \le t^4 \max_k \sigma_{nk}^2 \sum\nolimits_{k=1}^{k_n} \sigma_{nk}^2 = t^4 \max_k \sigma_{nk}^2 \to 0,\ n \to \infty$$

де враховане також означення стандартних серій$\boxdot$

Лема 2. (а) При кожному n функції

$$G_n(x) \equiv \sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2 1_{\{\zeta_{nk}\ <\ x\}} = \sum\nolimits_{k=1}^{k_n} \int_{(-\infty,\ x)} y^2\ dF_{nk}(y)$$

є функціями розподілу.

(б) Для всіх обмежених борелєвих функцій g виконується тотожність

$$\sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2\ g(\zeta_{nk}) = \int_{-\infty}^{\infty} g(y)\ dG_n(y).$$

Доведення.

(а) Монотонність $G_n(x)$ очевидна. Нормованість випливає з теореми Лебега про монотонну збіжність та нормованості (3) стандартної послідовності серій:

$$\lim\nolimits_{x\to\infty} G_n(x) = \sum\nolimits_{k=1}^{k_n} \lim\nolimits_{x\to\infty}\ M\zeta_{nk}^2 1_{\{\zeta_{nk}\ <\ x\}} =$$
$$\sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2 = \sum\nolimits_{k=1}^{k_n} D\zeta_{nk} = \sum\nolimits_{k=1}^{k_n} \sigma_{nk}^2 = 1.$$

Неперервність G_n зліва також виводиться з теореми Лебега про монотонну збіжність, з урахуванням неперервності зліва функції $1_{\xi\ <\ x}$:

$$\lim_{y\uparrow x} G_n(y) = \sum\nolimits_{k=1}^{k_n} \lim_{y\uparrow x}\ M\zeta_{nk}^2 1_{\{\zeta_{nk}\ <\ y\}} = \sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2 1_{\{\zeta_{nk}\ <\ x\}} = G_n(x).$$

(б) Позначимо через $\mathfrak{G}$ клас борелєвих функцій g, для яких виконується тотожність твердження (б):

$$\mathfrak{G} = \left\{ g : \sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2\ g(\zeta_{nk}) = \int_{-\infty}^{\infty} g(y)\ dG_n(y) \right\}.$$

З означення G_n випливає, що $g(y) = 1_{y\,<\,x} \in \mathfrak{G}$ при довільному $x \in \mathbb{R}$. З лінійності $\mathfrak{G}$ виводимо, що $1_A \in \mathfrak{G}$ 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 інтегралу Лебега – Стілтьєса клас $\mathfrak{G}$ замкнений відносно монотонної та мажорованої збіжності. Тому, зокрема, клас $\{B \in \mathfrak{B}(\mathbb{R}) : 1_B \in \mathfrak{G}\}$ замкнений відносно монотонної збіжності подій, звідки за теоремою про монотонний клас $1_B \in \mathfrak{G}$, $\forall B \in \mathfrak{B}(\mathbb{R})$. 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

$$f_t(x) = \begin{cases} (\exp(itx) - 1 - itx) \,/\, x^2, & x \neq 0, \\ -\,t^2 \,/\, 2, & x = 0. \end{cases}$$

Тоді

(а) $f_t(\cdot) \in C_b(\mathbb{R})$, $\forall t \in \mathbb{R}$.

(б) Для всіх t 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 $g(x) = f_t(x)$:

$$\int_{-\infty}^{\infty} f_t(x)\, dG_n(x) = \sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2\, f_t(\zeta_{nk}) =$$

$$\sum\nolimits_{k=1}^{k_n} M(\exp(it\zeta_{nk}) - 1 - it\zeta_{nk}) = \sum\nolimits_{k=1}^{k_n} (\varphi_{nk}(t) - 1),$$

оскільки за означенням $x^2 f_t(x) = \exp(itx) - 1 - itx$, і за властивостями стандартної послідовності серій $M\zeta_{nk} = 0$ $\boxdot$

31.3. Загальна 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*

$$G_n(x) \equiv \sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2 1_{\{\zeta_{nk}\ <\ x\}}$$

слабко збігається: $G_n \to^W G$ до деякої функції розподілу G.

Тоді послідовність сум $\zeta_n \equiv \sum_{k=1}^{k_n} \zeta_{nk} \to^W \zeta$ слабко збігається 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 $f_t \in C_b(\mathbb{R})$, то зі збіжності $G_n \to^W G$ та лем 3, 1 виводимо, що

$$\ln\varphi_n(t) - \int_{-\infty}^{\infty} f_t(x)\, dG(x) = \ln\varphi_n(t) - \sum\nolimits_{k=1}^{k_n} (\varphi_{nk}(t) - 1) +$$

$$+\int_{-\infty}^{\infty} f_t(x)\, dG_n(x) - \int_{-\infty}^{\infty} f_t(x)\, dG(x) \to 0,\ n \to \infty, \quad \forall t.$$

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 неперервний в нулі за t за теоремою Лебега про мажоровану збіжність, оскільки $|f_t(x)| \leq 1$ при $|t| \leq 1$ та $f_t(x) \to f_0(x) \equiv 0$ при $t \to 0$ для всіх x.Тому за теоремою Леві про критерій слабкої збіжності $\varphi_\zeta(t)$ є характеристичною функцією і $\zeta_n \to^W \zeta$ $\boxdot$

32. Центральні граничні теореми для послідовностей серій

32.1. Теорема Ліндеберга для стандартних серій

Теорема (центральна гранична теорема Ліндеберга для стандартних серій). *Нехай* $(\zeta_{nk},\ k = \overline{1, k_n},\ n \geq 1)$ – *стандартна послідовність серій випадкових величин, що задовольняє умову Ліндеберга:*

$$L_n(\varepsilon) \equiv \sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2\, 1_{\{|\zeta_{nk}| \geq \varepsilon\}} \to 0,\ n \to \infty,\ \forall\ \varepsilon > 0.$$

Тоді

$$\zeta_n \equiv \sum\nolimits_{k=1}^{k_n} \zeta_{nk} \to^W \zeta \simeq N(0, 1).$$

Зауваження. Можна довести, що умова Ліндеберга є необхідною для асимптотичної нормальності суми стандартних серій.

Доведення теореми. Досить оцінити в умовах загальної граничної теореми для стандартних серій при кожному $\varepsilon > 0$

$$G_n(-\varepsilon) + 1 - G_n(\varepsilon) = \sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2\, 1_{\{\zeta_{nk} < -\varepsilon\}} +$$

$$\sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2 - \sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2\, 1_{\{\zeta_{nk} < \varepsilon\}} = \sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2\, 1_{\{\zeta_{nk} < -\varepsilon\} \cup \{\zeta_{nk} \geq \varepsilon\}} \leq$$

$$\sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2\, 1_{\{|\zeta_{nk}| \geq \varepsilon\}} = L_n(\varepsilon) \to 0,$$

отже, $G_n(-\varepsilon) \to 0$, $G_n(\varepsilon) \to 1$, $\forall \varepsilon > 0$ і $G_n \to^O G$, де $G(x) = 1_{0 < x}$ є функцією розподілу.

Тому за теоремою про еквівалентність слабкої збіжності та в основному $G_n \to^W G$. Отже, згідно із загальною граничною теоремою для стандартних серій $\zeta_n \to^W \zeta$, де гранична характеристична функція дорівнює

$$\varphi_\zeta(t) = \exp(\int_{-\infty}^{\infty} f_t(x)\, dG(x)) = \exp(f_t(0)) = \exp(-t^2/2)$$

і є характеристичною функцією стандартного нормального розподілу. Звідси за теоремою про однозначну відповідність між функціями розподілу та характеристичними функціями $\zeta \simeq N(0,1)$ $\boxdot$

32.2. 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:*

$$\exists \delta > 0:\ L_n^{2+\delta} \equiv \sum_{k=1}^{k_n} M\,|\zeta_{nk}|^{2+\delta} \to 0,\ n \to \infty.$$

Тоді

$$\zeta_n \equiv \sum_{k=1}^{k_n} \zeta_{nk} \to^W \zeta \simeq N(0,1).$$

Доведення. Оцінимо при всіх $\varepsilon > 0$ показник Ліндеберга

$$L_n(\varepsilon) = \sum_{k=1}^{k_n} M\zeta_{nk}^2\, 1_{\{|\zeta_{nk}| \geq \varepsilon\}} \leq$$

$$\sum_{k=1}^{k_n} M\zeta_{nk}^2 \left(\frac{|\zeta_{nk}|}{\varepsilon}\right)^{\delta} 1_{\{|\zeta_{nk}| \geq \varepsilon\}} \leq \varepsilon^{-\delta} L_n^{2+\delta} \to 0,$$

тому з умови Ляпунова випливає умова Ліндеберга, а отже, і асимптотична нормальність $\boxdot$

32.3. Теорема Ліндеберга для загальних серій

Означення. *Подвійна послідовність випадкових величин* $(\xi_{nk},\ k=\overline{1,k_n},\ n\ge 1)$ *називається* загальною послідовністю серій, *якщо:*

(1) випадкові величини в кожній серії $(\xi_{nk},\ k=\overline{1,k_n})$ *незалежні в сукупності,*

(2) існують $M\xi_{nk}=\mu_{nk}$, $\mu_n\equiv\sum_{k=1}^{k_n}\mu_{nk}$,

(3) існують $D\xi_{nk}=\sigma^2_{nk}$, $\sigma^2_n\equiv\sum_{k=1}^{k_n}\sigma^2_{nk}$,

(4) $\max_{\,1\le\,k\,\le k_n}\ \sigma^2_{nk}\ /\ \sigma^2_n\to 0$ *при* $n\to\infty$.

Теорема (центральна гранична теорема Ліндеберга для загальних серій). *Нехай* $(\xi_{nk},\ k=\overline{1,k_n},\ n\ge 1)$ *– загальна послідовність серій випадкових величин, що задовольняє умову Ліндеберга:*

$$L_n(\varepsilon)\equiv\frac{1}{\sigma^2_n}\sum\nolimits_{k=1}^{k_n}M\left|\xi_{nk}-\mu_{nk}\right|^2 1_{\{|\xi_{nk}-\mu_{nk}|\,\ge\,\varepsilon\sigma_n\}}\to 0,\ n\to\infty,\forall\varepsilon>0.$$

Тоді суми $\xi_n\equiv\sum_{k=1}^{k_n}\xi_{nk}$ *після центрування та нормування є асимптотично нормальними:*

$$(\xi_n-\mu_n)\ /\ \sigma_n\to^W\zeta\simeq N(0,1).$$

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 загальних серій

$$M\zeta_{nk}=(M\xi_{nk}-\mu_{nk})/\sigma_n=0,\ \sum\nolimits_{k=1}^{k_n}D\zeta_{nk}=\sum\nolimits_{k=1}^{k_n}\sigma^2_{nk}/\sigma^2_n=1,$$
$$\max\nolimits_{1\le\,k\,\le k_n}D\zeta_{nk}\ =\max\nolimits_{1\le\,k\,\le k_n}\ \sigma^2_{nk}/\sigma^2_n\to 0,\ n\to\infty.$$

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 центральної граничної теореми Ліндеберга для стандартних серій $\boxdot$

Наслідок (центральна гранична теорема Ліндеберга для загальних серій). Якщо величини в кожній серії $(\xi_{nk}, \ k = \overline{1, k_n})$ однаково розподілені та

$$\lim_{n\to\infty} \ M \left|\xi_{n1} - \mu_{n1}\right|^2 \sigma_{n1}^{-2} \ 1_{\left\{|\xi_{n1}-\mu_{n1}| \geq \varepsilon\sqrt{n \ \sigma_{n1}^2}\right\}} = 0, \ \forall\varepsilon > 0,$$

то умова Ліндеберга виконується.

Дійсно, 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 $L_n(\varepsilon)$.

Вправа.Випадкові величини $(\xi_k, k \geq 1)$ незалежні, $P(\xi_k = \pm x_k) = 1/2$. За яких умов існують $b_n \to \infty$ такі, що $b_n^{-1} \sum_{k=1}^{n} \xi_k \to^W \zeta \simeq N(0,1)$?

32.4. 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:*

$$\exists\delta > 0 : \ L_n^{2+\delta} \equiv \frac{1}{\sigma_n^{2+\delta}} \sum_{k=1}^{k_n} M \left|\xi_{nk} - \mu_{nk}\right|^{2+\delta} \to 0, \ n \to \infty.$$

Тоді центровані та нормовані суми $\xi_n \equiv \sum_{k=1}^{k_n} \xi_{nk}$ асимптотично нормальні:

$$(\xi_n - \mu_n) \,/\, \sigma_n \to^W \zeta \simeq N(0,1).$$

Доведення, як і вище, спирається на стандартні серії $\zeta_{nk} = (\xi_{nk} - \mu_{nk}) \,/\, \sigma_n$ $\boxdot$

Наслідок. Умова Ляпунова виконується, коли величини в кожній серії $(\xi_{nk},\ k = \overline{1,k_n})$ однаково розподілені, а їх моменти $M\,|\xi_{nk}|^{2+\delta}$ порядку $2+\delta$ 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 $k_n^{2+\delta}$.

Вправа. Розглянемо послідовність серій $(\xi_{nk},\ k = \overline{1,n},\ n \ge 1)$ 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.

32.5. Гранична теорема Пуассона для стандартних серій

Граничним розподілом центрованих та нормованих сум незалежних величин може бути не тільки нормальний розподіл.

Теорема (гранична теорема Пуассона). *Нехай $(\zeta_{nk},\ k = \overline{1,k_n},\ n \ge 1)$ – стандартна послідовність серій випадкових величин, що при деякому $a > 0$ задовольняє умову*

$$P_n(\varepsilon) \equiv \sum\nolimits_{k=1}^{k_n} M\zeta_{nk}^2\, 1_{\{|\zeta_{nk}-a| \,\ge\, \varepsilon\}} \to 0,\ n \to \infty,\ \forall \varepsilon > 0.$$

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 $\lambda = 1/a^2$ та множиною значень $\{na - 1/a,\ n = 0, 1, ...\}$

Доведення.

Як і в центральній граничній теоремі Ліндеберга для стандартних серій, доводимо, що $G_n \to^O G$, де $G(x) = 1_{a\,<\,x}$ є функцією розподілу сталої a.

Звідси за загальною граничною теоремою для стандартних серій

$$\lim \varphi_{\zeta_n}(t) = \exp(\textstyle\int_{-\infty}^{\infty} f_t(x)\ dG(x)) = \exp(f_t(a)) = \varphi_\zeta(t),$$

де пуассонів розподіл границі ζ визначається з розкладу у ряд Тейлора експоненти від $f_t(a) = (\exp(ita) - 1 - ita)\ /\ a^2$:

$$\varphi_\zeta(t) = \sum\nolimits_{n\ge 0} \frac{\exp(it(an - 1/a))\ a^{-2n}}{n!} \exp(-a^{-2}) \boxdot$$

32.6. Центральна гранична теорема для випадкових векторів

Теорема (центральна гранична теорема для випадкових векторів). *Нехай $(\eta_{nk},\ k = \overline{1, k_n},\ n \ge 1)$ – 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 $m_{nk} = M\eta_{nk}$ та 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) $\max_{\,k\le k_n}\ t'V_{nk}t \to 0,\ \forall t \in \mathbb{R}^d$ при $n \to \infty$.

Якщо виконана умова Ліндеберга:

$$L_n(\varepsilon) \equiv \sum_{k=1}^{k_n} M\|\eta_{nk} - m_{nk}\|^2 \ 1_{\{\|\eta_{nk}-m_{nk}\| \geq \varepsilon\}} \to 0, \ n \to \infty, \forall \varepsilon > 0,$$

то суми всередині серій $\eta_n \equiv \sum_{k=1}^{k_n} \eta_{nk}$ *слабко збігаються до нормального вектора:*

$$\eta_n \to^W \xi \simeq N(m, V), \ n \to \infty.$$

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 послідовність випадкових величин

$$\xi_{nk} = t'\eta_{nk}, \ \xi_n = \sum_{k=1}^{k_n} \xi_{nk} = t'\eta_n.$$

Тоді $M\xi_n = t'm_n \to t'm$ та за теоремою про дисперсію лінійної форми від випадкового вектора $D\xi_{nk} = t'V_{nk}t$. Позначимо

$$\theta_n^2 = t'V_n t, \ \theta^2 = t'Vt.$$

Тоді за умовою

$$D\xi_n = \sum_{k=1}^{k_n} D\xi_{nk} = \sum_{k=1}^{k_n} t'V_{nk}t = \theta_n^2, \quad \theta_n^2 \to \theta^2, \ n \to \infty.$$

Припустимо, що $\theta = 0$. Тоді $D\xi_n \to 0$ і $\xi_n - M\xi_n \to^{L2} 0$. Тому $\xi_n \to^P t'm$. Оскільки 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 $\theta > 0$. Тоді послідовність (ξ_{nk}) є загальною послідовністю серій, оскільки $\theta_n^2 \geq \theta^2/2 > 0$, починаючи з деякого номера, та

$$\max_{k \le k_n} D\xi_{nk}/D\xi_n = \max_{k \le k_n} t'V_{nk}t\ /\theta_n^2 \to 0,\ n \to \infty,$$

Решта умов в означенні загальних серій випливає з умов (1-3) теореми.

Показник з умови Ліндеберга в центральній граничній теоремі Ліндеберга для загальних серій (ξ_{nk}) має вигляд

$$L_n(\varepsilon) = \frac{1}{\theta_n^2}\ \sum_{k=1}^{k_n} M\left|\xi_{nk} - M\xi_{nk}\right|^2\ 1_{\{|\xi_{nk} - M\xi_{nk}| \,\ge\, \varepsilon\,\theta_n\}} \le$$

$$\frac{1}{\theta_n^2}\ \sum_{k=1}^{k_n} M\|\eta_{nk} - m_{nk}\|^2 \|t\|^2\ 1_{\{\|\eta_{nk} - m_{nk}\|\ \ \|t\| \,\ge\, \varepsilon\,\theta_n\}} \le$$

$$\le \frac{\|t\|^2}{\theta^2/2} \sum_{k=1}^{k_n} M\|\eta_{nk} - m_{nk}\|^2\ 1_{\{\|\eta_{nk} - m_{nk}\| \,\ge\, \delta\}} \to 0,\ \ n \to \infty,$$

оскільки $0 < \delta \equiv \varepsilon\theta/2\|t\|\ \ \le \varepsilon\theta_n/\|t\|$, починаючи з деякого номера. Отже, за центральною граничною теоремою Ліндеберга для загальних серій $(\xi_n - t'm_n)/\theta_n\ \to^W\ \zeta\ \simeq\ N(0,1)$. Тому за теоремою про добуток слабко збіжної послідовності зі збіжною

$$\xi_n - t'm_n = \theta_n\ (\xi_n - t'm_n)/\theta_n \to^W \theta\zeta \simeq N(0, \theta^2).$$

Оскільки $t'm_n \to t'm$, то $\xi_n \to^W t'm + \theta\zeta \simeq N(t'm,\ \theta^2) = N(t'm,\ t'Vt)$.

Отже, і в даному випадку кожна лінійна форма від сум $\eta_n \equiv \sum_{k=1}^{k_n} \eta_{nk}$ слабко збігається:

$$t'\eta_n \to^W t'\xi \simeq N(t'm,\ t'Vt),\ n \to \infty,\ \forall t \in \mathbb{R}^d.$$

Тому за теоремою Леві про критерій слабкої збіжності випадкових векторів має місце слабка збіжність $\eta_n \to^W \xi$ $\boxdot$

Теорема (класична центральна гранична теорема для випадкових векторів). *Нехай* $(\gamma_n,\ \ n \ge 1)$ *послідовність незалежних у сукупності однаково розподілених випадкових векторів у* $\mathbb{R}^d$, $M\gamma_1 = m$, $Cov(\gamma_1) = V$, *а* $\eta_n \equiv \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (\gamma_k - m)$.

Тоді $\eta_n \to^W \zeta \simeq N(0,\ V)$.

Доведення. Покладемо $\eta_{nk} = \frac{1}{\sqrt{n}}(\gamma_k - m),\ k = \overline{1,n}$ 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). Далі, має місце і умова (4):

$$\max\ _{k \le k_n} t'V_{nk}t = t'Vt\ /n \to 0,\ n \to \infty.$$

Нарешті, показник з умови Ліндеберга прямує до нуля за теоремою Лебега про монотонну збіжність, оскільки з урахуванням однакової розподіленості

$$L_n(\varepsilon) = n\ M\|\eta_{n1}\|^2\ 1_{\{\|\eta_{n1}\|\ \ge\ \varepsilon\}} =$$
$$M\|\gamma_1 - m\|^2\ 1_{\{\|\gamma_1 - m\|\ \ge\ \varepsilon\sqrt{n}\}} \downarrow 0,\ n \to \infty.$$

Отже, шукана збіжність випливає з центральної граничної теореми для випадкових векторів $\boxdot$

33. Випадкові блукання та генератриси

Означення. Випадковим блуканням $(\zeta_n,\ n \geq 0)$ *на* $\mathbb{Z}$ *називаеться послідовність сум незалежних у сукупності однаково розподілених цілозначних випадкових величин* $(\xi_n,\ n \geq 1)$:

$$\zeta_n = \sum\nolimits_{k=1}^{n} \xi_k.$$

За означенням $\zeta_0 = 0$.

Величину ζ_n можна трактувати як положення частинки в момент часу n, якщо вона в кожний момент k стрибком зміщується з поточного положення ζ_{k-1} на величину ξ_k, тобто

$$\zeta_k = \zeta_{k-1} + \xi_k.$$

Послідовність точок (n, ζ_n) розглядається як траєкторія частинки в координатах час – простір.

Означення. *В момент* $n \geq 1$ *відбуваеться* повернення *в* 0, *якщо реалізуеться подія*

$$U_n = \{\zeta_n = 0\}.$$

У момент $n \geq 1$ *відбуваеться* перше повернення *в* 0, *якщо справджуеться подія*

$$F_n = \{\zeta_1 \neq 0, ..., \zeta_{n-1} \neq 0,\ \zeta_n = 0\}.$$

Означення. *Повернення в* 0 *відбудеться коли-небудь, якщо реалізуеться подія*

$$U = \bigcup\nolimits_{n \geq 1} U_n.$$

Теорема. Справедлива рівність

$$U = \bigcup_{n\geq 1} U_n \ = \bigcup_{n\geq 1} F_n,$$

тобто повернення коли-небудь є те саме, що повернення коли-небудь вперше.

Доведення очевидне ⊡

На події U визначений час до першого повернення

$$\tau \equiv \inf(n \geq 1 : \zeta_n = 0),$$

а саме: $\tau(\omega) = n$ на елементарних подіях $\omega \in F_n$. Будемо вважати, що $\tau = \infty$ на доповненні U. За означенням

$$U = \{\tau < \infty\}.$$

33.1. Рекурентність

Означення. *Випадкове блукання* $(\zeta_n,\ n \geq 0)$ *називається* рекурентним, *якщо* $P(U) = 1$, *тобто* $P(\tau < \infty) = 1$.

Для аналізу рекурентності блукання введемо такі позначення

$$u_n = P(U_n),\ f_n = P(F_n),\ u_0 = 1,\ f_0 = 0.$$

Теорема (про рівняння відновлення). *Ймовірності повернення* u_n *та ймовірності першого повернення* f_n *є розв'язками при* $n \geq 1$ *рівняння відновлення*

$$u_n = \sum_{k=1}^{n} f_k\, u_{n-k},$$

де за означенням $u_0 = 1$.

Означення. *Права частина рівняння відновлення називається* згорткою послідовностей u_n *та* f_n.

Доведення. Оскільки $\{\zeta_n = 0\} \subset \{\tau \leq n\}$, то

$$\{\zeta_n = 0\} = \{\zeta_n = 0,\ \tau \le n\}.$$

Враховуючи, що події $\{\tau = k\}$ попарно несумісні, звідси отримуємо

$$u_n = P(\zeta_n = 0) = P(\zeta_n = 0,\ \tau \le n) =$$
$$\sum_{k=1}^n P(\tau = k,\ \zeta_n = 0) = \sum_{k=1}^n P(\tau = k,\ \zeta_n - \zeta_k = 0) =$$
$$\sum_{k=1}^n P(\tau = k)\ P(\xi_{k+1} + ... + \xi_n = 0) =$$
$$\sum_{k=1}^n P(\tau = k)\ P(\xi_1 + ... + \xi_{n-k} = 0) =$$
$$\sum_{k=1}^n P(F_k)\ P(\zeta_{n-k} = 0) = \sum_{k=1}^n\ f_k\ u_{n-k}.$$

Тут використана незалежність подій
$\{\tau = k\} = \{\zeta_1 \ne 0, ..., \zeta_{k-1} \ne 0, \zeta_k = 0\}$ та

$\{\zeta_n - \zeta_k = 0\} = \{\xi_{k+1} + ... + \xi_n = 0\}$, яка випливає з теореми про векторні перетворення незалежних величин, а також однакова розподіленість випадкових векторів $(\xi_1, ..., \xi_{n-k})$ та $(\xi_{k+1}, ..., \xi_n)$ $\boxdot$

33.2. Генератриса послідовності

Означення. *Нехай* $(a_n,\ n \ge 0)$ *числова послідовність. Її* генератрисою *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 всередині одиничного кола $\{z : |z| < 1\}$.

Теорема (**про генератриси ймовірностей повернення**). *Для кожного $z \in \mathbb{Z}$ з $|z| < 1$ генератриси*

$$u(z) \equiv \sum_{n\geq 0} u_n z^n, \; f(z) \equiv \sum_{n\geq 0} f_n z^n$$

задовольняють рівняння

$$u(z) = 1 + u(z)f(z).$$

Доведення. Помножимо n-те рівняння відновлення на z^n та додамо результати:

$$\sum_{n\geq 1} u_n z^n = u(z) - 1 = \sum_{n\geq 1} z^n \sum_{k=1}^{n} f_k \, u_{n-k} =$$
$$\sum_{k=1}^{\infty} f_k z^k \sum_{n\geq k} z^{n-k} u_{n-k} = \sum_{k=1}^{\infty} f_k z^k \sum_{n\geq 0} z^n u_n = f(z)u(z)$$

Абсолютна збіжність рядів обумовлена обмеженістю послідовностей u_n, f_n $\boxdot$

33.3. Критерій рекурентності

Теорема (**про критерій рекурентності випадкового блукання**). *Випадкове блукання* $(\zeta_n, \; n \geq 1)$ *на* $\mathbb{Z}$ *рекурентне тоді й тільки тоді, коли*

$$\sum_{n\geq 0} P(\zeta_n = 0) = \infty.$$

Доведення. Оскільки $f(1) = \sum_{n\geq 0} f_n = P(\tau < \infty)$, то $f(1) = 1$ тоді й тільки тоді, коли блукання рекурентне.

З іншого боку, за теоремою Абеля про суму ряду із 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

$$u(1) = 1 + u(1)f(1).$$

Отже, твердження $f(1) < 1$ та $u(1) \equiv \sum_{n \ge 0} u_n < \infty$ еквівалентні $\boxdot$

33.4. Блукання Бернуллі

Означення. *Випадковим* блуканням Бернуллі $(\zeta_n, \ n \ge 1)$ *на* $\mathbb{Z}$ *називається блукання зі стрибками* $(\xi_n, \ n \ge 1)$ *такими, що*

$$P(\xi_n = 1) = p = 1 - P(\xi_n = -1) = 1 - q.$$

Теорема (про рекурентність блукання Бернуллі) *Блукання Бернуллі рекурентне тоді й тільки тоді, коли* $p = 1/2$.

Доведення. За формулою Стірлінга маємо

$$u_{2n} = C_{2n}^{\,n}\, p^n (1-p)^n \approx (4p(1-p))^n / \sqrt{\pi n}, \ n \to \infty,$$

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*

$$P(\tau < \infty) = 1 - |p - q|\,, \quad f_{2n} = \frac{2}{2n-1} C_{2n-1}^{\,n} (pq)^n.$$

Доведення. Обчислимо за формулою розкладу в ряд Тейлора при $|z| < 1$ генератрису

$$u(z) = \sum\nolimits_{n \ge 0} z^{2n} C_{2n}^{\,n} (pq)^n = (1 - 4pqz^2)^{-1/2},$$

звідки за теоремою про генератриси ймовірностей повернення

$$f(z) = 1 - 1 \;/\; u(z) = 1 - (1 - 4pqz^2)^{+1/2}.$$

Звідси при $z \uparrow 1$ дістанемо першу тотожність

$$P(\tau < \infty) = f(1-) = 1 - (1 - 4pq)^{+1/2} = 1 - |p - q|\,.$$

Другу тотожність обчислюємо за означенням генератриси за допомогою розкладу в ряд Тейлора функції
$(1 - 4pqz^2)^{+1/2}$.

34. 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 із 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*

$$\zeta_0 = 1, \quad \zeta_{n+1} = \sum\nolimits_{k=1}^{\zeta_n} \xi_k^n, \ n \ge 0.$$

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, які є дискретними аналогами характеристичних функцій.

34.1. Генератриси випадкових величин

Означення. *Нехай ν невід'ємна цілозначна випадкова величина з розподілом $p_n = P(\nu = n)$, $n = 0, 1, \ldots$* Генератрисою випадкової величини *ν називається* *генератриса* *послідовності p_n :*

$$\varphi_\nu(z) \equiv \sum\nolimits_{n\geq 0} z^n p_n = Mz^\nu.$$

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 $|z| \leq 1$ та однозначно визначає розподіл p_n,

(б) $\varphi_\nu(1) = 1$.

(в) $|\varphi_\nu(z)| \leq 1$ при $|z| \leq 1$.

(г) $\varphi_\nu'(1) = M\nu$.

(д) Якщо величини ξ, η – незалежні, то $\varphi_{\xi\,+\,\eta}(z) = \varphi_\xi(z)\varphi_\eta(z)$.

(е) Якщо величини $(\xi_n,\ n \geq 1)$ та ν – незалежні в сукупності, а величина ζ є сумою випадкового числа ν випадкових доданків ξ_n вигляду

$$\zeta = \sum\nolimits_{k=1}^{\nu} \xi_k = \sum\nolimits_{k\geq 1} \xi_k \, 1_{\{k\,\leq\,\nu\}}\ ,$$

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 $L_2[-\pi,\pi]$.

(б), (в) очевидні.

(г) при $z \in (0,1)$ похідну можна внести під знак суми ряду, оскільки ряд із похідних збігається абсолютно:

$$\varphi_\nu'(z) = \sum\nolimits_{n\ \geq 0} nz^{n-1}p_n = M\nu z^{\nu-1}.$$

Спрямовуючи $0 \leq z \uparrow 1$, за теоремою Лебега про монотонну збіжність дістанемо шукану рівність.

(д) Випливає з теореми про перетворення незалежних величин z^ξ та z^η, та з теореми про математичне сподівання добутку незалежних величин.

(е) За означенням

$$\varphi_\zeta(z) = Mz^\zeta = \sum\nolimits_{k\geq 0} Mz^\zeta 1_{\{\nu\,=\,k\}} = \sum\nolimits_{k\geq 0} Mz^{\xi_1+\ldots+\xi_\nu} 1_{\{\nu\,=\,k\}} =$$
$$\sum\nolimits_{k\geq 0} Mz^{\xi_1+\ldots+\xi_k} 1_{\{\nu\,=\,k\}} = \sum\nolimits_{k\geq 0} Mz^{\xi_1+\ldots+\xi_k}\ M1_{\{\nu\,=\,k\}} =$$
$$\sum\nolimits_{k\geq 0} (Mz^{\xi_1})^k\ P(\nu = k) = \sum\nolimits_{k\geq 0} (\varphi_\xi(z))^k\ P(\nu = k) =$$
$$\sum\nolimits_{k\geq 0} u^k\ P(\nu = k)\ |_{\,u\,=\,\varphi_\xi(z)} = \varphi_\nu(\varphi_\xi(z)),$$

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 одного доданку.

Зауваження. Твердження (е) залишається справедливим для дійснозначних величин ξ_n, якщо тільки коректно визначено перетворення

$$\varphi_\xi(z) = Mz^{\xi_1},$$

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 $\boxdot$

Вправа.

(1) Довести методом генератрис, що сума незалежних величин із розподілами Пуассона знову має розподіл Пуассона.

(2) 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.

34.2. Генератриса гіллястого процесу

Теорема (про генератриси гіллястого процесу). *Нехай* $(\zeta_n,\ n \geq 0)$ – *гіллястий процес*, $\varphi(z) = Mz^{\xi_1^1}$ – *генератриса числа нащадків однієї частинки*, $\varphi_n(z) = Mz^{\zeta_n}$ – *генератриса кількості частинок* n*-го покоління. Тоді для всіх* $n \geq 0$:

(а) $\varphi_{n+1}(z) = \varphi_n(\varphi(z)), \quad \varphi_0(z) = z,$

(б) $\varphi_{n+1}(z) = \varphi(\varphi_n(z)).$

Доведення.

(а) є очевидним наслідком означення гіллястого процесу та теореми про властивості генератрис, пункт (e). 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 $(\xi_k^m,\ k \geq 1, m \leq n-1)$, 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) $\boxdot$

34.3. Класифікація та властивості гіллястих процесів

Означення. Критичним показником μ *гіллястого процесу* $(\zeta_n, n \geq 0)$ *називається середня кількість нащадків однієї частинки:*

$$\mu = M\xi_1^1.$$

Гіллястий процес називається

(а) субкритичним, *якщо* $\mu < 1$,

(б) критичним, *якщо* $\mu = 1$,

(в) суперкритичним, *якщо* $\mu > 1$.

Означення. Ймовірністю виродження *у n-му поколінні називається число*

$$\pi_n = P(\zeta_n = 0),$$

а ймовірністю виродження – границя

$$\pi = \lim \pi_n = P(\cup_{n\,\geq\,0}\{\zeta_n = 0\}).$$

Зауваження. Оскільки за означенням гіллястого процесу $\{\zeta_n = 0\} \subset \{\zeta_{n+1} = 0\}$, то $\{\zeta_n = 0\} \uparrow \cup_{n\,\geq\,0}\{\zeta_n = 0\}$ і за неперервністю ймовірності $\pi_n \uparrow \pi$.

Теорема (про ймовірність виродження гіллястого процесу). *Якщо гіллястий процес*

(а) субкритичний або (б) критичний, то ймовірність виродження $\pi = 1$.

(в) Для суперкритичного процесу ймовірність $\pi < 1$ *є єдиним розв'язком рівняння*

$$\pi = \varphi(\pi)$$

на інтервалі $[0, 1)$.

Доведення. Оскільки за означенням генератриси $\varphi_n(0) = P(\zeta_n = 0) = \pi_n$, то з теореми про генератриси гіллястого процесу, пункт (б), отримуємо 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

$$p = \varphi(p),\ \ p \geq 0.$$

Оскільки $\varphi(1) = 1$, а функція φ як сума ряду із невід'ємними коефіцієнтами опукла донизу на $\mathbb{R}_+$, то дане рівняння завжди має не більше 2 коренів, один з яких дорівнює 1.

Доведемо, що ймовірність π є найменшим із можливих 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 $\varphi'(1) = \mu < 1$, тобто в лівому околі одиниці

$$\varphi(z) = 1 + (z-1)\mu + o(|z-1|) > z.$$

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 і $\pi = 1$.

У випадку (б) пряма 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 (1 – інший корінь) $\boxdot$

Теорема (про середню чисельність поколінь). *Нехай $(\zeta_n,\ n \geq 0)$ гіллястий процес, $\mu_n = M\zeta_n$. Тоді середня чисельність n-го покоління дорівнює*

$$\mu_n = \mu^n,$$

причому

(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) генератриси 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 $\mu_n = \mu^n$ $\boxdot$

35. Гранична теорема Реньї з теорії надійності

На відміну від центральної граничної теореми, в теорії надійності розглядаються 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,* $\mu = M\xi_1 < \infty$*, і не залежать від випадкової величини* ν*, що має геометричний розподіл із параметром* θ *:*

$P(\nu = n) = \theta(1-\theta)^{n-1}, n \geq 1.$ *Позначимо* $\zeta = \sum_{k=1}^{\nu} \xi_k$.

Тоді для всіх $x > 0$

$$P(\theta\ \zeta\ /\mu > x) \to \exp(-x),\ \theta \to 0.$$

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 дорівнює часу до першої відмови системи.

Доведення. Позначимо $\varphi(t)$ – характеристичну функцію випадкової 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 $z = \exp(it)$

$$M\exp(it\zeta) = \psi_\nu(M\exp(it\xi_1)) = \theta\varphi(t)/(1-\varphi(t)(1-\theta)).$$

Підставимо в цю формулу $t = s\theta/\mu \to 0$, $\theta \to 0$, та використаємо розклад $\varphi(t)$ в околі нуля

$$\lim_{\theta\to 0} M\exp(is\ \zeta\theta/\mu) = \lim_{\theta\to 0}\theta\varphi(s\theta/\mu)/(1-\varphi(s\theta/\mu)(1-\theta)) =$$
$$\lim_{\theta\to 0}\theta\cdot(1+O(\theta))\ /\ (1-(1+i\mu\ s\theta/\mu)(1-\theta)+o(\theta)) =$$
$$\lim_{\theta\to 0}\theta\ /\ (\theta - is\theta + o(\theta)) = 1/(1-is).$$

Гранична функція неперервна в нулі та є характеристичною функцією величини τ з показниковим розподілом та одиничним параметром. За теоремою Леві про критерій слабкої збіжності $\zeta\theta/\mu \to^W \tau$ $\boxdot$

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
$P(\zeta > 10 * 365) = P(0.01\frac{\zeta}{(350+380)/2} > 0.10) \simeq \exp(-0.10) \simeq 0.904.$

36. 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:

$$P(\text{Майбутнє} \cap \text{Минуле} \ / \ \text{Сучасне}) =$$
$$= P(\text{Майбутнє} \ / \ \text{Сучасне}) \cdot P(\text{Минуле} \ / \ \text{Сучасне})$$

З означення умовної ймовірності виводимо, що ця рівність еквівалентна

$$P(\text{Майбутнє} \ / \ \text{Минуле} \cap \text{Сучасне}) = P(\text{Майбутнє} \ / \ \text{Сучасне}).$$

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*

$$P(\zeta_{t+1} = j \ /\zeta_0 = i_0, ..., \zeta_{t-1} = i_{t-1}, \zeta_t = i) =$$
$$P(\zeta_{t+1} = j \ /\zeta_t = i) = P(\zeta_1 = j \ /\zeta_0 = i).$$

Перша з цих рівностей визначає власне марковську властивість, а друга – однорідність за часом.

Приклади.

1. *Незалежні однаково розподілені величини* – для них умовні ймовірності збігаються із 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 ξ_s, $s \leq t$. Тому умовна ймовірність

$$\begin{gathered} P(\zeta_{t+1}= j \;/\zeta_0= i_0, ..., \zeta_{t-1}= i_{t-1}, \zeta_t= i) = \\ P(i + \xi_{t+1}= j \;/\zeta_0= i_0, ..., \zeta_{t-1}= i_{t-1}, \zeta_t= i) = \\ P(i + \xi_{t+1}= j) = P(i + \xi_{t+1}= j \;/\; \zeta_t= i) \end{gathered}$$

збігається із безумовною.

3. *Рекурентні послідовності* вигляду

$$\zeta_{t+1} = g(\zeta_t, \xi_{t+1})$$

із незалежними $(\xi_t, t \geq 1)$ та борелєвими g є ланцюгами Маркова за таких самих міркувань, що і в попередньому пункті.

4. *Серії успіхів* у схемі Бернуллі. Якщо $\zeta_k \in \{0, 1\}$ результат k-го випробування Бернуллі, то довжина серії успіхів у момент n визначається як

$$\xi_n = \inf(k \geq 0 : \zeta_{n-k} = 0)$$

та задовольняє відповідне Марковське рекурентне рівняння.

5. *Модель черги.* Нехай ζ_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 ξ_n – кількість заявок у системі в момент n, то марковська властивість є наслідком рівняння

$$\xi_{n+1} = \max(\xi_n - 1, 0) + \zeta_{n+1}$$

6. *Гіллясті процеси.*

Вправа. Випадкові величини $(\xi_t, t \geq 0)$ незалежні і $P(\xi_t = \pm 1) = 1/2$. Довести, що послідовність $\zeta_t = (\xi_t + \xi_{t+1})/2, t \geq 0$, не є ланцюгом Маркова.

36.1. Перехідна матриця за один крок

Означення. Матрицею перехідних імовірностей за один крок *марковського ланцюга називається матриця*

$$P = (p_{ij},\ i, j \in E),\ \ p_{ij} = P(\zeta_1 = j\ /\zeta_0 = i).$$

Початковий розподіл *ланцюга дорівнює* $q_i = P(\zeta_0 = i)$.

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* $t \geq 1,\ i_0, i_1, ..., i_{t-1}, i_t \in E$

$$P(\zeta_0 = i_0, ..., \zeta_{t-1} = i_{t-1}, \zeta_t = i_t) = q_{i_0} p_{i_0 i_1} p_{i_1 i_2} ... p_{i_{t-1} i_t}.$$

Доведення випливає з теореми про ймовірність перерізу n подій (вираз через умовні ймовірності) та з марковської властивості ланцюга

$$P(\cap_{s=0}^{t}\{\zeta_s = i_s\}) = P(\zeta_0 = i_0)\prod_{s=1}^{t} P(\zeta_s = i_s / \cap_{r=0}^{s-1}\{\zeta_r = i_r\}) =$$

$$P(\zeta_0 = i_0)\prod_{s=1}^{t} P(\zeta_s = i_s/\zeta_{s-1} = i_{s-1}) \boxdot$$

36.2. Ймовірності переходу за t кроків

Означення. Ймовірностями переходу за t кроків *називаються умовні ймовірності*

$$p_{ij}^{(t)} = P(\zeta_t = j\ /\zeta_0 = i).$$

Теорема (про ймовірності переходу за t кроків).

(а) Для всіх $t, s \geq 1$ та $i, k \in E$ справедливі рівняння Колмогорова – Чепмена

$$p_{ik}^{(t+s)} = \sum_{j\ \in\ E} p_{ij}^{(t)} p_{jk}^{(s)}.$$

(б) Має місце тотожність

$$P^{(t)} = (p_{ij}^{(t)},\ i, j \in E) = (P)^t,$$

де права частина є ступенем порядку t матриці P.

Доведення. (а) Суму ймовірностей всіх траєкторій 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

$$p_{ik}^{(t+s)} = \sum_{i_1,...,.i_{t-1},\ j,\ j_1,...,.j_{s-1}} p_{ii_1}p_{i_1i_2}...p_{i_{t-1}j}\ p_{jj_1}...p_{j_{s-1}k} =$$

$$\sum_j \left(\sum_{i_1,...,.i_{t-1}} p_{ii_1}p_{i_1i_2}...p_{i_{t-1}j}\right)\left(\sum_{j_1,...,.j_{s-1}} p_{jj_1}...p_{j_{s-1}k}\right).$$

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 $P^{(1)} = P$ ⊡

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.

36.3. Ергодична теорема для ланцюгів Маркова

Теорема (ергодична теорема Колмогорова для ланцюгів Маркова). *Нехай ланцюг Маркова $(\zeta_t,\ t = 0, 1, \dots)$ 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 і 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

$$\pi_j = \sum\nolimits_{i\ \in\ E} \pi_i p_{ij},\ \forall j \in E.$$

Доведення.
Позначимо $m_j^{(s)} = \min_{i\in E} p_{ij}^{(s)} \leq \max_{i\in E} p_{ij}^{(s)} = M_j^{(s)}$.
З рівнянь Колмогорова – Чепмена при $t = 1$ виводимо, що

$$m_j^{(s)} = \min_{i\ \in\ E} p_{ij}^{(s)} = \min_{i\ \in\ E} \sum\nolimits_k p_{ik}p_{kj}^{(s-1)} \geq \min_{i\ \in\ E} \sum\nolimits_k p_{ik}m_j^{(s-1)} = m_j^{(s-1)},$$

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:

$$p_{ij}^{(t)} - p_{kj}^{(t)} = \sum_{h\in E}(p_{ih}^{(r)} - p_{kh}^{(r)})\, p_{hj}^{(t-r)} =$$

$$\sum_{h\in E}(p_{ih}^{(r)} - p_{kh}^{(r)})\, (p_{hj}^{(t-r)} - m_j^{(t-r)}) =$$

$$\sum_{h\in E}(p_{ih}^{(r)} - \varepsilon\delta_{hl} - p_{kh}^{(r)} + \varepsilon\delta_{hl})\, (p_{hj}^{(t-r)} - m_j^{(t-r)}) =$$

$$\sum_{h\in E}(q_{ih} - q_{kh})\, (p_{hj}^{(t-r)} - m_j^{(t-r)}) \leq \sum_{h\in E} q_{ih}\, (p_{hj}^{(t-r)} - m_j^{(t-r)}) \leq$$

$$\sum_{h\in E} q_{ih}\, (M_j^{(t-r)} - m_j^{(t-r)}) = (1-\varepsilon)(M_j^{(t-r)} - m_j^{(t-r)}).$$

Перейдемо в цій нерівності до верхньої межі за i, k, та знайдемо верхню оцінку

$$M_j^{(t)} - m_j^{(t)} \leq (1-\varepsilon)(M_j^{(t-r)} - m_j^{(t-r)}).$$

Аналогічно виводиться оцінка знизу

$$M_j^{(t)} - m_j^{(t)} \geq -(1-\varepsilon)(M_j^{(t-r)} - m_j^{(t-r)}).$$

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. За цією ж монотонністю $\pi_j \in [m_j^{(t)},\ M_j^{(t)}]$, тому з означення $m_j^{(t)} \le p_{ij}^{(t)} \le M_j^{(t)}$ виводимо, що

$$\left|p_{ij}^{(t)} - \pi_j\right| \le \left|M_j^{(t)} - m_j^{(t)}\right| \le (1-\varepsilon)^{[t\,/\,r]},$$

звідки $\lim_{t\to\infty} p_{ij}^{(t)} = \pi_j$.

Оскільки $\sum_{j\le N} \pi_j = \lim_{t\to\infty} \sum_{j\le N} p_{ij}^{(t)} \le 1$ для довільного N, то $\sum_{j\in E} \pi_j \le 1$.

Перейдемо до границі при $t \to \infty$ в нерівності

$$p_{ik}^{(t+1)} = \sum_{j\in E} p_{ij}^{(t)}\ p_{jk} \ge \sum_{j\le N} p_{ij}^{(t)}\ p_{jk},$$

та оберемо $N \to \infty$. Отримаємо нерівності $\pi_k \ge \sum_{j\in E} \pi_j\ p_{jk}$. Якби хоча б одна з цих нерівностей була строгою, то їх сума теж була б строгою нерівністю

$$\sum_{k\in E} \pi_k > \sum_{k\in E} \sum_{j\in E} \pi_j\ p_{jk} = \sum_{j\in E} \pi_j \sum_{k\in E}\ p_{jk} = \sum_{j\in E} \pi_j.$$

Звідси від супротивного дістанемо шукану систему рівнянь для π_j $\boxdot$

37. Процес Пуассона

Вище ми розглядали лише послідовності випадкових величин, що спостерігались у дискретні моменти 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 *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 в теорії ймовірностей, аргумент ω у позначенні процесу часто не вказується.

Звичайно опис випадкового процесу починається із 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) $\nu(0) = 0$,

(б) є кусково-сталим і неперервним зліва за t *при кожній* 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:* $0 < \tau_1 < \tau_2 < \ldots < \tau_n < \ldots,\ \lim \tau_n = \infty$ 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 відповід-*

ність задається формулами

$$\nu(t) = \sum_{n\geq 1} 1_{\{\tau_n < t\}} = |\{n \geq 1 : \tau_n < t\}|,$$

$$\tau_n = \inf(t > 0 : \nu(t) = n).$$

Доведення для скінчених послідовностей очевидне, для зліченних отримується граничним переходом зі скінчених $\boxdot$

Означення. *Нехай $\xi(t)$ – випадковий процес. Його* приростом на $[s,t)$ *називається випадкова величина*

$$\xi[s,t) \equiv \xi(t) - \xi(s).$$

Означення. *Випадковий процес $\xi(t)$ має* незалежні прирости, *якщо для довільних $0 \leq t_1 < t_2 < ... < t_n$ 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 $t \in \mathbb{R}_+$.*

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.

37.1. Процес Пуассона та його розподіл

Означення. *Випадковий процес $\nu(t)$, $t \in \mathbb{R}_+$, називається* процесом Пуассона, *якщо він*

(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 $t \geq 0$, $n \geq 0$*

$$P(\nu(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t},$$

$$M\nu(t) = \lambda t, \ \lambda = \lim_{h \to 0} \frac{1}{h} \ P(\nu(h) = 1).$$

Означення. *Параметр λ називається* інтенсивністю *процесу Пуассона.*

Зауваження. Твердження теореми залишаться справедливими, якщо замість умов (б),(в) на траєкторію в означенні лічильного процесу виконується така аналітична умова ординарності:

$$\text{(о)} \quad P(\nu(h) > 1) = o(h), \ h \to 0.$$

Доведення. Позначимо

$$p_n(t) = P(\nu(t) = n).$$

Нехай $t, s > 0$. Оскільки величини $\nu(t) = \nu[0, t)$ та $\nu[t, t+s)$ незалежні, а процес – лічильний, то

$$p_0(t+s) = P(\nu(t+s) = 0) =$$
$$= P(\nu(t) = 0, \ \nu[t, t+s) = 0) = P(\nu(t) = 0) \ P(\nu[t, t+s) = 0) =$$
$$= P(\nu(t) = 0) \ P(\nu[0, 0+s) = 0) = p_0(t)p_0(s),$$

де використана також однорідність приростів.

Отже, функція $p_0(t)$ мультиплікативна. З мультиплікативності 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, отже $p_0(1) < 1$.Якщо $p_0(1) = 0$, то $p_0(t) = 0$ при всіх t, і $P(\nu(h) \geq 1) = 1\ \forall h > 0$. У цьому випадку з незалежності та однорідності приростів отримуємо $P(\nu(1) \geq n) = 1$ для всіх n, що суперечить скінченності процесу.

Отже, стала $\lambda \equiv -\ln p_0(1)$ скінченна та додатна. Оскільки $p_0(r) = \exp(-\lambda r)$ для всіх раціональних r, 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. Переходячи до границі $r \downarrow t$, дістанемо $p_0(t) = \exp(-\lambda t)$ для всіх $t \geq 0$.

Доведемо, що з умов (б),(в) в означенні лічильного процесу випливає умова ординарності (о) зауваження. За доведеним вище $P(\nu(h) > 0) = 1 - p_0(h) = 1 - \exp(-\lambda h) \leq \lambda h$. Оскільки $\nu(h) = \sum_{k=0}^{2^n-1} \nu_k(h)$, де доданки $\nu_k(h) = \nu[kh2^{-n}, (k+1)h2^{-n})$ незалежні в сукупності, однаково розподілені і за умовою (в) не перевищують 1, починаючи з деякого n, то при виконанні умови $\nu(h) > 1$ принаймні 2 із них будуть додатними,

$$P(\nu(h) > 1) \leq \lim_{n\to\infty} P(\cup_{i\,<\,j}\{\nu_i(h) > 0,\ \nu_j(h) > 0\}) \leq$$

$$\lim_{n\to\infty} 2^n(2^n-1)\left(P(\nu_0(h2^{-n}) > 0)\right)^2 \leq \lambda^2 h^2 = o(h),\ \ h \to 0.$$

Отже, умова (о) виконана.

Враховуючи, що процес $\nu(t)$ – лічильний, запишемо для $n \geq 1$

$$\{\nu(t+h)=n\}=\{\nu(t)=n,\ \nu[t,t+h)=0\}\bigcup$$

$$\{\nu(t)=n-1,\ \nu[t,t+h)=1\}\bigcup\{\nu(t+h)=n,\ \nu[t,t+h)>1\}.$$

Так як 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 (o) ординарності. Після підстановки $p_0(h)=\exp(-\lambda h)=1-\lambda h+o(h)$ в останнє рівняння маємо

$$(p_n(t+h)-p_n(t))/h=-\lambda\, p_n(t)+\lambda\, p_{n-1}(t)+o(1),\ h\to 0.$$

Отже, функції p_n диференційовні та задовольняють рівняння

$$p_n'(t)=-\lambda\, p_n(t)+\lambda\, p_{n-1}(t),\ p_n(0)=0,\ \forall n\geq 1.$$

Використовуючи інтегруючий множник $\exp(\lambda t)$, для функцій $q_n(t)=\exp(\lambda t)p_n(t)$ виводимо 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 $q_n(t)=(\lambda t)^n/n!,\quad p_n(t)=(\lambda t)^n\, e^{-\lambda t}\,/n!$ $\boxdot$

37.2. 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 $\theta_n \equiv \tau_n - \tau_{n-1},\ n \ge 1$, 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.*

Доведення. Для випадкової величини $\theta_1 = \tau_1$ за теоремою про розподіл процесу Пуассона маємо

$$P(\theta_1 > t) = P(\tau_1 > t) = P(\nu(t) = 0) = \exp(-\lambda t),$$

отже, розподіл θ_1 – показниковий із параметром λ.

Для $0 < x < y$ обчислимо з урахуванням незалежності приростів

$$P(x \le \tau_1,\ \tau_2 < y) = P(\nu[0,x) = 0,\ \nu[x,y) \ge 2) =$$
$$P(\nu[0,x) = 0)\ P(\nu[x,y) \ge 2) =$$
$$\exp(-\lambda x)(1 - \exp(-\lambda(y-x)) - \lambda(y-x)\exp(-\lambda(y-x))) =$$
$$\exp(-\lambda x) - \exp(-\lambda y) - \lambda(y-x)\exp(-\lambda y)).$$

Звідси знаходимо сумісну щільність вектора (τ_1, τ_2)

$$f_{(\tau_1,\tau_2)}(x,y) = \frac{\partial}{\partial y}\frac{\partial}{\partial x}P(\tau_1 < x,\ \tau_2 < y) =$$
$$\frac{\partial}{\partial y}\frac{\partial}{\partial x}(P(\tau_2 < y) - P(x \le \tau_1,\ \tau_2 < y)) =$$
$$-\frac{\partial}{\partial y}\frac{\partial}{\partial x}P(x \le \tau_1,\ \tau_2 < y) = \lambda^2\exp(-\lambda y)\ 1_{\ 0 < x < y}.$$

Отже, сумісна характеристична функція інтервалів між стрибками дорівнює

$$\varphi_{(\theta_1,\theta_2)}(t,s) = M\exp(it\theta_1 + is\theta_2) = M\exp(i(t-s)\tau_1 + is\tau_2) =$$
$$\int_0^\infty dy \int_0^y dx \ \exp(i(t-s)x + isy)\lambda^2 \exp(-\lambda y) \ = \frac{\lambda}{\lambda - it} \ \frac{\lambda}{\lambda - is}.$$

Вираз у правій частині є сумісною характеристичною функцією двох незалежних однаково розподілених показникових величин із параметром λ. Тому за теоремою про однозначну відповідність між функціями розподілу та характеристичними функціями величини θ_1, θ_2 є показниковими однаково розподіленими. Доведення для наступних θ_n проводиться аналогічно $\boxdot$

Вправа.

(1) Доведіть, що лічильний процес, утворений стохастичним потоком сум незалежних однаково показниково розподілених випадкових величин, має незалежні та однорідні прирости, тобто є процесом Пуассона. Вказівка: скористайтесь властивістю відсутності післядії показникового розподілу.

(2) Виведіть посилений закон великих чисел: $\nu(t)\ /\ t \to^{P1} \lambda,\ t \to \infty$.

38. Процес народження та загибелі

Процес Пуассона є моделлю процесу чистого народження, якщо миттєвий приріст 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* $(X_t,\ t \in \mathbb{R}_+)$ *із 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*

$$P(X_{t+s} = j\ /X_{t_0} = i_0, ..., X_{t_{n-1}} = i_{n-1},\ X_t = i) =$$

$$P(X_{t+s} = j\ /X_t = i) = P(X_s = j\ /X_0 = i) \equiv p_{ij}(s).$$

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. Ймовірнісні властивості процесу визначаються поведінкою функцій $p_{ij}(s)$.

Означення. *Випадковий процес* $(X_t,\ t \in \mathbb{R}_+)$ *називається* процесом народження та загибелі, *якщо:*

(1) він є однорідним марковським процесом із множиною значень $E = \mathbb{Z}_+$*,*

(2) для деяких невід'ємних чисел λ_i *(*інтенсивностей народжень*) і* μ_i *(*інтенсивностей загибелі*),* $i \geq 0$*, справедливі інфінітезимальні зображення для ймовірностей одиничних стрибків*

$$P(X_h = i + 1\ /\ X_0 = i) = \lambda_i h + o(h),\ h \to 0,\ \forall i \geq 0,$$

$$P(X_h = i-1 \ / \ X_0 = i) = \mu_i h + o(h), \ \ h \to 0, \ \ \forall i > 0.$$

(3) ймовірності інших стрибків є нескінченно малими:

$$P(|X_h - i| > 1 \ / \ X_0 = i) = o(h), \ \ h \to 0$$

Наслідок (про ймовірність невиходу процесу народження та загибелі). *З умов (1)-(3) попереднього означення випливає зображення*

$$P(X_h = i \ / \ X_0 = i) = 1 - a_i h + o(h), \ \ h \to 0, \ \ \forall i \geq 0,$$

де сумарна інтенсивність

$$a_i \equiv \lambda_i + \mu_i,$$

і за означенням $\mu_0 = 0$.

Доведення.

$$P(X_h = i \ / \ X_0 = i) = 1 - P(|X_h - i| > 1 \ / \ X_0 = i) -$$
$$P(X_h = i+1 \ / \ X_0 = i) - P(X_h = i-1 \ / \ X_0 = i) =$$
$$1 - (\lambda_i + \mu_i)h + o(h), \ \ h \to 0.$$

Очевидно, що при $i = 0$ останній доданок у лівій частині останнього рівняння відсутній, тому для справедливості зображення в даному випадку слід покласти $\mu_0 = 0$.

Означення. Часом перебування *процесу в стані* $X_0 = i$ *називається випадкова величина*

$$\tau = \inf(t > 0 : X_t \neq X_0).$$

Теорема (про траєкторії процесу народження та загибелі). *Нехай* $(X_t, \ t \in \mathbb{R}_+)$ *процес народження та загибелі, що має неперервні справа траєкторії, а* τ *– час перебування в стані* $X_0 = i$*. Тоді для всіх* $i \geq 0, \ j \neq i$

$$P(\tau < t \;/\; X_0 = i) = 1 - \exp(-a_i t),$$

$$P(\tau < t,\; X_\tau = j \;/\; X_0 = i) = (1 - \exp(-a_i t))\; p_{ij},$$

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 $X_0 = i$ 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 $(\zeta_n,\; n \geq 0)$ ланцюг Маркова з матрицею перехідних імовірностей за один крок $(p_{ij},\; i, j \geq 0)$, випадкові величини $(\theta_n, n \geq 0)$ при фіксованих $(\zeta_n,\; n \geq 0)$ незалежні в сукупності і мають показниковий розподіл із параметром a_{ζ_n} відповідно, причому $\tau_n = \theta_0 + ... + \;\theta_n$. Тоді процес

$$X_t = \sum\nolimits_{n \geq 0} \zeta_n \; 1_{\tau_{n-1} \;\leq\; t \;<\; \tau_n},$$

є процесом народження та загибелі, що має вказані в означенні інфінітезимальні зображення.

Доведення теореми.

Позначимо

$$t_{nk} = k2^{-n},\;\; \tau_n = \inf\;\left(t_{nk} > 0 : X_{t_{nk}} \neq X_0\right).$$

Оскільки $\{t_{nk}\} \subset \{t_{n+1,k}\}$, то послідовність τ_n не зростає. Траєкторії процесу X_t набувають цілих значень та неперервні справа. Тому знайдеться м.н. додатне $\varepsilon(\omega) > 0$ таке, що $X_{\tau+s} = X_\tau$ для всіх $s \in [0, \varepsilon)$. З неперервності траєкторій процесу справа виводимо, що з імовірністю 1 $X_{\tau_n} = X_\tau$ починаючи з деякого номера, отже $\tau_n \downarrow \tau$, $X_{\tau_n} \to X_\tau$, $n \to \infty$. Тому за неперервністю ймовірності

$$P(\tau < t,\ X_\tau = j \ / \ X_0 = i) = \lim{}_{n\to\infty} P(\tau_n < t,\ X_{\tau_n} = j \ / \ X_0 = i).$$

Як і вище, позначимо

$$p_{ij}(t) = P(X_t = j \ / \ X_0 = i).$$

Нехай

$$A^i_{nk} = \{X_0 = i,\ X_{t_{n1}} = i, ...,\ X_{t_{nk}} = i\},$$

$$k_n(t) = [t\ 2^n] = \sup(k : t_{nk} < t).$$

Для часу перебування обчислимо за формулою про ймовірність перерізу n подій та за марковською властивістю процесу

$$P(\tau \ge t \ / \ X_0 = i) = \lim{}_{n\to\infty} P(\tau_n \ge t \ / \ X_0 = i) =$$

$$\lim{}_{n\to\infty} P(A^i_{n,\ k_n(t)} \ / \ X_0 = i) = \lim{}_{n\to\infty} \prod_{l=1}^{k_n(t)} P(X_{t_{nl}} = i \ / \ A^i_{n,\ l-1}) =$$

$$\lim{}_{n\to\infty} \prod_{l=1}^{k_n(t)} P(X_{t_{nl}} = i / X_{t_{n,l-1}} = i) = \lim{}_{n\to\infty} (p_{ii}(2^{-n}))^{k_n(t)} =$$

$$\lim{}_{n\to\infty} (1 - a_i 2^{-n} + o(2^{-n}))^{t2^{-n}} = \exp(-t\ a_i),$$

де використаний наслідок про ймовірність невиходу процесу народження та загибелі. Перше твердження доведене.

Якщо ж $j \ne i$, то ймовірність виходу дорівнює

$$P(\tau_n < t,\ X_{\tau_n} = j \ / \ X_0 = i) = \sum_{k\ =1}^{k_n(t)} P(\tau_n = t_{nk},\ X_{\tau_n} = j \ / \ X_0 = i) =$$
$$\sum_{k\ =1}^{k_n(t)} P(A^i_{n,k-1} \cap \{X_{t_{nk}} = j\} \ / \ X_0 = i) =$$
$$\sum_{k\ =1}^{k_n(t)} P(X_{t_{nk}} = j / A^i_{n,k-1}) \prod_{l=1}^{k-1} P(X_{t_{nl}} = i / A^i_{n,l-1}) =$$
$$\sum_{k\ =1}^{k_n(t)} P(X_{t_{nk}} = j / X_{t_{n,k-1}} = i) \prod_{l=1}^{k-1} P(X_{t_{nl}} = i / X_{t_{n,l-1}} = i) =$$
$$\sum_{k\ =1}^{k_n(t)} p_{ij}(2^{-n})(p_{ii}(2^{-n}))^{k-1} = p_{ij}(2^{-n}) \left(1 - (p_{ii}(2^{-n}))^{k_n(t)}\right) \ / \ (1 - p_{ii}(2^{-n})).$$

Границю другого множника в правій частині обчислено вище:

$$\lim_{\ n\to\infty} (1 - (p_{ii}(2^{-n}))^{k_n(t)}) = 1 - \exp(-t\ a_i).$$

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 $\boxdot$

Наступна теорема дає можливість обчислити ймовірності переходу через інфінітезимальні характеристики λ_i, μ_i. Зауважимо, що для будь-якого початкового розподілу $q_k = P(X_0 = k)$ безумовний розподіл процесу визначається за формулою повної ймовірності

$$P(X_t = j) = \sum_{\ k\geq 0} P(X_0 = k) P(X_t = j / X_0 = k) = \sum_{\ k\geq 0} q_k\ p_{kj}(t).$$

Теорема (про систему диференціальних рівнянь Колмогорова). *Нехай* $(X_t,\ t \in \mathbb{R}_+)$ *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 $p_k(t) = P(X_t = k)$ при довільних початкових умовах є диференційовною функцією часу t та задовольняє при $k \geq 0$ зворотну системи 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.

Для всіх j, k при $h > 0$ за означенням умовної ймовірності

$$P(X_t = j,\ X_{t+h} = k) = p_j(t)P(X_{t+h} = k/X_t = j) = p_j(t)p_{jk}(h).$$

Тому за формулою повної ймовірності

$$p_k(t+h) = P(X_{t+h} = k) =$$
$$P(X_t = k-1,\ X_{t+h} = k) + P(X_t = k+1,\ X_{t+h} = k) +$$
$$P(X_t = k,\ X_{t+h} = k) + P(|X_t - k| > 1,\ X_{t+h} = k) =$$
$$p_{k-1}(t)\ p_{k-1,k}(h) + p_{k+1}(t)\ p_{k+1,k}(h) + p_k(t)\ p_{kk}(h) + \sum_{j:\ |j-k|>1} p_j(t)p_{jk}(h).$$

За умовою

$$\sup{}_{j,k:|j-k|>1}\, p_{jk}(h) \leq \sup{}_j\, P(|X_h - j| > 1/X_0 = j) = o(h), h \to 0.$$

Тому останній доданок у попередній сумі дорівнює $o(h)$.

Підставимо в перші два доданки зображення з умови (2), а в третій – зображення з 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 $h \to 0$, 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 $\lambda_{-1} = 0, \mu_0 = 0$.

Означення. *Початковий розподіл* $q_k = P(X_0 = k)$ *називається* стаціонарним розподілом *процесу, якщо* $P(X_t = k) = q_k$ *для всіх* $t \geq 0$, $k \geq 0$.

Теорема (про рівняння для стаціонарних імовірностей). *Нехай* $(X_t,\, t \in \mathbb{R}_+)$ *є процесом народження та загибелі з інфінітезимальними характеристиками* (λ_i, μ_i), *причому зображення в умові (3) є рівномірним за* i.

Процес $(X_t,\ t \in \mathbb{R}_+)$ *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 ймовірностей. 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.

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

$$\delta_k = \pi_{k+1} \ \mu_{k+1} - \pi_k \lambda_k, \ \ k \geq 0.$$

Тоді з урахуванням крайових умов при $k = 0$ система рівнянь із теореми про рівняння для стаціонарних імовірностей еквівалентна системі $\delta_0 = 0, \ \delta_{k+1} = \delta_k, \ k \geq 0$.

Отже , з $\pi_k = \frac{\lambda_{k-1}}{\mu_k}\pi_{k-1}$ 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.

Теорема (**про ергодичність процесу народження та загибелі**). *За умови існування стаціонарного розподілу* $(\pi_k, k \geq 0)$ *процесу народження та загибелі для всіх* $i, k \geq 0$ *існує границя*

$$\lim\nolimits_{t\to\infty} \ p_{ik}(t) = \pi_k.$$

Доведення не наводиться.

39. Складний процес Пуассона

На відміну від звичайного, прирости складного процесу Пуассона не обов'язково є одиничними.

Означення. *Нехай* $(\xi_n,\ n \ge 1)$ *послідовність незалежних у сукупності однаково розподілених випадкових величин, а* $\nu(t)$ *– незалежний від них процес Пуассона з параметром* λ. Складним процесом Пуассона, *що породжується процесом* ν *та стрибками* ξ_n, *називається випадковий процес*

$$\xi(t) \equiv \sum\nolimits_{n=1}^{\nu(t)} \xi_n = \sum\nolimits_{n\ge 1} \xi_n \ 1_{n\ \le\ \nu(t)}, \quad t \ge 0.$$

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 об'ємів $(\xi_n, n \ge 1)$ 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* $u \in \mathbb{R}$

$$M \exp(iu\xi(t)) = \exp(\lambda t(\varphi(u) - 1)),$$

де $\varphi(u) = M\exp(iu\xi_1)$ *характеристична функція стрибка.*

Доведення. Незалежність та однорідність приростів є наслідком відповідної властивості процесу Пуассона, оскільки приріст процесу має вигляд

$$\xi[s,t) = \sum\nolimits_{n\ge 1}^{\nu[s,t)} \xi_{\nu(s)\ +\ n},\ t \ge s \ge 0,$$

де величини $\nu(s), \nu[s,\ t), \xi_n$ незалежні в сукупності.

Вираз для характеристичної функції виводимо із теореми про властивості генератрис, пункт (e), при $z = \exp(iu)$:

$$M\exp(iu\xi(t)) = \varphi_{\nu(t)}(M\exp(iu\xi_1)) = \varphi_{\nu(t)}(\varphi(u)) =$$

$$\sum\nolimits_{n\ge 0} (\varphi(u))^n \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \exp(\lambda t\varphi(u) - \lambda t),$$

де використана також теорема про розподіл процесу Пуассона $\boxdot$

Вправа. Знайти розподіл складного процесу Пуассона, якщо доданки ξ_n мають показниковий розподіл.

За допомогою граничного переходу зі складного процесу Пуассона можна отримати неперервний процес із незалежними та однорідними приростами. Для цього розглянемо процес із малими симетричними приростами: $P(\xi_n = \pm\varepsilon) = 1/2,\ \varepsilon \to 0$, 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

$$M\exp(iu\xi(t)) = \exp(\lambda t(\cos(\varepsilon u) - 1)).$$

Для існування її границі при $\varepsilon \to 0$, $\lambda \to \infty$ оберемо $\lambda \approx \varepsilon^{-2}$. Тоді

$$\lim\nolimits_{\varepsilon\to 0,\ \lambda\to\infty} M\exp(iu\xi(t)) = \exp(-u^2 t/2),$$

тобто слабка границя $\xi(t)$ повинна мати нормальний розподіл $N(0,t)$.

40. 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 середньоквадратичного відхилення частинок від початкового положення

Означення. Вінерівським, *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) $w(0) = 0$, $Mw(t) = 0$, $Dw(t) < \infty$.

40.1. Розподіл вінерівського процесу

Теорема (про розподіл вінерівського процесу). *Нехай* $w(t)$ *– вінерівський процес. Тоді існує стала* $\sigma \ge 0$, *така, що для всіх* $t > 0$

$$w(t) \simeq N(0, \sigma^2 t),\ Mw(t) = 0,\ Dw(t) = \sigma^2 t.$$

Зауваження. При виконанні решти припущень умову (а) неперервності траєкторій в означенні вінерівського процесу можна замінити на більш просту аналітичну умову

$$P(|w(h)| \ge \varepsilon) = o(h),\ h \to 0,\ \text{для кожного } \varepsilon > 0.$$

Доведення теореми. Нехай t фіксоване. Розглянемо прирости

$$w_{nk} \equiv w[t(k-1)/n,\ tk/n).$$

За умови (б) випадкові величини $(w_{nk,}\ k = \overline{1,n})$ незалежні в сукупності та однаково розподілені, $w_{n1} = w(t/n)$.

Лема 1. За припущень (б), (в) неперервність траєкторій процесу $w(t)$ еквівалентна умові, що сформульована в зауваженні.

Доведення. Неперервність траєкторій процесу на компактному інтервалі $[0, t]$ еквівалентна рівномірній неперервності, яка, в свою чергу, еквівалентна збіжності до нуля модуля неперервності: $\max_{1\le k\le n} |w_{nk}| \to^{P} 0$ при $n \to \infty$, тобто

$$P(\max{}_{1\le k\le n} |w_{nk}| \ge \varepsilon) \to 0,\ n \to \infty,\ \forall\varepsilon > 0.$$

Згідно з незалежністю та однорідністю приростів

$$P(\max_{1\le k\le n} |w_{nk}| < \varepsilon) = \exp(-n\ \ln P(|w_{n1}| < \varepsilon)).$$

Тому збіжність до нуля модуля неперервності еквівалентна співвідношенню $\ln P(|w_{n1}| < \varepsilon) = o(1/n),\ n \to \infty$, що внаслідок зображення $\ln(1 + \varepsilon) = \varepsilon + o(\varepsilon), \varepsilon \to 0$, еквівалентно співвідношенню:

$$P(|w_{n1}| \ge \varepsilon) \equiv P(|w(t/n)| \ge \varepsilon) = o(1/n),\ n \to \infty\boxdot$$

Лема 2. Існують $\varepsilon_n \to 0$ такі, що $nP(|w_{n1}| \ge \varepsilon_n) \to 0,\ n \to \infty$.

Доведення. Згідно з лемою 1 для кожного $\varepsilon > 0$ справедливе співвідношення

$$P(|w_{n1}| \ge \varepsilon) = o(1/n),\ n \to \infty.$$

Виходячи з нього, побудуємо неспадну послідовність $n_k \ge k$ таку, що

$$nP(|w_{n1}| \ge 1/k) \le 1/k$$

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 $n_k \leq n < n_{k+1}$

$$nP(|w_{n1}| \geq \varepsilon_n) = nP(|w_{n1}| \geq 1/k) \leq 1/k \to 0,\ n \to \infty \boxdot$$

Лема 3. За умов теореми
(а) Функція $\sigma^2(t) = Dw(t)$ 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

$$w(t+s) = w(t) + w[t, t+s),$$

де величини в правій частині незалежні. Тому адитивність випливає з теореми про дисперсію суми незалежних величин та з однорідності приростів. Монотонність є наслідком невід'ємності дисперсії

$$\sigma^2(t+s) = \sigma^2(t) + \sigma^2(s) \geq \sigma^2(t).$$

(б) Якби монотонна границя $\lim_{h \to 0} \sigma^2(h) \equiv \delta > 0$, то з монотонності випливало б, що $\sigma^2(h) \geq \delta$ при всіх $h > 0$ та $\sigma^2(1) = n\sigma^2(1/n) \geq n\delta \to \infty$, що суперечить скінченності дисперсії (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 $\sigma^2 = \sigma^2(1)$. З адитивності та нормованості виводимо, що

$\sigma^2(m/n) = m/n\ Dw(1) = (m/n)\sigma^2$ для натуральних m, n.Переходячи тут до границі $m/n \uparrow t$, з неперервності дістанемо шукану рівність $\boxdot$

Можна вважати, що $\sigma^2 > 0$ – інакше процес є тотожнім нулем майже напевне.

Позначимо

$$\xi_{nk} = w_{nk} 1_{|w_{nk}| \le \varepsilon_n}.$$

Випадкові величини $(\xi_{nk},\ k = \overline{1,n})$ незалежні за теоремою про перетворення незалежних величин, однаково розподілені та $|\xi_{nk}| \le \varepsilon_n$. Розглянемо суми

$$\xi_n = \sum\nolimits_{k=1}^{n} \xi_{nk}.$$

З нерівності

$$P(\xi_n \ne w(t)) \le P(\cup_{k=1}^{n} \{\xi_{nk} \ne w_{nk}\}) \le$$

$$\sum\nolimits_{k=1}^{n} P(|w_{nk}| > \varepsilon_n) = nP(|w_{n1}| > \varepsilon_n) \to 0,\ n \to \infty,$$

та леми 2 виводимо, що $\xi_n \to^P w(t),\ n \to \infty$.

Позначимо $\sigma_n^2 = D\xi_n$. Ця дисперсія існує та обмежена, оскільки

$$D\xi_n = \sum\nolimits_{k=1}^{n} D\xi_{nk} \le \sum\nolimits_{k=1}^{n} M\xi_{nk}^2 \le \sum\nolimits_{k=1}^{n} Mw_{nk}^2 = \sigma^2 t,$$

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 $\xi_n - M\xi_n \to^P 0$ по деякій підпослідовності, отже $w(t) - M\xi_n \to^P 0$. Оскільки величина $w(t)$ не залежить від n, то числова послідовність $M\xi_n$ повинна збігатися, тому $w(t)$ є сталою майже напевне. Це суперечить невиродженості $w(t)$ – як показано вище, $Dw(t) = \sigma^2 t > 0$.

(б) Отже, $\underline{\lim}_{\ n\to\infty}\ \sigma_n^2 > 0$, звідки $\sigma_n^2 \ge \theta^2 > 0$, починаючи з деякого номера. Враховуючи однакову розподіленість величин ξ_{nk}, підрахуємо показник з умови Ліндеберга для загальної послідовності серій $(\xi_{nk},$ $k=\overline{1,n},\ n\ge 1)$:

$$L_n(\varepsilon) = \frac{n}{\sigma_n^2} M(\xi_{n1} - M\xi_{n1})^2 1_{|\xi_{n1} - M\xi_{n1}|\,\ge\,\varepsilon\sigma_n} \le \frac{n}{\theta^2} 4\varepsilon_n^2 1_{2\varepsilon_n\,\ge\,\varepsilon\theta} = 0,$$

починаючи з деякого номера, оскільки $\varepsilon_n \to 0$. Аналогічно перевіряється умова рівномірної малості в означенні загальних послідовностей серій.

Отже, внаслідок центральної граничної теореми Ліндеберга для загальних серій має місце слабка збіжність $(\xi_n - \mu_n)/\sigma_n \to^W\ \zeta \simeq N(0,1)$, де $\mu_n = M\xi_n$. Обираючи підпослідовність номерів, для якої $\sigma_n \to b > 0$, за теоремою про добуток слабко збіжної послідовності зі збіжною дістанемо $\xi_n - \mu_n \to^W\ b\zeta$. Звідси за теоремою Леві про критерій слабкої збіжності виводимо існування границі характеристичних функцій

$$\lim_{n\to\infty} M\exp(is(\xi_n - \mu_n)) \ = \lim_{n\to\infty} \exp(-is\mu_n) M\exp(is\xi_n).$$

З іншого боку, границя

$$\lim_{n\to\infty} M\exp(is\xi_n) \ = M\exp(isw(t))$$

існує, оскільки $\xi_n \to^P\ w(t)$, як показано вище. Отже, існує границя $\lim_{n\to\infty}\mu_n = \mu$. Тому

$$M\exp(isw(t)) \ = \lim_{n\to\infty} M\exp(is\mu_n)\ \lim_{n\to\infty} M\exp(is(\xi_n - \mu_n))$$
$$= \exp(is\mu)\varphi_{b\,\zeta}(s) \ = \exp(is\mu - s^2b^2/2),$$

тобто $w(t) \simeq N(\mu, b^2)$. Оскільки $Mw(t) = 0, Dw(t) = \sigma^2 t$, то $w(t) \simeq N(0,\sigma^2 t)$ за теоремою про інтерпретацію параметрів нормального розподілу $\boxdot$

40.2. Властивості траєкторій вінерівського процесу

Теорема (про властивості траєкторій вінерівського процесу).
(а) $P(\,\exists\, w'(0)) = 0$,
(б) $Var_{[0,1]}w \equiv \lim\ _{n\to\infty} \sum_{k=1}^{2^n} \left| w[\frac{k-1}{2^n}, \frac{k}{2^n}) \right| = \infty$ *майже напевне*,
(в) $\sum_{k=1}^{2^n} \left(w[\frac{k-1}{2^n}, \frac{k}{2^n}) \right)^2 \to^P \sigma^2$.

Доведення. Вважатимемо, що $\sigma^2 = 1$.

(а) Позначимо $\zeta_n \equiv (w(2^{-n+1}) - w(2^{-n}))/2^{-n}$. Випадкові величини ζ_n незалежні за властивістю незалежності приростів. Оскільки за однорідністю приростів $\zeta_n \simeq w(2^{-n})/2^{-n} \simeq N(0, 2^{-n}/2^{-2n})$, то ряд

$$\sum_{n\ge1} P(|\zeta_n| > m) = \sum_{n\ge1} 2(1 - \Phi(m2^{-n/2})) = \infty$$

розбігається для кожного $m \ge 0$. За лемою Бореля – Кантеллі, пункт (б), $P(\overline{\lim}_{n\to\infty}\{|\zeta_n| > m\}) = 1,\ \forall m$. Отже, $\overline{\lim}\,|\zeta_n| = \infty$, і границя $\lim_{n\to\infty}\ \zeta_n$ не існує майже напевне. Оскільки

$$\zeta_n = 2w(2^{-n+1})/2^{-n+1} - w(2^{-n})/2^{-n}, \text{ то}$$
$$\{\exists\ \lim_{n\to\infty}\ \zeta_n\} \supset \{\exists\ \lim_{h\to0}\ w(h)/h\} = \{\exists\ w'(0)\},$$

і ймовірність останньої події дорівнює нулю.

(б) Позначимо

$$w_{nk} = w[(k-1)2^{-n}, k2^{-n}), \quad \zeta_n = \sum_{k=1}^{2^n} |w_{nk}|\,.$$

Тоді $\zeta_n \uparrow \zeta \le Var_{[0,1]}w \le \infty$. Крім того, $w_{nk} \simeq N(0, 2^{-n})$,тому випадкова величина $\eta \equiv 2^{n/2}w_{n1} \simeq N(0,1)$. Оскільки величини $(w_{nk}, k = 1, 2^n)$ незалежні в сукупності та однаково розподілені, то за теоремами про перетворення незалежних величин та про математичне сподівання добутку незалежних величин

$$M\exp(-s\zeta_n) = (M\exp(-s\,|w_{n1}|))^{2^n} =$$
$$(M\exp(-s2^{-n/2}\left|2^{n/2}w_{n1}\right|))^{2^n} = (M\exp(-s2^{-n/2}\,|\eta|)^{2^n} =$$
$$(1 - s2^{-n/2}M\,|\eta| + o(2^{-n/2}))^{2^n} \to 0\ \forall s > 0.$$

Тому $P(\zeta_n < \infty) \downarrow P(\zeta < \infty) = 0.$

(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

$$P(\left|\sum_{k=1}^{2^n}(w_{nk})^2 - 1\right| \geq \varepsilon) = P(\left|2^{-n}\sum_{k=1}^{2^n}(2^{n/2}w_{nk})^2 - 1\right| \geq \varepsilon)$$
$$= P(\left|2^{-n}\sum_{k=1}^{2^n}\eta_k^2 - 1\right| \geq \varepsilon) \to 0$$

за теоремою Чебишева про закон великих чисел, оскільки η_k^2 незалежні однаково розподілені, $M\eta_k^2 = 1$ $\boxdot$

Розділ 2

Математична статистика

Вступ

Первісна задача математичної статистики (МС) 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.

Умовно кожен висновок про ймовірнісний простір можна віднести до однієї з двох груп:

(1) висновок про *кількісне* значення деякої величини (параметра),

(2) *якісний* висновок про певний параметр чи властивість імовірнісного простору.

Відповідно всі задачі кожного розділу МС можна розбити на

(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* – як невідомі "параметри" виступають загальні функції розподілу;

(б) *теорія оптимальних незміщених оцінок*, де в класі незміщених оцінок дисперсія розглядається як міра якості і знаходиться вигляд оптимальних оцінок;

(в) *теорія оцінок максимальної вірогідності*, де пропонується універсальний метод для побудови досить якісних статистичних оцінок параметрів;

(г) *статистичні висновки для нормальних спостережень*, що грунтуються на спеціальних властивостях вибіркових статистик;

(д) *регресійний аналіз*, в якому вивчаються задачі встановлення функціональної залежності між числовими змінними в умовах наявності стохастичних похибок;

(е) *факторний 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 суттєво 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.

1. Статистичний простір, вибірка. Статистики та оцінки

Означення. Статистичним простором *називається трійка*

$$(\Omega,\ \mathcal{F},\ (P_\theta,\ \theta \in \Theta)),$$

що складається з таких елементів:
– Ω – *простір елементарних подій деякого стохастичного експерименту;*
– $\mathcal{F} \subset 2^\Omega$ – *сигма-алгебра випадкових подій – підмножин* Ω;
– $(P_\theta,\ \theta \in \Theta)$ – *деяка параметрична сім'я ймовірностей на* $\mathcal{F}$;
– Θ – параметричний простір – *множина довільної природи.*

Вважається, що вид залежності ймовірностей $P_\theta(A)$ від подій $A \in \mathcal{F}$ 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 $\mathcal{F}$ 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* $(\Omega,\ \mathcal{F},\ P_\theta)$, *символом*

$$M_\theta \xi = \int_\Omega \xi(\omega) P_\theta(d\omega)$$

будемо позначати математичне сподівання ξ *(абстрактний інтеграл Лебега) за ймовірністю* P_θ.

Статистичні висновки про ймовірність P_θ будемо робити на підставі спостережень, тобто значень певних випадкових величин, векторів та інших функцій від елементарних подій.

1.1. Статистична вибірка

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 (спостерігається) і може використовуватись для отримання статистичних висновків.

Надалі в якості вибіркового простору буде виступати 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 в якості міри λ виступає або точкова міра – відносно неї кожна одноточкова множина 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.

Приклад. Побудова статистичного висновку про симетричність монети за результатами серії з 1000 підкидань. Нехай вибірка містить результати кожного з підкидань, тоді вибірковий простір $S = \{A, P\}^{1000}$. Для параметризації задачі позначимо $\theta \in \Theta = (0, 1)$ ймовірність аверсу при одному підкиданні. Тоді ймовірність P_θ задає розподіл вектора індикаторів успіхів у схемі випробувань Бернуллі з параметрами $n = 1000,\ p = \theta$.

Для подальшого використання нагадаємо таке важливе поняття з теорії міри.

Означення. *Міра* μ *на деякому вимірному просторі* $(S,\ \Sigma)$ абсолютно неперервна *відносно сигма-скінченної 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 розподілу) відносно деякої фіксованої міри λ для кожного значення параметра θ :

$$P_\theta(X \in \cdot) \ll \lambda(\cdot), \ \forall\theta \in \Theta.$$

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

$$\Omega = \{\text{А,Р}\}, \ \ \mathcal{F} = 2^\Omega, \ S = \Omega, \ \Sigma = \mathcal{F}, \ \ \lambda - \text{точкова міра},$$
$$P(\{\text{А}\}) = 1 - P(\{\text{Р}\}) = \theta \in \Theta = [0, 1].$$

1.2. Функція вірогідності

Означення. *Нехай $X : \Omega \to S$ – вибірка зі значеннями у вимірному просторі (S, Σ, λ), яка задовольняє* умову підпорядкованості. Функцією вірогідності *називається щільність розподілу вибірки відносно фіксованої міри* λ *у вибірковому просторі:*

$$L(x,\theta) \equiv \frac{dP_\theta(X \in \cdot)}{d\lambda(\cdot)}(x),$$

тобто така вимірна за x *функція, що для всіх* $B \in \Sigma$ *і всіх* $\theta \in \Theta$

$$P_\theta(X \in B) = \int_B L(x,\theta)\ \lambda(dx).$$

Для абсолютно неперервної вибірки міра λ є мірою Лебега, функція вірогідності як функція x є щільністю розподілу:

$$P_\theta(X \in B) = \int_B L(x,\theta)\ dx,$$

а для дискретної вибірки λ – точкова міра, функція вірогідності є дискретним розподілом імовірностей:

$$P_\theta(X \in B) = \sum\nolimits_{x\ \in\ B\ \cap\ S}\ L(x,\theta).$$

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*

$$L(X,\theta) \equiv L(x,\theta)\ |_{\ x\ =\ X}\ .$$

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* кратною вибіркою, *якщо* $X = (\xi_1, ..., \xi_n)$ *складається з* *незалежних у сукупності однаково розподілених* *величин* ξ_k, $k = \overline{1, n}$, *зі значеннями у просторі* $(R, \mathfrak{B}, \upsilon)$, *тобто*

$$P_\theta(X \in B_1 \times ... \times B_n) = \prod_{k=1}^{n} P_\theta(\xi_1 \in B_k),\ \forall B_k \in \mathfrak{B},\ \theta \in \Theta.$$

Означення. *Число* n *називається* об'ємом вибірки X.

Якщо X – кратна вибірка, її функція вірогідності позначається як $L_n(X, \theta)$, де n – об'єм вибірки.

Означення. Функцією вірогідності спостережень для кратної вибірки називається щільність розподілу спостереження ξ_1 відносно міри υ

$$f(y, \theta) = \frac{dP_\theta(\xi_1 \in \cdot)}{d\upsilon(\cdot)}(y),$$

тобто така вимірна функція f, що

$$P_\theta(\xi_k \in B) = \int_B f(y,\theta)\ \upsilon(dy), \quad \forall B \in \mathfrak{B}, \forall \theta \in \Theta.$$

Теорема (про функцію вірогідності кратної вибірки). *Нехай $X = (\xi_1, ..., \xi_n)$ є кратною вибіркою, що 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

$$L(x,\theta) = \prod\nolimits_{k=1}^{n} f(x_k, \theta), \ x = (x_1, ..., x_n).$$

Доведення є прямим наслідком теореми про критерій незалежності абсолютно неперервних величин, оскільки вибірковий вектор $X = (\xi_1, ..., \xi_n)$ утворений саме незалежними однаково розподіленими величинами $\boxdot$

1.4. Статистики та оцінки

Означення. Статистикою *називається довільна вимірна функція (скалярна чи векторна) від вибірки: $T = T(X)$, яка не містить значень невідомого параметра θ.*

Множина значень статистики є довільним вимірним простором (Υ, Λ). Найчастіше це евклідів простір $(\Upsilon, \Lambda) = (\mathbb{R}^m, \mathfrak{B}(\mathbb{R}^m))$.

Зауваження. Словом "статистика" будемо одночасно визначати як саму функціональну залежність $T(x) : S \to \Upsilon$ від вибірки, так і її значення $T(\omega) = T(X(\omega)) : \Omega \to \Upsilon$, яке отримується після підстановки вибірки

$x = X(\omega)$. Точне тлумачення випливатиме зі змісту відповідного аргумента.

Означення. Оцінкою *невідомого параметра* $\theta \in \Theta$ *називається будь-яка статистика зі значеннями в 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.

Означення. *Нехай* $\Theta \subset \mathbb{R}^d$. Інтервальною оцінкою *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.*

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 як $\widehat{\theta}_n$.

Приклад. *Вибіркове середнє* є класичною оцінкою положення і дорівнює середньому арифметичному спостережень, які утворюють вибірку. Однак фактично це є послідовність оцінок, оскільки при кожному значенні об'єму вибірки маємо окрему статистику.

1.5. Властивості оцінок

Безперечно, якість тієї чи іншої оцінки потребує порівняльного аналізу. Для порівняння оцінок чи їх послідовностей будемо використовувати такі поняття.

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.

Означення. *Оцінка $\widehat{\theta}$ називається* незміщеною, *якщо її математичне сподівання збігається з точним значенням θ :*

$$M_\theta\,\widehat{\theta} = \theta, \quad \forall\theta \in \Theta.$$

Означення. *Оцінка $\widehat{\theta}_n$ називається* асимптотично незміщеною, *якщо має місце асимптотична збіжність середніх*

$$M_\theta\,\widehat{\theta}_n \to \theta,\ n \to \infty, \quad \forall\theta \in \Theta.$$

Означення. *Оцінка $\widehat{\theta}_n$ називається* консистентною, *якщо вона збігається за ймовірністю до істинного значення θ :*

$$\widehat{\theta}_n \to^{P_\theta} \theta,\ n \to \infty, \quad \forall\theta \in \Theta.$$

Означення. *Оцінка $\widehat{\theta}_n$ називається* строго консистентною, *якщо вона збігається з імовірністю 1 до істинного значення θ :*

$$\widehat{\theta}_n \to^{P_\theta 1} \theta,\ n \to \infty, \quad \forall\theta \in \Theta.$$

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 рівня p, *якщо для всіх $\theta \in \Theta$*

$$P_\theta(\widehat{\theta}_1 \le \theta \le \widehat{\theta}_2) = p.$$

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

$$c_n(\widehat{\theta}_n - \theta) \to^{W_\theta} \zeta \simeq N(0,1),\ n \to \infty,\ \forall\theta \in \Theta.$$

Зауваження (про побудову інтервальної оцінки). Якщо оцінка $\widehat{\theta}_n$ є асимптотично нормальною, а $c_n \approx \sqrt{n}/\sigma$, 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. Зі 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 вірогідного рівня $p \in (0,1)$ значення x_p так, щоб

$$P(|\zeta| \le x_p) = p.$$

Для знаходження x_p досить чисельно розв'язати рівняння

$$\Phi(x_p) = (1+p)/2,$$

де Φ – стандартна нормальна функція розподілу. Тоді при великих n наближено

$$P_\theta\left(\sqrt{n}\left|\widehat{\theta}_n - \theta\right| \ / \ \sigma(\theta) \le x_p\right) \approx P(|\zeta| \le x_p) = p.$$

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 $\{\sqrt{n}\left|\widehat{\theta}_n - \theta\right| / \sigma(\widehat{\theta}_n) \le x_p\}$ наближено теж має ймовірність p. Отже, властивості асимптотичної нормальності та консистентності дають можливість наближеної побудови інтервальних асимптотично незміщених оцінок рівня p для невідомого параметра, що мають вигляд

$$P_\theta\left(\widehat{\theta}_n - x_p\sigma(\widehat{\theta}_n)/\sqrt{n} \le \theta \le \widehat{\theta}_n + x_p\sigma(\widehat{\theta}_n)/\sqrt{n}\right) \approx p,\ \forall\theta \in \Theta.$$

Означення. *Оцінка* $\widehat{\theta}_n$ *називається* локально незміщеною, локально консистентною *тощо, якщо відповідна властивість виконується не для всіх* $\theta \in \Theta$, *а лише для всіх значень параметра* θ *із деякого околу істинного значення.*

2. Оцінювання ймовірності успіху у схемі Бернуллі

Розглянемо статистичний простір $(\Omega,\ \mathcal{F},\ (P_\theta,\ \theta \in \Theta))$, в якому

$$\Omega = \{\omega = (\omega_1, ..., \omega_n),\ \omega_k \in \{0,1\}\} = \{0,1\}^n,\ \mathcal{F} = 2^\Omega,$$
$$P_\theta(\{\omega\}) = \theta^{\nu_n(\omega)}(1-\theta)^{n-\nu_n(\omega)},$$

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 задачі, в якій проводяться n випробувань Бернуллі з невідомою ймовірністю $p = \theta$ успіху в окремому випробуванні.

Припустимо, що спостерігаються результати всіх випробувань: $(S, \Sigma) = (\Omega, \mathcal{F})$, λ – точкова міра на підмножинах Ω і кратна вибірка має вигляд $X = (\chi_1, ..., \chi_n)$, де $\chi_k(\omega) = 1_{\{\omega_k = 1\}} = \omega_k$ – індикатор успіху в k-му випробуванні.

За умовою випадкові величини $(\chi_1, ..., \chi_n)$ незалежні в сукупності та однаково розподілені,

$$P_\theta(\chi_1 = 1) = \theta,\ P_\theta(\chi_1 = 0) = 1-\theta,\quad M_\theta\, \chi_1 = \theta.$$

Розглянемо оцінку

$$\widehat{\theta}_n = \frac{\nu_n}{n} = \frac{1}{n}\sum\nolimits_{k=1}^n \chi_k,$$

яка збігається з відносною частотою успіхів в елементарній події.

Теорема (про властивості відносної частоти). *Відносна частота успіху у схемі випробувань Бернуллі:*

(1) має біноміальний розподіл:

$$P_\theta(\widehat{\theta}_n = k/n) = C_n^k \theta^k (1-\theta)^{n-k},\ k = \overline{0,n},$$

(2) є незміщеною оцінкою ймовірності успіху θ,

(3) є консистентною і строго консистентною оцінкою,

(4) є асимптотично нормальною з асимптотичною дисперсією $\theta(1-\theta)$.

Доведення.

(1) 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

$$M_\theta \widehat{\theta}_n = \frac{1}{n}\sum\nolimits_{k=1}^{n} M\chi_k = \frac{1}{n}\sum\nolimits_{k=1}^{n} \theta = \theta,\ \forall\theta\in\Theta.$$

(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

$$M_\theta \nu_n = n\theta,\ D_\theta \nu_n = n\theta(1-\theta).$$

Звідси $D_\theta \widehat{\theta}_n = M_\theta(\widehat{\theta}_n - \theta)^2 = n\theta(1-\theta)/n^2 \to 0,\ \ n \to \infty$. Тому $\widehat{\theta}_n \to^{L_2} \theta$, отже, $\widehat{\theta}_n \to^{P_\theta} \theta$. Більше того, за критерієм Колмогорова посиленого закону великих чисел $\widehat{\theta}_n \to^{P_\theta 1} M_\theta\, \chi_1 = \theta$.

(4) На підставі зображення пункту (3) із класичної центральної граничної теореми виводимо, що

$$\sqrt{n}(\widehat{\theta}_n - \theta)/\sqrt{\theta(1-\theta)} =$$
$$(\nu_n - M\nu_n)/\sqrt{D\nu_n} \to^W \zeta \simeq N(0,1),\ \ n \to \infty,$$

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 асимптотичною дисперсією $\theta(1-\theta)$ $\boxdot$

Зауваження. Властивість асимптотичної нормальності 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 $P(|\zeta| \le x_p) = p$, як це зроблено в розділі про властивості оцінок у зауваженні про побудову інтервальних оцінок.

(a) *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 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 інтервальною незміщеною оцінкою рівня p.

(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 інтервальну оцінку рівня p

$$P_\theta\left(\widehat{\theta}_n - x_p\sqrt{\widehat{\theta}_n(1-\widehat{\theta}_n)/n} \le \theta \le \widehat{\theta}_n + x_p\sqrt{\widehat{\theta}_n(1-\widehat{\theta}_n)/n}\right) \approx p, \ \forall\theta\in\Theta.$$

3. Емпірична функція розподілу

Розглянемо статистичний простір, в якому спостерігається кратна вибірка $X = (\xi_1, ..., \xi_n)$ зі спостереженнями ξ_k, $k = \overline{1, n}$, що мають невідому функцію розподілу $F(x) = P(\xi_k < x)$. У цьому випадку параметром θ статистичного простору є довільна функція розподілу F.

Означення. Емпіричною функцією розподілу *називається сім'я статистик*

$$\widehat{F}_n(x) = \frac{1}{n} \sum\nolimits_{k=1}^{n} 1_{\{\xi_k < x\}}\ ,\ x \in \mathbb{R}.$$

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 ξ_k, $k = \overline{1, n}$. При кожній фіксованій елементарній події ω вона є дискретною функцією розподілу.

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:

$$P_\theta(\widehat{F}_n(x) = k/n) = C_n^k F^k(x)(1 - F(x))^{n-k},\ k = \overline{0, n},$$

(2) є незміщеною оцінкою відповідного значення теоретичної функції розподілу:

$$M_\theta\ \widehat{F}_n(x) = F(x),$$

(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 $\theta=P(\xi_k<x)=F(x)$. За означенням величина $\widehat{F}_n(x)$ збігатиметься із 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 $\boxdot$

3.2. Рівномірні за x властивості

Емпірична функція розподілу $\widehat{F}_n(x)$ є випадковим процесом: вона одночасно є функцією елементарної події $\omega\in\Omega$ та числового аргумента $x\in\mathbb{R}$. У теорії випадкових процесів доведені такі теореми.

Теорема (теорема Глівенка – Кантеллі). *Якщо теоретична функція розподілу $F(x)$ неперервна, то*

$$\sup_{x\,\in\,\mathbb{R}}\left|\widehat{F}_n(x)-F(x)\right| \to^P 0,\ n\to\infty.$$

Доведення не наводиться.

Теорема (**теорема Колмогорова про відхилення емпіричної функції розподілу**). *Якщо теоретична функція розподілу $F(x)$ неперервна, то:*

(а) розподіл статистики Колмогорова:

$$\widehat{\varkappa}_n = \sqrt{n}\ \sup_{x\,\in\,\mathbb{R}} \left|\widehat{F}_n(x) - F(x)\right|$$

не залежить від вигляду невідомої функції F,

(б) має місце слабка збіжність

$$\widehat{\varkappa}_n \to^{W} \varkappa,\ n \to \infty,$$

причому розподіл граничної величини має вигляд

$$P_\theta(\varkappa < x) = K(x) = \sum_{k\,=\,-\infty}^{+\infty} (-1)^k \exp(-2k^2x^2),\ \ x > 0.$$

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 $F(F^{(-1)}(u)) = u$.(**Вправа**. Для нестрого монотонних F однозначно визначений неперервний зліва варіант $F^{(-1)}$.)

Зробимо заміну змінної $x = F^{(-1)}(u), 0 < u < 1$, в означенні емпіричної функції розподілу

$$\widehat{F}_n(F^{(-1)}(u)) = \frac{1}{n}\sum_{k=1}^{n} 1_{\{\xi_k\,<\,F^{(-1)}(u)\}} = \frac{1}{n}\sum_{k=1}^{n} 1_{\{F(\xi_k)\,<\,u\}},$$

оскільки за монотонністю та означенням оберненої функції має місце тотожність $\{\xi_k < F^{(-1)}(u)\} = \{F(\xi_k) < u\}$. Випадкові величини $\alpha_k = F(\xi_k)$ незалежні за теоремою про перетворення незалежних величин, та рівномірно розподілені на відрізку $[0,1]$:

$$P_\theta(F(\xi_k) < u) = P_\theta(\xi_k < F^{(-1)}(u)) = F(F^{(-1)}(u)) = u, \ \forall u \in [0,1].$$

Після аналогічної заміни в означенні статистики Колмогорова дістанемо

$$\widehat{\varkappa}_n = \sqrt{n}\ \sup_{u\,\in\,(0,1)} \left|\widehat{F}_n(F^{(-1)}(u)) - F(F^{(-1)}(u))\right| =$$

$$\sqrt{n}\sup_{u\,\in\,(0,1)} \left|\frac{1}{n}\sum_{k=1}^{n}\ 1_{\{\alpha_k\,<\,u\}} - u\right|.$$

Отже, розподіл $\widehat{\varkappa}_n$ однозначно визначається рівномірно розподіленим випадковим вектором $(\alpha_1, ..., \alpha_n)$ і не залежить від F.

(б) Обчислення граничного розподілу статистики Колмогорова спирається на центральну граничну теорему теорії випадкових процесів – сучасного розділу теорії ймовірностей і виходить за межі даного курсу $\boxdot$

Теорема Колмогорова дає можливість обчислити наближену інтервальну незміщену оцінку рівня p для функції розподілу

$$P_\theta(\{\widehat{F}_n(x) - x_p\,/\sqrt{n} \le F(x) \le \widehat{F}_n(x) + x_p\,/\sqrt{n},\ \forall x \in \mathbb{R}\}) \approx p,$$

де x_p – розв'язок рівняння $K(x_p) = p$.

Для більш точного оцінювання замість K слід використати точний вираз для функції розподілу статистики $\widehat{\varkappa}_n$. Він був знайдений В. С. Королюком.

Вправа. Знайти розподіл статистик $\widehat{\varkappa}_1, \widehat{\varkappa}_2$.

4. Варіаційний ряд. Квантилі

У процесі побудови графіка емпіричної функції розподілу ми насамперед маємо впорядкувати точки її стрибків – вибіркові значення $(\xi_k,\ k = \overline{1,n})$.

Означення. Варіаційним рядом *вибірки* $X = (\xi_1, ..., \xi_n)$ *називаються вибіркові значення* $\{\xi_{(k)},\ k = \overline{1,n}\} = \{\xi_k,\ k = \overline{1,n}\}$, *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 $\{2, 3, 5\}$.

Зауваження. Якщо вибірка складається з незалежних у сукупності однаково розподілених величин, які мають неперервну функцію розподілу F, то з імовірністю 1 всі нерівності в означенні варіаційного ряду є строгими:

$$\xi_{(1)} < \xi_{(2)} < ... < \xi_{(n)}.$$

Дійсно, за напівадитивністю ймовірності

$$P_\theta(\bigcup_{i \ne j}\{\xi_{(i)} = \xi_{(j)}\}) = P_\theta(\bigcup_{i \ne j}\{\xi_i = \xi_j\}) \le \sum_{i \ne j} P_\theta(\xi_i = \xi_j) = 0,$$

оскільки при $h > 0$

$$P_\theta(\xi_i = \xi_j) \le \sum_{n \in \mathbb{Z}} P_\theta([\xi_i/h] = [\xi_j/h] = n) =$$
$$\sum_{n \in \mathbb{Z}} (F(nh + h) - F(nh))^2 \le \sup_{n \in \mathbb{Z}} (F(nh + h) - F(nh)) \to 0,\ \ h \to 0.$$

Надалі будемо припускати виконання *умови неперервності функції розподілу.*

4.1. Розподіл порядкових статистик

Теорема (**про функцію розподілу порядкових статистик**). *Нехай вибірка* $X = (\xi_1, ..., \xi_n)$ *складається з незалежних у сукупності, однаково розподілених величин, які мають неперервну функцію розподілу* $F(x)$. *Тоді*

$$P_\theta(\xi_{(k)} < x) = \sum_{i=k}^{n} C_n^i \, F^i(x)(1 - F(x))^{n-i}, \; 1 \le k \le n.$$

Доведення.
Розглянемо цілозначну випадкову величину

$$\nu_n(x) = n\widehat{F}_n(x) = \sum_{k=1}^{n} 1_{\{\xi_k < x\}}.$$

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 $\{\xi_k < x\}$. За означенням порядкових статистик

$$\{\xi_{(k)} < x\} = \{\nu_n(x) \ge k\},$$

оскільки інтервал $(-\infty, x)$ містить k-ту порядкову статистику тоді й тільки тоді, коли він містить щонайменше k спостережень. Тому шуканий вираз випливає з формули для біноміального розподілу величини $\nu_n(x)$ та рівняння

$$P_\theta(\xi_{(k)} < x) = P_\theta(\nu_n(x) \ge k) = \sum_{i=k}^{n} P_\theta(\nu_n(x) = i) =$$
$$\sum_{i=k}^{n} C_n^i \, F^i(x)(1 - F(x))^{n-i} \; \boxdot$$

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) $P_\theta(\xi_{(k)} < \xi \le \xi_{(k+1)}) = 1/(n+1),\ k = \overline{0,n},$

(б) $P_\theta(\xi \le \xi_{(k)}) = k/(n+1),\ k = \overline{0,n},$

де за означенням $\xi_{(0)} = -\infty,\ \xi_{(n+1)} = \infty.$

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

$$P_\theta(\xi_{(k)} < x \le \xi_{(k+1)}) = P_\theta(\{\xi_{(k)} < x\} \backslash \{\xi_{(k+1)} < x\}) =$$
$$P_\theta(\xi_{(k)} < x) - P_\theta(\xi_{(k+1)} < x) = C_n^k\ F^k(x)(1-F(x))^{n-k}.$$

Оскільки випадкові вектори ξ та $(\xi_{(k)}, \xi_{(k+1)})$ незалежні за теоремою про векторні перетворення незалежних величин, то їх сумісна функція розподілу розпадається в добуток і внаслідок теореми Фубіні про кратний та повторні інтеграли

$$P_\theta(\xi_{(k)} < \xi \le \xi_{(k+1)}) = \int_{-\infty}^{\infty} dF_\xi(x) \int_{x_k < x,\ x \le x_{k+1}} dF_{(\xi_{(k)},\xi_{(k+1)})}(x_k, x_{k+1}) =$$
$$\int_{-\infty}^{\infty} dF_\xi(x) P_\theta(\xi_{(k)} < x \le \xi_{(k+1)}).$$

Отже, за теоремою про функцію розподілу порядкових статистик

$$P_\theta(\xi_{(k)} < \xi \le \xi_{(k+1)}) = \int_{-\infty}^{\infty} \left(C_n^k \, F^k(x)(1-F(x))^{n-k}\right) \, dF(x) =$$
$$\int_0^1 C_n^k \, u^k(1-u)^{n-k} du = C_n^k \, B(k, n-k) =$$
$$C_n^k \, \Gamma(k+1)\Gamma(n-k+1)/\Gamma(k+1+n-k+1) = 1/(n+1),$$

де використана заміна змінної $u = F(x)$ та означення повної бета-функції.

(б) Дана рівність випливає з попереднього твердження та з формули

$$P_\theta(\xi \le \xi_{(k)}) = \sum_{i=0}^{k-1} P_\theta(\xi_{(i)} < \xi \le \xi_{(i+1)}) = \frac{k}{n+1} \; \boxdot$$

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, тобто задовольняє умову:*

$$(\xi_{(\nu_1)}, ..., \xi_{(\nu_n)}) = (\xi_1, ..., \xi_n).$$

Іншими словами, величина ν_k збігається з місцем k-го спостереження у вибірці за порядком зростання, де найменший є першим. **Наприклад**, для вибірки $(3, 9, 2, 5)$ величина $\nu_1 = 2$, оскільки 1-е спостереження 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 одиниця, оскільки всі значення порядкових статистик є різними майже напевне. 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. Тоді за означенням справедливі тотожності

$$\xi_k = \xi_{(\nu_k)}, \quad \xi_{\nu_k^-} = \xi_{(k)}, \ \{\nu = \pi^-\} = \{\xi_{\pi_1} < \xi_{\pi_2} < ... < \xi_{\pi_n}\}.$$

Теорема (про розподіл вектора рангів). *Вектор рангів рівномірно розподілений на* Π_n.

Доведення.

Як указано вище в доведенні теореми Колмогорова про відхилення емпіричної функції розподілу, випадкові величини $\alpha_k = F(\xi_k)$ незалежні в сукупності та рівномірно розподілені на відрізку $[0, 1]$.

Припустимо для спрощення, що функція F строго монотонна. Нехай $\pi = (\pi_1, ..., \pi_n) \in \Pi_n$. Тоді

$$P_\theta(\nu = \pi^-) = P_\theta(\xi_{\pi_1} < \xi_{\pi_2} < ... < \xi_{\pi_n}) =$$
$$P_\theta(F(\xi_{\pi_1}) < F(\xi_{\pi_2}) < ... < F(\xi_{\pi_n})) =$$
$$P_\theta(\alpha_{\pi_1} < \alpha_{\pi_2} < ... < \alpha_{\pi_n}) = P_\theta(\alpha_1 < \alpha_2 < ... < \alpha_n) =$$
$$\int_0^1 dx_n \int_0^{x_n} dx_{n-1} ... \int_0^{x_2} dx_1 = 1 \,/\, n!,$$

де передостання рівність випливає з того, що випадкові вектори $(\alpha_{\pi_1}, ..., \alpha_{\pi_n})$ та $(\alpha_1, ..., \alpha_n)$ мають однакові сумісні функції розподілу, що дорівнюють добуткові n рівномірних функцій розподілу, та з теореми про обчислення ймовірностей, пов'язаних із випадковим вектором. Очевидно, що множина всіх обернених перестановок $\{\pi^-, \pi \in \Pi_n\}$ збігається з Π_n $\boxdot$

Теорема (про сумісний розподіл порядкових статистик). *Припустимо, що функція розподілу спостережень* F *має щільність* f*. Тоді*

Старт
Початок
Зміст

Назад
Екран
Закрити
Выхід

сумісна щільність вектора варіаційного ряду $(\xi_{(1)},...,\xi_{(n)})$ *дорівнює*

$$f_{\xi_{(1)},...,\xi_{(n)}}(x_1,...,x_n) = n!\ f(x_1)...f(x_n)1_{x_1\,<\,x_2\,<...<\,x_n}.$$

Доведення.
За властивостями вектора рангів

$$P_\theta(\xi_{(1)} < x_1,...,\xi_{(n)} < x_n) =$$
$$\sum\nolimits_{\pi\in\ \Pi_n} P_\theta(\xi_{(1)} < x_1,...,\xi_{(n)} < x_n,\ \nu = \pi^-) =$$
$$\sum\nolimits_{\pi\in\ \Pi_n} P_\theta(\xi_{\pi_1} < x_1,...,\xi_{\pi_n} < x_n,\ \xi_{\pi_1} < \xi_{\pi_2} < ... < \xi_{\pi_n}) =$$
$$n!\ P_\theta(\xi_1 < x_1,...,\xi_n < x_n,\ \xi_1 < \xi_2 < ... < \xi_n) =$$
$$\int..\int_{y_1<x_1,...,y_n<x_n} n!\ f(y_1)...f(y_n)1_{y_1\,<\,y_2\,<..<\,y_n}dy_1...dy_n,$$

де передостання рівність випливає з того, що 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 випадковим вектором, через його сумісну щільність. За означенням сумісна щільність вектора порядкових статистик дорівнює підінтегральній функції в правій частині, що збігається зі вказаною в формулюванні теореми ⊡

Вправа.

(1) Знайти сумісну щільність порядкових статистик $\xi_{(j)},\xi_{(k)}, j < k$.

(2) Знайти функцію розподілу, математичне сподівання та дисперсію розмаху $\xi_{(n)} - \xi_{(1)}$.

(3) 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.

4.3. Теоретичні квантилі

При побудові вірогідних інтервалів широко застосовуються такі числові характеристики функцій розподілу.

Означення. *Теоретичним квантилем рівня $p \in (0, 1)$ для випадкової величини ξ з функцією розподілу F називається число x_p, яке є розв'язком рівняння*

$$F(x_p) = p = P(\xi < x_p).$$

Якщо функція F неперервна і строго монотонно зростає на $\mathbb{R}$, то квантиль будь-якого рівня визначений однозначно. Далі припускатимемо, що ця умова виконана.

Частковими випадками квантилей є поняття медіани та квартилей.

Означення. *Теоретичною* медіаною, нижнім та верхнім квартилем *називаються відповідно квантилі $x_{1/2}$, $x_{1/4}$, $x_{3/4}$.*

Приклад. Для стандартного нормального розподілу $(x_{1/4},\ x_{1/2},\ x_{3/4}) \approx (-0.674,\ 0,\ +0.674)$.

Вправа. Довести, що теоретична медіана $m = x_{1/2}$ є абсолютним центром положення інтегровної випадкової величини ξ, тобто

$$m = \arg\min_{a\,\in\,\mathbb{R}} M\left|\xi - a\right|.$$

4.4. Емпіричні квантилі

Враховуючи теорему про емпіричний розподіл порядкових статистик, приходимо до вибіркових аналогів квантилів.

Означення. *Нехай $X = (\xi_1, ..., \xi_n)$ – вибірка з незалежних у сукупності, однаково розподілених величин, 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 *дорівнює* $\xi_{([n/4])}$ *та* $\xi_{([3n/4])}$ *відповідно.*

Розмахом вибірки *називається відстань між найбільшим та найменшим спостереженням, тобто статистика* $\xi_{(n)} - \xi_{(1)}$.

Зауваження. У прикладній статистиці для скороченого опису вибірки $X = (\xi_1, ..., \xi_n)$ використовується поняття ящик із вусами (whisker-box). Це п'ятірка статистик

$$\xi_{(1)} \le \xi_{([n/4])} \le \widehat{m}_n \le \xi_{([3n/4])} \le \xi_{(n)},$$

які зображують у вигляді горизонтально розташованого ящика шириною $\xi_{([3n/4])} - \xi_{([n/4])}$ (міжквартильний розмах) із виділеним центром $\widehat{m}_n$ та вусами зліва й справа шириною $\xi_{([n/4])} - \xi_{(1)}$ та $\xi_{(n)} - \xi_{([3n/4])}$ відповідно.

4.5. Консистентність вибіркових квантилей

Теорема (про строгу консистентність вибіркових квантилей). *Нехай функція розподілу* $F(x)$ *неперервна і строго монотонно зростає на* $\mathbb{R}$*. 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 означенням порядкових статистик

$$\{\zeta_{n\alpha} < x\} = \{\xi_{([n\alpha])} < x\} = \{\nu_n(x) \ge [n\alpha]\} = \{\widehat{F}_n(x) \ge [n\alpha]/n\}.$$

Оскільки при кожному $\varepsilon > 0$ внаслідок зауваження про верхню границю величин та подій

$$\{ \underline{\lim}_{n\to\infty}\zeta_{n\alpha} < x_\alpha - \varepsilon\} \subset \overline{\lim}_{n\to\infty} \{\zeta_{n\alpha} < x_\alpha - \varepsilon\},$$
$$\overline{\lim}_{n\to\infty}\{\eta_n \ge \varepsilon\} \subset \{\overline{\lim}_{n\to\infty}\eta_n \ge \varepsilon\},$$

то за монотонністю ймовірності

$$P_\theta\left(\{ \underline{\lim}_{n\to\infty}\zeta_{n\alpha} < x_\alpha - \varepsilon\}\right) \le P_\theta\left(\overline{\lim}_{n\to\infty} \{\zeta_{n\alpha} < x_\alpha - \varepsilon\}\right) =$$

$$= P_\theta\left(\overline{\lim_{n\to\infty}}\{\widehat{F}_n(x_\alpha-\varepsilon) \ge \frac{[n\alpha]}{n}\}\right) = P_\theta\left(\overline{\lim_{n\to\infty}}\{\widehat{F}_n(x_\alpha-\varepsilon)\frac{n}{[n\alpha]} \ge 1\}\right) \le$$

$$\le P_\theta\left(\{\overline{\lim}_{n\to\infty}\widehat{F}_n(x_\alpha-\varepsilon)\frac{n}{[n\alpha]} \ge 1\}\right) = P_\theta\left(F(x_\alpha-\varepsilon)/\alpha \ge 1\right) = 0.$$

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 умовою 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 $\boxdot$

4.6. Асимптотична нормальність вибіркових квантилей

Теорема (про асимптотичну нормальність вибіркових квантилей). *Нехай функція розподілу* $F(x)$ *має щільність* $f(x)$, *строго монотонно зростає на* $\mathbb{R}$, *має квантиль* x_α *рівня* $\alpha \in (0,1)$, *причому функція* f *неперервна та додатна в точці* x_α. *Тоді вибірковий квантиль* $\zeta_{n\alpha} \equiv \xi_{([n\alpha])}$ *є асимптотично нормальною оцінкою для* x_α :

$$\sqrt{n}(\zeta_{n\alpha} - x_\alpha) \to^W \ \eta \simeq N(0,\sigma^2),\ n \to \infty,$$
$$\text{де } \sigma^2 = \alpha(1-\alpha) \ / \ f^2(x_\alpha).$$

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, $\nu_n(x) = n\widehat{F}_n(x) = \sum_{k=1}^n 1_{\{\xi_{\ k} < x\}}$.

Як показано в доведенні теореми про функцію розподілу порядкових статистик, має місце тотожність $\{\xi_{(k)} < x\} = \{\nu_n(x) \ge k\}$. Тому

$$P_\theta(\beta_n < y) = P_\theta(\zeta_{n\alpha} < x_\alpha + y\sigma/\sqrt{n}) =$$
$$P_\theta(\xi_{([n\alpha])} < x_\alpha + y\sigma/\sqrt{n}) = P_\theta(\xi_{([n\alpha])} < z_n) = P_\theta(\nu_n(z_n) \ge [n\alpha]),$$
$$\text{де} \ \ z_n = x_\alpha + y\sigma/\sqrt{n} = .x_\alpha + o(1), n \to \infty.$$

Послідовність $\nu_n(z_n)$ є сумою незалежних у сукупності, однаково розподілених індикаторних випадкових величин $\chi_{nk} \equiv 1_{\{\xi_k \ < \ z_n\}}$, 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

$$\gamma_n \equiv (\nu_n(z_n) - m_n)/s_n \to^W \beta \simeq N(0,1),$$
$$\text{де } m_n \equiv M_\theta \nu_n(z_n) = nF(z_n), \;\; s_n^2 \equiv D_\theta \nu_n(z_n) = nF(z_n)(1 - F(z_n)).$$

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

$$m_n = n\alpha + y\sqrt{n\alpha(1-\alpha)} + o(\sqrt{n}), \quad s_n^2 = n\alpha(1-\alpha) + o(n).$$

Зокрема, послідовність $y_n \equiv ([n\alpha] - m_n)/s_n \to -y, \;\; n \to \infty$.

Звідси виводимо граничне співвідношення

$$P_\theta(\beta_n < y) = P_\theta(\nu_n(z_n) \geq [n\alpha]) = P_\theta(\gamma_n \geq ([n\alpha] - m_n)/s_n) =$$
$$P_\theta(\gamma_n - y_n \geq 0) \to_{n\to\infty} P_\theta(\beta + y \geq 0) = P_\theta(\beta \geq -y) = P_\theta(\beta < y).$$

Остання рівність випливає із 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 $\boxdot$

Вправа. Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з неперервною функцією розподілу F спостережень, а $\xi_{(1)}, \xi_{(n)}$ – найменша та найбільша порядкова статистика. Довести, що величина $2n\sqrt{F(\xi_{(1)})(1 - F(\xi_{(n)}))}$ має граничний розподіл при $n \to \infty$ із середнім $\pi/2$ та дисперсією $4 - \pi^2/4$.

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.

5.1. Теоретичні та вибіркові моменти

Означення. *Нехай ξ – випадкова величина. Її (нецентральним)* теоретичним моментом *порядку $k \in \mathbb{N}$ називається число*

$$\mu_k \equiv M\,\xi^k.$$

Центральним теоретичним моментом *порядку k називається число*

$$\mu_k^0 \equiv M\,(\xi - \mu)^k,$$

де $\mu \equiv \mu_1$ – математичне сподівання величини ξ. Зокрема, $\mu_2^0 \equiv \sigma^2$ – дисперсія ξ.

Зауваження. Значення центральних моментів використовуються в теорії розподілів для означення таких спеціальних характеристик:

$k_a = \mu_3^0\,/\sigma^3$ – коефіцієнт асиметрії (нульовий для симетричних розподілів),

$k_e = \mu_4^0\,/\sigma^4 - 3$ – коефіцієнт ексцесу (нульовий для нормальних спостережень).

Означення. *Нехай $X = (\xi_1, ..., \xi_n)$ – вибірка. (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*

$$\widehat{\mu}^0_{kn} \equiv \frac{1}{n}\sum\nolimits_{i=1}^{n}(\xi_i - \widehat{\mu}_n)^k,$$

де $\widehat{\mu}_n \equiv \widehat{\mu}_{1n}$ – вибіркове середнє. *Зокрема,* $\widehat{\mu}^0_{2n} \equiv \widehat{\sigma}^2_n$ – вибіркова дисперсія.

Зауваження. Важливою властивістю центральних моментів є інваріантність відносно зсувів: вони не змінюються при одночасному зсуві всіх спостережень на сталу:

$$\widehat{\mu}^0_{kn} = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left(\xi_i - \frac{1}{n}\sum\nolimits_{j=1}^{n}\xi_j\right)^k = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left(\xi_i - c - \frac{1}{n}\sum\nolimits_{j=1}^{n}(\xi_j - c)\right)^k.$$

Зауваження. Як і у випадку теоретичних моментів, справедлива тотожність

$$\widehat{\sigma}^2_n = \widehat{\mu}_{2,n} - (\widehat{\mu}_n)^2.$$

Дійсно,

$$\widehat{\sigma}^2_n = \frac{1}{n}\sum\nolimits_{i=1}^{n}(\xi_i - \widehat{\mu}_n)^2 = \frac{1}{n}\sum\nolimits_{i=1}^{n}\xi_i^2 - 2(\widehat{\mu}_n)^2 + (\widehat{\mu}_n)^2.$$

Теорема (про моменти вибіркових моментів). *Нехай* $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка. Тоді*

$$M\widehat{\mu}_{kn} = \mu_k,\ D\widehat{\mu}_{kn} = \frac{1}{n}(\mu_{2k} - \mu_k^2).$$

Зокрема, $M\widehat{\mu}_n = \mu$, $D\widehat{\mu}_n = \sigma^2/n$. *Крім того,*

$$M\,\widehat{\sigma}^2_n = \frac{n-1}{n}\sigma^2,\quad D\,\widehat{\sigma}^2_n = \frac{n-1}{n^3}((n-1)\mu_4^0 - (n-3)\sigma^4).$$

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:

$$D\left(\frac{1}{n}\sum\nolimits_{i=1}^{n} \xi_i^k\right) = \frac{1}{n^2}\sum\nolimits_{i=1}^{n} D(\xi_i^k) = \frac{1}{n^2} n D(\xi_1^k) = \frac{1}{n}(M\xi_1^{2k} - (M\xi_1^k)^2).$$

Середнє для вибіркової дисперсії обчислюється з урахуванням останнього зауваження шляхом піднесення до квадрату під знаком суми:

$$M\,\widehat{\sigma}_n^2 = M(\widehat{\mu}_{n2} - \widehat{\mu}_n^2) = M\widehat{\mu}_{n2} - M\,\widehat{\mu}_n^2 =$$
$$\mu_2 - \left(M\sum\nolimits_{i=1}^{n} \xi_i^2 + M\sum\nolimits_{i\neq j} \xi_i\xi_j\right)\,/\,n^2 =$$
$$\mu_2 - \mu_2/n - \mu^2(n-1)/n = (\mu_2 - \mu^2)(n-1)/n = \sigma^2(n-1)/n.$$

При обчисленні дисперсії $\widehat{\sigma}_n^2$ можна вважати, що $\mu = 0$, оскільки оцінка $\widehat{\sigma}_n^2$ інваріантна відносно зсувів, зокрема відносно віднімання від спостережень їх спільного середнього μ. Тому:

$$n^4 M\left(\widehat{\sigma}_n^2\right)^2 = M\left(n\sum \xi_k^2 - (\sum \xi_k)^2\right)^2 = M\left((n-1)\sum \xi_k^2 - 2\sum\nolimits_{i<j} \xi_i\xi_j\right)^2 =$$
$$(n-1)^2 M\left(\sum \xi_k^2\right)^2 + 4M(\sum\nolimits_{i\,<\,j} \xi_i\xi_j)^2 - 4(n-1)M(\sum \xi_k^2\ \sum\nolimits_{i\,<\,j} \xi_i\xi_j) =$$
$$(n-1)^2\left(n\,\mu_4^0 + n(n-1)\sigma^4\right) + 2n(n-1)\sigma^4 - 0 =$$
$$(n-1)\left[(n-1)\mu_4^0 + (n^2 - 2n + 3)\sigma^4\right],$$
$$D\,\widehat{\sigma}_n^2 = M\left(\widehat{\sigma}_n^2\right)^2 - \left(M\widehat{\sigma}_n^2\right)^2 = M\left(\widehat{\sigma}_n^2\right)^2 - \sigma^4(n-1)^2/n^2 =$$
$$(n-1)\left((n-1)\mu_4^0 - (n-3)\sigma^4\right)/\,n^3,$$

де використане припущення $M\xi_j = 0$ $\boxdot$

Вправа.

(1) Для нормальної випадкової величини $\xi \simeq N(\mu, \sigma^2)$

$$\mu_{2k}^0 = \sigma^{2k} 2^k \pi^{-1/2} \Gamma(k + 1/2) = \sigma^{2k} 1 \cdot 3 \cdot ...(2k - 1) \equiv \sigma^{2k}(2k - 1)!!.$$

(2) Довести, що

$$D(\widehat{\mu}_{kn}^0) = (\mu_{2k} - \mu_k^2 + k\mu_{k-1}(k\mu_{k-1}\mu_2 - 2\mu_{k+1}))/n \ + \ O(1/n^2), \ n \to \infty.$$

З останньої теореми 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 *називається статистика*

$$\widehat{s}_n^2 \equiv \frac{1}{n-1} \sum\nolimits_{i=1}^n (\xi_i - \widehat{\mu}_n)^2 = \frac{n}{n-1} \widehat{\sigma}_n^2 \ ,$$

яка, очевидно, є незміщеною оцінкою дисперсії:

$$M\widehat{s}_n^2 = \tfrac{n}{n-1} M\widehat{\sigma}_n^2 = \sigma^2 \ \boxdot$$

Теорема (про властивості вибіркових моментів). *Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка.*

(а) Якщо $M\left|\xi_1\right|^k < \infty$, то нецентральний вибірковий момент $\widehat{\mu}_{kn}$ є незміщеною строго консистентною оцінкою теоретичного моменту μ_k

(б) Якщо $M\xi_1^{2k} < \infty$, 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 одного доданка наведені в теоремі про моменти вибіркових моментів. Тому асимптотична нормальність є наслідком класичної центральної граничної теореми $\boxdot$

Вправа. Довести, що вибіркове середнє для кратної вибірки з щільністю Коші $1/\pi(1+(x-\theta)^2)$ не є консистентною оцінкою центра симетрії θ, оскільки розподіл статистики не залежить від n.

Вибіркові моменти можна розглядати одночасно у вигляді випадкового вектора вибіркових моментів наперед заданої розмірності d:

$$\widehat{\mu}_n^{(d)} \equiv (\widehat{\mu}_{kn},\ k=\overline{1,d}), \qquad M\widehat{\mu}_n^{(d)} = \mu^{(d)} \equiv (\mu_k,\ k=\overline{1,d}).$$

Теорема (про асимптотику вектора вибіркових моментів). *Нехай*
$X=(\xi_1,...,\xi_n)$ – *кратна вибірка.*

(а) Якщо $M\left|\xi_1\right|^d<\infty$, то векторна статистика $\widehat{\mu}_n^{(d)}$ є строго консистентною оцінкою для вектора $\mu^{(d)}$.

(б) Якщо $M\xi_1^{2d}<\infty$, то вектор вибіркових моментів є асимптотично нормальним:

$$\sqrt{n}(\widehat{\mu}_n^{(d)}-\mu^{(d)}) \to^W \eta \simeq N_d(0,\ V^{(d)}),\ n\to\infty,$$

з коваріаційною матрицею $V^{(d)} = (\mu_{k+l} - \mu_k\, \mu_l,\ k, l = \overline{1, d})$.

Доведення.

(a). З теореми про властивості вибіркових моментів виводимо, що $\widehat{\mu}_{kn} \to^{P_\theta 1} \mu_k$ для всіх $k = \overline{1, d},\ \theta \in \Theta$. Тому

$$P_\theta(\lim \widehat{\mu}_n^{(d)} = \mu^{(d)}(\theta)) = P_\theta(\cap_{k=1}^d \{\lim \widehat{\mu}_{kn} = \mu_k\}) = 1.$$

(б). Розглянемо послідовність випадкових векторів

$$\gamma_j \equiv (\xi_j^k - \mu_k,\ k = \overline{1, d}),\quad j \geq 1.$$

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

$$M\gamma_j = 0,\ \ \mathrm{Cov}(\gamma_j) = (M(\xi_j^k - \mu_k)(\xi_j^l - \mu_l), k, l = \overline{1, d}) =$$
$$((\mu_{k+l} - \mu_k\mu_l),\ k, l = \overline{1, d}) = V^{(d)}.$$

Тому за вказаною теоремою

$$\sqrt{n}(\widehat{\mu}_n^{(d)} - \mu^{(d)}) = \tfrac{1}{\sqrt{n}} \textstyle\sum_{j=1}^n \gamma_j = \eta_n \to^W \eta \simeq N_d(0,\ V^{(d)})\ \boxdot$$

5.2. Метод моментів

Метод моментів є спеціальним методом побудови оцінок невідомих числових 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 функції

$$\mu_k(\theta) \equiv M_\theta \, \xi_1^k$$

теж повністю відомі.

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 виконана, зокрема, за теоремою про обернене відображення з 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 називається статистика*

$$\widehat{\theta}_n \equiv T_d(\widehat{\mu}_n^{(d)}),$$

де вектор вибіркових моментів $\widehat{\mu}_n^{(d)} = (\widehat{\mu}_{nk},\ k = \overline{1,d})$ містить значення перших d вибіркових моментів, *а T_d – обернена до $\mu^{(d)}(\theta)$ функція.*

Зауваження. З означення оберненої функції випливає, що оцінка методу моментів $\widehat{\theta}_n$ є єдиним розв'язком системи рівнянь методу моментів

$$\mu^{(d)}(\widehat{\theta}_n) = \widehat{\mu}_n^{(d)}.$$

Теорема (**про строгу консистентність оцінок методу моментів**). *Якщо* $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка*, $M_\theta |\xi_1|^d < \infty$, *і функція* T_d *неперервна, то оцінка методу моментів є строго консистентною оцінкою для параметра* θ.

Доведення. За теоремою про асимптотику вектора вибіркових моментів $\widehat{\mu}_n^{(d)} \to^{P_\theta 1} \mu^{(d)}(\theta)$ для всіх $\theta \in \Theta$. Тому з неперервності та означення T_d випливає, що

$$\widehat{\theta}_n = T_d(\widehat{\mu}_n^{(d)}) \to^{P_\theta 1} T_d(\mu^{(d)}(\theta)) = \theta, \ n \to \infty, \ \forall \theta \in \Theta \boxdot$$

Зауваження. Якщо функція T_d неперервно диференційовна і $M_\theta \xi_1^{2d} < \infty$, то можна довести, що оцінка методу моментів є асимптотично нормальною:

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to^W \eta \simeq N_d(0, \ t_d V^{(d)} t_d'), \ n \to \infty,$$

де матриця $t_d = \frac{\partial}{\partial \mu} T_d(\mu), \ \mu = \mu^{(d)}(\theta)$.

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

$$\mu_1(\theta) = \alpha/\lambda, \quad \mu_2(\theta) = \alpha/\lambda^2 + (\alpha/\lambda)^2.$$

Оцінку методу моментів $(\widehat{\alpha}, \widehat{\lambda})$ знаходимо із 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

$$\mu_1(\theta) = (a + b)/2, \quad \mu_2(\theta) = \mu_1^2(\theta) + (b - a)^2/12.$$

Оцінку методу моментів $(\widehat{a}, \widehat{b})$ знаходимо із 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 з відповідними параметрами: $\ln \xi \simeq N(\mu, \sigma^2)$.

Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з логнормальним розподілом: $\xi_1 \simeq LN(\mu, \sigma^2)$ Для оцінки параметрів методом моментів можна скористатися перетворенням вибірки за допомогою логарифмічної функції. За теоремою про перетворення незалежних величин векторна статистика $(\eta_1, ..., \eta_n)$ з координатами $\eta_k = \ln \xi_k, k = \overline{1, n}$, знову є кратною вибіркою і має нормальний розподіл. Внаслідок теореми про інтерпретацію параметрів нормального розподілу вибіркові середнє та дисперсія логарифмічного перетворення

$$\widehat{\mu}_n = \frac{1}{n} \sum\nolimits_{k=1}^{n} \ln \xi_k, \quad \widehat{\sigma}_n^2 = \frac{1}{n} \sum\nolimits_{k=1}^{n} (\ln \xi_k - \widehat{\mu}_n)^2$$

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 в залежності від властивостей розподілу спостережень.

Вправа. Знайти щільність, математичне сподівання та дисперсію випадкової величини з логнормальним розподілом.

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.

6.1. Незміщені оцінки

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 невідомого параметра* $\theta \in \Theta$, *якщо* $\dim T = \dim \tau$ *і*

$$M_\theta T = \tau(\theta), \quad \forall \theta \in \Theta.$$

Приклад. Для кратної вибірки $X = (\xi_1, ..., \xi_n)$
(а) вибіркова дисперсія $\widehat{\sigma}_n^2$ є незміщеною оцінкою функції $\frac{n-1}{n}\sigma^2$ від невідомої дисперсії σ^2, а нормована вибіркова дисперсія $\widehat{s}_n^2$ є незміщеною оцінкою σ^2,
(б) статистики ξ_1 та $(\xi_1 - \xi_2)^2/2$ є незміщеними оцінками середнього та дисперсії – хоча не зовсім ефективними.

6.2. Оптимальні оцінки

Якщо існують незміщені оцінки, то їх можна порівнювати між собою за величиною середньоквадратичного відхилення від значення, яке оцінюється, та одночасно збігається із середнім цих оцінок.

Позначимо через

$$\Gamma_\tau = \{T = T(X) : M_\theta T = \tau(\theta), \quad \forall\theta \in \Theta\}$$

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 дисперсію у класі всіх лінійних незміщених оцінок:*

$$D_\theta T^* \le D_\theta T, \quad \forall T \in \Gamma_\tau, \ \forall\theta \in \Theta.$$

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 $\sigma^2(\theta) = D_\theta T_1 = D_\theta T_2$. Розглянемо оцінку $T = (T_1 + T_2)/2$. Оскільки $M_\theta T = (M_\theta T_1 + M_\theta T_2)/2 = \tau$, то $T \in \Gamma_\tau$, і за означенням $\sigma^2(\theta)$

$$\sigma^2(\theta) \le D_\theta T = M_\theta(T_1 - \tau + T_2 - \tau)^2/\ 4 =$$

$$(D_\theta T_1 + D_\theta T_2 + 2\,\mathrm{Cov}(T_1, T_2))\ /\ 4 =$$

$$(2\sigma^2(\theta) + 2\ M(T_1 - \tau)(T_2 - \tau))\ /\ 4 \le \sigma^2(\theta),$$

де остання нерівність випливає з нерівності Коші. Отже, за означенням оптимальної оцінки $\sigma^2(\theta) = D_\theta T$, і нерівність Коші має перетворитися на рівність. За зауваженням про рівність у нерівності Коші існує невипадкова стала b така, що $T_2 - \tau = b(T_1 - \tau)$ м.н. Тому

$$\sigma^2(\theta) = \mathrm{Cov}(T_1, T_2) = b\ \mathrm{Cov}(T_1, T_1) = bD_\theta T_1 = b\sigma^2(\theta).$$

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 $G(1-\theta)$ із невідомою ймовірністю неуспіху $\theta \in (0,1)$. Для довільної статистики $T = T(X)$ умова незміщеності при оцінюванні функції $\tau = \tau(\theta)$ має вигляд

$$M_\theta T(X) = \sum\nolimits_{n\ \geq 1} (1-\theta)\theta^{n-1} T(n) = \tau(\theta),\ \forall\theta \in (0,1).$$

Нехай $\tau(\theta) = \theta$. Тоді з єдиності розкладу функції $\theta/(1-\theta)$ у ряд Тейлора виводимо, що існує єдина незміщена оцінка для θ, до того ж не дуже змістовна: $T(X) = 0$ при $X = 1$ та $T(X) = 1$ для інших X.

Якщо ж $\tau(\theta) = 1\ /\ \theta$, то незміщених оцінок взагалі не існує, оскільки функція $1/\theta(1-\theta)$ не припускає розклад у збіжний ряд Тейлора на відрізку $(0,1)$.

2. *Оцінювання дисперсії у загальній нормальній моделі.* 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 відповідне середньоквадратична похибка дорівнює

$$D_\theta \widehat{s}^2_n = (\frac{n}{n-1})^2 D_\theta \widehat{\sigma}^2_n = \frac{1}{n(n-1)}((n-1)\mu^0_4 - (n-3)\sigma^4) =$$

$$\frac{1}{n(n-1)}((n-1)3\sigma^4 - (n-3)\sigma^4) = \frac{2\sigma^4}{n-1},$$

де враховане співвідношення $\mu_4^0 = 3\sigma^4$ для моментів нормального розподілу (**Вправа**. Отримати це співвідношення диференціюванням характеристичної функції нормального розподілу).

Одночасно розглянемо зміщену оцінку для σ^2 вигляду $\widehat{\tau}_n^2 = t\,\widehat{s}_n^2$, де t – деяка стала. Обчислимо середньоквадратичну похибку

$$M_\theta(\widehat{\tau}_n^2 - \sigma^2)^2 = M_\theta(t(\widehat{s}_n^2 - \sigma^2) - (1-t)\sigma^2)^2 =$$
$$t^2 D_\theta \widehat{s}_n^2 + (1-t)^2\sigma^4 = (2t^2/(n-1) + (1-t)^2)\sigma^4.$$

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 середньоквадратичного відхилення від оцінюваного значення σ^2 гірша за зміщену оцінку $\widehat{\tau}_n^2$.Це і не дивно – адже мінімум у класі всіх оцінок часто є меншим від мінімуму в певному підкласі.

Не зважаючи на негативний зміст наведених прикладів, у багатьох схемах незміщені оцінки все-таки є досить ефективними.

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 $f(x_k, \theta)$ – щільність одного спостереження.

У даному розділі припускатимемо, що параметричний простір евклідів: $\Theta \subset \mathbb{R}^d$, а функція вірогідності $L(x, \theta)$ – диференційовна за параметром θ.

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 $\theta \in \Theta \subset \mathbb{R}^d$, то похідну слід розуміти як вектор, що складений із частинних похідних (градієнт), а функція впливу буде 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 невідомого параметра θ (параметр не впливає на розподіл), то відповідна 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). *Якщо $X = (\xi_1, ..., \xi_n)$ – кратна вибірка, то 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 $P(\lambda)$ та невідомим параметром $\theta = \lambda$

$$u(y,\theta) = \frac{\partial}{\partial\theta}\ln(\theta^y\exp(-\theta)/y!) = y/\theta - 1,\ \ y\in\mathbb{Z}_+,$$
$$U(X,\theta) = \sum_{k=1}^{n}(\xi_k/\theta - 1) = n(\widehat{\mu}_n/\theta - 1).$$

2. *Вибірка з гама-розподілу.* Для кратної вибірки з гама-розподілом $\Gamma(\lambda,\alpha)$ та невідомим параметром $\theta = (\lambda,\alpha)$

$$\ln f(y,\theta) = \alpha\ln\lambda + (\alpha-1)\ln y - \lambda y - \ln\Gamma(\alpha),$$
$$u_1(y,\theta) = \alpha/\lambda - y,$$
$$u_2(y,\theta) = \ln\lambda + \ln y - \psi(\alpha),\quad \psi(\alpha) = \frac{\partial}{\partial\alpha}\ln\Gamma(\alpha),$$
$$U_1(X,\theta) = n(\alpha/\lambda - \widehat{\mu}_n),$$
$$U_2(X,\theta) = (n\ln\lambda + \ln\prod_{k=1}^{n}\xi_k - n\psi(\alpha)).$$

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 $L(X,\theta)$ строго додатна, не залежить від θ.

2. Функція вірогідності $L(X,\theta)$ двічі неперервно диференційовна за параметром θ.

3. Функція впливу $U(X,\theta)$ інтегровна у квадраті: $M_\theta U^2(X,\theta) < \infty\ \forall\theta$.

4. Знак похідної за параметром θ можна внести під знак інтегралів $M_\theta L(X,\theta)$ за аргументом $x\in S$ від функції вірогідності $L(x,\theta)$, а також від її похідних за θ.

7.3. Властивості функції впливу, інформація за Фішером

Теорема (про центрованість функції впливу). *За умов регулярності функція впливу центрована:*

$$M_\theta\, U(X,\theta) = 0, \quad \forall \theta \in \Theta.$$

Доведення. За означенням функції вірогідності

$$P_\theta(X \in S) = 1 = \int_S L(x,\theta)\,\lambda(dx),$$

оскільки вибірка X набуває значень у вибірковому просторі S, а $L(x,\theta)$ – її 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 інтегралу Лебега від нульової функції, з наведеної вище тотожності дістанемо

$$0 = \frac{\partial}{\partial\theta}\,1 = \frac{\partial}{\partial\theta}\int_S L(x,\theta)\,\lambda(dx) = \frac{\partial}{\partial\theta}\int_{S_0} L(x,\theta)\,\lambda(dx) =$$

$$\int_{S_0} \frac{\partial}{\partial\theta}L(x,\theta)\,\lambda(dx) = \int_{S_0} \left(\frac{\partial}{\partial\theta}\ln L(x,\theta)\right) L(x,\theta)\,\lambda(dx) =$$

$$\int_S \left(\frac{\partial}{\partial\theta}\ln L(x,\theta)\right) L(x,\theta)\,\lambda(dx) = M_\theta\, U(X,\theta),$$

де остання рівність випливає з теореми про обчислення математичного сподівання функції від випадкової величини, і використовується очевидна тотожність

$$\tfrac{\partial}{\partial\theta}L(x,\theta) = \left(\tfrac{\partial}{\partial\theta}\ln L(x,\theta)\right) L(x,\theta)$$

на множині S_0, де $L(x,\theta) > 0$ $\boxdot$

Означення. *Нехай параметр $\theta \in \Theta \subset \mathbb{R}$ – скалярний.* Інформацією за Фішером *у вибірці X називається функція*

$$I(\theta) \equiv D_\theta\, U(X,\theta) = M_\theta U^2(X,\theta).$$

Друга рівність в означенні випливає з теореми про центрованість функції впливу та з властивості дисперсії.

Теорема (про обчислення інформації за Фішером). *За умов регулярності справедлива тотожність*

$$I(\theta) = -M_\theta \frac{\partial^2}{\partial\theta^2}\ln L(X,\theta) = -M_\theta \frac{\partial}{\partial\theta}U(X,\theta), \quad \forall\theta \in \Theta.$$

Доведення. Продиференціюємо тотожність із теореми про центрованість функції впливу та використаємо умови регулярності:

$$0 = \frac{\partial}{\partial\theta}M_\theta\, U(X,\theta) = \frac{\partial}{\partial\theta}\int_S U(x,\theta)L(x,\theta)\lambda(dx) =$$

$$\int_S \frac{\partial}{\partial\theta}\left(U(x,\theta)L(x,\theta)\right)\,\lambda(dx) =$$

$$\int_S \left(\frac{\partial}{\partial\theta}U(x,\theta)\right) L(x,\theta)\,\lambda(dx) + \int_S U(x,\theta)\frac{\partial}{\partial\theta}L(x,\theta)\,\lambda(dx) =$$

$$M_\theta\frac{\partial}{\partial\theta}U(X,\theta) + M_\theta U^2(X,\theta) = M_\theta\frac{\partial^2}{\partial\theta^2}\ln L(X,\theta) + I(\theta)\boxdot$$

Теорема (про адитивність інформації за Фішером). *Якщо* $X = (\xi_1, ..., \xi_n)$ – *кратна вибірка, то*

$$I(\theta) = n\, i(\theta),$$

де

$$i(\theta) = D_\theta\, u(\xi_1, \theta)$$

інформація за Фішером в одному спостереженні.

Доведення. За теоремою про функцію впливу кратної вибірки та за теоремою про дисперсію суми незалежних величин

$$I(\theta) = D_\theta U(X, \theta) = D_\theta \sum\nolimits_{k=1}^{n} u(\xi_k, \theta) = \sum\nolimits_{k=1}^{n} D_\theta u(\xi_k, \theta) = n\, i(\theta),$$

де незалежність величин $u(\xi_k, \theta)$ випливає з теореми про перетворення незалежних величин і 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

$$f(y, \theta) = \theta 1_{y=1} + (1 - \theta) 1_{y=0} = \theta^y (1 - \theta)^{1-y},\ y \in \{0, 1\},$$

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

$$L(X,\theta) = \prod_{k=1}^{n} \theta^{\chi_k}(1-\theta)^{1-\chi_k} = \theta^{\nu_n}(1-\theta)^{n-\nu_n}, \ \nu_n = \sum_{k=1}^{n} \chi_k,$$

$$U(X,\theta) = \sum_{k=1}^{n} u(\chi_k,\theta) = \frac{n}{\theta(1-\theta)}(\widehat{\theta}_n - \theta), \quad \widehat{\theta}_n = \frac{\nu_n}{n},$$

функції інформації за Фішером мають вигляд

$$i(\theta) = M_\theta\left(\frac{\chi_1-\theta}{\theta(1-\theta)}\right)^2 = \frac{D_\theta\chi_1}{\theta^2(1-\theta)^2} = \frac{1}{\theta(1-\theta)}, \ I(\theta) = \frac{n}{\theta(1-\theta)}.$$

2. *Показниковий розподіл.* Розглянемо кратну вибірку $X = (\xi_1,...,\xi_n)$ із показниковим розподілом спостережень $\xi_k \simeq Exp(\theta)$. Функції впливу мають вигляд

$$u(y,\theta) = \frac{\partial}{\partial\theta}(\ln\theta - \theta y) = 1/\theta - y,$$

$$U(X,\theta) = \sum_{k=1}^{n} u(\xi_k,\theta) = n/\theta - n\widehat{\mu}_n,$$

а функції інформації за Фішером у спостереженні та у вибірці дорівнюють

$$i(\theta) = M_\theta(1/\theta - \xi_1)^2 = D_\theta\xi_1 = 1/\theta^2, \quad I(\theta) = n/\theta^2.$$

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* $\Theta \subset \mathbb{R}$. *За умов регулярності:*

(а) для довільної оцінки $T = T(X) \in \Gamma_\tau$ при всіх $\theta \in \Theta$ справедлива нерівність Крамера – Рао

$$M_\theta(T-\tau)^2 \equiv D_\theta T \geq \frac{\tau_\theta^2(\theta)}{I(\theta)},$$

де $\tau_\theta(\theta) = \frac{d}{d\theta}\tau(\theta)$, $I(\theta)$ – інформація за Фішером у вибірці X,

(б) рівність у нерівності (а) має місце 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 $c(\theta)$, причому у випадку рівності

$$c(\theta) = \tau_\theta(\theta)/\, I(\theta).$$

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 $D_\theta T$ має місце також і для зміщених оцінок T, якщо в правій частині τ_θ^2 замінити на $(\tau_\theta + b_\theta)^2$, де $b_\theta(\theta) = \frac{d}{d\theta}b(\theta)$, $b(\theta) = M_\theta T - \tau(\theta)$. 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

$$M_\theta T(X) = \int_S T(x)L(x,\theta)\ \mu(dx) = \tau(\theta),\ \forall\theta\in\Theta.$$

За умови регулярності з урахуванням означення функції впливу та теореми про обчислення математичного сподівання функції від випадкової величини знайдемо

$$\tau_\theta(\theta) = \tfrac{\partial}{\partial\theta}\textstyle\int_S T(x)L(x,\theta)\ \mu(dx) = \int_S T(x)\tfrac{\partial}{\partial\theta}L(x,\theta)\ \mu(dx) =$$
$$\textstyle\int_S T(x)U(x,\theta)L(x,\theta)\ \mu(dx) =$$
$$M_\theta T(X)U(X,\theta) = M_\theta(T(X)-\tau(\theta))U(X,\theta),$$

де остання рівність є наслідком теореми про центрованість функції впливу.

Згідно з нерівністю Коші з отриманої тотожності виводимо

$$\tau_\theta^2(\theta) = (\ M_\theta(T(X)-\tau(\theta))U(X,\theta)\ )^2 \le$$
$$M_\theta(T(X)-\tau(\theta))^2\ M_\theta U^2(X,\theta) = D_\theta T\cdot I(\theta),$$

за означенням $I(\theta)$. Це доводить твердження (а).

Для того, щоб нерівність Коші перетворювалася на рівність, необхідно й достатньо, щоб множники під знаком математичного сподівання були лінійно пов'язані майже напевне при кожному θ. Звідси виводимо твердження (б) теореми. Вираз для значення відповідної сталої знаходимо після підстановки тотожності критерію рівності в отриману вище тотожність:

$$\tau_\theta(\theta) = M_\theta(T(X) - \tau(\theta))U(X,\theta) = M_\theta c(\theta)U^2(X,\theta) = c(\theta)I(\theta) \boxdot$$

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

$$D_\theta(\sqrt{n}(T - \tau)) = n\ D_\theta T \geq \frac{\tau_\theta^2(\theta)}{i(\theta)} \equiv \widetilde{\sigma}_{opt}^2.$$

Означення. *Якщо послідовність оцінок* $T_n \in \Gamma_\tau$ *є асимптотично нормальною з асимптотичною дисперсію* σ_a^2, *тобто*

$$\sqrt{n}(T_n - \tau) \to^W \eta \simeq N(0, \sigma_a^2),$$

причому $\sigma_a^2 = \widetilde{\sigma}_{opt}^2$ *збігається з найменшим можливим за нерівністю Крамера – Рао значенням, то оцінка* (T_n) *називається* асимптотично ефективною.

7.5. Нерівність Крамера – Рао для векторного параметра

Означення. *Нехай параметр* $\theta \in \Theta \subset \mathbb{R}^d$ *– векторний, тоді функція впливу також є вектором-стовпчиком*

$$U(X,\theta) \equiv \left(U_i(X,\theta) \equiv \frac{\partial}{\partial\theta_i}U(X,\theta),\ i = \overline{1,d}\right).$$

Інформаційною матрицею за Фішером *вибірки* X *називається коваріаційна матриця векторної функції впливу:*

$$I(\theta) \equiv \mathrm{Cov}_\theta\, U(X,\theta) = M_\theta U(X,\theta)U'(X,\theta) =$$
$$\big(\mathrm{Cov}(U_i(X,\theta),\ U_j(X,\theta)),\ i,j=\overline{1,d}\big)\,.$$

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

$$\ln L(X,\theta) = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum\nolimits_{k=1}^{n}(\xi_k - \widehat{\mu}_n + \widehat{\mu}_n - \mu)^2 =$$

$$-\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}(\widehat{\sigma}_n^2 + (\widehat{\mu}_n - \mu)^2).$$

Звідси знаходимо координати функції впливу

$$U_1(X,\theta) = \frac{n(\widehat{\mu}_n - \mu)}{\sigma^2},$$

$$U_2(X,\theta) = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}(\widehat{\sigma}_n^2 + (\widehat{\mu}_n - \mu)^2).$$

За теоремою про моменти вибіркових моментів обчислюємо

$$I_{11}(\theta) = \frac{n^2}{\sigma^4}D_\theta\widehat{\mu}_n = \frac{n}{\sigma^2}.$$

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). *Нехай* $\dim\Theta = d > 1$, *виконані умови регулярності, інформаційна матриця за Фішером* $I(\theta)$ *невироджена, а* $T = T(X) \in \Gamma_\tau$ *– незміщена оцінка скалярної функції* $\tau(\theta)$.

(а) Тоді для всіх $\theta \in \Theta$ *справедлива нерівність*

$$M_\theta(T-\tau)^2 = D_\theta\, T \geq B(I(\theta), \tau_\theta) \equiv \tau'_\theta I^{-1}(\theta)\tau_\theta,$$

де $\tau_\theta(\theta) = \frac{d}{d\theta}\tau(\theta)$ *– вектор-стовпчик градієнта,* τ'_θ *– спряжений вектор-рядок.*

(б) Рівність у нерівності (а) має місце тоді й тільки тоді, коли T *є лінійною функцією функції впливу вибірки*

$$T(X) - \tau(\theta) = c'(\theta)\, U(X,\theta) \text{ м.н.}, \quad \forall\theta \in \Theta.$$

Зауваження. Твердження теореми залишаються справедливими і для виродженої матриці $I(\theta)$, якщо замінити квадратичну форму в правій частині (а) на

$$B(I(\theta), \tau_\theta) = \sup_{b\in\mathbb{R}^d} \frac{(b'\tau_\theta)^2}{b'I(\theta)b}.$$

Доведення.

Як і у скалярному випадку, з умов нормованості функції вірогідності та незміщеності на підставі регулярності доводимо тотожності

$$M_\theta\, U(X,\theta) = 0,\ M_\theta(T(X) - \tau(\theta))U(X,\theta) = \tau_\theta(\theta).$$

Нехай $b \in \mathbb{R}^d$. Скалярно помножимо другу тотожність на b і скористаємося нерівністю Коші та теоремою про дисперсію лінійної форми від випадкового вектора:

$$(b'\tau_\theta)^2 = (b'M_\theta(T-\tau)U(X,\theta))^2 = (M_\theta(T-\tau)b'U(X,\theta))^2 \leq$$

$$M_\theta(T-\tau)^2\ M_\theta(b'U(X,\theta))^2 = D_\theta T\ D_\theta(b'U(X,\theta)) = D_\theta T\ \ b'I(\theta)b.$$

Звідси

$$D_\theta T \geq \frac{(b'\tau_\theta)^2}{b'I(\theta)b} = \tau'_\theta I^{-1}(\theta)\tau_\theta,$$

де остання рівність досягається при $b = I^{-1}(\theta)\tau_\theta$. Умова рівності випливає з відповідної умови в нерівності Коші. Твердження зауваження для виродженої інформаційної матриці дістанемо після переходу до верхньої межі за b в останній нерівності ⊡

Вправа. Довести, що у випадку векторної функції $\tau(\theta)$ в умовах попередньої теореми для 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.

7.6. Приклад оцінювання 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

$$I(\theta) = D_\theta \nu_n(X)/\theta^2(1-\theta)^2 = n \;/\; \theta(1-\theta),$$

нижня границя для дисперсії незміщеної оцінки параметра θ має вигляд

$$D_\theta T \geq \theta(1-\theta) \;/\; n.$$

За критерієм ефективності Крамера – Рао ефективна оцінка повинна мати вигляд

$$T^* = \theta + c\left(\frac{\nu_n(X)}{\theta(1-\theta)} - \frac{n}{1-\theta}\right).$$

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 відносна частота успіхів $T^* = \frac{1}{n}\nu_n(X) = \frac{1}{n}\sum_{k=1}^n \chi_k$ є оптимальною незміщеною оцінкою ймовірності успіху θ.

7.7. Приклад оцінювання параметрів нормальних спостережень

7.7.1. 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

$$\ln L(X, \theta) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{k=1}^n (\xi_k - \mu)^2$$

$$= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}(\widehat{\mu}_{2n} - 2\widehat{\mu}_n\mu + \mu^2).$$

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, $T^* = \widehat{\mu}_n$ є оптимальною оцінкою середнього при відомій дисперсії.

7.7.2. Оцінювання дисперсії при відомому середньому

Розглянемо таку ж 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

$$T^* = \widetilde{\sigma}_n^2 = \widehat{\mu}_{2n} - 2\widehat{\mu}_n\mu + \mu^2 = \frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mu)^2.$$

7.7.3. 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 векторної 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

$$T^* - \mu = c_1 U_1(X, \theta) + c_2 U_2(X, \theta) = \widehat{\mu}_n - \mu$$

має місце рівність для $T^* = \widehat{\mu}_n$, тобто вибіркове середнє є оптимальною оцінкою і при невідомій дисперсії.

7.8. Експоненційна модель

Нехай $\Omega \subset \mathbb{R}^d$, $X = (\xi_1, ..., \xi_n)$ – кратна вибірка, причому щільність одного спостереження має вигляд

$$f(y, \theta) = \exp(a(y)b(\theta) + c(y) + d(\theta)),$$

де b, d – диференційовні функції параметра θ. За теоремою про функцію вірогідності кратної вибірки

$$\ln L(X, \theta) = A(X)nb(\theta) + C(X) + nd(\theta),$$

$$A(X) = \tfrac{1}{n}\sum_{k=1}^{n} a(\xi_k),\ C(X) = \sum_{k=1}^{n} c(\xi_k),$$

отже,

$$U(X,\theta) = A(X)n\beta(\theta) + nd_\theta(\theta),$$

де $\beta(\theta) = db(\theta)/d\theta$.

Припустимо, що виконані умови регулярності (основна умова автоматично випливає з додатності експоненційної функції на всій числовій осі). Внаслідок теореми про центрованість функції впливу звідси отримуємо

$$U(X,\theta) = n\beta(\theta)(A(X) - M_\theta A(X)).$$

Оскільки скалярна випадкова величина $A(X)$ є сумою незалежних у сукупності, однаково розподілених величин, то інформаційна матриця за Фішером має вигляд

$$I(\theta) = n^2\beta(\theta)\beta'(\theta)D_\theta A(X) = n\ \sigma^2(\theta)\beta(\theta)\beta'(\theta), \quad \sigma^2(\theta) = D_\theta a(\xi_1).$$

Рівність у критерії ефективності Крамера – Рао для оцінювання невідомої функції $\tau(\theta)$ має вигляд

$$T(X) - \tau(\theta) = nc'(\theta)\beta(\theta)(A(X) - M_\theta A(X)),$$

та виконується тоді й тільки тоді, коли $nc'(\theta)\beta(\theta) = 1$ для всіх θ. Очевидно, що відповідна функція $c(\theta)$ існує тоді й тільки тоді, коли $\beta(\theta) \neq 0\ \forall\theta$. У цьому випадку ефективна оцінка (і одночасно оптимальна) значення $\tau(\theta)$ існує, дорівнює

$$T^*(X) = A(X) = \frac{1}{n}\sum\nolimits_{k=1}^{n} a(\xi_k), \quad \tau(\theta) = M_\theta T^*(X) = M_\theta a(\xi_1),$$

і має дисперсію

$$D_\theta T^*(X) = D_\theta a(\xi_1)/n = \sigma^2(\theta)/n.$$

8. Достатні статистики та оптимальність

Означення. *Статистика* $T = T(X)$ *називаєтьcя* достатньою статистикою *для вибірки* X, *якщо* *вибіркова функція вірогідності* *має вигляд*

$$L(X,\theta) = g(T(X),\theta)\ h(X)\ j(\theta)$$

для деяких вимірних невід'ємних функцій g, h, j.

Зауважимо, що на розмірність достатньої статистики обмеження не накладаються, однак при виборі цієї статистики доцільно мінімізувати її розмірність. В усякому випадку, достатні статистики існують: наприклад, статистика $T(X) = X$ завжди є достатньою, якщо обрати $g = L, h \equiv 1, j \equiv 1$.

8.1. Приклади достатніх статистик

1. *Пуассонівська вибірка.* Якщо X – кратна вибірка з розподілом Пуассона $\xi_1 \simeq \Pi(\lambda)$, то

$$\ln L(X,\theta) = \ln \prod_{k=1}^{n} \frac{\lambda^{\xi_k}}{\xi_k!} \exp(-\lambda) =$$

$$(\ln\lambda) \sum_{k=1}^{n} \xi_k - \sum_{k=1}^{n} \ln(\xi_k!) - n\lambda,$$

тому існує скалярна достатня статистика $T(X) = \widehat{\mu}_n \equiv \frac{1}{n}\sum_{k=1}^{n} \xi_k$.

2. *Рівномірна вибірка.* Якщо X – кратна вибірка з $\xi_1 \simeq U(a,b)$, то

$$L(X,\theta) = (b-a)^{-n}\ 1_{\{a\ <\ \min\xi_k,\ \max\xi_k\ <\ b\}}.$$

Отже, двовимірна статистика
$T(X) = (\min_{1\le k\le n} \xi_k,\ \max_{1\le k\le n} \xi_k) = (\xi_{(1)}, \xi_{(n)})$ є достатньою.

3. *Нормальна вибірка.* Якщо X – 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.

8.2. Умовний розподіл вибірки

Теорема (про властивість достатньої статистики). *Якщо $T(X)$ – достатня статистика, то умовний розподіл*

$$P_\theta(X = x\ /T(X) = t) = l(x,t)$$

не залежить від θ.

Доведення проведемо в припущенні, що вибірка є дискретним випадковим вектором. Зауважимо, що дана умовна ймовірність дорівнює нулю (і не залежить від θ), якщо $T(x) \ne t$. Отже, можна припустити, що $T(x) = t$. За означенням умовної ймовірності

$$P_\theta(X = x\ /T(X) = t) = \frac{P_\theta(X = x,\ T(X) = t)}{P_\theta(T(X) = t)} =$$

$$\frac{P_\theta(X = x,\ T(x) = t)}{P_\theta(T(X) = t)} = L(x,\theta)\ / \sum\nolimits_{y:\ T(y)=t} L(y,\theta) =$$

$$g(T(x),\theta)\ h(x)\ j(\theta)\ / \sum\nolimits_{y:\ T(y)=t} g(T(y),\theta)\ h(y)\ j(\theta) =$$

$$g(t,\theta)h(x)\ / \sum\nolimits_{y:\ T(y)=t} g(t,\theta)\ h(y) = h(x)\ / \sum\nolimits_{y:\ T(y)=t} h(y) = l(x,t)\boxdot$$

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 $T_1(X) = H(T(X))$ від достатньої статистики $T(X)$, яка також є незміщеною 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

$$H(t) = M_\theta(T_0(X)\ /T(X) = t) \equiv \frac{M_\theta\left(T_0(X)\ 1_{\{T(X)=t\}}\right)}{P_\theta(T(X) = t)} =$$

$$\sum\nolimits_{x\,\in\,S} T_0(x)\ P_\theta(X = x\ /T(X) = t) = \sum\nolimits_{x\,\in\,S} T_0(x)\ l(x,t),$$

яка за теоремою про властивість достатньої статистики не залежить від параметра θ.

Визначимо статистику

$$T_1(X) = H(T(X)).$$

Вона є незміщеною, оскільки за формулою повної ймовірності та за означенням функцій $H(t)$ і $l(x,t)$

$$M_\theta T_1(X) = M_\theta H(T(X)) = \sum\nolimits_t H(t)\ P_\theta(T(X) = t) =$$
$$\sum\nolimits_t \left(\sum\nolimits_{x\,\in\,S} T_0(x)\ l(x,t)\right)\ P_\theta(T(X) = t) =$$
$$\sum\nolimits_{x\,\in\,S} T_0(x) \sum\nolimits_t P_\theta(X = x\ /T(X) = t)\ P_\theta(T(X) = t) =$$
$$\sum\nolimits_{x\,\in\,S} T_0(x) P_\theta(X = x) = M_\theta T_0(X) = \tau(\theta),\ \forall\theta \in \Theta.$$

Далі, обчислимо дисперсію

$$D_\theta T_0 = M_\theta (T_0 - H(T) + H(T) - \tau(\theta))^2 =$$
$$M_\theta (T_0 - H(T))^2 + D_\theta H(T) \geq D_\theta H(T) = D_\theta T_1,$$

оскільки за означенням $H(t)$ математичне сподівання добутку дорівнює нулю:

$$M_\theta (T_0 - H(T))(H(T) - \tau(\theta)) =$$
$$\sum{}_t M_\theta \left(T_0 - H(T)\right) \ \left(H(T) - \tau(\theta)\right) \ 1_{\{T\,=\,t\}} =$$
$$\sum{}_t M_\theta \left(T_0 - H(t)\right) \ \left(H(t) - \tau(\theta)\right) \ 1_{\{T\,=\,t\}} =$$
$$\sum{}_t \left(M_\theta T_0 \ 1_{\{T\,=\,t\}} - H(t) P_\theta (T = t)\right) \ \left(H(t) - \tau(\theta)\right) \ =$$
$$\sum{}_t \left(M_\theta (T_0 / T = t) - H(t)\right) \ P_\theta (T = t) \ \left(H(t) - \tau(\theta)\right) = 0.$$

Отже, $D_\theta T_0 \geq D_\theta T_1$ для довільної незміщеної оцінки T_0, що й доводить теорему $\boxdot$

Теорема (теорема Рао – Блекуела – Колмогорова). *Якщо існує оптимальна оцінка $T^* \in \Gamma_\tau$, а $T = T(X)$ – деяка достатня статистика, то $T^* = H^*(T)$ є вимірною функцією від цієї достатньої статистики.*

Доведення.

Оберемо в теоремі про дисперсію функції від достатньої статистики незміщену оцінку $T_0 \equiv T^*$, звідки $D_\theta T^* \geq D_\theta H(T)$ для вимірної функції H, що побудована в цій теоремі, причому $H(T) \in \Gamma_\tau$. Однак за означенням оптимальної оцінки $D_\theta H(T) \geq D_\theta T^*$. Отже, обидві нерівності є рівностями для всіх θ і кожна з оцінок $T^*, H(T)$ є оптимальною. З теореми про єдиність оптимальної оцінки виводимо, що $T^* = H^*(T)$ м.н. для $H^* \equiv H$ $\boxdot$

Означення. *Достатня статистика* $T = T(X)$ *називається* повною *статистикою, якщо для кожної* *вимірної* *функції* $\varphi(\cdot)$ *із тотожностей*

$$M_\theta\, \varphi(T) = 0, \quad \forall\theta \in \Theta,$$

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 $\tau(\theta) = M_\theta g(T(X))$. Якщо $T_0 \in \Gamma_\tau$ – довільна незміщена оцінка для $\tau = \tau(\theta)$, то внаслідок теореми про дисперсію функції від достатньої статистики знайдеться вимірна функція H така, що статистика $T_1 \equiv H(T)$ знову є незміщеною, причому $D_\theta T_1 \le D_\theta T_0$, $\forall\theta \in \Theta$. За умовою незміщеності $M_\theta T_1 = M_\theta H(T) = \tau(\theta)$, де за означенням $\tau(\theta) = M_\theta g(T)$. Звідси $M_\theta(H(T) - g(T)) = 0$, $\forall\theta \in \Theta$. Тому з повноти T випливає $H(T) - g(T) = 0$ м.н., тобто $T_1 \equiv H(T) = g(T)$ м.н., причому функція g вже не залежить від вибору T_0.

Отже, $D_\theta g(T) \le D_\theta T_0$, $\forall\theta \in \Theta$, $\forall T_0 \in \Gamma_\tau$, тобто $g(T)$ – оптимальна незміщена оцінка для $\tau(\theta)$ $\boxdot$

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. Ця статистика є повною, оскільки має функцію розподілу

$P_\theta(T < x) = (x/\theta)^n$, $0 \le x \le \theta$, та з рівностей

$M_\theta\ \varphi(T) = \theta^{-n}\int_0^\theta \varphi(x) n x^{n-1}dx = 0,\ \forall\theta > 0,$
випливає, що $\varphi(x) = 0$ майже для всіх $x \ge 0$.

Отже, статистика $T(X) = \max \xi_k$ є оптимальною незміщеною оцінкою свого математичного сподівання $M_\theta\ T(X) = \tau(\theta) = \frac{n\theta}{n+1}$. Слід зазначити, що дисперсія цієї оцінки дорівнює (**вправа**)

$$D_\theta T = \theta^2 n/(n+2)(n+1)^2 = O(1/n^2),\ n \to \infty,$$

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.

Вправа.

(1) Для кратної нормальної вибірки з невідомим середнім та відомою дисперсією вибіркове середнє є повною достатньою статистикою.

(2) Для пуассонівської кратної вибірки з невідомим параметром θ вибіркова сума $\widehat{S}_n$ є повною достатньою статистикою. Зокрема, оптимальною оцінкою для $\sum_{k=1}^m c_k\theta^k$ є статистика $\sum_{k=1}^m c_k\pi_k(\widehat{S}_n)n^{-k}$, де $\pi_k(s) = s(s-1)..(s-k+1)$.

(3) Для кратної вибірки з (а) геометричним розподілом, (б) показниковим розподілом вибіркова сума є повною достатньою статистикою.

(4) 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) λ/α, (б) α^k при заданому k.

(5) Для кратної вибірки X із неперервною функцією розподілу спостережень її варіаційний ряд є повною достатньою статистикою.

9. Оцінки максимальної вірогідності

Визначення оцінки максимальної вірогідності грунтується 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 цієї умови повністю визначена вибіркова функція вірогідності $L(X,\ \theta)$. Значення цієї функції і дають критерій "найбільшої ймовірності".

Означення. Оцінкою максимальної вірогідності (ОМВ, *EML*) *параметру* θ *за вибіркою* X *називається статистика*

$$\widehat{\theta} = \widehat{\theta}(X) \equiv \arg\max_{\ \theta \in \Theta} L(X,\ \theta),$$

де $L(X,\theta)$ *– вибіркова функція вірогідності.*

У випадку кратної вибірки ОМВ позначається як $\widehat{\theta}_n$, де n – об'єм вибірки.

Необхідною умовою існування ОМВ є припущення існування точки максимуму. ОМВ визначена однозначно за умови єдиності цієї точки.

Теорема (про рівняння максимальної вірогідності). *Якщо* $\Theta \subset \mathbb{R}^d$, *максимум в означенні оцінки максимальної вірогідності досягається всередині параметричної множини, а функція вірогідності диференційовна, то ОМВ задовольняє рівняння максимальної вірогідності*

$$\frac{\partial}{\partial\theta} \ln L(X,\theta)\ |_{\theta\,=\,\widehat{\theta}} = 0,$$

тобто ОМВ є коренем функції впливу $U(X,\theta)$:

$$U(X,\widehat{\theta}) = 0.$$

У випадку векторного параметра 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 $U(X,\widehat{\theta}) = 0$ часто використовують метод Ньютона – Рафсона, який складається з двох етапів.

(1) Обирають якусь консистентну оцінку $\widehat{\theta}_0 = \widehat{\theta}_0(X)$.

(2) Рекурентно обчислюють

$$\widehat{\theta}_{m+1} = \widehat{\theta}_m + U(X,\widehat{\theta}_m) \;/\; I(\widehat{\theta}_m),$$

де $I(\theta)$ – функція інформації за Фішером.

Доведено, що за певних умов границя $\lim \widehat{\theta}_m$ існує та збігається з ОМВ.

9.1. Співвідношення з ефективними оцінками. Інваріантність

Теорема (про співвідношення ОМВ з ефективними оцінками).

(а) Якщо параметр θ – скалярний і існує його ефективна оцінка θ^*, то існує оцінка максимальної вірогідності $\widehat{\theta}$, причому $\widehat{\theta} = \theta^*$.

(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 θ^*) похідна за аргументом θ від логарифмічної функції вірогідності має вигляд

$$\frac{\partial}{\partial\theta} \ln L(X,\theta) \equiv U(X,\theta) = (\theta^* - \theta)/c(\theta),$$

де функція $c(\theta) = \tau_\theta / I(\theta) = 1/I(\theta)$ невід'ємна. Тому логарифмічна функція вірогідності зростає при $\theta < \theta^*$ і спадає при $\theta > \theta^*$. Отже, вона набуває максимального значення при $\theta = \theta^*$ і θ^* 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 $T(X)$ $\boxdot$

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 ОМВ для значення* $q(\theta)$.

Доведення.

Позначимо через $\theta(q)$ функцію, обернену до $q(\theta)$. Функція вірогідності для значення параметра q має вигляд $L_q(X,q) = L(X,\theta(q))$ і за умови однозначності

$$L_q(X,q(\theta)) = L(X,\theta(q(\theta))) = L(X,\theta),\ \forall\theta\in\Theta.$$

Тому з означення ОМВ дістаємо

$$L_q(X,q(\theta)) = L(X,\theta) \le L(X,\widehat{\theta}) = L_q(X,q(\widehat{\theta})),$$
$$L_q(X,q) \le \max_{\theta\in\Theta} L_q(X,q(\theta)) \le L_q(X,q(\widehat{\theta})),$$

оскільки множина значень $q(\theta), \theta\in\Theta$, збігається з Q. Останні нерівності є рівностями при $q = q(\widehat{\theta})$. Отже, $q(\widehat{\theta})$ є ОМВ для $q(\theta)$ $\boxdot$

9.2. Приклади обчислення оцінок максимальної вірогідності

1. *Схема Бернуллі:*

$$\arg\max \ln L(X,\theta) = \arg\max \ln \theta^{\nu_n(X)}(1-\theta)^{n-\nu_n(X)} =$$

$$\arg\max(\theta \ln \nu_n(X) + (1-\theta)\ln(n-\nu_n(X))) = \nu_n(X)/n = \widehat{\theta}_n.$$

2. *Пуассонівська вибірка:*

$$\ln L(X,\theta) = \ln \prod_{k=1}^{n} \frac{\theta^{\xi_k}}{\xi_k!} e^{-\theta} = \ln\theta \sum_{k=1}^{n} \xi_k - n\theta + \ln h(X),$$

$$\widehat{\theta}_n = \arg\max \ln L(X,\theta) = \frac{1}{n}\sum_{k=1}^{n} \xi_k = \widehat{\mu}_n.$$

3. *Показникова вибірка* (зміщеність ОМВ):

$$\ln L(X,\theta) = \ln \prod_{k=1}^{n} \theta \exp(-\theta\xi_k) = n\ln\theta - \theta\sum_{k=1}^{n} \xi_k,$$

$$\widehat{\theta}_n = \arg\max \ln L(X,\theta) = \frac{n}{\sum_{k=1}^{n}\xi_k} = \frac{1}{\widehat{\mu}_n}.$$

Зауважимо, що сума $\sum_{k=1}^{n}\xi_k$ має розподіл Ерланга з параметрами n, θ, тому

$$M_\theta \widehat{\theta}_n = \int_0^\infty \frac{n}{x}\theta\frac{(\theta x)^{n-1}}{(n-1)!}\exp(-\theta x)dx = \theta\frac{n}{n-1} \neq \theta \ \forall n \geq 1 \ \forall \theta > 0.$$

Отже, у загальному випадку не можна розраховувати на незміщеність ОМВ.

4. *Нормальна вибірка.* Як показано вище в розділі про оцінювання середнього для нормальних спостережень,

$$\ln L(X,\theta) - \frac{n}{2}\ln 2\pi = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}(\widehat{\mu}_{2n} - 2\widehat{\mu}_n\mu + \mu^2) =$$

$$-\frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}(\widehat{\mu}_{2n} - (\widehat{\mu}_n)^2 + (\widehat{\mu}_n - \mu)^2) =$$

$$-\frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}(\widehat{\sigma}_n^2 + (\widehat{\mu}_n - \mu)^2) \le -\frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}\widehat{\sigma}_n^2 =$$

$$-\frac{n}{2}\left(\ln\sigma^2 + \frac{1}{\sigma^2}\widehat{\sigma}_n^2\right) \le -\frac{n}{2}\left(\ln\widehat{\sigma}_n^2 + \frac{1}{\widehat{\sigma}_n^2}\widehat{\sigma}_n^2\right) = -\frac{n}{2}(\ln\widehat{\sigma}_n^2 + 1),$$

оскільки функція $\ln s\ + \frac{a}{s}$ набуває найменшого значення при $s = a$. Дані нерівності є рівностями при $\mu = \widehat{\mu}_n, \sigma^2 = \widehat{\sigma}_n^2$. Тому оцінка максимальної вірогідності параметра $\theta = (\mu, \sigma^2)$ дорівнює

$$\widehat{\theta}_n \equiv \arg\max \ln L(X,\theta) = (\widehat{\mu}_n,\ \widehat{\sigma}_n^2).$$

Зокрема, за теоремою про інваріантність оцінки максимальної вірогідності для границь вірогідного інтервалу з правила 3 сигма має вигляд

$$\widehat{\mu \pm 3\sigma} = \widehat{\mu}_n \pm 3\sqrt{\widehat{\sigma}_n^2}.$$

Вправа.

(1) Довести, що ОМВ центра симетрії θ розподілу із щільністю 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 нормальним розподілом $N(\theta, 2\theta)$. Довести, що ОМВ параметра θ має вигляд $\sqrt{1 + \sum_{k=1}^n \xi_k^2\ /n} - 1$ і є консистентною.

(4) Знайти ОМВ значення функції $\tau(\theta) = \Phi(-\mu/\sigma) \equiv P_\theta(\xi_1 < 0)$ для нормальної кратної вибірки $X = (\xi_1, ..., \xi_n)$, $\xi_k \simeq N(\mu, \sigma^2)$ із невідомим параметром $\theta = (\mu, \sigma^2)$.

(5) Знайти ОМВ для невідомої кількості спостережень n за біноміальною одноелементною вибіркою $X \simeq B(n, p)$.

(6) Довести, що ОМВ для вибірки з рівномірним розподілом $U(\theta, 1 + \theta)$ утворюють цілий інтервал.

9.3. Умови консистентності ОМВ

Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка зі щільністю $f(y, \theta)$ одного спостереження.

У даному розділі символом θ_0 будемо позначати істинне значення параметра θ.

Розглянемо такі умови консистентності ОМВ:

(c1) Для всіх $\theta \neq \theta_0$ відношення $f(y, \theta)/f(y, \theta_0)$ не дорівнює майже скрізь одиниці на просторі значень спостережень $y \in R$.

(c2) Параметрична множина Θ є компактною підмножиною евклідового простору $\mathbb{R}^d$.

(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 :

$$|\ln f(\xi_1, \theta)| \le \eta, \quad \forall \theta \in \Theta,$$

$$|\partial \ln f(\xi_1, \theta)/\partial\theta| \le \eta, \ M_\theta \eta < \infty, \ \forall \theta \in \Theta.$$

9.4. Інформація за Кульбаком

Означення. *Функцією* інформації за Кульбаком *називається числова функція*

$$I(\theta,\theta_0) = -\ M_{\theta_0} \ln \frac{f(\xi_1,\theta)}{f(\xi_1,\theta_0)}.$$

Теорема (про властивості інформації за Кульбаком). *За умов консистентності ОМВ:*

(а) $I(\theta,\theta_0) \ge 0$,

(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 теореми про обчислення математичного сподівання функції від випадкової величини та з означення щільності $f(y,\theta)$ обчислимо

$$M_{\theta_0}\zeta = \int \frac{f(y,\theta)}{f(y,\theta_0)} f(y,\theta_0)\nu(dy) = \int f(y,\theta)\nu(dy) = 1.$$

З опуклості функції $\ln x$ виводимо елементарну нерівність $\ln x \le x-1$ при $x \in \mathbb{R}$, звідки за монотонністю математичного сподівання

$$I(\theta,\theta_0) = -\ M_{\theta_0} \ln\zeta \ge -\ M_{\theta_0}(\zeta - 1) = 0.$$

Так як нерівність $\ln x \le x-1$ є строгою при $x \ne 1$, то за властивістю позитивності математичного сподівання від невід'ємної та ненульової м.н. випадкової величини $\zeta - 1 - \ln\zeta$ інформація за Кульбаком дорівнює нулю тоді й тільки тоді, коли $\zeta = 1$ м.н., тобто коли $f(y,\theta) = f(y,\theta_0)$ майже скрізь на просторі значень спостережень.

Неперервність $I(\theta,\theta_0)$ випливає з неперервності щільності спостережень та з теореми Лебега про мажоровану збіжність:

$$\lim_{n\to\infty} I(\theta_n,\theta_0) = -\lim_{n\to\infty} M_{\theta_0} \ln \frac{f(\xi_1,\theta_n)}{f(\xi_1,\theta_0)} = -M_{\theta_0} \lim_{n\to\infty} \ln \frac{f(\xi_1,\theta_n)}{f(\xi_1,\theta_0)} = I(\theta,\theta_0),$$

оскільки величини під знаком математичного сподівання за умов консистентності не перевищують інтегровну величину 2η $\boxdot$

9.5. Консистентність ОМВ

Теорема (про консистентність ОМВ). *За умов консистентності ОМВ (c1)-(c3) оцінка максимальної вірогідності $\widehat{\theta}_n$ параметра θ за кратною вибіркою X є строго консистентною.*

Доведення. Позначимо

$$\chi_k(\theta) = \ln \frac{f(\xi_k,\theta_0)}{f(\xi_k,\theta)},\ \chi_k(U) = \inf_{\theta\in U} \chi_k(\theta),$$

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

$$I(\theta,\theta_0) = M_{\theta_0}\chi_1(\theta) \ge 0.$$

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 $\theta \in U$

$$|\chi_1(U) - \chi_1(\theta)| \le (diamU) \sup_{\theta \in U} |\partial \ln f(\xi_1, \theta)/\partial\theta| \le \eta\ diamU.$$

Тоді за теоремою Лебега про мажоровану збіжність

$$M_{\theta_0}\ \chi_1(U) \to M_{\theta_0}\ \chi_1(\theta) = I(\theta, \theta_0) > 0, \quad \forall\theta \ne \theta_0.$$

Більше того, з умови мажорованості похідної випливає, що

$$|M_{\theta_0}\ \chi_1(U) - M_{\theta_0}\ \chi_1(\theta)| \le M_{\theta_0}\ |\chi_1(U) - \chi_1(\theta)| \le$$
$$(diamU)\ M_{\theta_0} \sup\nolimits_{\theta \in U}\ |\partial \ln f(\xi_1, \theta)/\partial\theta| \le K\ diamU,$$

де стала $K = M_{\theta_0}\eta < \infty$ за умовою (с3).

Нехай $\varepsilon > 0$. Визначимо замкнену і компактну внаслідок (с2) множину

$$\Theta(\varepsilon) \equiv \Theta \cap \{\theta : |\theta - \theta_0| \ge \varepsilon\}.$$

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 порушенні цієї умови в множині $\Theta(\varepsilon)$ знайдеться точка θ (гранична для підпослідовності 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 $diam\ U_j(\delta) \le \delta$. Оберемо $\delta < \alpha/2K$. Тоді за наведеною вище нерівністю та за означенням нижньої межі при виборі довільних $\theta_j \in U_j$

$$M_{\theta_0}\ \chi_1(U_j(\delta)) \ge M_{\theta_0}\ \chi_1(\theta_j) - |M_{\theta_0}\chi_1(U_j(\delta)) - M_{\theta_0}\chi_1(\theta_j)| \ge$$
$$I(\theta_j,\theta_0) - K\ diam U_j(\delta) \ge \alpha - K\delta \ge \alpha/2 > 0, \quad \forall j = \overline{1,m}.$$

Очевидно, що з виконання для всіх $\theta \in \Theta(\varepsilon)$ нерівностей $L(X,\theta_0) > L(X,\theta)$ 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 включення

$$\{\left|\widehat{\theta}_n - \theta_0\right| < \varepsilon\} = \{\ \arg\max_{\theta\in\Theta} L(X,\theta) \notin \Theta(\varepsilon)\} \supset \left\{\inf_{\theta\,\in\,\Theta(\varepsilon)} \tfrac{L(X,\theta_0)}{L(X,\theta)} > 1\right\} =$$
$$\left\{\inf_{\theta\,\in\,\Theta(\varepsilon)} \ln \tfrac{L(X,\theta_0)}{L(X,\theta)} > 0\right\} = \left\{\inf_{\theta\,\in\,\Theta(\varepsilon)} \ln \prod_{k=1}^n \tfrac{f(\xi_k,\theta_0)}{f(\xi_k,\theta)} > 0\right\} =$$
$$\left\{\inf_{\theta\,\in\,\Theta(\varepsilon)} \sum_{k=1}^n \chi_k(\theta) > 0\right\} = \left\{\min_{1\le j\le.m} \inf_{\theta\,\in\,U_j(\delta)} \sum_{k=1}^n \chi_k(\theta) > 0\right\}.$$

Отже, за монотонністю ймовірності

$$P_{\theta_0}\left(\left|\widehat{\theta}_n - \theta_0\right| < \varepsilon\right) \ge P_{\theta_0}\left(\min_{1\le j\le.m} \inf_{\theta\,\in\,U_j(\delta)} \sum_{k=1}^n \chi_k(\theta) > 0\right) \ge$$
$$P_{\theta_0}\left(\min_{1\le j\le.m} \sum_{k=1}^n \chi_k(U_j(\delta)) > 0\right) =$$
$$P_{\theta_0}\left(\bigcap_{j=1}^m \left\{\tfrac{1}{n}\sum_{k=1}^n \chi_k(U_j(\delta)) > 0\right\}\right) \to 1,\ n\to\infty,$$

оскільки за критерієм Колмогорова посиленого закону великих чисел при кожному j

$$\frac{1}{n}\sum_{k=1}^n \chi_k(U_j(\delta)) \to^{P1} M_{\theta_0}\ \chi_1(U_j(\delta)) \ge \alpha/2 > 0,\ n\to\infty,$$

а скінченний переріз подій одиничної ймовірності має ймовірність 1.

Отже, ОМВ – консистентна.

Застосовуючи наведені вище міркування до подій

$$\underline{\lim}_{\ n\to\infty} \left\{\left|\widehat{\theta}_n - \theta_0\right| < \varepsilon\right\} \supset \underline{\lim}_{\ n\to\infty} \bigcap_{j=1}^{m} \left\{\frac{1}{n}\sum_{k=1}^{n} \chi_k(U_j(\delta)) > 0\right\},$$

аналогічно виводимо нерівність

$$P_{\theta_0}\left(\underline{\lim}_{\ n\to\infty}\{\left|\widehat{\theta}_n - \theta_0\right| < \varepsilon\}\right) \geq$$
$$P_{\theta_0}\left(\underline{\lim}_{\ n\to\infty} \bigcap_{j=1}^{m} \left\{\frac{1}{n}\sum_{k=1}^{n} \chi_k(U_j(\delta)) > 0\right\}\right),$$

де права частина дорівнює 1 за критерієм Колмогорова посиленого закону великих чисел. Що і доводить строгу консистентність ОМВ ⊡

9.6. Асимптотична нормальність та 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 $f(y, \cdot)$ задовольняє умови регулярності, умови консистентності ОМВ (c1-c3) та тричі неперервно диференційовна за θ, причому відповідні похідні мажоруються за модулем інтегровною величиною η :*

$$\left|\partial^{(i)} \ln f(\xi_1, \theta)/\partial^{(i)}\theta\right| \leq \eta, \quad i = \overline{1,3}, \qquad M_{\theta_0}\eta < \infty.$$

Якщо існує строго консистентна ОМВ $\widehat{\theta}_n$ для значення невідомого параметра θ_0, яка 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:

$$i(\theta) = M_\theta \left(\frac{\partial}{\partial\theta} \ln f(\xi_1, \theta) \right)^2 .$$

Зауваження. Оскільки повна інформація за Фішером у вибірці за теоремою про адитивність.інформації за Фішером дорівнює $I_n(\theta) = n\, i(\theta)$, то з даної теореми випливає асимптотична ефективність ОМВ, оскільки величина $1/n\ i(\theta_0) = 1/I_n(\theta)$ 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:

$$U^{(i)}(X,\theta) = \partial^{(i)} \ln L_n(X,\theta)/\partial^{(i)}\theta,$$
$$u^{(i)}(y,\theta) = \partial^{(i)} \ln f(y,\theta)/\partial^{(i)}\theta,$$
$$\zeta_k^{(i)}(\theta) = u^{(i)}(\xi_k,\theta).$$

За теоремою про функцію вірогідності кратної вибірки має місце адитивність:

$$U^{(i)}(X,\theta) = \sum\nolimits_{k=1}^{n} \zeta_k^{(i)}(\theta), \quad \forall j = \overline{1,3},$$

де доданки в сумі є незалежними у сукупності та однаково розподіленими випадковими величинами, як невипадкові однакові функції від незалежних та однаково розподілених випадкових величин ξ_k. За зробленим у теоремі припущенням всі ці доданки мажоруються інтегровною величиною η.

Зауважимо, що за теоремою про рівняння максимальної вірогідності ОМВ є коренем функції впливу $U = U^{(1)}$. Звідси за формулою розкладу в ряд Тейлора

$$0 = U(X, \widehat{\theta}_n) \equiv U^{(1)}(X, \widehat{\theta}_n) = U^{(1)}(X, \theta_0) +$$

$$(\widehat{\theta}_n - \theta_0) U^{(2)}(X, \theta_0) + 0.5 \cdot (\widehat{\theta}_n - \theta_0)^2 U^{(3)}(X, \widetilde{\theta}_n),$$

де $\widetilde{\theta}_n \in [\widehat{\theta}_n, \theta_0]$ і $\widetilde{\theta}_n \to^{P1} \theta_0$ внаслідок строгої консистентності ОМВ.

Розв'язуючи останню тотожність відносно $\widehat{\theta}_n - \theta_0$, дістанемо зображення

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = -\frac{\alpha_n}{\beta_n + (\widehat{\theta}_n - \theta_0)\ \gamma_n},$$

де

$$\alpha_n = \frac{1}{\sqrt{n}} U^{(1)}(X, \theta_0) = \frac{1}{\sqrt{n}} \sum\nolimits_{k=1}^{n} \zeta_k^{(1)}(\theta_0),$$

$$\beta_n = \frac{1}{n} \sum\nolimits_{k=1}^{n} \zeta_k^{(2)}(\theta_0), \quad \gamma_n = \frac{1}{2n} \sum\nolimits_{k=1}^{n} \zeta_k^{(3)}(\widetilde{\theta}_n).$$

За теоремою про центрованість функції впливу та про обчислення інформації за Фішером (в умовах регулярності)

$$M_{\theta_0}\ \zeta_k^{(1)}(\theta_0) = M_{\theta_0}\ u(\xi_1, \theta_0) = 0,$$

$$D_{\theta_0}\ \zeta_k^{(1)}(\theta_0) = D_{\theta_0}\ u(\xi_1, \theta_0) = i(\theta_0).$$

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

$$M_{\theta_0}\zeta_k^{(2)}(\theta_0) = M_{\theta_0}\partial u(\xi_1, \theta_0)/\partial\theta_0 = -M_{\theta_0}(u(\xi_1, \theta_0))^2 = -i(\theta_0),$$

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 похідних логарифмічної функції вірогідності та з посиленого закону великих чисел випливає також обмеженість майже напевне послідовності γ_n. Тому зі строгої консистентності ОМВ виводимо, що

$$(\widehat{\theta}_n - \theta_0)\,\gamma_n \to^{P1} 0,\ n \to \infty.$$

Отже, знаменник у наведеному вище зображенні для $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ має границю

$$\beta_n + (\widehat{\theta}_n - \theta_0)\,\gamma_n \to^{P1} -i(\theta_0),\ n \to \infty.$$

Тому за теоремою про добуток слабко збіжної послідовності зі збіжною

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to^{W} -\frac{\alpha}{-i(\theta_0)} \simeq N(0,\ i^{-1}(\theta_0)),\ n \to \infty \boxdot$$

Вправа. Для кратної вибірки з гама-розподілом $\Gamma(1, \theta)$ спостережень вивести рівняння ОМВ та довести, що ОМВ є асимптотично нормальною та асимптотично ефективною. Знайти оцінку методу моментів та довести, що її ефективність менша за 1.

10. Статистики від нормальних вибірок

Важлива особливість нормальних спостережень полягає в тому, що вибіркове середнє та нормована вибіркова дисперсія незалежні й мають наперед відомі розподіли.

Нагадаємо, що ортонормованою називається матриця U така, що $U^{-1} = U'$.Усі вектори розглядатимемо як вектори-стовпчики, 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.

10.1. Стандартна нормальна вибірка

Теорема (про вибіркові моменти стандартної нормальної вибірки). *Нехай вибірка $\zeta = (\zeta_1, ..., \zeta_n)$ є стандартним нормальним вектором. Тоді вибіркові моменти*

$$\widehat{\mu}_n = \frac{1}{n}\sum\nolimits_{k=1}^{n} \zeta_k, \quad \widehat{s}^2_n = \frac{1}{n-1}\sum\nolimits_{k=1}^{n} (\zeta_k - \widehat{\mu}_n)^2$$

незалежні, причому

$$\sqrt{n}\, \widehat{\mu}_n \simeq N(0,1), \quad (n-1)\, \widehat{s}^2_n \simeq \chi^2_{n-1},$$

де χ^2_{n-1} – випадкова величина з хі-квадрат розподілом, що має $n-1$ ступінь свободи.

Доведення.

Визначимо вектор $q = (1/\sqrt{n},\ k = \overline{1,n})$, що має одиничну евклідову норму. Тоді $\widehat{\mu}_n = \zeta' q/\sqrt{n}$ і

$$(n-1)\widehat{s}^2_n = \sum\nolimits_{k=1}^{n} (\zeta_k - \widehat{\mu}_n)^2 = \sum\nolimits_{k=1}^{n} \zeta^2_k - n\, \widehat{\mu}^2_n = \zeta'\zeta - (\zeta' q)(q'\zeta).$$

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. Тому

$$(n-1)\widehat{s}_n^2 = \zeta'\zeta - (\zeta' q)(q'\zeta) = \zeta'(I - q \cdot q')\zeta =$$

$$\eta' U(I - q \cdot q')U'\eta = \eta'(I - e \cdot e')\eta = \sum\nolimits_{k=1}^{n-1} \eta_k^2 \simeq \chi_{n-1}^2$$

за означенням хі-квадрат розподілу, оскільки $\eta_k, \ k = \overline{1, n-1}$, є незалежними у сукупності величинами із стандартним нормальним розподілом.

За теоремою про векторні перетворення незалежних величин статистика χ_{n-1}^2 не залежить від $\eta_n = q'\zeta = \sqrt{n}\widehat{\mu}_n$, оскільки величини $(\eta_1, ..., \eta_{n-1})$ та η_n незалежні внаслідок означення стандартного нормального вектора $\eta = (\eta_1, ..., \eta_n)$.

Нарешті, середнє $\widehat{\mu}_n$ як лінійне перетворення за теоремою про лінійні перетворення нормальних векторів має нормальний розподіл, із середнім 0 і дисперсією $D\widehat{\mu}_n = n/n^2 = 1/n$ $\boxdot$

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, причому*

$$\sqrt{n}\,\frac{\widehat{\mu}_n - \mu}{\sigma} \simeq N(0, 1), \quad \frac{(n-1)\widehat{s}_n^2}{\sigma^2} \simeq \chi_{n-1}^2,$$

$$n\,\frac{\widehat{\sigma}_n^2}{\sigma^2} = \frac{1}{\sigma^2}\sum\nolimits_{k=1}^{n}(\xi_k-\mu)^2 \simeq \chi_n^2.$$

Доведення.

Розглянемо лінійне перетворення $\zeta = (\zeta_k = (\xi_k-\mu)/\sigma,\ \ k=\overline{1,n})$. Вектор ζ є стандартним нормальним вектором, тому що за означенням величини $(\xi_k-\mu)/\sigma$ є незалежними стандартними нормальними. 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 із сумою $\sum_{k=1}^{n}\zeta_k^2$ і має хі-квадрат розподіл за означенням $\boxdot$

Вправа.

(1) Вивести з центральної граничної теореми, що
$(\chi_n^2-n)/\sqrt{2n} \to^W \zeta \simeq N(0,1),\ \ n\to\infty.$

(2) Довести апроксимацію Фішера:
$\sqrt{2\chi_n^2}-\sqrt{2n} \to^W \zeta \simeq N(0,1),\ \ n\to\infty.$

10.3. Статистики Стьюдента та Фішера

Означення. *Випадкова величина τ_n має* t-розподіл, *або* розподіл Стьюдента, *із 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 із хі-квадрат розподілом та n ступенями свободи.

Вправа.

(1) Виходячи з виразу для щільності хі-квадрат як часткового випадку гама-щільності, довести, що щільність τ_n має вигляд

$$f_{\tau_n}(x) = c_n(1 + x^2/n)^{-(n+1)/2}, \;\; c_n = (\pi n)^{-1/2}\Gamma((n+1)/2)/\Gamma(n/2),$$

причому $M\tau_n = 0, \;\; D\tau_n = n/(n-2), n > 2.$

(2) Довести, що $\tau_n \to^W \zeta \simeq N(0,1), \;\; n \to \infty.$

Означення. *Випадкова величина* $\phi_{n,m}$ *має* розподіл Фішера, *або ж* Снедекора – Фішера, *із* 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 із* хі-квадрат розподілом *та* n, m *ступенями свободи відповідно.*

Вправа. Довести, що щільність $\phi_{n,m}$ має вигляд

$$f_{nm}(x) = c_{nm}x^{n/2-1}\left(1 + \tfrac{n}{m}x\right)^{-(n+m)/2}, \; c_{nm} = (\tfrac{n}{m})^{n/2}B(\tfrac{n}{2}, \tfrac{m}{2}).$$

Теорема (про статистику Стьюдента від нормальної вибірки). *Нехай* $X = (\xi_1, ..., \xi_n)$ *–* кратна вибірка *з нормальним розподілом* $\xi_k \simeq N(\mu, \sigma^2)$*. Тоді* статистика

$$\tau_{n-1} = \sqrt{n}\,\frac{\widehat{\mu}_n - \mu}{\widehat{s}_n}$$

має розподіл Стьюдента *із* $n-1$ *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 $\boxdot$

Старт

Початок

Зміст

◀◀ ▶▶

◀ ▶

Назад

Екран

Закрити

Выхід

11. Інтервальні оцінки параметрів нормальних спостережень

Розглянемо лише двосторонні інтервальні оцінки для параметрів кратної нормальної вибірки $X = (\xi_1, ..., \xi_n)$, $\xi_k \simeq N(\mu, \sigma^2)$, односторонні оцінки виводяться аналогічно.

Нижче x_α означатиме двосторонній квантиль вірогідного рівня $1 - \alpha$ для стандартного нормального розподілу:

$$P(|\zeta| \leq x_\alpha) = 1 - \alpha, \ \Phi(x_\alpha) = 1 - \alpha/2.$$

Аналогічний зміст мають квантилі $y_{n\alpha}$ для розподілу Стьюдента, $z_{n\alpha}$ – для хі-квадрат розподілу:

$$P(|\tau_n| \leq y_{n\alpha}) = 1 - \alpha, \quad P(\chi_n^2 \leq z_{n\alpha}) = 1 - \alpha/2.$$

Доведення наведених нижче оцінок є очевидним наслідком теореми про вибіркові середнє та дисперсію нормальної вибірки, та означення відповідних квантилей.

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

$$P\left(\widehat{\mu}_n - \frac{\sigma}{\sqrt{n}}x_\alpha \leq \mu \leq \widehat{\mu}_n + \frac{\sigma}{\sqrt{n}}x_\alpha\right) = P(|\zeta| \leq x_\alpha) = 1 - \alpha.$$

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

$$P(n\widehat{\sigma}_n^2 \;/\; z_{n\alpha} \le \sigma^2 \le n\widehat{\sigma}_n^2 \;/\; z_{n,2-\alpha}) =$$
$$P(z_{n,2-\alpha} \le \chi_n^2 \le z_{n\alpha}) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha.$$

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

$$P\left(\widehat{\mu}_n - \frac{\widehat{s}_n}{\sqrt{n}} y_{n-1,\alpha} \le \mu \le \widehat{\mu}_n + \frac{\widehat{s}_n}{\sqrt{n}} y_{n-1,\alpha}\right) = P(|\tau_n| \le y_{n-1,\alpha}) = 1 - \alpha.$$

11.4. Оцінка дисперсії при невідомому середньому

Оскільки $(n-1)\;\widehat{s}_n^2/\sigma^2 \simeq \chi_{n-1}^2$, то

$$P((n-1)\widehat{s}_n^2 \;/\; z_{n-1,\alpha} \le \sigma^2 \le (n-1)\widehat{s}_n^2 \;/\; z_{n-1,2-\alpha}) =$$
$$P(z_{n-1,2-\alpha} \le \chi_{n-1}^2 \le z_{n-1,\alpha}) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha.$$

Вправа. Нехай $X = (\xi_1, ..., \xi_n)$ кратна нормальна вибірка з розподілом $\xi_k \simeq N(\mu, \sigma^2)$, $\Phi_{\mu\sigma}$ нормальна функція розподілу з параметрами μ, σ^2, а $\widehat{\mu}_n, \widehat{\sigma}_n^2$ – вибіркові середнє та дисперсія. 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 $M\Delta_t = \alpha$.

12. Задачі перевірки 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 P_θ. Тут θ – істинне значення параметра, а $(P_\theta,\ \theta \in \Theta)$ – параметрична сім'я розподілів з означення статистичного простору.

Довільне припущення про розподіл вибірки при істинному значенні параметра θ називається статистичною гіпотезою. Часто статистичні гіпотези формулюють у вигляді

$$H_0 : \theta \in \Theta_0,$$

де $\Theta_0 \subset \Theta$ – певна параметрична підмножина. Якщо параметричний простір збігається з множиною всіх можливих розподілів вибірки, то будь-яка гіпотеза може бути зображена у наведеному вигляді.

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 називається таке припущення про розподіл вибірки, яке вважається завжди виконаним у випадку, коли не справджується основна статистична гіпотеза. Статистичні альтернативи формулюють у вигляді

$$H_1 : \theta \in \Theta_1,$$

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 $P_{\theta_0}(X \in \cdot)$ відомий повністю.*

Приклади

1. *Нормальна вибірка з відомою дисперсією.* В якості невідомого параметра нормального розподілу виступає середнє: $\theta = \mu$, дисперсія σ^2 вважається відомою. При фіксації μ щільність спостережень повністю відома, тому гіпотеза $H_0 : \theta = \mu_0$ є простою.

2. *Нормальна вибірка з невідомою дисперсією.* Якщо параметр $\theta = (\mu, \sigma^2)$, то при виконанні гіпотези $H_0 : \theta = \mu_0$ вибіркові щільності можуть мати різні дисперсії, тому гіпотеза не є простою.

12.1. Статистика критерію, критична область

Звичайно статистичний висновок щодо вибірки робиться на підставі аналізу 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. Для визначення результату статистичного висновку щодо якісної (дихотомічної) властивості розподілу спостережень на основі значення $\widehat{\varkappa}$ досить розбити множину $D = D_0 \cup D_1$ на дві частини й у випадку

(0) включення $\widehat{\varkappa} \in D_0$ – приймати нульову гіпотезу,

(1) при $\widehat{\varkappa} \in D_1$ – приймати альтернативу, тобто 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 число x_0 визначає критичний рівень статистики критерію.

У загальному випадку з критичною областю статистики D_1 можна пов'язати критичну область вибірки

$$W = \{x \in S : \widehat{\varkappa}(x) \in D_1\}.$$

При потраплянні вибіркового значення X у критичну область нульова гіпотеза відкидається, в протилежному випадку – відкидається альтернатива.

Означення. Статистичним критерієм *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,

(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.

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 випадковими подіями. Ймовірності цих подій 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 *статистичного критерію* $(\widehat{\varkappa}, D_1)$ *називається функція від* θ, *що задає ймовірності похибок першого роду:*

$$P_\theta(\widehat{\varkappa}(X) \in D_1) = P_\theta(X \in W),\ \theta \in \Theta_0.$$

Потужністю *критерію називається доповнення ймовірності похибки другого роду до одиниці (тобто ймовірність правильного – альтернативного – висновку при альтернативі), що задається функцією:*

$$1 - P_\theta(\widehat{\varkappa}(X) \in D_0) = P_\theta(\widehat{\varkappa}(X) \in D_1) = P_\theta(X \in W),\ \theta \in \Theta_1.$$

Зауваження. У випадку, коли нульова гіпотеза є простою гіпотезою, вірогідний рівень задається одним числом: $P_{\theta_0}(X \in W)$, оскільки при $\theta \in \Theta_0 = \{\theta_0\}$ розподіл вибірки відомий повністю.

Ймовірність $P_\theta(X \in W)$ попадання вибірки у критичну область як функція від усіх $\theta \in \Theta_0 \cup \Theta_1$ 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 невідомого параметра:

– при виконанні нульової гіпотези набувала б переважно значень із множини D_0 (наприклад, "досить помірних" значень) – що зменшує ймовірність похибки першого роду,

– при виконанні альтернативи – переважно значень із доповнення $D_1 = D \setminus D_0$ (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 $W = \emptyset$ імовірність похибки першого роду нульова, другого роду – дорівнює одиниці, а потужність нульова. При $W = S$ має місце протилежна ситуація – похибка першого роду і потужність одночасно одиничні. Тому задачу відшукання оптимального критерію зводять до задачі умовної оптимізації – знаходження максимуму потужності при фіксації рівня – максимальної похибки першого роду.

Означення. *Критерій* $(\widehat{\varkappa}, D_1)$ *має* вірогідний рівень α, *якщо ймо-*

вірності похибок першого роду не перевищують α :

$$P_\theta(\widehat{\varkappa}(X) \in D_1) = P_\theta(X \in W) \leq \alpha, \ \forall \theta \in \Theta_0.$$

і хоча б одна з цих нерівностей є рівністю.

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 :

$$P_\theta(\widehat{\varkappa}(X) \in D_1) = P_\theta(X \in W) \geq \alpha, \ \forall \theta \in \Theta_1.$$

Означення. *Критерій (точніше, послідовність критеріїв)* $(\widehat{\varkappa}_n, D_1)$ *вірогідного рівня* α, *що побудований для кожного* n *за кратною вибіркою об'єму* n, *є* консистентним *критерієм, якщо його потужність прямує до одиниці:*

$$P_\theta(\widehat{\varkappa}_n \in D_1) \to 1, \ n \to \infty, \ \forall \theta \in \Theta_1.$$

Очевидно, що консистентні критерії є асимптотичними розв'язками умовної задачі оптимізації критерію за рівнем та потужністю.

13. Непараметричні критерії для функції розподілу

Розглянемо кратну вибірку $X = (\xi_1, ..., \xi_n)$, в якій окремі спостереження мають невідому неперервну функцію розподілу $F(x) = P(\xi_1 < x)$. З формального погляду дана схема відповідає параметричному просторові Θ, що містить усі функції розподілу, а нульова гіпотеза може мати вигляд $H_0 : \theta = F$. для заданої F.

13.1. Критерій узгодженості Колмогорова

Перевіряється гіпотеза узгодженості кратної вибірки із заданою функцією розподілу F, тобто припущення, 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. Критерій є непараметричним, адже невідомою є вся функція розподілу, а не окремі параметри.

Побудова критерію грунтується на статистиці Колмогорова:

$$\widehat{\varkappa}_n(X) = \sqrt{n}\ \sup_{x\,\in\,\mathbb{R}} \left| \widehat{F}_n(x) - F(x) \right|,$$

де

$$\widehat{F}_n(x) = \frac{1}{n} \sum\nolimits_{k=1}^{n} 1_{\{\xi_k\,<\,x\}}$$

– емпірична функція розподілу.

За теоремою Колмогорова про відхилення емпіричної функції розподілу (що наведена вище в 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

$$P(\widehat{\varkappa}_n < x) \to P(\varkappa < x) = K(x),\ n \to \infty,$$

де функція розподілу $K(x)$ повністю відома:

$$K(x) \equiv \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2x^2).$$

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 $\widehat{\varkappa}_n \geq x_\alpha$. Рівень критерію наближено (і тим точніше, чим більший обсяг вибірки) дорівнює

$$P(\widehat{\varkappa}_n \geq x_\alpha) \approx P(\varkappa \geq x_\alpha) = 1 - K(x_\alpha) = \alpha.$$

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 $H_1 : \theta = \theta_1 \equiv F_1$. Тоді потужність критерію дорівнює

$$P_{\theta_1}\left(\sqrt{n}\sup_{x\,\in\,\mathbb{R}}\left|\widehat{F}_n(x) - F(x)\right| \geq x_\alpha\right) =$$

$$P_{\theta_1}\left(\sqrt{n}\sup_{x\,\in\,\mathbb{R}}\left|\widehat{F}_n(x) - F_1(x) + F_1(x) - F(x)\right| \geq x_\alpha\right) \geq$$

$$P_{\theta_1}(\sqrt{n}\Delta - \widetilde{\varkappa}_n \geq x_\alpha),\ \text{де } \Delta = \sup_{x\,\in\,\mathbb{R}}|F(x) - F_1(x)| > 0,$$

$$\widetilde{\varkappa}_n = \sqrt{n}\ \sup_{x\,\in\,\mathbb{R}}\left|\widehat{F}_n(x) - F_1(x)\right| \to^W \varkappa$$

за теоремою Колмогорова в припущенні H_1. Тому $P_{\theta_1}(\sqrt{n}\Delta - \widetilde{\varkappa}_n \geq x_\alpha) \to 1$, отже, потужність критерію Колмогорова прямує до 1 при кожній простій альтернативі. Критерій є консистентним.

13.2. Критерій однорідності Смірнова

Одночасно з вибіркою X розглянемо іншу кратну вибірку $Y = (\eta_1, ..., \eta_m)$ об'ємом m, що не залежить від X, із неперервною функцією розподілу окремих спостережень $G(y) = P(\eta_1 < y)$. Параметризуючи розподіл повного вектора спостережень (X, Y) як пару функцій розподілу $\theta = (F, G)$, розглянемо складну нульову гіпотезу однорідності

$$H_0 : F = G, \; F \text{ – неперервна}, \; X \text{ і } Y \text{ – незалежні}$$

проти альтернативи

$$H_1 : F \neq G, \; F, G \text{ – неперервні}, \; X \text{ і } Y \text{ – незалежні}.$$

Позначимо емпіричну функцію розподілу другої вибірки

$$\widehat{G}_m(y) = \frac{1}{m} \sum\nolimits_{k=1}^{m} 1_{\{\eta_k \, < \, y\}}.$$

Теорема (теорема Смірнова про відхилення емпіричних функцій розподілу). *За нульової гіпотези H_0 розподіл статистики Смірнова:*

$$\widehat{\varkappa}_{nm} = \sqrt{\frac{mn}{m+n}} \sup\nolimits_{x \, \in \, \mathbb{R}} \left| \widehat{F}_n(x) - \widehat{G}_m(x) \right|$$

не залежить від вигляду невідомої функції розподілу $F = G$, і для всіх $x \in \mathbb{R}$ має місце збіжність

$$P(\widehat{\varkappa}_{nm} < x) \to P(\varkappa < x) = K(x), \; n, m \to \infty,$$

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 $K(x_\alpha) = 1-\alpha$. Рівень критерію наближено дорівнює

$$P(\widehat{\varkappa}_{nm} \geq x_\alpha) \approx P(\varkappa \geq x_\alpha) = 1 - K(x_\alpha) = \alpha.$$

Для дослідження потужності критерію припустимо, що виконується альтернатива: $\theta \in H_1$. Тоді потужність критерію дорівнює

$$P_\theta \left(\sqrt{\tfrac{mn}{m+n}} \sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - \widehat{G}_m(x) \right| \geq x_\alpha \right) =$$

$$P_\theta \left(\sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F(x) + G(x) - \widehat{G}_m(x) + F(x) - G(x) \right| \sqrt{\tfrac{mn}{m+n}} \geq x_\alpha \right)$$

$$\geq P_\theta \left(\sqrt{\tfrac{mn}{m+n}} \Delta - \widetilde{\varkappa}_n - \widetilde{\varkappa}_m \geq x_\alpha \right), \text{ де } \Delta = \sup_{x \in \mathbb{R}} |F(x) - G(x)| > 0,$$

$$\widetilde{\varkappa}_n = \sqrt{n}\ \sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F(x) \right| \to^W \varkappa,$$

$$\widetilde{\varkappa}_m = \sqrt{m}\ \sup_{x \in \mathbb{R}} \left| \widehat{G}_m(x) - G(x) \right| \to^W \varkappa$$

за теоремою Колмогорова про відхилення емпіричної функції розподілу. Тому $\sqrt{\frac{mn}{m+n}}\Delta - \widetilde{\varkappa}_n - \widetilde{\varkappa}_m \to^P \infty$, отже, потужність критерію Смірнова прямує до одиниці при кожній простій альтернативі. Критерій є конзистентним.

14. Деякі рангові критерії

Значна кількість стійких до випадкових збурень критеріїв будується з використанням рангових статистик.

Означення. Ранговими статистиками *називаються координати* вектора рангів $R = (R_1, ..., R_n)$, *де значення* R_k *задає номер* k*-го спостереження у* *варіаційному ряді:*

$$\xi_k = \xi_{(R_k)}, \quad (\xi_1, ..., \xi_n) = (\xi_{(R_1)}, ..., \xi_{(R_n)}).$$

Приклад. Нехай спостерігається вибірка $X = (2, 7, 4, 5)$. Тоді варіаційний ряд має вигляд $(2, 4, 5, 7)$, а вектор рангів дорівнює $(1, 4, 2, 3)$.

Як показано в розділі про рангові статистики, для кратної вибірки з неперервною функцією розподілу вектор R рівномірно розподілений на множині всіх перестановок Π_n порядку n.

14.1. Критерій однорідності Вілкоксона

Розглянемо кратну вибірку $X = (\xi_1, ..., \xi_n)$, в якій окремі спостереження мають невідому неперервну функцію розподілу $F(x) = P(\xi_1 < x)$ і одночасно незалежну кратну вибірку $Y = (\eta_1, ..., \eta_m)$ зі спільною неперервною функцією розподілу $G(y) = P(\eta_1 < y)$. Для перевірки нульової гіпотези однорідності

$$H_0 : G = F, \quad X \text{ та } Y \text{ незалежні},$$

використаємо об'єднану вибірку

$$Z = (\xi_1, ..., \xi_n, \; \eta_1, ..., \eta_m)$$

та позначимо через R_j ранг j-го спостереження у вибірці Z. За нульової гіпотези вибірка Z є кратною об'єма $N = n+m$, тому вектор R рівномірно розподілений на множині перестановок Π_N порядку N.

Визначимо статистику Вілкоксона

$$S_{nm} = \sum_{j=1}^{n} R_j,$$

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 величину $\Delta > 0$, тобто

$$H_1 : G(y) = F(y + \Delta),\ \forall y \in \mathbb{R}, \quad X \text{ та } Y \text{ незалежні.}$$

У цьому разі $\eta_1 \simeq \xi_1 - \Delta$, отже, значення статистики S_{nm} будуть переважно більшими у випадку 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 рівень $x_{nm}(\alpha)$ визначається умовою $P(S_{mn} \geq x_{nm}(\alpha)) \approx \alpha$. Знаходження рівня можливе з використанням таблиць або спеціальних комп'ютерних програм.

14.2. Зауваження щодо рандомізованих критеріїв.

Оскільки статистика S_{mn} має дискретний розподіл, то її функція розподілу має лише 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 використовують таку процедуру рандомізації. Нехай $\alpha_1 \leq \alpha \leq \alpha_2$ – найближчі до α ймовірності, для яких рівняння

$$P(S_{mn} \geq x_i) = \alpha_i, \quad i = 1, 2,$$

мають точні розв'язки x_i. Згенеруємо незалежну від S_{mn} випадкову величину $\upsilon \in \{1, 2\}$ таку, що $P(\upsilon = 1) = \varepsilon$, $P(\upsilon = 2) = 1 - \varepsilon$. Визначимо випадковий (рандомізований) критичний рівень $x(\alpha) = x_\upsilon \equiv x_1 1_{\{\upsilon=1\}} + x_2 1_{\{\upsilon=2\}}$.Тоді за формулою повної ймовірності

$$P(S_{mn} \geq x(\alpha)) = \varepsilon P(S_{mn} \geq x_1) + (1-\varepsilon)P(S_{mn} \geq x_2) =$$
$$\varepsilon\alpha_1 + (1-\varepsilon)\alpha_2 = \alpha,$$

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.

14.3. Асимптотична нормальність статистики Вілкоксона

При великих значеннях вибірки можна скористатись асимптотичною нормальністю статистики Вілкоксона. Обчислимо

$$MS_{nm} = \textstyle\sum_{j=1}^n MR_j = n(N+1)/2 = n(n+m+1)/2,$$
$$DS_{nm} = MS_{nm}^2 - (MS_{nm})^2 =$$
$$\textstyle\sum_{j=1}^n MR_j^2 + \sum_{i\neq j=1}^n MR_jR_i - (MS_{nm})^2 =$$

$$\sum_{j=1}^{n}\sum_{r\in\Pi_N} r_j^2/N! + \sum_{i\neq j=1}^{n}\sum_{r\in\Pi_N} r_i r_j/N! - (MS_{nm})^2 =$$
$$n\ (N+1)(2N+1)/6 + n(n-1)(N+1)(3N+2)/12 - (MS_{nm})^2 =$$
$$n(N+1)(N-n)/12 = n\ m\ (n+m+1)/12.$$

Можна довести, що за нульової гіпотези має місце слабка збіжність при $n, m \to \infty$ центрованої та нормованої статистики Вілкоксона до величини зі стандартним нормальним розподілом

$$\varkappa_{nm} = \frac{S_{nm} - n(n+m+1)/2}{\sqrt{n\ m\ (n+m+1)/12}} = \frac{S_{nm} - MS_{nm}}{\sqrt{DS_{nm}}} \to^W \zeta \simeq N(0,1).$$

Звідси знаходимо критичну область для статистики S_{nm} так, щоб вірогідний рівень критерію наближено збігався із 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

$$MS_{nm} = \sum_{j=1}^{n} M(1 + \sum_{k=1,k\neq j}^{n} 1_{\{\xi_j\ \geq\ \xi_k\}} + \sum_{k=1}^{m} 1_{\{\xi_j\ \geq\ \eta_k\}}) =$$
$$n\ (1 + (n-1)/2 + m\ P(\xi_1 \geq \eta_1)) =$$
$$n(n+m+1)/2 - nm\ \delta, \text{ де } \delta = P(\xi_1 < \eta_1) - 1/2.$$

За умови, що $\delta \equiv \int F(x)dG(x) - 1/2 < 0$, звідси виводимо, що

$$\varkappa_{nm} = \frac{S_{nm} - n(n+m+1)/2}{\sqrt{n\ m\ (n+m+1)/12}} = \frac{S_{nm} - MS_{nm} - nm\ \delta}{\sqrt{n\ m\ (n+m+1)/12}} \to^P \infty,\ \ n, m \to \infty,$$

тобто критерій є консистентним.

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 $R = (R_1, ..., R_n)$ та $S = (S_1, ..., S_n)$ вектори рангів цих вибірок. При виконанні нульової гіпотези H_0 про незалежність X і Y та неперервність спільної функції розподілу вектори R і S незалежні в сукупності та рівномірно розподілені на множині Π_n перестановок порядку n. Зокрема,

$$MR_k = MS_k = (n+1)/2,$$
$$MR_k^2 = \sum_{i=1}^n i^2(n-1)!/n! = (n+1)(2n+1)/6, \quad DR_k = (n^2-1)/12,$$
$$MR_kR_j = \sum_{i\neq l} il(n-2)!\ /n! = (n+1)(3n+2)/12, \quad j \neq k,$$
$$\mathrm{Cov}(R_k, R_j) = MR_kR_j - MR_kMR_j = -(n+1)/12, \quad j \neq k.$$

Використаємо для перевірки H_0 статистику, що визначає вибірковий коефіцієнт кореляції між векторами рангів:

$$\widehat{\rho}_n = \sum_{k=1}^n (R_k - MR_k)(S_k - MS_k)\ /\ \widehat{\sigma}_R\, \widehat{\sigma}_S,$$
$$\widehat{\sigma}_R^2 = \sum_{k=1}^n (R_k - MR_k)^2, \quad \widehat{\sigma}_S^2 = \sum_{k=1}^n (S_k - MS_k)^2.$$

За нерівністю Коші статистика $\widehat{\rho}_n$ набуває значень із відрізку $[-1, 1]$. При повній тотожності векторів рангів $R = S$, 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 вказують на залежність векторів рангів та відповідних вибірок.

З використанням теореми про розподіл вектора рангів можна підрахувати (**Вправа**), що

$$\widehat{\sigma}_R^2 = \widehat{\sigma}_S^2 = n(n^2-1)\ /\ 12.$$

Звідси знаходимо еквівалентне зображення для статистики Спірмена

$$\widehat{\rho}_n = \frac{12}{n(n^2-1)} \sum\nolimits_{k=1}^{n} \left(R_k - \frac{n+1}{2}\right)\left(S_k - \frac{n+1}{2}\right) =$$

$$1 - \frac{6}{n(n^2-1)} \sum\nolimits_{k=1}^{n} (R_k - S_k)^2.$$

За нульової гіпотези

$$M\widehat{\rho}_n = \frac{12}{n(n^2-1)} \sum\nolimits_{k=1}^{n} M\left(R_k - \frac{n+1}{2}\right)\left(S_k - \frac{n+1}{2}\right) = 0,$$

$$D\widehat{\rho}_n = M\widehat{\rho}_n^2 = \frac{144}{n^2(n^2-1)^2} \sum\nolimits_{i,\ j=1}^{n} \mathrm{Cov}(R_i,\ R_j)\,\mathrm{Cov}(S_i,\ S_j) = \frac{1}{n-1}.$$

Вправа. Перевірити останню рівність на підставі наведених вище виразів для коваріацій рангових статистик.

Як відзначено вище, критичними для гіпотези незалежності є великі 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.

15. Критерій хі-квадрат Пірсона у поліноміальній схемі Бернуллі

Критерій хі-квадрат – найбільш універсальний з відомих критеріїв і може бути застосований до перевірки гіпотез для великої кількості статистичних моделей.

Використовуючи групування спостережень, або ж підрахунок емпіричних частот для певної групи подій, дуже часто вибірковий вектор можна звести до вигляду $X = (\zeta_1, ..., \zeta_n)$, де окремі спостереження незалежні у сукупності, однаково розподілені й можуть набувати скінченну кількість із k різних значень $\{x_1, ..., x_k\}$:

$$P_\theta(\zeta_1 = x_i) = p_i > 0,\ i = \overline{1,k},\ \sum\nolimits_{i=1}^{k} p_i = 1.$$

Така схема випробувань є узагальненням біноміальної схеми випробувань Бернуллі, де $k = 2$, і називається поліноміальною схемою випробувань Бернуллі. 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

$$\widehat{\nu}_{ni} = \sum\nolimits_{j=1}^{n} 1_{\{\zeta_j\, =\, x_i\}} = \left|\{j : \zeta_j = x_i\}\right|,\ i = \overline{1,k},$$

дорівнює кількості тих випробувань із загальної кількості n, що завершилися результатом x_i.

Оскільки функція вірогідності вибірки X із вектором параметрів $\theta = (p_i, i = \overline{1,k})$ має вигляд

$$L(X, \theta) = \prod\nolimits_{j=1}^{n} p_{\zeta_j} = \prod\nolimits_{i=1}^{k} p_i^{\widehat{\nu}_{ni}},$$

то вектор відносних емпіричних частот $\widehat{\nu}_n/n$ збігається з оцінкою максимальної вірогідності параметра θ :

$$\arg\max{}_{\theta} L(X,\theta) = \arg\max{}_{\theta} \sum\nolimits_{i=1}^{k} \widehat{\nu}_{ni} \ln p_i = (\widehat{\nu}_{ni}/n, i=\overline{1,k}) = \widehat{\nu}_n/n.$$

Для обчислення умовного максимуму на множині $\{\theta = (p_i) : p_i > 0, \sum p_i = 1\}$ 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,

$$M_\theta \widehat{\nu}_{ni} = np_i,\ D_\theta \widehat{\nu}_{ni} = np_i(1-p_i),$$

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 $\theta = (p_i,\ i=\overline{1,k})$:

$$P_\theta(\lim{}_{n\to\infty} \widehat{\nu}_n/n = \theta) = P_\theta(\bigcap\nolimits_{i=1}^{k} \{\lim{}_{n\to\infty} \widehat{\nu}_{ni}/n = p_i\}) = 1.$$

Емпіричні частоти залежні, причому вектор $\widehat{\nu}_n \in \mathbb{R}^k$ лежить у $(k-1)$-вимірному підпросторі: $\sum_{i=1}^{k} \widehat{\nu}_{ni} = n$.

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 з вектором імовірностей $\theta = (p_1, ..., p_k) > 0$ результатів в окремих спостереженнях. 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, який не залежить від розподілу 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 спостерігає статистик, та їх очікувані (середні) значення $np_i = M\widehat{\nu}_{ni}$. Тому використовується така мнемонічна форма для обчислення статистики:

$$\chi^2 = \sum \frac{(O-E)^2}{E},$$

де O означає Observed (що спостерігається), а E – Expected.(що очікується)

Доведення.

Розглянемо k-вимірний випадковий вектор

$$\eta(n) = \left(\frac{\widehat{\nu}_{ni} - np_i}{\sqrt{np_i}},\ i = \overline{1,k} \right)$$

та для кожного $j = \overline{1,n}$ незалежні однаково розподілені вектори

$$\gamma_j = \left((1_{\{\xi_j\, =\, x_i\}} - p_i)\ /\ \sqrt{p_i},\ i = \overline{1,k} \right).$$

Тоді за означенням частот $\widehat{\nu}_{ni}$

$$\eta(n) = \left(\sum_{j=1}^{n}(1_{\{\xi_j\,=\,x_i\}} - p_i)\frac{1}{\sqrt{np_i}},\ i = \overline{1,k}\right) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\gamma_j,$$

$$\widehat{\varkappa}_{k-1}^2(n) = \eta^2(n) = \eta'(n)\eta(n).$$

Обчислимо середнє та коваріацію одного доданку в сумі для $\eta(n)$:

$$M_\theta\gamma_1 = \left((P_\theta(\xi_1 = x_i) - p_i)\,/\sqrt{p_i},\ i = \overline{1,k}\right) = 0,$$

$$\mathrm{Cov}(\gamma_1) = \left(M_\theta(1_{\{\xi_1\,=\,x_i\}} - p_i)(1_{\{\xi_1\,=\,x_l\}} - p_l)\,/\sqrt{p_ip_l},\ i,l = \overline{1,k}\right) =$$

$$\left((p_i\delta_{il} - p_ip_l)\ /\ \sqrt{p_ip_l},\ i,l = \overline{1,k}\right) = I - q\cdot q',$$

де I – одинична матриця, а матриця $q\cdot q' = (\sqrt{p_ip_l},\ i,l = \overline{1,k})$ розміру $k\times k$ утворена прямим 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

$$\eta(n) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\gamma_j \to^W \xi \simeq N_k(0,\ I - q\cdot q').$$

З неперервності квадратичної функції та з означення слабкої збіжності векторів звідси випливає збіжність випадкових величин

$$\widehat{\chi}_{k-1}^2(n) = \eta^2(n) \to^W \xi^2,\ n\to\infty.$$

Зауважимо, що вектор $q = (\sqrt{p_i},\ i = \overline{1,k})'$ ортонормований. 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

$$M_\theta \beta = 0, \quad \mathrm{Cov}(\beta) = U\mathrm{Cov}(\xi)U' = U(I - q \cdot q')U' =$$
$$UU' - \ Uq \cdot (Uq)' = I - e \cdot e' = (\delta_{ij} 1_{i<k}, \ i, j = \overline{1, k}),$$

то випадкові величини $(\beta_1, ..., \beta_{k-1})$ є незалежними у сукупності стандартними нормальними величинами, а $\beta_k = 0$, оскільки $D\beta_k = \mathrm{Cov}(\beta)_{kk} = 0$. Обчислимо

$$\xi^2 = \xi'\xi = (U'\beta)'U'\beta = \ \beta'UU'\beta = \beta^2 = \sum_{i=1}^{k-1} \beta_i^2 \ \simeq \chi^2_{k-1},$$

звідки за означенням хі-квадрат розподілу $\xi^2 \simeq \chi^2_{k-1}$.

Отже, $\widehat{\chi}^2_{k-1}(n) \to^W \chi^2_{k-1}$ $\boxdot$

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 $(p_1, ..., p_k)$.

Критерій узгодженості хі-квадрат для перевірки гіпотези H_0 має вигляд $(\widehat{\chi}^2_{k-1}(n), \ [x_\alpha, \infty))$, де статистика хі-квадрат

$$\widehat{\chi}^2_{k-1}(n) = \sum_{i=1}^{k} \frac{(\widehat{\nu}_{ni} - np_i)^2}{np_i}$$

визначена в теоремі Пірсона про асимптотику статистики хі-квадрат, а критичний рівень x_α знаходиться з умови $P(\chi^2_{k-1} < x_\alpha) = 1-\alpha$. З теореми Пірсона випливає, що рівень критерію наближено (при великих об'ємах вибірки) дорівнює α.

Відоме емпіричне (позаматематичне) правило: при обмеженій кількості спостережень критерію хі-квадрат можна довіряти, коли *на кожний рівень припадає не менше 6 (12, 20, ...) спостережень* (тобто $\widehat{\nu}_{ni} \geq 6\ (12, 20, ..)$ для всіх i).

Критерій хі-квадрат є консистентним для кожної простої альтернативи. Дійсно, якщо гіпотетичне значення $\widetilde{\theta} = (\widetilde{p}_i, i = \overline{1,k})$ не дорівнює істинному $\theta = (p_i,\ i = \overline{1,k})$, то

$$\widetilde{\chi}^2_{k-1}(n) = \sum\nolimits_{i=1}^{k} \frac{(\widehat{\nu}_{ni} - n\widetilde{p}_i)^2}{n\widetilde{p}_i} = \sum\nolimits_{i=1}^{k} \frac{(np_i - n\widetilde{p}_i + \widehat{\nu}_{ni} - np_i)^2}{n\widetilde{p}_i} \geq$$

$$\sum\nolimits_{i=1}^{k} \frac{(np_i - n\widetilde{p}_i)^2/2 - (\widehat{\nu}_{ni} - np_i)^2}{n\widetilde{p}_i} \geq n\Delta - c\widehat{\chi}^2_{k-1}(n),$$

де

$$\Delta = \sum\nolimits_{i=1}^{k} (p_i - \widetilde{p}_i)^2/2\ \widetilde{p}_i > 0, \quad c = \max{}_i(p_i/\widetilde{p}_i),$$

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 $P_\theta(\widetilde{\chi}^2_{k-1}(n) \geq x_\alpha) \to 1, n \to \infty$. 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 $P(\chi^2_{599} > 9.24) \approx 0.01$. Оскільки значення $1.70 < 9.24$, то альтернатива про асиметрію повинна бути відкинута на рівні 0.01, і приймається нульова гіпотеза про симетрію.

Вправа.Результати підрахунку частот цифр $0, 1, ..., 9$ у 10002 десяткових знаках числа $\pi - 3$ наведені в таблиці

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.

15.3. Групування, гістограма

Значну кількість статистичних випробувань можна звести до поліноміальної схеми випробувань. Для цього застосовується метод 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 $g(x) = \sum_{i=1}^{k} i\ 1_{x \in H_i}$, 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.

Відповідний вектор емпіричних частот $\widehat{\nu}_n = (\widehat{\nu}_{n1}, ..., \widehat{\nu}_{nk})$ для групованої вибірки (ζ_j) має вигляд

$$\widehat{\nu}_{ni} = \sum\nolimits_{j=1}^{n} 1_{\{\zeta_j = i\}} = \sum\nolimits_{j=1}^{n} 1_{\{\xi_j\ \in\ H_i\}} = \left|\{j : \xi_j\ \in\ H_i\}\right|,\ i = \overline{1,k},$$

і називається гістограмою вибірки X при розбитті H. Гістограми часто зображують у вигляді ряду вертикальних стовпчиків, висоти яких пропорційні відповідним емпіричним частотам.

Відповідний вектор теоретичних частот обчислюється через розподіл спостережень:

$$p_i = P_\theta(\xi_1\ \in\ H_i),\ i = \overline{1,k}.$$

Слід мати на увазі, що при застосуванні групування відбувається підміна формулювання нульової гіпотези. Так, у задачі про узгодженість вибірки із заданою функцією розподілу нульова гіпотеза після групування зводиться до твердження, що ймовірності попадання одного спостереження в інтервали розбиття збігаються з наперед заданими. Останню умову задовольняють багато *різних* теоретичних функцій розподілу.

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 інтервалів $\mathbb{R} = \cup_{i=1}^{k} \Delta_i$, що не перетинаються, та обчислюють гістограму

$$\widehat{\nu}_{ni} = \sum\nolimits_{j=1}^{n} 1_{\{\xi_j \in \Delta_i\}} = \left|\{j : \xi_j \in \Delta_i\}\right|, \quad i = \overline{1,k}.$$

Статистика критерію дорівнює

$$\widehat{\chi}^2_{k-1}(n) = \sum\nolimits_{i=1}^{k} \frac{(\widehat{\nu}_{ni} - nF(\Delta_i))^2}{nF(\Delta_i)}$$

і має асимптотичний хі-квадрат розподіл χ^2_{k-1}. Теоретичні ймовірності $F(\Delta_i) = P(\xi_j \in \Delta_i)$ дорівнюють приростам функції F на відповідних інтервалах.

Критерій узгодженості хі-квадрат із функцією розподілу має вигляд $(\widehat{\chi}^2_{k-1}(n), \quad [z_\alpha, \infty))$, де критичний рівень z_α визначається умовою $P(\chi^2_{k-1} < z_\alpha) = 1 - \alpha$. З теореми Пірсона про асимптотику статистики хі-квадрат випливає, що вірогідний рівень критерію $P(\widehat{\chi}^2_{k-1}(n) \geq \quad z_\alpha)$ наближено (при великих об'ємах вибірки) дорівнює α.

15.5. Критерій хі-квадрат для складних гіпотез

Розглянемо поліноміальну схему випробувань, в якій розподіл одного спостереження залежить від значення векторного параметра $\theta \in \Theta \subset \mathbb{R}^d$:

$$P_\theta(\zeta_1 = x_i) = p_i(\theta), \; i = \overline{1,k}.$$

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*

(а) $\inf_{\theta,\, i}\, p_i(\theta) > 0$,

(б) $p_i(\theta) \in C^2(\Theta)$, $\forall i = \overline{1,k}$,

(в) $rang(\partial p_i(\theta)/\partial\theta,\ i = \overline{1,k}) = d \equiv \dim\Theta < k$.

Якщо оцінка $\widehat{\theta}_n$ є оцінкою максимальної вірогідності за емпіричними частотами $\widehat{\nu}_n = (\widehat{\nu}_{n1}, ..., \widehat{\nu}_{nk})$, тобто

$$\widehat{\theta}_n = \arg\max{}_\theta\, L(\widehat{\nu}_n, \theta) = \arg\max{}_\theta \prod_{i=1}^k p_i^{\widehat{\nu}_{ni}}(\theta),$$

то модифікована статистика хі-квадрат

$$\widehat{\chi}^2(\widehat{\theta}_n) \equiv \sum_{i=1}^k \frac{(\widehat{\nu}_{ni} - np_i(\widehat{\theta}_n))^2}{np_i(\widehat{\theta}_n)}$$

слабко збігається до величини з хі-квадрат розподілом та кількістю ступенів свободи, яка скоригована на число оцінюваних параметрів:

$$\widehat{\chi}^2(\widehat{\theta}_n) \to^W \chi^2_{k-1-d},\ n \to \infty.$$

Доведення теореми не наводиться.

Модифікований критерій узгодженості хі-квадрат має вигляд $(\widehat{\chi}^2_{k-d-1}(n),\quad [z_\alpha, \infty))$, де критичний рівень z_α визначається умовою $P(\chi^2_{k-d-1} < z_\alpha) = 1 - \alpha$. З теореми про асимптотику модифікованої статистики хі-квадрат випливає, що вірогідний рівень критерію наближено (при великих об'ємах вибірки) дорівнює α.

15.6. Критерій хі-квадрат однорідності

Розглянемо статистичні випробування, що складаються із m серій поліноміальних випробувань $X_s = (\zeta_{sj},\ j = \overline{1, n_s})$ по n_s спостережень у s-й серії, $s = \overline{1, m}$. Окреме випробування в кожній серії може закінчитися одним із k результатів $\{x_1, ..., x_k\}$. Спостереження зображені подвійним вектором емпіричних частот

$$\widehat{\nu}_n = (\widehat{\nu}_s(n_s), s = \overline{1, m}) = ((\widehat{\nu}_{si},\ i = \overline{1, k}),\ s = \overline{1, m}),$$

$$\widehat{\nu}_{si} = \sum\nolimits_{j=1}^{n_s} 1_{\{\zeta_{sj} = x_i\}} = \left|\{j : \zeta_{sj} = x_i\}\right|,\quad n = \sum\nolimits_{s=1}^{m} n_s.$$

Усі випробування незалежні у сукупності. Нульова гіпотеза полягає в тому, що розподіли всіх спостережень однакові, однак припущення щодо спільного розподілу не формулюються. Якби спільний розподіл спостережень $\theta = (p_i, i = \overline{1, k})$ був відомий, то для перевірки нульової простої гіпотези

$$H_0 : P_\theta(\zeta_{s1} = x_i) = p_i,\ \forall i = \overline{1, k},\ \forall s = \overline{1, m},$$

можна було б використати при $n_s \to \infty$, $s = \overline{1, m}$, статистику хі-квадрат

$$\widehat{\chi}^2(\theta) = \sum\nolimits_{s=1}^{m} \sum\nolimits_{i=1}^{k} \frac{(\widehat{\nu}_{si} - n_s p_i)^2}{n_s p_i} \to^W \sum\nolimits_{s=1}^{m} \chi^2_{k-1,\ s} \simeq \chi^2_{mk-m}.$$

Остання рівність тут випливає з адитивності хі-квадрат розподілів відносно додавання незалежних величин, що є наслідком теореми про інваріантність гама-розподілів відносно згортки.

При невідомому розподілі вибірки, тобто за нульової гіпотези

$$H_0' : P_\theta(\zeta_{sj} = x_i) = P_\theta(\zeta_{11} = x_i),\ \forall i = \overline{1, k},\ \forall s = \overline{1, m},\ \forall \theta,$$

для цього розподілу $\theta = (p_i, i = \overline{1, k})$ можна побудувати оцінку максимальної вірогідності

$$\widehat{\theta}_n = \arg\max{}_\theta \prod\nolimits_{s=1}^{m} \prod\nolimits_{i=1}^{k} p_i^{\widehat{\nu}_{si}} =$$

$$\arg\max{}_\theta \sum\nolimits_{i=1}^{k} \widehat{\nu}_{\bullet i} \ln p_i = (\widehat{\nu}_{\bullet i} \ / \ n,\ i = \overline{1, k}),$$

де

$$\widehat{\nu}_{\bullet i} = \sum\nolimits_{s=1}^{m} \widehat{\nu}_{si},\ n = \sum\nolimits_{s=1}^{m} n_s.$$

Зауважимо, що фактична розмірність параметричного простору $\dim \Theta = k - 1$, оскільки сума координат k-вимірного вектора θ завжди дорівнює одиниці.

Отже, модифікована статистика хі-квадрат для перевірки однорідності має вигляд

$$\widehat{\chi}^2(\widehat{\theta}_n) = \sum\nolimits_{s=1}^{m} \sum\nolimits_{i=1}^{k} \frac{(\widehat{\nu}_{si} - n_s \widehat{\nu}_{\bullet i}/n)^2}{n_s \widehat{\nu}_{\bullet i}/n}$$

і за теоремою про асимптотику модифікованої статистики хі-квадрат

$$\widehat{\chi}^2(\widehat{\theta}_n) \to^W \chi^2_{mk-m-(k-1)} = \chi^2_{(m-1)(k-1)}.$$

Критерій хі-квадрат однорідності має вигляд $\left(\widehat{\chi}^2(\widehat{\theta}_n),\ [z_\alpha, \infty)\right)$, де критичний рівень z_α визначається умовою $P(\chi^2_{(m-1)(k-1)} < z_\alpha) = 1 - \alpha$. З теореми про асимптотику модифікованої статистики хі-квадрат випливає, що вірогідний рівень критерію наближено (при великих об'ємах вибірки) дорівнює α.

15.7. Критерій хі-квадрат незалежності

Розглянемо кратну вибірку, що містить двовимірні спостереження (пари спостережень):

$$X = ((\xi_s, \eta_s),\ s = \overline{1, n}).$$

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

$$H_0 : P_\theta(\xi_1 = x_i, \eta_1 = y_j) = p_i q_j, \forall i = \overline{1, k},\ \forall j = \overline{1, m},$$

де розподіли координат p_i, q_j невідомі і підлягають оцінці. Якби ці розподіли були відомими, то критерій хі-квадрат містив би сумісні емпіричні частоти

$$\nu_{ij} = \sum_{s=1}^{n} 1_{\{\xi_s = x_i, \eta_s = y_j\}} = |\{s : \xi_s = x_i,\ \eta_s = y_j\}|$$

і мав використовувати статистику хі-квадрат

$$\widehat{\chi}^2_{km-1}(\theta) = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(\nu_{ij} - np_i q_j)^2}{np_i q_j}.$$

Таблиця із частотами $(\nu_{ij}, i = \overline{1,k},\ j = \overline{1,m})$ називається таблицею спряженості факторів, що впливають на значення спостережень у рядках та стовпчиках.

При невідомому розподілі $\theta = (p_i q_j, i = \overline{1,k},\ j = \overline{1,m})$ нульова гіпотеза має вигляд

$$H_0' : P_\theta(\xi_1 = x_i,\ \eta_1 = y_j) =$$
$$P_\theta(\xi_1 = x_i) P_\theta(\eta_1 = y_j),\ \forall i = \overline{1,k},\ \forall j = \overline{1,m},$$

а оцінка максимальної вірогідності для цього розподілу дорівнює

$$\widehat{\theta}_n = \operatorname{argmax}_{\theta} \prod_{i=1}^{k} \prod_{j=1}^{m} (p_i q_j)^{\nu_{ij}} =$$
$$\operatorname{argmax}_{\theta} \left(\sum_{i=1}^{k} \widehat{\nu}_{i\bullet} \ln p_i + \sum_{j=1}^{m} \widehat{\nu}_{\bullet j} \ln q_j\right) =$$
$$(\widehat{\nu}_{i\bullet}\, \widehat{\nu}_{\bullet j} \;/\; nm,\ i = \overline{1,k},\ j = \overline{1,m}),$$

де

$$\widehat{\nu}_{i\bullet} = \sum_{j=1}^{m} \widehat{\nu}_{ij},\ \widehat{\nu}_{\bullet j} = \sum_{i=1}^{k} \widehat{\nu}_{ij}.$$

Отже, модифікована статистика хі-квадрат для перевірки незалежності парних спостережень має вигляд

$$\widehat{\chi}^2(\widehat{\theta}_n) \equiv \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(\widehat{\nu}_{ij} - \widehat{\nu}_{i\bullet}\widehat{\nu}_{\bullet j}/n)^2}{\widehat{\nu}_{i\bullet}\widehat{\nu}_{\bullet j}/n}.$$

Фактична 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

$$\widehat{\chi}^2(\widehat{\theta}_n) \to^W \chi^2_{mk-1-(m-1+k-1)} = \chi^2_{(m-1)(k-1)}.$$

Критерій хі-квадрат незалежності парних спостережень має вигляд $(\widehat{\chi}^2(\widehat{\theta}_n), \quad [z_\alpha, \infty))$, де критичний рівень z_α визначається умовою $P(\chi^2_{(m-1)(k-1)} < z_\alpha) = 1-\alpha$. З теореми про асимптотику модифікованої статистики хі-квадрат випливає, що вірогідний рівень критерію наближено (при великих об'ємах вибірки) дорівнює α.

Приклад. Соціологи провели опитування серед трьох груп населення: із 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	$75 = 100 \cdot 150/600$	-10	1.33
Σ		600	600	0	8.00

Оскільки значення статистики $8.00 > 5.99 = x_{0.01}$, то нульова гіпотеза про незалежність фактора освіти від думки щодо необхідності реформ відкидається на рівні 0.01 – тобто приймається альтернатива про залежність (у даному випадку – позитивну залежність від рівня освіти).

16. Перевірка гіпотез про параметри нормальних спостережень

Критерії узгодженості для нормальних спостережень грунтуються на спеціальних властивостях вибіркових моментів, які використовувались вище при побудові інтервальних оцінок.

Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з нормальним розподілом спостережень: $\xi_k \simeq N(\mu, \sigma^2)$. Викладені далі критерії містять такі статистики: $\widehat{\mu}_n$ – вибіркове середнє, $\widehat{\sigma}_n^2$ – вибіркова дисперсія, та $\widehat{s}_n^2$ – нормована вибіркова дисперсія.

Нижче x_α означає двосторонній квантиль вірогідного рівня $1-\alpha$ для стандартного нормального розподілу: $P(|\zeta| < x_\alpha) = 1-\alpha$. Аналогічний зміст мають $y_{n\alpha}$ для розподілу Стьюдента з n ступенями свободи, $z_{n\alpha}$ – для хі-квадрат розподілу з n ступенями свободи.

16.1. Перевірка гіпотез про параметри однієї вибірки

16.1.1. Гіпотеза про середнє при відомій дисперсії

Якщо дисперсія σ^2 відома, то статистика

$$\widehat{\varkappa}_n = \sqrt{n}\,(\widehat{\mu}_n - \mu)/\sigma$$

має стандартний нормальний розподіл за теоремою про вибіркові середнє та дисперсію нормальної вибірки. Тому двосторонній критерій $(|\widehat{\varkappa}_n|, [x_\alpha, \infty))$ перевірки гіпотези про значення μ має вірогідний рівень α. Цей критерій є консистентним, оскільки при $\theta_0 \neq \theta$ статистика

$$\widetilde{\varkappa}_n = \sqrt{n}(\widehat{\mu}_n - \mu_0)/\sigma \ = \sqrt{n}(\widehat{\mu}_n - \mu + \mu - \mu_0)/\sigma \to^P \pm\infty,\ n \to \infty.$$

16.1.2. Гіпотеза про дисперсію при відомому середньому

При відомому середньому μ вибіркова дисперсія $\widehat{\sigma}_n^2$ є статистикою, а величина

$$\widehat{\chi}_n^2 = n\widehat{\sigma}_n^2 \,/\, \sigma^2 = \sum\nolimits_{k=1}^{n} (\xi_k - \mu)^2 \,/\, \sigma^2$$

має хі-квадрат розподіл із n ступенями свободи за теоремою про вибіркові середнє та дисперсію нормальної вибірки. Відповідно критерій $(\widehat{\chi}_n^2,\ [z_{n\alpha}, \infty))$ має вірогідний рівень α. Критерій не є консистентним.

16.1.3. Гіпотеза про середнє при невідомій дисперсії

За теоремою про статистику Стьюдента від нормальної вибірки статистика

$$\widehat{\tau}_{n-1} = \sqrt{n}\,\frac{\widehat{\mu}_n - \mu}{\widehat{s}_n}$$

має розподіл Стьюдента з $n-1$ ступенем свободи. Тому критерій $(|\widehat{\tau}_{n-1}|\,,\ [y_{n-1,\alpha}, \infty))$ для перевірки гіпотези про значення μ 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 значення статистики $\widehat{\tau}_9 = \sqrt{10}(3.1 - 0)/4.75 = 2.1 > y_{\,10-1,\ 0.01} = 1.83$.

Отже, на рівні 0.01 гіпотеза про поліпшення не може бути відхилена – приймається.

16.1.4. Гіпотеза про дисперсію при невідомому середньому

За теоремою про вибіркові середнє та дисперсію нормальної вибірки статистика

$$\widehat{\chi}^2_{n-1} = (n-1)\widehat{s}^2_n \,/\, \sigma^2$$

має хі-квадрат розподіл із $n-1$ ступенем свободи. Відповідно критерій $\left(\widehat{\chi}^2_{n-1},\ [z_{n-1,\alpha}, \infty)\right)$ має вірогідний рівень α. Критерій не є консистентним.

16.2. Перевірка гіпотез про параметри двох вибірок

16.2.1. Гіпотеза про різницю середніх при відомих дисперсіях

Нехай одночасно спостерігаються дві незалежні кратні нормальні вибірки $X = (\xi_1, ..., \xi_n)$ та $Y = (\eta_1, ..., \eta_m)$ з розподілами

$$\xi_1 \simeq N(\mu_1, \sigma_1^2), \quad \eta_1 \simeq N(\mu_2, \sigma_2^2)$$

з невідомими середніми μ_1, μ_2 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, $\widehat{\mu}_n - \mu_1 \simeq N(0, \sigma_1^2/n)$, $\widehat{\mu}_m - \mu_2 \simeq N(0, \sigma_2^2/m)$. Тому за нульової гіпотези

$$\widehat{\mu}_m - \widehat{\mu}_n - \Delta \simeq N(0, \sigma_1^2/n + \sigma_2^2/m),$$

внаслідок теореми про нормальність суми незалежних нормальних векторів.

Отже, статистика критерію

$$\widehat{\zeta}_{nm} = \left(\widehat{\mu}_m - \widehat{\mu}_n - \Delta\right) / \sqrt{\sigma_1^2/n + \sigma_2^2/m}$$

має за нульової гіпотези стандартний нормальний розподіл, а критична область відповідає великим її значенням.

Критерій $\left(\left|\widehat{\zeta}_{nm}\right|,\ [x_\alpha, \infty)\right)$ для перевірки гіпотези про різницю середніх має вірогідний рівень α та є консистентним.

16.2.2. 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_1, μ_2 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 і Y, а $\widehat{s}_n^2$, $\widehat{s}_m^2$ – нормовані вибіркові дисперсії. Тоді з незалежності вибірок X і 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

$$\sqrt{\frac{nm}{n+m}}(\widehat{\mu}_n - \widehat{\mu}_m - \Delta)/\sigma,$$

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 хі-квадрат розподілу.

Отже, за нульової гіпотези H_0 внаслідок означення розподілу Стьюдента статистика, що утворена відношенням:

$$\widehat{\tau}_{n+m-2} = \sqrt{\frac{nm(n+m-2)}{n+m}}\;\frac{\widehat{\mu}_n - \widehat{\mu}_m - \Delta}{\sqrt{(n-1)\widehat{s}_n^2 + (m-1)\widehat{s}_m^2}}$$

має розподіл Стьюдента з $n+m-2$ ступенями свободи.

Критерій Стьюдента $(\left|\widehat{\tau}_{n+m-2}\right|, \; [y_{n+m-2,\alpha}, \infty))$ для перевірки гіпотези про різницю середніх має вірогідний рівень α. Критерій є консистентним.

16.2.3. Гіпотеза про відношення дисперсій при відомих середніх

Нехай одночасно спостерігаються дві 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_1, μ_2 та невідомими дисперсіями σ_k^2.

Нульова гіпотеза формулюється як

$$H_0 : \sigma_2^2 = \rho\sigma_1^2,$$

де відношення ρ відоме (зокрема, часто $\rho = 1$). Тоді за нульової гіпотези статистика

$$\widehat{\phi}_{n,m} = \frac{n\widehat{\sigma}_n^2 \ / \ \sigma_1^2}{m\widehat{\sigma}_m^2 \ / \ \sigma_2^2} = \rho \frac{n\widehat{\sigma}_n^2}{m\widehat{\sigma}_m^2}$$

має розподіл Фішера із 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 $P(\phi_{n,m} < w_\alpha) = \alpha$.

16.2.4. Гіпотеза про відношення дисперсій при невідомих середніх

При невідомих середніх слід обрати нормовані вибіркові дисперсії. У цьому випадку статистика

$$\widehat{\phi}_{n-1,m-1} = \frac{(n-1)\widehat{s}_n^2 \ / \ \sigma_1^2}{(m-1)\widehat{s}_m^2 \ / \ \sigma_2^2} = \rho \frac{(n-1)\widehat{s}_n^2}{(m-1)\widehat{s}_m^2}$$

має розподіл Фішера з $n-1, m-1$ ступенями свободи. Відповідно критерій

$$(\widehat{\phi}_{n-1,m-1}, \ [0, w_{\alpha/2}/\rho] \cup [w_{1-\alpha/2}/\rho, \infty))$$

має рівень α.

Зауваження. На практиці всі наведені критерії використовуються також для вибірок, які не мають нормального розподілу (наприклад, для цілозначних спостережень). Правомірність такого використання може бути обгрунтована 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.

Приклад. 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 $[0, 1]$

Вправа. Одночасно спостерігаються дві незалежні кратні нормальні вибірки $X_k = (\xi_{k1}, ..., \xi_{kn})$ з розподілами $\xi_{k1} \simeq N(\mu_k, \sigma_k^2),\ k = 1, 2$ з невідомими середніми μ_k та дисперсіями σ_k^2. Нехай $\widehat{\mu}_{kn},\ \widehat{\sigma}_{kn}^2$ вибіркові середнє та дисперсія для k-ї вибірки. Позначимо

$$\widehat{\rho}_n = \frac{1}{(n-1)\widehat{\sigma}_{1n}\widehat{\sigma}_{2n}} \sum_{j=1}^{n} (\xi_{1j} - \widehat{\mu}_{1n})(\xi_{2j} - \widehat{\mu}_{2n})$$

вибірковий коефіцієнт кореляції між вибірками X_1, X_2. Довести, що статистика

$$\sqrt{n}\frac{\widehat{\mu}_{1n} - \widehat{\mu}_{2n} - \mu_1 + \mu_2}{\sqrt{\widehat{\sigma}_{1n}^2 + \widehat{\sigma}_{2n}^2 - 2\widehat{\rho}_n\widehat{\sigma}_{1n}\widehat{\sigma}_{2n}}}$$

має розподіл Стьюдента з $n - 1$ ступенем свободи.

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 $P_\theta(X \in W), \ \theta \in \Theta_0$.

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^*:*

$$P_\theta(X \in W^*) \geq P_\theta(X \in W), \ \forall \theta \in \Theta_1.$$

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 $F_i(B) = P_{\theta_i}(X \in B)$, $B \in \Sigma$, 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. 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*

$$P_{\theta_0}(l_{01}(X) \geq l_\alpha) = \alpha,$$

то критерій відношення вірогідностей $(l_{01}(X), [l_\alpha, \infty))$ *з критичною областю*

$$W^* = \{x \in S : l_{01}(x) \geq l_\alpha\}$$

є найбільш потужним рівня α.

Зауваження. Функція

$$P_{\theta_0}(l_{01}(X) \geq l) \;=\; F_0(\{x \in S : l_{01}(x) \geq l\})$$

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 скористатись рандомізованим критерієм, як це описано вище в розділі про рангові критерії.

Доведення.

Розглянемо довільний критерій із критичною областю W рівня α, тобто

$$P_{\theta_0}(X \in W) = F_0(W) \leq \alpha.$$

Обчислимо його потужність

$$P_{\theta_1}(X \in W) = F_1(W) = F_1(W \setminus W^*) + F_1(W \cap W^*) =$$
$$F_1(W^*) + F_1(W \setminus W^*) - F_1(W^* \setminus W) =$$
$$F_1(W^*) + \int_{W \setminus W^*} l_{01}(x)F_0(dx) - \int_{W^* \setminus W} l_{01}(x)F_0(dx) \leq$$
$$F_1(W^*) + l_\alpha F_0(W \setminus W^*) - l_\alpha F_0(W^* \setminus W) =$$
$$F_1(W^*) + l_\alpha F_0(W) - l_\alpha F_0(W^*) \leq F_1(W^*) + l_\alpha \alpha - l_\alpha \alpha = F_1(W^*),$$

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 рівень W^* дорівнює $F_0(W^*) = \alpha$ $\boxdot$

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 нульової гіпотези за теоремою про вибіркові середнє та дисперсію нормальної вибірки ліва частина має стандартний нормальний розподіл. Позначаючи через x_α праву частину останньої нерівності, звідси для довільного рівня α знаходимо критичний рівень x_α з рівняння $\Phi(x_\alpha) = 1 - \alpha$ 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

$$\left\{ \frac{\sqrt{n}}{\sigma}(\widehat{\mu}_n - \mu_0) \geq x_\alpha \right\}.$$

Для обчислення потужності критерію зауважимо, що в умовах альтернативи вибіркове середнє має розподіл $\widehat{\mu}_n \simeq N(\mu_1, \sigma^2/n)$. Тому потужність дорівнює

$$P_{\theta_1}\left(\tfrac{\sqrt{n}}{\sigma}(\widehat{\mu}_n - \mu_0) \geq x_\alpha\right) =$$
$$P_{\theta_1}\left(\tfrac{\sqrt{n}}{\sigma}(\widehat{\mu}_n - \mu_1) - \sqrt{n}(\mu_0 - \mu_1)/\sigma \geq x_\alpha\right) =$$
$$1 - \Phi(x_\alpha + \sqrt{n}(\mu_0 - \mu_1)/\sigma) =$$
$$\Phi(-x_\alpha + \sqrt{n}(\mu_1 - \mu_0)/\sigma) \to 1, \quad n \to \infty,$$

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 передостаннього 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

$$n \geq \sigma^2(x_\alpha + x_\beta)^2 \ / \ (\mu_1 - \mu_0)^2.$$

Отже, кількість необхідних спостережень зростає зі зменшенням імовірностей похибок першого та другого роду α, β, пропорційна дисперсії спостережень та обернено пропорційна квадрату різниці між гіпотетичними середніми.

Вправа. Обчислити критерій відношення вірогідностей для перевірки гіпотези про значення параметра для кратної вибірки з (а) пуассоновим, (б) показниковим розподілом. Знайти потужність критерію, перевірити його незміщеність та консистентність.

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 функцією $M_\theta \pi(\widehat{\varkappa})$ при $\theta \in \Theta_0$, а потужність – значеннями $M_\theta \pi(\widehat{\varkappa})$ при $\theta \in \Theta_1$.

Ці значення обчислюються за теоремою про обчислення математичного сподівання функції від випадкової величини

$$M_\theta \pi(\widehat{\varkappa}) = \int \pi(t) P_\theta(\widehat{\varkappa} \in dt) = \int_S \pi(\widehat{\varkappa}(x)) P_\theta(X \in dx).$$

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.

На відміну від звичайних критеріїв, для рандомізованих справедливе таке твердження.

Лема (про побудову рандомізованого критерію). Для довільної скалярної статистики $\widehat{\varkappa}$ та для довільного рівня $\alpha \in (0, 1)$ знайдеться простий рандомізований критерій з деяким критичним рівнем t_α такий, що його вірогідний рівень при заданому θ збігається з α.

Дійсно, за наведеною вище формулою цей рівень має вигляд

$$M_\theta \pi(\widehat{\varkappa}) = P_\theta(\widehat{\varkappa} = t_\alpha)\pi(t_\alpha) + P_\theta(\widehat{\varkappa} > t_\alpha) = \alpha,$$

якщо обрати

$$t_\alpha = \inf(t : P_\theta(\widehat{\varkappa} > t) \le \alpha),$$

$$\pi(t_\alpha) = (\alpha - P_\theta(\widehat{\varkappa} > t_\alpha))\ /\ P_\theta(\widehat{\varkappa} = t_\alpha).$$

У випадку, коли t_α є точкою неперервності: $P_\theta(\widehat{\varkappa} = t_\alpha) = 0$, можна обрати $\pi(t_\alpha) = 1$. У такому випадку простий рандомізований критерій із критичним рівнем збігається зі звичайним $(\widehat{\varkappa}(X), [t_\alpha, \infty))$. Отже, у випадку неперервності розподілу статистики $\widehat{\varkappa}$ прості рандомізовані критерії збігаються зі звичайними.

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 вірогідного 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*

$$M_{\theta_0}\pi^*(l_{01}(X)) = \alpha.$$

Крім того, потужність цього критерію не менша за рівень α :

$$M_{\theta_1}\pi^*(l_{01}(X)) \geq \alpha.$$

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

$$M_{\theta_1}\pi^*(l_{01}(X)) - M_{\theta_1}\pi(T(X)) = \int_{S^+\,\cup\,S^-}(\pi^*(l_{01}(x)) - \pi(T(x)))F_1(dx) =$$
$$\int_{S^+\,\cup\,S^-}(\pi^*(l_{01}(x)) - \pi(T(x)))\ l_{01}(x)F_0(dx) \geq$$
$$\int_{S^+\,\cup\,S^-}(\pi^*(l_{01}(x)) - \pi(T(x)))\ t^*_\alpha F_0(dx) =$$
$$t^*_\alpha\left(M_{\theta_0}\pi^*(l_{01}(X)) - M_{\theta_0}\pi(T(X))\right) = t^*_\alpha(\alpha - M_{\theta_0}\pi(T(X)) \geq 0,$$

оскільки вірогідний рівень критерію не більший за α. Передостання нерівність тут є наслідком означення простого рандомізованого критерію відношення вірогідностей: з кожного включення $x \in S^{\pm}$ випливає відповідно $\pi^*(l_{01}(x)) \geqq 0$, тобто $l_{01}(x)) \gtreqless t_\alpha^*$, і одночасно різниця під знаком інтегралу додатна чи від'ємна відповідно.

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 дійсна статистика $T(X)$, що для кожної пари $\theta_0 < \theta_1 \in \Theta$ знайдеться дійсна 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 відносно статистики $T(X)$. Тоді для кожного вірогідного рівня $\alpha \in (0, 1)$ простий рандомізований критерій $(T(X),\ \pi^*(t),\ t_\alpha^*)$, який однозначно визначається з рівняння $M_{\theta_0} \pi^*(T(X)) = \alpha$, є рівномірно найбільш потужним рівня α для* *перевірки складної гіпотези* $H_0 : \theta \le \theta_0$ *проти складної альтернативи* $H_1 : \theta > \theta_0$.

Доведення.

Нехай $\theta_0 < \theta_1$. Розглянемо спочатку просту гіпотезу $H_0' : \theta = \theta_0$ проти простої альтернативи $H_1' : \theta = \theta_1$. Визначимо простий рандомізований критерій відношення вірогідностей $(l_{01}(X),\ \widetilde{\pi}(t),\ g_{01}(t^*_\alpha))$, де $\widetilde{\pi}(g_{01}(t^*_\alpha)) = \pi^*(t^*_\alpha)$. Тоді з тотожностей

$$\{l_{01}(X) \gtreqless g_{01}(t^*_\alpha)\} = \{g_{01}(T(X)) \gtreqless g_{01}(t^*_\alpha)\} = \{T(X) \gtreqless t^*_\alpha\}$$

випливає, що цей критерій збігається з побудованим у теоремі критерієм $(T(X), \pi^*(t), t^*_\alpha)$ і, зокрема, має такий же рівень α.

За теоремою Неймана – Пірсона для рандомізованих критеріїв побудований критерій $(T(X), \pi^*(t), t^*_\alpha)$ є найбільш потужним рівня α для перевірки H_0' проти H_1', а його потужність не менша за рівень:

$$M_{\theta_1}\pi^*(T(X)) \ge \alpha = M_{\theta_0}\pi^*(T(X)).$$

Отже, функція $M_\theta \pi^*(T(X))$ не спадає при $\theta > \theta_0$.

Аналогічно доводимо, що $M_\theta \pi^*(T(X))$ не спадає при $\theta < \theta_0$.

Звідси $M_\theta \pi^*(T(X)) \le \alpha$ для всіх $\theta \le \theta_0$, де при $\theta = \theta_0$ має місце рівність. Тому критерій $(T(X), \pi^*(t), t^*_\alpha)$ має рівень α для перевірки складної гіпотези $H_0 : \theta \le \theta_0$.

Нехай $(\widehat{\varkappa}(X),\ \pi(t))$ довільний рандомізований критерій рівня α для перевірки H_0, а $\theta_1 > \theta_0$.Тоді, зокрема, $M_{\theta_0}\pi(\widehat{\varkappa}(X)) \le \alpha$ і внаслідок доведеної вище властивості найбільшої потужності критерію $(T(X), \pi^*(t), t^*_\alpha)$ отримуємо

$$M_{\theta_1}\pi(\widehat{\varkappa}(X)) \le M_{\theta_1}\pi^*(T(X)),$$

тобто при кожному альтернативному значенні $\theta = \theta_1 > \theta_0$ потужність критерію $(\widehat{\varkappa}(X),\ \pi(t))$ не більша за потужність простого критерію $(T(X), \pi^*(t), t^*_\alpha)$ $\boxdot$

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 $H_0 : \theta \le \theta_0$ проти $H_1 : \theta > \theta_1$ є звичайний критерій $(\widehat{\mu}_n,\ [t_\alpha, \infty))$.

Вправа.

(1) Спостерігаються дві незалежні кратні нормальні вибірки з відомими середніми та невідомими дисперсіями $X = (\xi_{ik}, i = \overline{1, n_k}, k = 1, 2)$, $\xi_{ik} \simeq N(\mu_k, \sigma_k^2)$. Побудувати рівномірно найбільш потужний критерій відношення вірогідностей для перевірки гіпотези $H_0 : \sigma_1^2 \le \sigma_2^2$ проти альтернативи $H_0 : \sigma_1^2 > \sigma_2^2$.

(2) Знайти рівномірно найбільш потужний критерій відношення вірогідностей для перевірки гіпотези про значення ймовірності успіху для вибірки з біноміального розподілу.

(3) Знайти рівномірно найбільш потужний критерій для перевірки гіпотези $H_0 : \theta = \theta_0$ проти альтернативи $H_1 : \theta > \theta_0$ для вибірки з рівномірного розподілу $U(0, \theta)$.

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 розглянемо задачу перевірки простих гіпотез $H_i : \theta = \theta_i, i = 0, 1$ для кратної вибірки $X = (\xi_1, ..., \xi_n)$. Позначимо $f(y, \theta)$ щільність одного 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 конзистентності ОМВ $M_{\theta_0} h(\xi_1) < 0$. Аналогічно можна довести, що $M_{\theta_1} h(\xi_1) > 0$.

Найбільш потужний критерій відношення вірогідностей має критичну область вигляду $\{l_{01}(X) \geq l_\alpha\} = \{S_n \geq \ln l_\alpha\}$.

Нехай для кожного $k = \overline{1, n}$ задано сталі $c_{0k} < 0 < c_{1k}$. Позначимо $S_k = \sum_{j=1}^{k} h(\xi_j)$. Розглянемо узагальнений послідовний критерій відношення вірогідностей:

(0) якщо $\{S_1 \in (c_{01}, c_{11}), ..., S_{k-1} \in (c_{0,k-1}, c_{1,k-1}),\ S_k \le c_{0k}\}$ для деякого $k \le n$, то спостереження припиняються в такий момент k, і приймається нульова гіпотеза H_0,

(1) якщо $\{S_1 \in (c_{01}, c_{11}), ..., S_{k-1} \in (c_{0,k-1}, c_{1,k-1}),\ S_k \ge c_{1k}\}$ для деякого $k \le n$, то спостереження припиняються в такий момент k, і приймається альтернативна гіпотеза H_1,

(n) інакше приймається рішення на користь H_0.

Вибір знаків сталих обгрунтовується 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 $n \to \infty$.

На відміну від критерію відношення вірогідностей послідовний критерій спирається на випадкову кількість спостережень

$$\tau = \min(n, \inf(k \ge 1 : S_k \notin (c_{0k}, c_{1k}))).$$

Якщо обрати $c_{0k} = -\infty, k \le n$, $c_{1k} = \infty, k < n$, і $c_{1n} = \ln l_\alpha$, то отримаємо критерій відношення вірогідностей як частковий випадок узагальненого критерію. Тому 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: $\inf M\tau \le n$. 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 A_{0k} подію в умові (0) критерію, A_{1k} – в умові (1) та 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 $L_k(X_k, \theta_1) \ / \ L_k(X_k, \theta_0) = \exp\left(\sum_{j=1}^k h(\xi_j)\right) = \exp(S_k)$. 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 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 $A_{ik} = \{X_k \in B_{ik}\}$. Із включення $x_k \in B_{1k}$ випливає, що $L_k(x_k, \theta_1) \ / \ L_k(x_k, \theta_0) = \exp(\sum_{j=1}^k h(y_j)) \geq \exp(c_1) = C_1$. Аналогічно, при $x_k \in B_{0k}$ маємо $L_k(x_k, \theta_1) \ / \ L_k(x_k, \theta_0) \leq \exp(c_0) = C_0$. Звідси

$$P(A_{1k}/H_0) = \int_{B_{1k}} L(x_k, \theta_0)\mu_k(dx_k) \leq \int_{B_{1k}} \frac{L(x_k,\theta_1)}{C_1}\mu_k(dx_k) = \frac{P(A_{1k}/H_1)}{C_1},$$
$$P(A_{0k}/H_1) = \int_{B_{0k}} L(x_k, \theta_1)\mu_k(dx_k) \leq \int_{B_{0k}} C_0 L(x_k, \theta_0)\mu_k(dx_k) =$$
$$C_0 P(A_{1k}/H_0).$$

Нехай α, β – ймовірності похибок першого та другого роду для побудованого вище послідовного критерію. Тоді

$$\alpha = P(\cup_{k=1}^n A_{1k}/H_0) = \sum_{k=1}^n P(A_{1k}/H_0) \leq \sum_{k=1}^n P(A_{1k}/H_1)/C_1 =$$
$$(1 - \sum_{k=1}^n P(A_{0k}/H_1) - P(A_n/H_1))/C_1 = (1-\beta)/C_1,$$
$$\beta = \sum_{k=1}^n P(A_{0k}/H_1) + P(A_n/H_1) \leq \sum_{k=1}^n C_0 P(A_{1k}/H_0) + C_0 P(A_n/H_0) =$$
$$(1-\alpha)C_0.$$

Отже, критичні значення $c_i = \ln C_i$ та похибки в послідовному критерії відношення вірогідностей пов'язані системою нерівностей

$$\alpha \leq (1-\beta)/C_1, \quad \beta \leq (1-\alpha)C_0.$$

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 – ймовірності похибок для цього критерію. Тоді з нерівностей для ймовірностей похибок 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:

$$M_\theta\ S_\tau = m_\theta\ M_\theta \tau, \quad m_\theta = M_\theta S_1 = M_\theta h(\xi_1).$$

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

$$M_\theta\ S_\tau = \sum_{k\ge 1} M_\theta \left(h(\xi_k)\ 1_{\ \{k\le\tau\}} \right) = \sum_{k\ge 1} M_\theta h(\xi_k)\ P_\theta(k \le \tau) =$$
$$m_\theta \sum_{k\ge 1} \sum_{j\ge k} P_\theta(\tau = j) = m_\theta \sum_{j\ge 1} j\ P_\theta(\tau = j) = m_\theta\ M_\theta \tau.$$

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 $n \to \infty$. Ймовірність P_{θ_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) маємо $S_\tau \geq c_1$, а за умови (0) – $S_\tau \leq c_0$. Припустимо, що останні нерівності є наближеними рівностями, тобто суми S_n не занадто перестрибують відповідні рівні. Тоді $M_{\theta_0} S_\tau \approx \alpha c_1 + (1-\alpha)c_0$. Отже, внаслідок тотожності Вальда

$$M_{\theta_0}\tau \approx (\alpha c_1 + (1-\alpha)c_0)/m_{\theta_0}.$$

Якщо цю формулу застосувати до схеми нормальних спостережень із невідомим середнім, що викладена у попередньому прикладі, то можна довести, що при малих значеннях α, β має місце *подвійна економія числа спостережень* у порівнянні з критерієм відношення вірогідностей.

Вправа. Нехай кратна вибірка $X = (\xi_1, ..., \xi_n)$ утворена спостереженнями з показниковими розподілами $Exp(\theta)$. Знайти розподіл випадкової величини τ для одностороннього критерію з $c_0 = 0$.

18. Метод найменших квадратів для лінійної регресії

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.

18.1. Модель лінійної регресії

Зокрема, в теорії лінійної регресії розглядається модель лінійної залежності

$$\xi_j = \sum\nolimits_{i=1}^{k} \theta_i \, t_{ij} + \varepsilon_j, \ j = \overline{1,n},$$

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 (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

$$M_\theta \ \varepsilon = 0, \quad \mathrm{Cov}_\theta \ \varepsilon = \sigma^2 I$$

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

$$\widehat{\theta} = \arg\min_{\theta \in \mathbb{R}^k} (X - T'\theta)^2.$$

Очевидно, що функція $\widehat{\theta}$ обчислюється за X і T, тобто є оцінкою. Існування оцінки МНК очевидне, оскільки квадратична функція неперервна та обмежена знизу.

Теорема (про співвідношення між оцінкою МНК та ОМВ). *Припустимо, що похибки ε_i 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, тому логарифмічна функція вірогідності дорівнює

$$\ln L(X, \theta) = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2}(X - T'\theta)^2.$$

Очевидно, що точка максимуму $L(X, \theta)$ за θ збігається з оцінкою МНК

$\boxdot$

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

$$V \equiv TT'$$

– симетрична матриця $k \times k$.

(б) Якщо rang $T = k$, то $\det V \neq 0$ і єдина оцінка МНК має вигляд

$$\widehat{\theta} = V^{-1}TX.$$

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 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:*

$$(X - T'\theta)^2 = (X - T'\widehat{\theta})^2 + (T'(\widehat{\theta} - \theta))^2,$$

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.

Зауваження. Відношення нормованих статистик із пункту (в) має розподіл Фішера після нормування на кількості ступенів свободи і використовується для перевірки гіпотези про суттєвість залежності теоретичного середнього спостережень від параметра, що оцінюється.

Доведення.

За теоремою про лінійні перетворення нормальних векторів статистика $\widehat{\theta} = V^{-1}TX$ має нормальний розподіл, тому (а) випливає з рівнянь

$$M_\theta \widehat{\theta} = MV^{-1}T(T'\theta + \varepsilon) = \theta + V^{-1}TM\varepsilon = \theta,$$

$$\mathrm{Cov}_\theta(\widehat{\theta}) = V^{-1}T\mathrm{Cov}(X)(V^{-1}T)' = \sigma^2 V^{-1}TT'V^{-1} = \sigma^2 V^{-1},$$

де використано теорему про коваріаційну матрицю лінійного перетворення.

Твердження (б) є наслідком (в), оскільки

$$M_\theta \widehat{\sigma}^2_{n-k} = M_\theta\ SS_v\ /(n-k) = \sigma^2 M_\theta\ (SS_v\ /\sigma^2)/(n-k) =$$

$$\sigma^2 M_\theta\ (\chi^2_{n-k})/(n-k) = \sigma^2\ (n-k)/(n-k) = \sigma^2.$$

Для доведення (в) зауважимо, що $n \times n$-матриця $\Pi \equiv T'V^{-1}T$ симетрична, ідемпотентна: $\Pi^2 = T'V^{-1}TT'V^{-1}T = T'V^{-1}VV^{-1}T = T'V^{-1}T = \Pi$ та має ранг k. Тому існує ортонормована матриця U, що приводить її до діагональної матриці вигляду

$$U\Pi U' = I_n(k) = (\delta_{ij} 1_{i \le k},\ i, j = \overline{1, n}).$$

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. Далі, за теоремою про рівняння методу найменших квадратів

$$\widehat{\theta} = V^{-1}TX = V^{-1}T(T'\theta + \varepsilon) = \theta + V^{-1}T\varepsilon.$$

Тому внаслідок теореми про розклад суми квадратів відхилень

$$\varepsilon^2 = (X - T'\theta)^2 = SS_v + SS_r = (X - T'\widehat{\theta})^2 + (T'(\widehat{\theta} - \theta))^2 =$$
$$(T'\theta + \varepsilon - T'(\theta + V^{-1}T\varepsilon))^2 + (T'(\theta + V^{-1}T\varepsilon) - T'\theta)^2 =$$
$$((I - T'V^{-1}T)\varepsilon)^2 + (T'V^{-1}T\varepsilon)^2 = ((I - \Pi)\varepsilon)^2 + (\Pi\varepsilon)^2 =$$
$$\varepsilon'(I - \Pi)'(I - \Pi)\varepsilon + \varepsilon'\Pi'\Pi\varepsilon = \varepsilon'(I - \Pi)\varepsilon + \varepsilon'\Pi\varepsilon.$$

Звідси

$$SS_v/\sigma^2 + SS_r/\sigma^2 = \varepsilon'(I - \Pi)\varepsilon/\sigma^2 + \varepsilon'\Pi\varepsilon/\sigma^2 =$$
$$\zeta'U(I - \Pi)U'\zeta + \zeta'U\Pi U'\zeta =$$
$$\zeta'(I - I_n(k))\zeta + \zeta' I_n(k)\zeta = \sum_{i=k+1}^{n} \zeta_i^2 + \sum_{i=1}^{k} \zeta_i^2 = \chi^2_{n-k} + \chi^2_k,$$

де величини в правій частині незалежні за теоремою про векторні перетворення незалежних величин ζ_i і мають за означенням відповідні хі-квадрат розподіли $\boxdot$

Теорема (теорема Гаусса – Маркова). *Нехай* $\operatorname{rang} T = k$, $\widehat{\theta}$ – *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 $\widehat{z} = Z\widehat{\theta}$ є лінійною незміщеною оцінкою для z, і має найменшу дисперсію (тобто найменші діагональні елементи коваріаційної матриці) у класі всіх лінійних за X незміщених оцінок вектора z.

Доведення.

Нехай $\widetilde{z} = CX$ є довільною лінійною незміщеною оцінкою для z. Тоді

$$M_\theta \widetilde{z} = C\ M_\theta X = CT'\theta = z = Z\theta,\ \forall \theta \in \mathbb{R}^k,$$

звідки виводимо тотожність

$$Z = CT'.$$

За теоремою про коваріаційну матрицю лінійного перетворення

$$\begin{gathered}\mathrm{Cov}_\theta(\widetilde{z}) = C\mathrm{Cov}_\theta(X)C' = C\mathrm{Cov}_\theta(\varepsilon)C' = \sigma^2 CC' = \\ \sigma^2(ZV^{-1}T + C - ZV^{-1}T)(ZV^{-1}T + C - ZV^{-1}T)' = \\ \sigma^2 ZV^{-1}Z' + \sigma^2(C - ZV^{-1}T)(C - ZV^{-1}T)' \geq \\ \sigma^2 ZV^{-1}Z' = \mathrm{Cov}_\theta(Z\widehat{\theta}) = \mathrm{Cov}_\theta(\widehat{z}),\end{gathered}$$

де нерівність матриць означає одночасну нерівність всіх відповідних діагональних елементів і використано ортогональність матриць $ZV^{-1}T$ і $C - ZV^{-1}T$. Остання випливає з наведеної вище тотожності $Z = CT'$ та рівностей

$$\begin{gathered}ZV^{-1}T(C - ZV^{-1}T)' = ZV^{-1}(TC' - TT'V^{-1}Z') = \\ ZV^{-1}(TC' - Z') = ZV^{-1}(TC' - (CT')') = 0\ \boxdot\end{gathered}$$

Вправа. Нехай $X = (\xi_j,\ j = \overline{1,n})$ – довільна вибірка така, що $M\xi_j = \mu$, $D\xi_j = \sigma^2$, $\mathrm{Cov}(\xi_j, \xi_k) = \rho\sigma^2$, $j \neq k$. Довести, що в класі лінійних за спостереженнями ξ_j найменшу дисперсію має вибіркове середнє $\widehat{\mu}_n$.

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 регресії матиме вигляд

$$\xi_j = a + bx_j + \varepsilon_j, \quad j = \overline{1, n}.$$

Матриці T, V задаються рівностями

$$T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & \ldots & x_n \end{pmatrix},$$

$$V = TT' = n \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix}, \quad \overline{x} = \frac{1}{n}\sum_{j=1}^n x_j,\ \overline{x^2} = \frac{1}{n}\sum_{j=1}^n x_j^2.$$

Звідси обчислюємо обернену матрицю V^{-1} та оцінки МНК

$$V^{-1}T = \frac{1}{n\overline{\sigma}^2} \begin{pmatrix} \overline{x^2} - x_1\overline{x} & \overline{x^2} - x_2\overline{x} & & \overline{x^2} - x_n\overline{x} \\ x_1 - \overline{x} & x_2 - \overline{x} & \ldots & x_n - \overline{x} \end{pmatrix},$$

$$\overline{\sigma}^2 = \overline{x^2} - (\overline{x})^2 = \frac{1}{n}\sum_{j=1}^n (x_j - \overline{x})^2,$$

$$\widehat{b} = \frac{1}{n\overline{\sigma}^2}\sum_{j=1}^n \xi_j (x_j - \overline{x}),$$

$$\widehat{a} = \frac{1}{n\overline{\sigma}^2}\sum_{j=1}^n \xi_j (\overline{x^2} - x_j\overline{x}) = -\overline{x}\widehat{b} + \frac{1}{n\overline{\sigma}^2}\sum_{j=1}^n \xi_j (\overline{x^2} - (\overline{x})^2) =$$

$$\frac{1}{n}\sum_{j=1}^n \xi_j - \overline{x}\widehat{b} = \widehat{\mu}_n - \overline{x}\widehat{b}\ \boxdot$$

Оцінка $\widehat{b}$ називається вибірковим коефіцієнтом кореляції векторів (ξ_j), (x_j). За теоремою про незміщеність і розподіл оцінки МНК $D(\widehat{b}) = \sigma^2(V^{-1})_{22} = \sigma^2/n\overline{\sigma}^2$. Якщо спостереження ξ_j отримано з активного експерименту, тобто статистик має можливість обирати точки спостережень (x_j), то для мінімізації дисперсії оцінки $\widehat{b}$ слід максимізувати величину $\overline{\sigma}^2$.

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 $V = \left(\sum_j x_j^i x_j^l, \; 0 \le i, l \le k\right) \approx (1/(i+l+1), \; 0 \le i, l \le k)$. Можна довести, що визначник цієї матриці дуже близький до нуля навіть при великих k. Отже, обчислення оцінки МНК стикатиметься з великими труднощами.

Тому для поліпшення якості оцінювання в поліноміальній схемі доцільно розглянути її в еквівалентному вигляді

$$\xi_j = \sum_{i=0}^{k} \theta_i \; \varphi_i(x_j) + \varepsilon_j, \; j = \overline{1, n},$$

де $\varphi_i(x)$ – деякі поліноми ступеня i від x. Матриця V у такій схемі набуває вигляду

$$V = \left(\sum_{j=1}^{n} \varphi_i(x_j) \; \varphi_l(x_j), \; 0 \le i, l \le k \right).$$

Для поліпшення якостей оцінки МНК поліноми φ_i обирають так, щоб матриця V була діагональною. З огляду на її визначення розглянемо таку білінійну форму (скалярний добуток) від функцій f, g:

$$(f, g) \equiv \sum_{j=0}^{k} f(x_j) \; g(x_j).$$

Визначимо рекурентно послідовність поліномів Чебишева

$$\varphi_0(x) \equiv 1, \quad \varphi_k(x) \equiv x^k - \sum_{i=0}^{k-1} \varphi_i(x) \frac{(x^k, \varphi_i)}{(\varphi_i, \varphi_i)}, \; k \ge 1.$$

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, матриця $V = ((\varphi_i, \varphi_i)\delta_{ik})$ є діагональною і оцінка МНК дорівнює

$$\widehat{\theta} = V^{-1}TX = \left(\sum_{j=1}^{n} \xi_j \; \varphi_i(x_j) \; / \; (\varphi_i, \varphi_i), \quad i = \overline{0, k} \right).$$

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 у випадку нормальних похибок – незалежні.

Вправа. Розглянемо лінійну двохфакторну модель із вибіркою $X = (\xi_{ij}, i = \overline{1,n}.j = \overline{1,m})$, що 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 $\sum_{i=1}^n \alpha_i = 0, \sum_{j=1}^m \beta_j = 0$.Довести, що вектор

$$\widehat{\theta} = (\widehat{X}_{\bullet\bullet},\ (\widehat{X}_{i\bullet} - \widehat{X}_{\bullet\bullet}, i = \overline{1,n}),\ (\widehat{X}_{\bullet j} - \widehat{X}_{\bullet\bullet}, j = \overline{1,m}),\ \widehat{\sigma}^2)$$

є оптимальною оцінкою для θ, що складається з 4 незалежних груп. Тут групові середні статистики мають вигляд

$$\widehat{X}_{\bullet\bullet} = \sum\nolimits_{i=1}^n \sum\nolimits_{j=1}^m \xi_{ij}/nm,$$

$$\widehat{X}_{i\bullet} = \sum\nolimits_{j=1}^m \xi_{ij}/m, \quad \widehat{X}_{\bullet j} = \sum\nolimits_{i=1}^n \xi_{ij}/n,$$

$$\widehat{\sigma}^2 = \sum\nolimits_{i=1}^n \sum\nolimits_{j=1}^m (\xi_{ij} - \widehat{X}_{i\bullet} - \widehat{X}_{\bullet j} + \widehat{X}_{\bullet\bullet})^2/(n-1)(m-1).$$

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

$$L_2(P) = \{\xi : \Omega \to \mathbb{C},\ M\,|\xi|^2 < \infty\}$$

гільбертів простір квадратично інтегровних комплекснозначних випадкових величин зі скалярним добутком $(\xi, \eta) = M\xi\overline{\eta}$, де $\overline{x}$ комплексно спряжене до x, із нормою $\|\xi\|^2 = (\xi, \xi) = M\,|\xi|^2$ та відповідною середньоквадратичною (с.к.) границею $\xi = l.i.m\ \xi_n$.

Нехай $G \subset L_2(P)$ – довільна підмножина. Позначимо породжений G лінійний підпростір

$$\pounds(G) = \{\eta = \sum\nolimits_{k=1}^{n} c_k\xi_k,\ \xi_k \in G,\ c_k \in \mathbb{C}\}$$

та його с.к. замикання

$$\aleph(G) = \{\xi \in L_2(P),\ \xi = l.i.m\ \eta_n,\ \eta_n \in \pounds(G)\}.$$

Назвемо ізометрією підмножин гільбертових просторів таку їх лінійну біекцію, що зберігає норму (а, отже, і скалярний добуток).

Теорема (про продовження ізометрії). *Нехай* $(H_k, \|\cdot\|_k),\ k = 1, 2,$ *пара гільбертових просторів, кожний з яких містить щільний лінійний підпростір* $H_{0k} \subset H_k$, $H_k = \aleph(H_{0k})$. *Для довільної ізометрії підпросторів*

$\Im_0 : H_{01} \to H_{02}$ *існує 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 $h_{2n} = \Im_0 h_{1n}$. З фундаментальності h_{1n} виводимо фундаментальність h_{2n}

$$\|h_{2n} - h_{2m}\|_2 = \|\Im_0(h_{1n} - h_{1m})\|_2 = \|h_{1n} - h_{1m}\|_1 \to 0,$$

а отже, і збіжність цієї послідовності. 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 $\boxdot$

19.1. Стаціонарні випадкові процеси

Означення. *Нехай параметрична множина* $T = \mathbb{Z}$ *або* $T = \mathbb{R}$. Квадратично інтегровний *комплекснозначний випадковий процес*

$$(\xi_t,\ t \in T)\ \subset\ L_2(P)$$

називається стаціонарним (стаціонарним процесом другого порядку, *або ж* стаціонарним у широкому розумінні), *якщо*

$$M\xi_t = m,\ \mathrm{Cov}(\xi_t,\ \overline{\xi_s}) = r(t-s),\ \ \forall t, s \in T$$

для деяких комплексних m, $r(t)$, $t \in T$. Функція $r(t)$ називаєтьcя коваріаційною функцією *процесу.*

Приклади

(а) Стохастична гармоніка. Нехай ζ_0 квадратично інтегровна центрована випадкова величина: $M\zeta_0 = 0, M\left|\zeta_0\right|^2 < \infty$, а λ дійсна стала. Тоді процес $\xi_t = \zeta_0 \exp(i\lambda t)$ є стаціонарним із нульовим середнім та коваріаційною функцією $r(t) = M\left|\zeta_0\right|^2 \exp(i\lambda t)$.

(б) Процес зі скінченним спектром. Нехай $(\zeta_k, k = \overline{1,n})$ набір із центрованих некорельованих випадкових величин: $M\zeta_k = 0$, $M\zeta_k\overline{\zeta_l} = 0, k \neq l$, а λ_k – дійсні сталі. Тоді процес

$$\xi_t = \sum\nolimits_{k=1}^{n} \zeta_k \exp(i\lambda_k t)$$

є стаціонарним із нульовим середнім та коваріаційною функцією

$$\mathrm{Cov}(\xi_t,\ \overline{\xi_s}) = M \sum\nolimits_{k=1}^{n} \zeta_k \exp(i\lambda_k t) \sum\nolimits_{j=1}^{n} \overline{\zeta_j} \exp(-i\lambda_j s) =$$

$$\sum\nolimits_{k=1}^{n} M\left|\zeta_k\right|^2 \exp(i\lambda_k(t-s)) = r(t-s).$$

Вправа.Нехай $(\varepsilon_n, n \in \mathbb{Z})$ послідовність центрованих некорельованих нормованих величин: $M\varepsilon_n = 0$, $M\varepsilon_n\overline{\varepsilon_k} = \delta_{nk}$,а $(c_k, k = \overline{0,m})$ – довільні сталі. Тоді послідовність $\xi_n = \sum_{k=0}^{m} c_k\varepsilon_{n-k}$ є стаціонарною другого порядку. Знайти її коваріаційну функцію.

Означення. *Процес $(\xi_t,\ t \in T)$ є* с.к. неперервним, *якщо*

$$M\left|\xi_t - \xi_s\right|^2 \equiv \left\|\xi_t - \xi_s\right\|^2 \to 0,\ t \to s,\ \forall s.$$

Зауваження (про критерій с.к.неперервності). *С.к. неперервність стаціонарного процесу еквівалентна неперервності його коваріаційної функції в нулі.*

Дійсно, після центрування процесу

$$\|\xi_t - \xi_s\|^2 = M(\xi_t - \xi_s)(\overline{\xi_t - \xi_s}) = M(|\xi_t|^2 - \xi_t\overline{\xi_s} - \xi_s\overline{\xi_t} + |\xi_s|^2) =$$
$$2(r(0) - \operatorname{Re} r(t-s)),$$
$$|r(t) - r(0)|^2 = \left|M(\xi_t - \xi_0)\overline{\xi_0}\right|^2 \le \|\xi_t - \xi_0\|^2 r(0).$$

Теорема (про властивості коваріаційної функції). *Функція $r(t)$ є коваріаційною функцією деякого стаціонарного процесу тоді й тільки тоді, коли вона є додатно визначеною, тобто для довільних n, $t_k \in T$, $c_k \in \mathbb{C}$*

$$\overline{r(t_1)} = r(-t_1),\ \sum_{j,k=1}^{n} c_k\overline{c_j}\, r(t_k - t_j) \ge 0.$$

Доведення.
Необхідність виводиться обчисленням наведених форм через ξ_t

$$\overline{r(t)} = \overline{M\xi_t\overline{\xi_0}} = M\overline{\xi_t}\xi_0 = M\overline{\xi_0}\xi_{-t} = r(-t),$$
$$\sum_{j,k=1}^{n} c_k\overline{c_j}\, r(t_k - t_j) = \sum_{j,k=1}^{n} c_k\overline{c_j}\, M\xi_{t_k}\overline{\xi_{t_j}} =$$
$$M\sum_{j,k=1}^{n} c_k\overline{c_j}\, \xi_{t_k}\overline{\xi_{t_j}} = M\left|\sum_{k=1}^{n} c_k\, \xi_{t_k}\right|^2 \ge 0.$$

Достатність випливає з існування відповідних нормальних випадкових послідовностей, доведення не наводиться ⊡

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 $\mathbb{R}$ 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 $F(x) \equiv F((-\infty, x))$ – 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 $\lambda \in [-\pi, \pi]$

$$f_N(\lambda) = \frac{1}{2\pi N} \sum\nolimits_{k,l=1}^{N} r(k-l) \exp(-ik\lambda + il\lambda) =$$

$$\frac{1}{2\pi} \sum\nolimits_{|n| \leq N} \left(1 - \frac{|n|}{N}\right) r(n) \exp(-in\lambda),$$

де зроблено заміну змінної $n = k - l$.

Ці функції дійсні та невід'ємні за умовою додатної визначеності, в якій слід обрати $t_k = k$, $c_k = \exp(-ik\lambda)$.

Позначивши

$$F_N(B) = \int_{B \,\cap\, [-\pi,\pi]} f_N(\lambda)\, d\lambda,$$

з урахуванням ортогональності гармонік

$$\int_{-\pi}^{\pi} \exp(i\lambda k) \exp(-i\lambda j) d\lambda = 2\pi\ \delta_{kj},$$

обчислюємо

$$\int_{-\pi}^{\pi} \exp(i\lambda n) dF_N(\lambda) = \left(1 - \frac{|n|}{N}\right)^+ r(n),$$

де $x^+ = \max(x, 0)$.

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 $\{F_N(-\infty, x)/r(0)\}$ слабко компактна, оскільки верхня межа в критерії нульова вже при $c > \pi$. Тому за означенням слабкої компактності для деякої скінченної міри F та деякої підпослідовності $F_{N_k} \to^W F$, $N_k \to \infty$. Використовуючи цю збіжність у попередній рівності при $N = N_k$, дістанемо шукане зображення:

$$\int_{-\pi}^{\pi} \exp(i\lambda n) dF(\lambda) = \lim{}_{k\to\infty} \int_{-\pi}^{\pi} \exp(i\lambda n) dF_{N_k}(\lambda) =$$

$$\lim{}_{k\to\infty}(1 - \frac{|n|}{N_k})^+ r(n) = r(n), \ \forall n \in \mathbb{Z} \boxdot$$

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 $(\xi_t, \ t \in T)$ тоді й тільки тоді, коли справедливе 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 $T = \mathbb{Z}$ спектральна міра зосереджена на відрізку $[-\pi, \pi]$.

Зауваження.За формулою обертання для характеристичних функцій справедлива тотожність

$$F(t) - F(s) = \frac{1}{2\pi} \lim_{\varepsilon \downarrow 0} \int_s^t dx \int_{-\infty}^{\infty} \exp(-iux - \varepsilon^2 u^2)\, r(u) du.$$

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 $\boxdot$

19.2. Міри з ортогональними значеннями, стохастичні інтеграли

Означення. *Сім'я комплекснозначних випадкових величин* $(\mu(B),\ \ B \in \mathfrak{B}(\mathbb{R})) \subset L_2(P)$ *називається* випадковою мірою з ортогональними значеннями, *якщо для деякої скінченної міри* m *на* $\mathfrak{B}(\mathbb{R})$ (структурної міри) *виконуються умови*

(а) $M\ \mu(A)\ \overline{\mu(B)} = m(A \cap B),\ \forall A, B \in \mathfrak{B}(\mathbb{R}),$

(б) $\mu(A \cup B) = \mu(A) + \mu(B)$ *м.н. для всіх несумісних* $A, B \in \mathfrak{B}(\mathbb{R}),$

(в) $M\ \mu(A) = 0.$

Приклад. Нехай $(\zeta_k,\ \ k \in \mathbb{Z})$ – послідовність попарно некорельованих випадкових величин із нульовим середнім. Тоді $\mu(A) = \sum_{k \in A} \zeta_k$ є мірою з ортогональними значеннями та зі структурною мірою $m(A) = \sum_{k \in A} M\,|\zeta_k|^2$:

$$M\ \mu(A)\ \overline{\mu(B)} = \sum_{k \in A} \sum_{j \in B} M \zeta_k \overline{\zeta_j} = \sum_{k = j \in A \cap B} M\,|\zeta_k|^2 = m(A \cap B),$$

адитивність та центрованість μ очевидні.

Означення. *Нехай* $(\mu(B),\ B \in \mathfrak{B}(\mathbb{R}))$ – *випадкова міра з ортогональними значеннями зі структурною мірою* m. *Позначимо через:*

$H_2(\mu) = \aleph(\mu(B),\ B \in \mathfrak{B}(\mathbb{R}))$ – *замкнений лінійний підпростір* $L_2(P)$, *що породжений її значеннями, та через*

$L_2(m)$ – *гільбертів простір комплекснозначних борелевих функцій, що інтегровні з квадратом за мірою* m : $\|f\|_m^2 = \int |f|^2\, dm < \infty$, *зі скалярним добутком* $(f,g)_m = \int f\overline{g}dm$.

Теорема (про ізометрію просторів H2 і L2). *Існує єдина лінійна ізометрія* $\Im : L_2(m) \to H_2(\mu)$, *така, що*

$$\Im(1_B) = \mu(B),\ \forall B \in \mathfrak{B}(\mathbb{R}).$$

Доведення. Розглянемо лінійні підпростори

$$H_{01} = \pounds(1_B,\ B \in \mathfrak{B}(\mathbb{R})), \quad H_{02} = \pounds(\mu(B),\ B \in \mathfrak{B}(\mathbb{R})).$$

Вказане в теоремі відображення породжує ізометрію H_{01} на H_{02}, оскільки за умовою

$$\|1_B\|_m^2 = m(B) = M\ \mu(B)\overline{\mu(B)} = \|\mu(B)\|^2 = \|\Im(1_B)\|^2,$$

і внаслідок ортогональності значень для попарно несумісних B_k

$$\left\|\sum c_k 1_{B_k}\right\|_m^2 = \int \left|\sum c_k 1_{B_k}(x)\right|^2 m(dx) =$$

$$\int \sum\nolimits_k c_k 1_{B_k}(x) \sum\nolimits_j \overline{c_j} 1_{B_j}(x)\ m(dx) =$$

$$\int \sum\nolimits_{k,j} c_k \overline{c_j} 1_{B_k \cap B_j}(x) m(dx) = \sum |c_k|^2\, m(B_k) =$$

$$\sum\nolimits_{k,j} c_k \overline{c_j}\ M\mu(B_k)\overline{\mu(B_j)} = M\left|\sum c_k\mu(B_k)\right|^2 = \left\|\sum c_k\mu(B_k)\right\|^2.$$

За означенням інтегралу Лебега – Стілтьєса за мірою m та за означенням математичного сподівання вкладення $H_{01} \subset L_2(m)$ та $H_{02} \subset H_2(\mu)$ є щільними, оскільки ці інтеграли збігаються з границями інтегралів від відповідних простих функцій. Тому твердження теореми випливає з теореми про продовження ізометрії $\boxdot$

Означення. *Нехай $f \in L_2(m)$. Стохастичним інтегралом від функції f за випадковою мірою μ з ортогональними значеннями називається значення ізометрії $\Im(f) \in H_2(\mu)$, що існує, єдине за теоремою про ізометрію просторів H2 і L2 та позначається як*

$$\Im(f) = \int f\ d\mu.$$

Зауваження. Стохастичний інтеграл є лінійним, неперервним за f у метриці простору $L_2(m)$ та задовольняє умову

$$M\left(\int f\ d\mu\right)\left(\overline{\int g\ d\mu}\right) = \int f\overline{g}\ dm.$$

Доведення.

Лінійність та неперервність є властивостями лінійної ізометрії.

Остання рівність полягає в збіжності скалярних добутків ізометричних елементів, яка випливає з ізометричної збіжності норм відповідних елементів та з тотожності

$$4(f,g) = (f+g,\ f+g) - (f-g,\ f-g) = \|f+g\|^2 - \|f-g\|^2 \ \boxdot$$

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 міра з ортогональними значеннями* $(\mu^*(B),\ B \in \mathfrak{B}(\mathbb{R}))$ *зі структурною мірою* m *така, що* $\mu^*(B) = \mu(B)$ *м.н.,* $\forall B \in \mathfrak{A}(\mathbb{R})$.

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 $\mathfrak{A}(\mathbb{R})$. 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

$$\|\mu(A_n) - \mu(A_m)\|^2 = M(\mu(A_n) - \mu(A_m))\overline{(\mu(A_n) - \mu(A_m))} =$$
$$m(A_n) + m(A_m) - 2m(A_n \cap A_m) = m(A_n \Delta A_m) \to 0,\ n.m \to \infty.$$

Отже, послідовність $\mu(A_n)$ фундаментальна в $L_2(P)$, тому існує її с.к. границя $\mu^*(B) = l.i.m\ \mu(A_n)$. Ця границя визначена коректно і задовольняє умови (а),(б),(в), оскільки ці умови зберігаються при переході до с.к. границі $\boxdot$

19.3. Процеси з ортогональними приростами

Означення. Квадратично інтегровний *процес* $(\zeta(t),\ t \in T)$ *називається процесом з* ортогональними приростами, *якщо:*

(а) $M\zeta[s,\ t)\overline{\zeta[u,\ v)} = 0$ *при всіх* $s < t < u < v$,
де приріст процеса $\zeta[a,b) \equiv \zeta(b) - \zeta(a)$,

(б) $\sup_{s\,<\,t}\ M\left|\zeta[s,t)\right|^2 < \infty$,

(в) $M\zeta[s,\ t) = 0$.

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*

$$M\left|\zeta[s,t)\right|^2 = F(t) - F(s),$$
$$M\zeta(I)\overline{\zeta(J)} = F(I \cap J),$$

для всіх інтервалів I, J*, де значення функції від інтервалу інтерпретується як відповідний приріст.*

Означення. *Функція* F *називається* спектральною функцією *процесу* ζ.

Доведення.

Функція $M\left|\zeta(I)\right|^2$ є адитивною на алгебрі об'єднань інтервалів, що не перетинаються:

$$M\left|\zeta(I \cup J)\right|^2 = M(\zeta(I) + \zeta(J))(\overline{\zeta(I) + \zeta(J)}) =$$
$$M\left|\zeta(I)\right|^2 + M\left|\zeta(J)\right|^2 + 2\mathrm{Re}M\zeta(I)\overline{\zeta(J)} = M\left|\zeta(I)\right|^2 + M\left|\zeta(J)\right|^2,$$

отже, є монотонно неспадною функцією інтервалу $I = [s,t)$. Тому з (б) випливає існування монотонної обмеженої функції

$$\lim_{s \to\ -\infty} M\left|\zeta[s,t)\right|^2 = F(t),$$

звідки з тотожності

$$\lim_{s\to\ -\infty} M\left|\zeta[s,t)\right|^2 + M\left|\zeta[t,u)\right|^2 = \lim_{s\to\ -\infty} M\left|\zeta[s,u)\right|^2, \quad t<u,$$

виводиться перше твердження теореми.

Друге виводиться після підстановки $I=(I\cap J)\cup(I\backslash J)$ із білінійності коваріації та з ортогональності приростів:

$$M\zeta(I)\overline{\zeta(J)} = M\zeta(I\cap J)\overline{\zeta(J)} + M\zeta(I\backslash J)\overline{\zeta(J)} = M\zeta(I\cap J)\overline{\zeta(J)} =$$
$$M\zeta(I\cap J)\overline{\zeta(J\cap I)} + M\zeta(I\cap J)\overline{\zeta(J\backslash I)} = M\zeta(I\cap J)\overline{\zeta(J\cap I)} = F(I\cap J)\ \boxdot$$

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 $F(t)=\sum_{k\,<\,t} M\left|\eta_k\right|^2$.

Вправа. Довести, що вінерівський процес $(w(t),t\in T)$ на інтервалі $T=[0,a]$ є процесом з ортогональними приростами. Знайти його спектральну функцію.

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:

$$m([s,t)) = M\left|\mu([s,t))\right|^2 = M\left|\zeta(t)-\zeta(s)\right|^2 = F(t)-F(s).$$

Доведення.

Випливає з теореми про продовження міри з ортогональними значеннями. Міра μ на множинах $A = \cup_{k=1}^{n}[a_k, b_k)$ з алгебри $\mathfrak{A}(\mathbb{R})$ 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:

$$M\mu(A)\,\overline{\mu(B)} = M\sum\nolimits_{I\subset A}\sum\nolimits_{J\subset B}\zeta(I)\overline{\zeta(J)} =$$

$$\sum\nolimits_{I\subset A}\sum\nolimits_{J\subset B}F(I\cap J) = \sum\nolimits_{I=J\subset A\cap B}F(I) = F(A\cap B)\boxdot$$

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*

$$\int f\, d\zeta \equiv \int f\, d\mu.$$

Зауваження. Згідно з властивостями інтегралу за випадковою мірою інтеграл за стохастичним процесом є лінійним та неперервним за f і задовольняє умову

$$M\left(\int f\, d\zeta\right)\left(\overline{\int g\, d\zeta}\right) = \int f\,\overline{g}\, dF.$$

19.4. Спектральне зображення стаціонарного процесу

Теорема (теорема Карунена про спектральне зображення стаціонарного процесу). *Нехай* $(\xi_t,\ t \in T) \subset L_2(P)$ – *стаціонарний процес із нульовим середнім:* $M\xi_t = 0$, *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 :

$$H_2(\xi) = \aleph(\xi_t,\ t \in T) \subset\ L_2(P).$$

Існує процес із ортогональними приростами $(\zeta(\lambda),\ \lambda \in \mathbb{R})$ *зі значеннями в* $H_2(\xi)$, *із тією ж спектральною мірою* F :

$$M\left|\zeta[s,t)\right|^2 = F(t) - F(s),$$

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, то її щільність f називається спектральною щільністю процесу.

Доведення.

Позначимо через $L_2(F)$ гільбертів простір комплекснозначних борелєвих функцій, що інтегровні з квадратом за мірою F :

$$L_2(F) = \{f : \|f\|_F^2 = \int |f|^2 \, dF < \infty\}.$$

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 $H_2(\xi)$ підпростір

$$H_{02} = \pounds(\xi_t,\ t \in T) \subset H_2(\xi)$$

є також щільно вкладеним.

Розглянемо відображення підпростору H_{01} на H_{02}, що задається формулою

$$\Im(\exp(it\bullet)) = \xi_t,\ t \in T,$$

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

$$\left\|\sum c_k \exp(i\lambda t_k)\right\|_F^2 = \int \left|\sum\nolimits_k c_k \exp(i\lambda t_k)\right|^2 dF(\lambda) =$$

$$\int \sum\nolimits_k c_k \exp(i\lambda t_k) \sum\nolimits_j \overline{c_l} \exp(-i\lambda t_l)\ dF(\lambda) =$$

$$\int \sum\nolimits_{k,j} c_k \overline{c_l} \exp(i\lambda(t_k - t_l))\ dF(\lambda) = \sum\nolimits_{k,j} c_k \overline{c_l}\, r(t_k - t_l) =$$

$$\sum_{k,j} c_k \overline{c_l}\ M\xi(t_k)\overline{\xi(t_l)} =$$

$$M\left|\sum c_k \xi(t_k)\right|^2 = \left\|\sum c_k \xi(t_k)\right\|^2,$$

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 $1_{(-\infty,\, t)} \in L_2(F)$. Позначимо

$$\zeta(t) \equiv \Im(1_{(-\infty,\, t)}) \in H_2(\xi).$$

Тоді за ізометрією

$$M\zeta[s,\ t)\overline{\zeta[u,\ v)} = \ (1_{[s,\, t)}, 1_{[u,\, v)})\ _F = 0, \quad \forall s < t < u < v,$$

$$M\left|\zeta[s,t)\right|^2 = \left\|1_{[s,\, t)}\right\|_F^2 = F(t) - F(s).$$

Оскільки $M\xi(t) = 0$ і за нерівністю Коші математичне сподівання є неперервною функцією на $L_2(P)$, то $M\eta = 0$ для всіх $\eta \in H_2(\xi)$. Зокрема, $M\zeta(t) = 0$.

Отже, $\zeta(t)$ – випадковий процес із ортогональними приростами та зі спектральною мірою F.

Розглянемо лінійний підпростір

$$L_2^0(F) \equiv \left\{ g \in L_2(F) : \Im(g) = \int g(\lambda) d\zeta(\lambda) \right\}.$$

Для довільної простої функції

$$g(\lambda) = \sum c_k 1_{[t_k,\, s_k)}(\lambda)$$

з лінійності $\Im$ отримуємо

$$\Im(g) = \sum c_k \zeta[t_k, s_k) = \int \sum c_k 1_{[t_k,\, s_k)}(\lambda) d\zeta(\lambda) = \int g(\lambda) d\zeta(\lambda),$$

тобто прості функції $g \in L_2^0(F)$.

Оскільки клас простих функцій щільний у $L_2(F)$, а простір $L_2^0(F)$ замкнений через неперервність ізометрії $\Im$ та неперервність стохастичного інтегралу, то $L_2^0(F) = L_2(F)$.

Тому підстановкою $g(\lambda) = \exp(i\lambda t)$ у тотожність $\Im(g) = \int g(\lambda) d\zeta(\lambda)$ дістанемо спектральне зображення для процесу ξ_t :

$$\xi_t = \Im(\exp(it\lambda)) = \int_{\mathbb{R}} \exp(i\lambda t)\, d\zeta(\lambda),\ \forall t \in T \boxdot$$

Зауваження. У випадку $T = \mathbb{Z}$ спектральна міра F та процес $\zeta(\lambda)$ зосереджені на відрізку $[-\pi, \pi]$.

Вправа. Розклад Карунена – Лоева – Пугачова.

Нехай $\xi(t), t \in [0,1]$ с.к. неперервний дійсний випадковий процес з нульовим середнім та коваріаційною функцією $r(t,s) = M\xi(t)\xi(s)$. Ця функція неперервна і інтегровна з квадратом на $[0,1]^2$. Тому інтегральне рівняння Фредгольма

$$\varphi(t) = \lambda \int_0^1 r(t,s)\varphi(s) ds$$

має для деяких $\lambda = \lambda_n \geq 0$ в $L_2[0,1]$ розв'язки $\varphi_n \in L_2[0,1], n \geq 1$, що утворюють ортонормований базис. Зокрема, справедливе зображення

$$r(t,s) = \sum_{n=1}^{\infty} \lambda_n^{-1} \varphi_n(t) \varphi_n(s).$$

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

$$\xi(t) = \sum\nolimits_{n=1}^{\infty} \lambda_n^{-1/2} \varphi_n(t)\ \zeta_n$$

з попарно ортогональними та нормованими $\zeta_n \in L_2(P)$.

Зокрема, для вінерівського процесу $w(t)$ маємо $r(t,s) = \min(t,s)$, $\lambda_n = \pi^2(n-1/2)^2$, $\varphi_n(t) = \sqrt{2}\sin(n-1/2)\pi t$, та

$$w(t) = \sqrt{2} \sum\nolimits_{n=1}^{\infty} \zeta_n\ \pi^{-1}(n-1/2)^{-1} \sin(n-1/2)\pi t.$$

20. Оптимальний лінійний прогноз стаціонарних послідовностей

За допомогою спектрального зображення можна розв'язати задачу знаходження лінійного прогнозу з найменшою дисперсією.

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

$$L_n = \pounds(\xi_k,\ |k| \le n),\ n \ge 1.\ L_\infty = \cup L_n.$$

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 $\xi - \eta_n \perp L_n$.

Позначимо $a = \|\xi\|^2$, $\varepsilon_n = \|\xi - \eta_n\|^2$. Тоді з $\xi = \xi - \eta_n + \eta_n$ отримуємо

$$\|\eta_n\|^2 = a - \varepsilon_n, \ (\xi, \ \eta_n) = \|\eta_n\|^2 = \ a - \varepsilon_n.$$

Оскільки $L_n \subset L_m$ при $m > n$, то $\varepsilon_n \downarrow \varepsilon = \lim \varepsilon_n$, $\ \xi - \eta_m \perp \eta_n$. Тому

$$\|\eta_n - \eta_m\|^2 = \|\eta_n - \xi\|^2 +$$

$$\|\eta_m - \xi\|^2 - 2(\eta_n - \xi, \eta_m - \xi) = \varepsilon_n + \varepsilon_m + 2(\xi, \eta_m - \xi) =$$

$$\varepsilon_n + \varepsilon_m + 2(a - \varepsilon_m) - 2a = \varepsilon_n - \varepsilon_m \to 0.$$

Отже, послідовність η_n фундаментальна й має границю $\zeta = l.i.m \ \eta_n$, 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

$$\|\xi - \eta\|^2 = \varepsilon_n \geq \varepsilon = \|\xi - \zeta\|^2 \quad \forall \eta \in \ L_\infty.$$

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 $\xi - \eta_n \perp L_k$ при $n \geq k$. 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 $(\xi - \zeta,\ \xi_k) = 0,\ \forall k \in \mathbb{Z}$. Тоді за лінійністю та неперервністю $(\xi - \zeta,\ \eta) = 0,\ \forall \eta \in H_0$. Отже, для довільного $\eta \in H_0$

$$\|\xi - \eta\|^2 = \|\xi - \zeta\|^2 + \|\zeta - \eta\|^2 + \\ 2(\xi - \zeta,\ \zeta - \eta) \geq \|\xi - \zeta\|^2,$$

де рівність $(\xi - \zeta,\ \zeta - \eta) = 0$ випливає з включення $\zeta - \eta \in H_0$.

Лінійність (г) є наслідком (в): якщо $\zeta(i) = \Pi_0\xi(i),\ i = 1, 2$, то

$$(\xi(1)+\xi(2)-\zeta(1)-\zeta(2),\ \xi_k)=$$
$$(\xi(1)-\zeta(1),\ \xi_k)+(\xi(2)-\zeta(2),\ \xi_k)=0,\ \forall k\in\mathbb{Z},$$

звідки $\Pi_0\xi(1)+\Pi_0\xi(2)=\zeta(1)+\zeta(2)=\Pi_0(\xi(1)+\xi(2))$.

Ортогональність (д) виводиться з тотожності $\Pi_0\eta=\eta,\ \eta\in H_0$, за якою $\Pi_0^2\eta=\Pi_0(\Pi_0\eta)=\Pi_0\eta=\eta$ $\boxdot$

20.2. Регулярні стаціонарні послідовності

Розглянемо стаціонарну квадратично інтегровну послідовність $(\xi_n,\ n\in\mathbb{Z})$ з нульовим середнім і коваріаційною функцією $r(n)$.

Позначимо лінійні підпростори

$$L_n=\ \pounds(\xi_k,\ k\le n),\ H_n=\aleph(L_n)\subset H_{n+1},$$

$$L_\infty=\cup L_n,\quad H_\infty=\cup H_n,\ H_{-\infty}=\cap H_n.$$

Нехай лінійне відображення $T:L_\infty\to L_\infty$ визначається рівністю

$$T\xi_n=\xi_{n+1},\ \forall n\in\mathbb{Z},$$

на твірних елементах L_∞ та продовжується за лінійністю на всі лінійні форми з L_∞ :

$$T\sum c_n\xi_n=\sum c_n\xi_{n+1}.$$

Оскільки послідовність $(\xi_n,\ n\in\mathbb{Z})$ – стаціонарна, то відображення T є ізометрією простору L_n на L_{n+1}, а за неперервністю внаслідок теореми про продовження ізометрії – 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* $(\xi_n,\ n \in \mathbb{Z})$ *називається* регулярною, *якщо* $\cap_{n \in \mathbb{Z}} H_n = \{0\}$.

Теорема (про структуру регулярної стаціонарної послідовності). *Нехай стаціонарна послідовність* $(\xi_n,\ n \in \mathbb{Z})$ *є регулярною. Тоді:*

(а) Включення $H_n \subset H_{n+1}$ *є строгим (не щільним).*

(б) Величини $\varepsilon_{n+1} = \xi_{n+1} - \Pi_n \xi_{n+1} \in H_{n+1}$ *невироджені, попарно ортогональні,* $\varepsilon_{n+1} \perp H_n$.

(в) Має місце рівність $\varepsilon_{n+1} = T\varepsilon_n$.

(г) Послідовність $(\varepsilon_k,\ k \le n)$ *є ортогональним базисом у* H_n.

(д) Існують комплексні $c_k,\ k \le 0,\ c_0 = 1$, *такі, що справедливе зображення рухомого середнього*

$$\xi_n = \sum_{k \le 0} c_k\, \varepsilon_{k+n} = \sum_{k \le n} c_{k-n}\, \varepsilon_k .$$

(е) Спектральна міра послідовності $(\xi_n,\ n \in \mathbb{Z})$ *має спектральну щільність на* $[-\pi, \pi]$

$$f(\lambda) = \frac{M\,|\varepsilon_0|^2}{2\pi} \left| \sum_{k \le 0} c_k \exp(ik\lambda) \right|^2 .$$

Доведення.

(а) Припустимо, що $H_m = H_{m+1}$ для деякого m.

Тоді існують лінійні форми $\eta_m^n = \sum_{k\le m} c_{nk}\xi_k \in L_m$ такі, що

$$\left\|\xi_{m+1} - \eta_m^n\right\| \to 0,\ \ n \to \infty.$$

Так як T – ізометрія, то для величин $\eta_{m-1}^n = T^{-1}\eta_m^n \in L_{m-1}$ справедливі співвідношення

$$\left\|\xi_m - \eta_{m-1}^n\right\| = \left\|T(\xi_m - \eta_{m-1}^n)\right\| = \left\|\xi_{m+1} - \eta_m^n\right\| \to 0.$$

Отже, $H_{m-1} = H_m$, звідки за індукцією $H_{-\infty} = H_m$, що суперечить 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, з включення $\varepsilon_n \in H_n$ виводимо попарну ортогональність $\varepsilon_{n+1} \perp \varepsilon_n$.

(в) Для довільного $\eta_{n-1} \in L_{n-1}$

$$\left\|\xi_n - \eta_{n-1}\right\| = \left\|T(\xi_n - \eta_{n-1})\right\| = \left\|\xi_{n+1} - T\eta_{n-1}\right\|.$$

Оскільки T – ізометрія L_{n-1} на L_n, то нижні границі за η_{n-1} в обох частинах цієї рівності є однаковими, а відповідні аргументи збігаються до векторів, пов'язаних відображенням T. Тому $T\Pi_{n-1}\xi_n = \Pi_n\xi_{n+1}$ і

$$\varepsilon_{n+1} = \xi_{n+1} - \Pi_n\xi_{n+1} = T(\xi_n - \Pi_{n-1}\xi_n) = T\varepsilon_n.$$

(г) Нехай $\zeta \in H_n,\ \zeta \perp \varepsilon_k,\ k \le n$. Тоді для деякої послідовності лінійних форм

$$\eta_n(m) = \sum_{k\le n} c_k(m)\xi_k \in L_n$$

має місце збіжність $\eta_n(m) \to \zeta, \ m \to \infty$, причому

$$c_n(m) \left\| \varepsilon_n \right\|^2 = (\eta_n(m), \ \varepsilon_n) \to (\zeta, \varepsilon_n) = 0.$$

Звідси

$$\eta_n(m) - \Pi_{n-1}\eta_n(m) = c_n(m)\varepsilon_n \to 0,$$

отже,

$$\zeta = l.i.m_{\ m\to\infty}\eta_n(m) = l.i.m_{\ m\to\infty}\Pi_{n-1}\eta_n(m) \in H_{n-1}.$$

За індукцією $\zeta \in H_{-\infty} = \{0\}$, тобто $\zeta = 0$. Отже, послідовність $(\varepsilon_k, \ k \le n)$ є базисом. Ортогональність доведено в (б).

(д) З попереднього пункту 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 T дістанемо шукане зображення для довільного n.

(е) Вираз для спектральної щільності виводиться з (д) та з означення спектральної міри

$$\int_{-\pi}^{\pi} \exp(in\lambda) f(\lambda) d\lambda = r(n) = M\xi_n\overline{\xi_0} =$$

$$M \sum\nolimits_{k \ \le \ 0} c_k \, \varepsilon_{k \ +n} \overline{\sum\nolimits_{j \ \le \ 0} c_j \varepsilon_j} =$$

$$\sum_{k\ \le\ 0}\sum_{j\ \le\ 0} c_k\overline{c_j}\ M\varepsilon_{k+n}\overline{\varepsilon_j} = M\left|\varepsilon_0\right|^2 \sum_{k\ \le\ 0} c_k\,\overline{c_{k+n}} =$$

$$\frac{M\left|\varepsilon_0\right|^2}{2\pi}\int_{-\pi}^{\pi}\exp(in\lambda)\left|\sum_{k\ \le\ 0} c_k\exp(ik\lambda)\right|^2 d\lambda,\ \ \forall n\in\mathbb{Z}\boxdot$$

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* $\xi_k,\ \ k\le 0$. *Для регулярної послідовності*

$$\widehat{\xi}_1 = \sum_{k\ <\ 0} c_k\ \varepsilon_{\ k+1},\quad M\left|\xi_1-\widehat{\xi}_1\right|^2 = M\left|\varepsilon_0\right|^2.$$

Доведення очевидне. Зауважимо, що побудова прогнозу складається з 2-х етапів: відшукання базису (ε_k), що пов'язаний зі спектральним розкладом послідовності, і обчислення коефіцієнтів c_k, які визначаються коваріаційною функцією $\boxdot$

20.3. Спектральний розв'язок задачі лінійного прогнозу

Нехай $(\xi_n,\ n\in\mathbb{Z})$ – довільна стаціонарна послідовність із нульовим середнім, F – її спектральна міра на $[-\pi,\pi]$. Позначимо через

$$\widehat{\xi}_1 = \Pi_0\xi_1 \equiv \arg\min_{\ \eta\in H_2^-(\xi)}\left\|\xi_1-\eta\right\|^2 \in H_2^-(\xi)\equiv\aleph(\xi_n, n\le 0)$$

оптимальний лінійний прогноз значення ξ_1 за спостереженнями з $H_2^-(\xi)$.

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 $L_2(F)$.

Позначимо також через L_2, $L_2^{\pm}$ аналогічні $L_2(F)$, $L_2^{\pm}(F)$ простори функцій для випадку, коли F – міра Лебега на $[-\pi, \pi]$. Як відомо з теорії рядів Фур'є,

$$L_2 = \{g(\lambda) = \sum\nolimits_{n\,\in\,\mathbb{Z}} c_n \exp(in\lambda) : \sum |c_n|^2 < \infty\},$$

$$L_2^{\pm} = \{g(\lambda) = \sum\nolimits_{n\,\in\,\mathbb{Z}_\pm} c_n \exp(in\lambda) : \sum\nolimits_{n\,\in\,\mathbb{Z}_\pm} |c_n|^2 < \infty\}.$$

Теорема (про ізоморфізм спектральних просторів). *Нехай виконується умова:*

(L) спектральна міра F має щільність f таку, що

$$0 < \inf f,\ \sup f < \infty.$$

Тоді

$$L_2(F) = L_2, \ \ L_2^{\pm}(F) = L_2^{\pm}.$$

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:

$$\|g\|_F^2 = \int |g|^2 f(\lambda) d\lambda \le \sup f \int |g|^2 d\lambda = \sup f \ \ \|g\|^2,$$

$$\|g\|_F^2 = \int |g|^2 f(\lambda) d\lambda \ge \inf f \int |g|^2 d\lambda = \inf f \ \ \|g\|^2 \ \boxdot$$

Як і вище, символом $\Im$ позначатимемо ізометрію $\Im : L_2(F) \to H_2(\xi)$, що визначена в теоремі Карунена про спектральне зображення стаціонарного процесу як $\Im(\exp(it\bullet)) = \xi_t$ та породжена стохастичним інтегралом за відповідним процесом із ортогональними приростами: $\Im(g) = \int g(\lambda) d\zeta(\lambda)$.

Лема (про звуження).

За умови (L) кожне зі звужень $\Im : L_2^{\pm}(F) \to H_2^{\pm}(\xi) = \aleph(\xi_n, n \in \mathbb{Z}_{\pm})$ є ізометрією.

Доведення очевидне, оскільки $\Im(\exp(it\lambda)) = \xi_t$, а послідовності $(\exp(in\lambda), n \in \mathbb{Z}_{\pm})$ є базисами в $L_2^{\pm}(F) = L_2^{\pm}$ $\boxdot$

Теорема (про критерій оптимальності спектрального зображення). *За умови (L) функція $\widehat{g} \in L_2$ відповідає оптимальному прогнозу:*

$$\Im(\widehat{g}) = \widehat{\xi}_1$$

тоді й тільки тоді, коли $\widehat{g} \in L_2^-$ і

$$\int_{-\pi}^{\pi} (\exp(i\lambda) - \widehat{g}(\lambda)) \, \overline{\exp(in\lambda)} \, f(\lambda) d\lambda = 0, \quad \forall n \le 0.$$

Доведення.

З леми про звуження виводимо, що включення $\widehat{g} \in L_2^-$ еквівалентне $\widehat{\xi}_1 \in H_2^-(\xi)$.

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

$$(\xi_1 - \zeta, \ \xi_n) \equiv M(\xi_1 - \zeta)\overline{\xi_n} = 0, \ \forall n \le 0.$$

Оскільки має місце ізометрія $\Im(\exp(i\lambda n)) = \xi_n$, $\Im(\widehat{g}) = \widehat{\xi}_1$, то

$$(\xi_1 - \zeta, \ \xi_n) = \int_{-\pi}^{\pi} (\exp(i\lambda) - \widehat{g}(\lambda)) \, \overline{\exp(in\lambda)} \, f(\lambda) d\lambda,$$

отже, наведена вище необхідна і достатня умова еквівалентна тотожності теореми $\boxdot$

Теорема (**про побудову оптимального прогнозу**). *Нехай виконана умова (L). Функція* $\widehat{g} \in L_2^-$ *відповідає оптимальному 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*

$$f(\lambda) = \frac{f_+(\lambda)}{f_-(\lambda)},\ \lambda \in [-\pi, \pi],$$

де знаменник $f_-(\lambda) = 1 - \exp(-i\lambda)\widehat{g}(\lambda)$.

Доведення.

Необхідність. Якщо $\widehat{\xi}_1$ – оптимальний прогноз із відповідною спектральною функцією $\widehat{g}$, то з теореми про критерій оптимальності спектрального зображення виводимо, що $\widehat{g} \in L_2^-(F)$ і

$$\int_{-\pi}^{\pi} (\exp(i\lambda) - \widehat{g}(\lambda)) \exp(-in\lambda)\ f(\lambda) d\lambda =$$

$$\int_{-\pi}^{\pi} f_-(\lambda) \exp(-i(n-1)\lambda)\ f(\lambda) d\lambda = 0,\ \forall n \le 0.$$

Визначимо $f_+(\lambda) = f_-(\lambda)\ f(\lambda)$. Тоді $f_+ \in L_2$, тому що $f_- \in L_2$,а f обмежена. Оскільки $\exp(in\lambda)$ – базис у L_2, то з отриманих рівностей випливає, що

$$f_+(\lambda) = \sum\nolimits_{n\ \ge 0} c_n \exp(in\lambda)\ \in L_2^+.$$

Достатність. Якщо для спектральної функції $\widehat{g} \in L_2^-(F)$ та функції $f_\pm \in L_2^\pm$ справджуються умови теореми, то за теоремою Карунена про спектральне зображення стаціонарного процесу $\widehat{\xi}_1 = \Im(\widehat{g}) \in H_0$ задовольняє умову (б) теореми про лінійний прогноз у гільбертовому просторі $\boxdot$

Теорема (про обчислення оптимального прогнозу). *За умов теореми про побудову оптимального прогнозу оптимальний прогноз має вигляд*

$$\widehat{\xi}_1 = \sum\nolimits_{n\ <\ 0} b_n \xi_{n+1},$$

де коефіцієнти b_n визначаються з рівності

$$f_-(\lambda) = 1 - \exp(-i\lambda)\widehat{g}(\lambda) = 1 - \sum\nolimits_{n\ <\ 0} b_n \exp(in\lambda).$$

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

$$\widehat{\xi}_1 = \int_{-\pi}^{\pi} \widehat{g}(\lambda)\, d\zeta(\lambda) =$$

$$\int_{-\pi}^{\pi} \sum\nolimits_{n\ <\ 0} b_n \exp(i(n+1)\lambda)\, d\zeta(\lambda) = \sum\nolimits_{n\ <\ 0} b_n \xi_{n+1} \boxdot$$

Теорема (про натуральну факторизацію спектральної щільності). *Нехай виконується умова:*

(R) логарифм спектральної щільності $f(\lambda)$ розкладається в абсолютно збіжний ряд Фур'є

$$\ln f(\lambda) = \sum a_k \exp(ik\lambda),\ A \equiv \sum |a_k| < \infty.$$

Тоді функції

$$f_+(\lambda) \equiv \exp(\sum_{k\ \geq 0} a_k \exp(ik\lambda)),$$
$$f_-(\lambda) \equiv \exp(-\sum_{k\ <\ 0} a_k \exp(ik\lambda))$$

задовольняють умови теореми про побудову оптимального прогнозу, причому вказані в ній елементи факторизації $f_\pm(\lambda)$ визначені однозначно.

Доведення.

Зауважимо, що з (R) випливає умова (L):

$$\exp(-A) = \exp(-\sum |a_k|) \leq f(\lambda) \leq \exp(\sum |a_k|) = \exp(A),$$

а також обмеженість сум

$$f_\pm(\lambda) = \sum_{m\ \geq 0} \left(\pm \sum_{k\ (\geq)(<)\ 0} a_k \exp(ik\lambda) \right)^m /m! \leq$$

$$\sum_{m\ \geq 0} A^m/m! = \exp(A) < \infty,$$

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 елементів канонічної факторизації та з тотожності $L_2^+ \cap L_2^- = \{c, c \in \mathbb{R}\}$ $\boxdot$

Теорема (про похибку лінійного прогнозу). *Нехай виконується умова (R). Тоді*

$$M\left|\xi_1 - \widehat{\xi}_1\right|^2 = 2\pi \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln f(\lambda) d\lambda\right).$$

Доведення.

За означенням ізометрії та теоремою про побудову оптимального прогнозу

$$M\left|\xi_1-\widehat{\xi}_1\right|^2 = M(\xi_1-\widehat{\xi}_1)\overline{\xi_1} =$$
$$\int_{-\pi}^{\pi}(\exp(i\lambda)-\widehat{g}(\lambda))\exp(-i\lambda)f(\lambda)d\lambda =$$
$$\int_{-\pi}^{\pi} f_-(\lambda)f(\lambda)d\lambda = \int_{-\pi}^{\pi} f_+(\lambda)d\lambda = \int_{-\pi}^{\pi}\exp\left(\sum_{k\ge 0} a_k\exp(ik\lambda)\right)d\lambda =$$
$$\int_{-\pi}^{\pi}\exp(a_0)\sum_{m\ge 0}\left(\sum_{k>0} a_k\exp(ik\lambda)\right)^m / \; m!\; d\lambda = 2\pi\exp(a_0)\ \boxdot$$

20.4. Приклади обчислення оптимального прогнозу

1. Процес рухомого середнього (MA: moving average)**.**

Нехай спектральна щільність стаціонарної послідовності $(\xi_n,\ n\in\mathbb{Z})$ має вигляд

$$f(\lambda) = \frac{1}{2\pi}\left|P(\exp(-i\lambda))\right|^2,\ P(z) = \sum_{k=0}^{m} p_k z^k,$$

де дійсний поліном P не має коренів на одиничному колі і $p_0 = 1$.

Позначимо через ζ відповідний процес із ортогональними приростами. Тоді випадкові величини

$$\varepsilon_n = \int_{-\pi}^{\pi}\frac{\exp(in\lambda)}{P(\exp(-i\lambda))}\,d\zeta(\lambda),\ n\in\mathbb{Z},$$

ортонормовані:

$$M\varepsilon_n\overline{\varepsilon_k} = \int_{-\pi}^{\pi}\frac{\exp(i(n-k)\lambda)}{\left|P(\exp(-i\lambda))\right|^2}\,f(\lambda)d\lambda = \delta_{nk}.$$

Обчислимо

$$\sum_{k=0}^{m} p_k\varepsilon_{n-k} = \int_{-\pi}^{\pi}\exp(in\lambda)d\zeta(\lambda) = \xi_n,$$

$$\xi_n = \sum\nolimits_{k=-m}^{0} p_{-k}\varepsilon_{n+k}.$$

Отже, відповідна стаціонарна послідовність $(\xi_n,\ n \in \mathbb{Z})$ є послідовністю рухомого середнього.

Вправа. Припустимо, що величини в послідовності $(\varepsilon_n,\ n \in \mathbb{Z})$ незалежні однаково розподілені. Доведіть, 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. 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.

Розглянемо приклад $Q(z) = 1 - qz,\ |q| < 1$, звідки $r(n) = q^{|n|}/(1-q^2)$. У цьому випадку

$$f(\lambda) = 1\ /\ (2\pi\, |1 - q\exp(-i\lambda)|^2),$$

рівняння авторегресії має вигляд

$$\xi_n - q\, \xi_{n-1} = \varepsilon_n.$$

Елементи факторизації спектральної щільності дорівнюють

$$f_+(\lambda) = 1\ /\ (2\pi Q(\exp(i\lambda))),$$
$$f_-(\lambda) = Q(\exp(-i\lambda))\ = 1 - q\exp(-i\lambda).$$

Отже, оптимальний прогноз

$$\widehat{\xi}_1 = q\xi_0\ \boxdot$$

3. Процес із дробово-раціональною спектральною щільністю (ARMA: AutoRegression and Moving Average).

Нехай спектральна щільність стаціонарної послідовності $(\xi_n,\ n \in \mathbb{Z})$ є дробово-раціональною функцією

$$f(\lambda) = \frac{P(\exp(i\lambda))}{Q(\exp(i\lambda))},$$

де P, Q – комплексні поліноми, що не мають коренів на одиничному колі. Якщо поліном P або Q має корінь $\alpha,\ |\alpha| < 1$, то в розкладі f є множник $\exp(i\lambda) - \alpha$. Оскільки функція $f = \overline{f}$ дійсна, то у цьому розкладі знайдеться також і множник $\exp(-i\lambda) - \overline{\alpha}$. За наявності кореня $|\beta| > 1$ відповідні нормовані множники також утворюють внесок вигляду

$$(\beta - \exp(i\lambda))(\overline{\beta} - \exp(-i\lambda)) = |\beta|^2 (\exp(i\lambda) - \alpha)(\exp(-i\lambda) - \overline{\alpha}).$$

Групуючи такі множники попарно, можна привести щільність до вигляду

$$f(\lambda) = c\frac{\prod_{k=1}^{l}(1 - \alpha_k \exp(-i\lambda))(1 - \overline{\alpha_k} \exp(i\lambda))}{\prod_{j=1}^{m}(1 - \beta_j \exp(-i\lambda))(1 - \overline{\beta_j} \exp(i\lambda))},$$

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 $\Delta = T^{-1}$ зворотного часового зсуву $(\Delta\xi_n = \xi_{n-1},\ \Delta\varepsilon_n = \varepsilon_{n-1})$ та з ортогональними величинами ε_n, що визначаються як

$$\varepsilon_n = \int_{-\pi}^{\pi} \exp(in\lambda)\ f_-(\lambda) d\zeta(\lambda).$$

Тут $f_-(\lambda)$ відповідна компонента факторизації

$$f_-(\lambda) = \frac{\prod_{j=1}^{m}(1-\beta_j \exp(-i\lambda))}{\prod_{k=1}^{l}(1-\alpha_k \exp(-i\lambda))} =$$

$$\prod_{j=1}^{m}(1-\beta_j \exp(-i\lambda)) \prod_{k=1}^{l} \sum_{n\geq 0} \alpha_k^n \exp(-in\lambda).$$

Оптимальний лінійний прогноз задається спектральною функцією $\widehat{g}(\lambda) = (1 - f_-(\lambda)) \exp(i\lambda)$ $\boxdot$

Розділ 3

Додаток

Список рекомендованої літератури

Основна

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 - Москва, Наука,1980

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. Конспект лекцій з курсу теорії ймовірностей (Семестровий курс для математиків і статистиків) - Київ, КНУ,2001.

11. Карташов М.В. Конспект лекцій з курсу математичної статистики (Семестровий курс для математиків і статистиків) - Київ, КНУ,2003.

12. Дороговцев А.Я.,Сільвестров Д.С.,Скороход А.В.,Ядренко М.Й. Теория вероятностей Сборник задач - Киев, Вища школа,1976

13. Теория вероятностей. Методические указания к ведению практических занятий, Киев, КГУ, 1983

14. Завдання до лабораторних занять з курсу "Математична статистика", Київ, механіко-математичний факультет Київського національного університету, 2003

Додаткова

1. Гнеденко Б.В. Курс теории вероятностей - Москва, Наука,1987
2. Севастьянов Б.А. Курс теории вероятностей и математической статистики - Москва, Наука,1982
3. Боровков А.А. Теория вероятностей - Москва, Наука,1987
4. Коваленко И.Н., Гнеденко Б.В. Теория вероятностей - Киев, Вища школа,1990
5. Розанов А.А. Теория вероятностей, случайные процессы и математическая статистика. - Москва, Наука,1985.
6. Лоев М. Теория вероятностей - Москва, ИЛ,1962.
7. Неве Ж. Математические основы теории вероятностей - Москва, 1969.
8. Ламперти Дж. Вероятность - Москва, Наука,1973
9. Феллер В. Введение в теорию вероятностей и ее приложения - В двух томах, Москва, Мир,1984
10. Уиттл П. Вероятность. - М.: Наука, 1982.
11. Секей Г. Парадоксы в теории вероятностей и математической статистике. - Москва, Мир,1990.
12. Крамер Г. Математические методы статистики - Москва, Мир,1975
13. Закс Ш. Теория статистических выводов - Москва, Мир,1975
14. Уилкс С. Математическая статистика - Москва, Наука,1967
15. Леман Э. Проверка статистических гипотез - Москва, Наука,1979
16. Леман Э. Теория точечного оценивания - Москва, Наука,1991

17. Питмен Э. Основы теории статистических выводов - Москва, Мир, 1986

18. Гихман И. И., Скороход А. В. Введение в теорию случайных процессов, Москва, Наука,1977

19. Карлин С. Основы теории случайных процессов, Москва, Мир,1971

20. Мешалкин Л.Д. Сборник задач по теории вероятностей. - Москва, МГУ

21. Зубков А.М.,Севастьянов Б.А.,Чистяков В.П. Сборник задач по теории вероятностей - Москва, Наука,1989

22. Климов Г.П., Кузьмин А.Д. Вероятность, процессы, статистика. Задачи с решениями - Москва, МГУ, 1986

23. Чибисов Д.М., Пагурова В.И. Задачи по математической статистике - Москва, МГУ, 1990

24. Турчин В.М. Теорія ймовірностей, Основні поняття, приклади, задачі - Київ, А.С.К., 2004

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
Ξ, 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.

$a_n \approx b_n$ – відношення послідовностей $a_n.b_n$ прямує до 1 при $n \to \infty$.

$\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{R}_+, \mathbb{C}$ – множини натуральних, цілих, дійсних, невід'ємних дійсних та комплексних чисел.

$C(I), C_b(I)$ – простори неперервних та обмежених неперервних функцій на множині I.

$\operatorname{rang} A$ – ранг матриці A.

$P(A/B)$ – 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 кутами.

$\mathfrak{B}(\mathbb{R})$, $\mathfrak{B}(\mathbb{R}^n)$ – борелєві сигма-алгебри в $\mathbb{R}$ та $\mathbb{R}^n$.

$1_{\{A\}}(\omega), 1_B$ – індикаторна величина події A та індикатор висловлювання B.

$\delta_{kj} = 1_{k=j}$ – 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, що має розподіл Пуассона з параметром λ.

$B_1 \times B_2$ – декартів добуток множин.

$x^+ = \max(0,x)$ – додатна частина числа x.

ξ^+, ξ^- – додатна та від'ємна частини випадкової величини ξ.

м.н. – майже напевне.

$M\xi$ – математичне сподівання випадкової величини ξ.

$D\xi$ – дисперсія випадкової величини ξ.

$\operatorname{argmin} F$, $\operatorname{argmax} F$ – значення аргументу, для якого досягається найменше чи найбільше значення функції F.

$L_2(P)$ – простір квадратично інтегровних випадкових величин.

$N(\mu, \sigma^2)$ – величина із нормальним розподілом, середнім μ і 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, що має гама-розподіл із параметрами λ, α.

χ_n^2 – випадкова величина, що має хі-квадрат розподіл із n ступенями свободи.

τ_n – випадкова величина, що має розподіл Стьюдента із n ступенями свободи.

f_{nm} – випадкова величина, що має розподіл Фішера із 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.
Θ – параметрична множина довільної природи.
$P_\theta(\cdot)$ – ймовірнісний розподіл спостережень при значенні параметра θ.
X – статистична вибірка із значеннями в S.
$(S,\ \Sigma,\ \lambda)$ – вибірковий простір із сигма-алгеброю Σ та мірою λ.
$F_1 \ll F_0$ – міра F_1 абсолютно неперервна відносно F_0.
$X = (\xi_1, ..., \xi_n)$ – кратна вибірка.
$\widehat{\theta}$ – оцінка параметра θ.
$L(X, \theta)$ – вибіркова функція вірогідності.
$f(\xi, \theta)$ – функція вірогідності одного спостереження.
$\widehat{F}_n(x)$ – емпірична функція розподілу.
$\xi_{(k)}$ – k-та порядкова статистика.
$\widehat{\mu}_n,\ \widehat{\sigma}^2_n$ – вибіркове середнє та вибіркова дисперсія.
$\widehat{s}^2_n$ – нормована вибіркова дисперсія.
$\mu_k,\ \mu^0_k$ – нецентральний теоретичний момент та центральний теоретичний момент k-го порядку.
$\widehat{\mu}_{kn},\ \widehat{\mu}^0_{kn}$ – нецентральний вибірковий момент та центральний вибірковий момент k-го порядку.
$U(X, \theta)$ – вибіркова функція впливу.
$u(\xi, \theta)$ – функція впливу одного спостереження.
ОМВ – оцінка максимальної вірогідності.
$H_0 : \theta \in \Theta_0$ – нульова гіпотеза щодо значення невідомого параметра θ.
$H_1 : \theta \in \Theta_1$ – альтернативна гіпотеза.

$(\widehat{\varkappa}(X),\ D_1)$ – статистичний критерій з критичною областю D_1.
$(\widehat{\varkappa}(X),\ \pi(t))$ – рандомізований статистичний критерій.
МНК – метод найменших квадратів.
$\pounds(G)$ – лінійна оболонка множини G.
$\aleph(G)$ – замикання лінійної оболонки множини G.

Предметний покажчик

Зміст