П.В. Слюсарчук

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА

Затверджено Міністерством освіти і науки України як підручник для студентів вищих навчальних закладів

Ужгород
Вид-во “Карпати”
2005

УДК 519.21 (075.8)
ББК 22.17я73
С 49

Затверджено Міністерством освіти і науки України як підручник для студентів вищих навчальних закладів (від 20.12.04 №14/18.2-2726)

Слюсарчук П.В. Теорія ймовірностей та математична статистика. –Ужгород: Вид-во 2005р.

Розглядаються основні відомості з теорії ймовірностей, теорії випадкових процесів та математичної статистики. Відомості з теорії ймовірностей викладені в таких розділах: ймовірність випадкової події; схема Бернуллі; випадкові величини та їх розподіли; числові характеристики випадкових величин; закон великих чисел; характеристичні функції; центральна гранична теорема. В розділі елементи випадкових процесів наведені ланцюги Маркова, означення основних понять теорії випадкових процесів, розглядаються пуассонівський, вінерівський, стаціонарні та марківські процеси. Визначені основні задачі математичної статистики, викладаються теорія статистичного оцінювання параметрів, методи перевірки гіпотез, основні поняття вибіркової кореляції і регресії. Наведені приклади, що ілюструють основні поняття, та вправи для самостійного розв’язування.

Розрахований на студентів університетів із спеціальностей математика, прикладна математика, статистика та студентів технічних вузів.

Список літ.: 61 назв

Рецензенти.

Ю.В. Козаченко, доктор фіз.-мат. наук, професор кафедри теорії ймовірностей та математичної статистики Київського національного університету;

Д.В. Гусак, провідний науковий співробітник Інституту математики НАН України, доктор фізико-математичних наук, професор;

Ю.Ю. Червак, завідувач кафедри системного аналізу і методів оптимізації, доктор фіз.-мат. наук, професор.

ISBN 966-7400-28-X

Зміст

Вступ

Розв'язання багатьох практичних задач, що виникають в різних галузях діяльності людини, є неможливим без використання математичних методів. Зокрема, вивчення реальних процесів, в яких необхідно враховувати випадкові фактори, вплив яких неможливо наперед передбачувати, вимагає використання теорії ймовірностей і математичної статистики.

В теорії ймовірностей одним із основних понять є поняття випадкового експерименту. Так називають експерименти, результати яких не можна передбачити наперед. Прикладом такого експерименту є контроль якості продукції, що випускається деяким підприємством. Нехай перевіряється якість виробів серіями по n штук, при цьому для контролю відбирають m ($m \leq n$) виробів. Експеримент полягає у виборі m виробів та перевірці їх якості. Результатом експерименту буде число бракованих виробів серед відібраних. Описаний експеримент – випадковий. Відзначимо, що такий метод контролю використовується у виробництві і називається вибірковим. Можна навести і інші приклади випадкових експериментів: розігрування лотереї, поширення епідемії, перепис населення та інші.

З математичної точки зору в ймовірнісних експериментах нас цікавлять їх можливі результати – випадкові події. Важливою властивістю випадкових експериментів є можливість повторювати їх велику кількість разів. Закономірності, що проявляються в масових випадкових експериментах, і є предметом теорії ймовірностей. Математичний підхід до вивчення випадкових явищ намагалися знайти ще з давніх часів. Ще у стародавньому Китаї і у стародавньому Римі були відомі факти стійкості частот випадкових подій, пов'язаних з демографічними даними і даними про постачання великих міст. Думка про те, що закони природи проявляються через масу випадкових явищ виникла ще у філософів стародавньої Греції.

Виникнення теорії ймовірностей, як науки, відносять до середини 17–го століття. Ймовірнісні питання, що виникали у статистичній практиці, практиці страхових товариств, при обробці результатів спостережень і в інших областях, привернули увагу вчених. У першу чергу це відноситься до Б.Паскаля (1623–1662), П.Ферма (1601–1665), Г.Гюйгенса (1629–1695) і особливо Я.Бернуллі (1654–1705). В цей період виникають перші ймовірнісні поняття. Справжня теорія ймовірностей починається з робіт Бернуллі, Муавра, Лапласа, Гауса, коли в центрі уваги стали граничні теореми, теорія ймовірностей широко застосовується в різних галузях природознавства.

Наступний період, починаючи з другої половини 19 століття, у розвитку теорії ймовірностей перш за все пов'язаний з іменами вчених В.Я.Буняковського (1804–1889), П.Л.Чебишова (1821–1894), А.А.Маркова (1856–1922), О.М.Ляпунова (1857–1918). Ці чудові традиції були продовжені радянськими вченими. Школа теорії ймовірностей радянського періоду займає у світовій науці провідне місце. Великі досягнення в теорії ймовірностей пов'язані з іменами багатьох радянських вчених, зокрема С.Н.Бернштейна, О.Я.Хінчіна, А.М.Колгоморова, В.І.Романовського, Б.В.Гнєденко, а також ряду зарубіжних вчених. Особливе місце в теорії ймовірностей займають роботи А.М.Колгоморова, він дав найбільш вдалу аксіоматичну побудову теорії ймовірностей. Значний вклад в розвиток теорії ймовірностей внесли і українські математики.

Теорія ймовірностей виникла з практичних потреб людини, постійно знаходила джерела свого розвитку в практиці, і наш час характерний інтенсивним впровадженням теоретико-ймовірнісних методів у практику. Застосування ймовірнісно-статистичних методів є традиційним у природознавстві та техніці, фінансовій та страховій справах, біології та метеорології, економіці та соціально-гуманітарних науках.

Розділ 1. Ймовірність випадкової події

1.1. Випадкові події

Основним в теорії ймовірностей є поняття *випадкового (стохастичного) експерименту* – експерименту, результати якого передбачити неможливо. Можливі наслідки випадкового експерименту називаються *випадковими подіями*. Важливою умовою є можливість повторювати експеримент велику кількість разів. Експеримент визначається певним комплексом умов і можливими його наслідками. Випадкові події позначають частіше великими латинськими буквами.

Подія Ω, яка настає при кожній реалізації експерименту, називається *достовірною*. Подія $\varnothing$, яка не може настати ні при одній реалізації експерименту, називається *неможливою*.

Приклади випадкових експериментів	Приклади подій, що пов’язані з даними експериментами
1. Кидання монети	Випадає цифра, випадає герб
2. Кидання грального кубика	E_i – випадає і – очок (і=1,...,6), A – випадає парна кількість очок, B – випадає непарна кількість очок, C – випадає число очок, не менше трьох
3. Розиграш лотереї	Випаде виграш на даний квиток, на даний квиток випаде максималь–ний виграш,...
4. Спостереження за частинкою у броунівському русі	Можливі траєкторії частинки, частинка досягне певного рівня.

Із кожної подією A можна пов’язати подію, яка полягає у тому, що A не настає. Цю подію називають *протилежною* до A і позначають $\overline{A}$. Протилежною до $\overline{A}$ є подія A **(**$\overline{\overline{A}} = A$**)**.

Сумою двох подій A і B (позначається $A \bigcup B$ або $A + B$) називається така подія, яка полягає в настанні принаймні однієї із цих подій.

Добутком двох подій A і B (позначається $A \bigcap B$ або $A\ B$) називається така подія, яка полягає в тому, що відбуваються обидві події – і A і B.

Дві події A і B називаються *несумісними*, якщо їх сумісне настання неможливе, $A \bigcap B = \varnothing$.

Подія A *спричиняє* подію B $(B$ є наслідком $A)$, якщо із того, що подія A настала випливає, що настає подія B, позначається $A \subset B$.

Сукупність подій $A_1, ..., A_n$ утворюють *повну групу*, якщо одна і тільки одна із цих подій в результаті експерименту обов’язково настає: $A_1 \bigcup A_2 \bigcup ... \bigcup A_n = \Omega$, $A_i \bigcap A_j = \varnothing, \quad i \neq j$.

Операції додавання і множення подій мають такі властивості:

1. $A \bigcup B = B \bigcup A$, $A \bigcap B = B \bigcap A$;
2. $(A \bigcup B) \bigcup C = A \bigcup (B \bigcup C)$, $\quad (A \bigcap B) \bigcap C = A \bigcap (B \bigcap C)$;
3. $A \bigcap (B \bigcup C) = (A \bigcap B) \bigcup (A \bigcap C)$.

4. Достовірна подія при множенні відіграє роль одиниці $A\bigcap\Omega = A$, а неможлива – нуля $A\bigcap\varnothing = \varnothing$.

5. $A\bigcup\Omega = \Omega$, $A\bigcup\varnothing = A$, $A\bigcap A = A$, $A\bigcup A = A$.

6. $\overline{A\bigcup B} = \overline{A}\bigcap\overline{B}$, $\overline{A\bigcap B} = \overline{A}\bigcup\overline{B}$.

Для випадкових подій вводиться також *різниця* подій: $A\setminus B = A\bigcap\overline{B}$.

Подія ω буде *елементарною*, якщо для довільної події *A* вона спричиняє або подію A або $\overline{A}$. Тобто елементарні події є найпростішими наслідками випадкового експерименту. Множина $\Omega = \{\omega\}$ всіх елементарних подій називається простором елементарних подій. Випадкові події розглядаються як підмножини Ω. Звідси стає зрозумілим використання позначень для операцій із випадковими подіями.

Множина $\mathfrak{A}$ підмножин Ω називається *алгеброю* подій, якщо:

А1. $\Omega\in\mathfrak{A}$;

А2. $\forall A\in\mathfrak{A} \Rightarrow \overline{A}\in\mathfrak{A}$;

А3. $\forall A\in\mathfrak{A}$ і $\forall B\in\mathfrak{A} \Rightarrow A\bigcup B\in\mathfrak{A}$.

У багатьох випадкових експериментах необхідно розглядати нескінченні послідовності подій і дії над ними. Якщо в означенні алгебри подій третю умову А3 замінити на умову

А3'. Для довільної послідовності $A_1,\ldots,A_n,\ldots$, $A_i\in\mathfrak{A}$ випливає, що $\bigcup\limits_{i=1}^{\infty}A_i\in\mathfrak{A}$, тоді $\mathfrak{A}$ називається $\sigma-$*алгеброю* подій.

$\mathfrak{A} = \{\Omega,\varnothing\}$ – найпростіша алгебра подій. Із даних умов випливає, що $\varnothing\in\mathfrak{A}$. Якщо $A\in\mathfrak{A}$ і $B\in\mathfrak{A}$, то $A\bigcap B\in\mathfrak{A}$. Це випливає із того, що $A\bigcap B = \overline{\overline{A}\bigcup\overline{B}}\in\mathfrak{A}$. Надалі тільки елементи алгебри або $\sigma-$алгебри ми будемо називати випадковими подіями. Очевидно, будь–яку алгебру подій можна доповнити підмножинами Ω до сігма-алгебри. Переріз всіх сігма-алгебр, що містять деякий клас подій *K*, називають мінімальною сігма–алгеброю, що породжена *K*, і позначають $\sigma(K)$.

Пара $\{\Omega,\mathfrak{A}\}$, де Ω – простір елементарних подій, а $\mathfrak{A}$ – алгебра або $\sigma-$алгебра його підмножин, називається *вимірним простором*.

Вправи.

1. Монету підкидають двічі. Описати простір елементарних подій. Описати події: *A* – принаймні один раз випаде герб; *B* – при другому підкиданні випаде герб.
2. Монету підкидають до першого випадання герба. Описати простір елементарних подій. Описати події: *A* – буде здійснено парне число підкидань; *B* – експеримент закінчиться до шостого підкидання.
3. Експеримент полягає у вимірюванні двох величин, які набувають значення із відрізка $[0,1]$. Описати простір елементарних подій. Описати подію: *A* – значення величин відрізняються не більше, ніж на $1/3$.

1.2. Ймовірність випадкової події

1.2.1. Статистичне означення ймовірності. Коли ми проводимо деякий випадковий експеримент, то недостатньо знати, які події можуть настати в цьому експерименті, а необхідно ще знати деяку величину, яка б характеризувала можливість настання кожної події. Нехай в однакових умовах проводиться серія із n випадкових експериментів (випробувань), у кожному із яких може настати деяка подія *A*. Якщо $\mu(A)$– число експериментів, у яких подія A настала, то відношення $\nu(A)=\frac{\mu(A)}{n}$ (відношення числа експериментів, у яких подія A настала, до числа всіх проведених експериментів) називається *відносною частотою* настання події.

Відносна частота має такі властивості: $0\le\nu(A)\le 1$; $\nu(\Omega)=1$; $\nu(\varnothing)=0$; якщо A і B – несумісні події ($A\cap B=\varnothing$), то $\nu(A+B)=\nu(A)+\nu(B)$.

Спостереження показали, що при проведенні різних серій із великої кількості експериментів, відносні частоти події для цих серій мало відрізняються від певного числа. Ця закономірність називається властивістю статистичної стійкості відносних частот. Таким чином, з кожною випадковою подією можна пов'язати деяке стале число, біля якого групуються відносні частоти події, яке є характеристикою об'єктивного зв'язку між експериментом і випадковою подією, є об'єктивною характеристикою випадкової події.

Наприклад, статистичні спостереження показали, що відносна частота народження дівчаток близька до 0,486, відносна частота випадання герба, при киданні монети досить велику кількість раз, близька до числа 0,5.

Означення. Число, біля якого групуються відносні частоти випадкової події A (до якого наближається відносна частота події A) при зростанні числа експериментів, називається ймовірністю події A і позначається $P(A)$.

Це означення ймовірності називається статистичним. На практиці, при великій кількості експериментів, за ймовірність наближено приймають відносну частоту.

Із даного означення і властивостей відносної частоти випливають такі властивості ймовірності: $0\le P(A)\le 1$; $P(\Omega)=1$; $P(\varnothing)=0$; якщо $A\cap B=\varnothing$, то $P(A+B)=P(A)+P(B)$.

Дане означення ймовірності грунтується на реальному експерименті, але недолік його в тому, що для надійного визначення ймовірності необхідно провести велику кількість експериментів, які часто пов'язані з матеріальними витратами.

1.2.2. Ймовірнісна модель експерименту із скінченним або зліченним числом наслідків. Класичне означення ймовірності. Нехай простір елементарних подій скінченний або зліченний:

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

Кожній елементарній події ω_i поставимо у відповідність деяке число $p_i=P(\omega_i)$, що задовольняє умовам: $p_i\ge 0$, $\sum_{i=1}^{\infty}p_i=1$. Нехай подія $A=\{\omega_{k_1},...,\omega_{k_n},...\}$. Покладемо

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

Тоді, очевидно, визначена таким чином ймовірність має такі властивості: $0\le P(A)\le 1$; $P(\Omega)=1$; $P(\varnothing)=0$; якщо $A\cap B=\varnothing$, то $P(A+B)=P(A)+P(B)$.

Теорія ймовірностей не вивчає методів визначення ймовірностей елементарних подій. При визначенні ймовірностей елементарних подій приймається до уваги інтуїтивне представлення про ймовірність, основане на статистичному означенні. Теорія ймовірностей вивчає методи знаходження ймовірностей різних складних подій.

Розглянемо частинний випадок розглянутої моделі, а саме, класичну модель, коли простір елементарних подій скінченний і елементарні події рівноможливі (немає підстав вважати, що одна із них настає частіше іншої). Нехай $\Omega=\{\omega_1,...,\omega_n\}$, а випадкова подія $A\subset\Omega$, $A=\{\omega_{i_1},...,\omega_{i_m}\}$. Виходячи із рівноможливості елементарних подій, одержуємо

$p_i = P(\omega_i) = \dfrac{1}{n}$. Тоді $P(A) = \sum\limits_{i:\omega_i \in A} p_i = m\dfrac{1}{n} = \dfrac{m}{n}$. Отже, можна сформулювати наступне означення, яке називають класичним означенням ймовірності.

Означення. Нехай із експериментом пов’язані *n* рівноможливих елементарних подій, *m* із яких спричиняють подію A, тоді

$$P(A) = \frac{m}{n}$$

(відношення числа *m*, наслідків експерименту – елементарних подій, які спричиняють подію A, до числа *n*, всіх рівноможливих наслідків експерименту, називають ймовірністю події A).

Для розв’язування задач із використанням класичного означення ймовірності необхідно: встановити суть випадкового експерименту; чи наслідки його є рівноможливі; підрахувати число всіх можливих наслідків експерименту і число наслідків, що спричиняють дану подію. Тому застосування класичного означення ймовірності частіше всього пов’язане із використанням комбінаторики.

Правило добутку. Нехай маємо дві множини із m і n елементів, тоді число різних упорядкованих пар, які можна утворити, взявши один елемент із першої множини, а другий – із другої, дорівнює $m \cdot n$. Якщо першу дію можна виконати *m* способами, а другу – *n* способами, то одну за другою дві дії можна виконати *mn* способами.

Множина разом із зазначеним порядком її елементів називається упорядкованою. Встановлений у скінченній множині порядок називається перестановкою її елементів. Число перестановок із *n* елементів дорівнює

$$P_n = n!.$$

Упорядковані *k*–елементні підмножини множини із *n* елементів називаються розміщеннями із *n* елементів по *k*. Число розміщень із *n* по *k* дорівнює

$$A_n^k = n(n-1)\cdots(n-(k-1)) = \frac{n!}{(n-k)!}.$$

Довільні *k*–елементні підмножини множини із *n* елементів називаються комбінаціями із *n* елементів по *k*. Число комбінацій із *n* по *k* дорівнює

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

Приклад. В урні є 10 куль: 3 білих і 7 чорних. Із урни випадково вибирають дві кулі. Знайти ймовірність того, що вибрані кулі будуть білі.

Розв’язування. Експеримент, що розглядається у даній задачі, полягає у виборі двох куль із десяти. Так як вибір випадковий, то наслідки експерименту рівноможливі, тому *n* дорівнює числу способів вибору двох куль із десяти: $n = C_{10}^2 = \dfrac{10!}{2!8!} = 45$. Нехай *A* – подія, яка полягає у тому, що вибрані кулі будуть білі. Тоді число наслідків експерименту, що спричиняють подію *A*, дорівнює числу способів вибору двох куль із трьох білих: $m = C_3^2 = 3$. Отже, $P(A) = \dfrac{m}{n} = \dfrac{3}{45} = \dfrac{1}{15}$.

Вправи.

4. Навмання взятий телефонний номер складається із п’яти цифр. Знайти ймовірність того, що у ньому а) всі цифри різні; б) всі цифри парні.
5. На семи карточках написані букви: л, л, о, о, о, т, т. Випадково вибирають чотири букви і кладуть по порядку (зліва на право). Знайти ймовірність того, що утвориться слово “лото”.
6. Із колоди в 36 карт випадково вибирають три карти. Знайти ймовірність того, що: а) вони виявляться однакової масті; б) різної масті; в) рівно одна карта буде тузом; г) принаймні одна карта буде тузом.

7. Із 10 лотерейних білетів є 4 виграшні. Знайти ймовірність того, що із 5 випадково взятих білетів виграшними будуть: а) рівно 2; б) жодного; в) принаймні один.
8. Є 6 відрізків, довжини яких відповідно дорівнюють: 2, 4, 6, 8, 10, 12 одиниць. Яка ймовірність того, що із трьох випадково вибраних відрізків можна утворити трикутник?
9. Студент підготував 20 питань із 25. Знайти ймовірність того, що студент відповість на два питання із трьох йому випадково заданих.

1.2.3. Геометричне означення ймовірності. Нехай випадковий експеримент полягає у випадковому виборі точки із відрізка $[a,b]$. Очевидно, множина елементарних подій даного експерименту незліченна. Оскільки вибір випадковий, то елементарні події рівноможливі, але застосувати раніше введену схему неможливо. Тому будемо говорити про ймовірність вибору точки із проміжку. Нехай A – подія, яка полягає у тому, що точка вибирається із відрізка $[x,y]\subset[a,b]$. Припущення, що вибір випадковий, дає можливість вважати, що ймовірність вибору точки із відрізка $[x,y]$ не залежить від положення цього відрізка, а залежить тільки від його довжини і пропорційна його довжині. Враховуючи умову $P([a,b])=1$, одержимо: $P([x,y])=\dfrac{y-x}{b-a}$. Розглядаючи загальний випадок, приходимо до наступного означення.

Означення. Нехай простір елементарних подій є область $\Omega\subset R^n$, що має міру Лебега $m(\Omega)$, а експеримент полягає у випадковому виборі точок із Ω. Вимірні за Лебегом підмножини $A\subset\Omega$ розглядаємо як випадкові події. Тому для $A\subset\Omega$ покладемо

$$P(A)=\frac{m(A)}{m(\Omega)}.$$

Задача про зустріч. Двоє осіб домовилися зустрітися в певному місці у проміжок часу від 0 до T. Моменти приходу кожного – випадкові і рівноможливі у проміжку $[0,T]$. Перший, що приходить, чекає на другого протягом часу T_1 і, коли не дочекається, йде ($0<T_1<T$). Знайти ймовірність події A – зустріч настане.

Розв'язування. Нехай x – момент приходу першого, y – момент приходу другого, тоді множина всіх наслідків експерименту $\Omega=\{(x,y): x\in[0,T],\, y\in[0,T]\}$ – множина пар чисел із квадрата, $(x,y)\in[0,T]\times[0,T]$ і $m(\Omega)=T^2$. Зустріч відбудеться тільки тоді, коли $|x-y|\le T_1$, тому $A=\{(x,y)\in\Omega: |x-y|\le T_1\}$ – частина внутрішності квадрата, що розміщена між прямими $y=x+T_1$ і $y=x-T_1$. Частина Ω, що розміщена поза вказаними прямими є два рівнобедренні прямокутні трикутники із катетами $T-T_1$, тому їх площа разом дорівнює $(T-T_1)^2$, а $m(A)=T^2-(T-T_1)^2$. Отже,

$$P(A)=\frac{m(A)}{m(\Omega)}=\frac{T^2-(T-T_1)^2}{T^2}=1-\left(1-\frac{T_1}{T}\right)^2.$$

Вправи.
10. Задача Бюфона. Площина розграфлена паралельними прямими на віддалі $2a$. На неї випадково кидається голка довжиною $2l$ $(l<a)$. Знайти ймовірність того, що голка перетне одну із прямих.
11. Стержень довжиною l навмання розламали на три частини. Знайти ймовірність того, що із цих частин можна утворити трикутник.

12. Випадково вибрано три відрізки, довжина кожного із яких не перевищує l. Знайти ймовірність того, що із цих відрізків можна утворити трикутник.
13. Із відрізка [0,2] випадково вибирають два числа. Знайти ймовірність того, що їх сума буде менша одиниці.
14. На колі із радіусом R навмання вибираються дві точки. Знайти ймовірність того, що відстань між ними не перевищує R.
15. Числа a і b вибираються навмання із відрізка $[-1,1]$. Знайти ймовірність того, що корені рівняння $x^2+ax+b=0$ будуть комплексні.

1.2.4. Аксіоматичне означення ймовірності. При введенні ймовірності за допомогою аксіом необхідно їх вибирати так, щоб ймовірність зберігала властивості відносної частоти. Тільки у цьому випадку теорія буде узгоджуватись із практикою. Найбільш поширеною є аксіоматична побудова теорії ймовірностей, що запропонована в 1933 році А.М.Колмогоровим.

Означення. Нехай задано простір елементарних подій Ω і алгебру подій $\mathfrak{A}$. Ймовірністю події A називають функцію $P(A)$ визначену на алгебрі подій $\mathfrak{A}$ (кожній події A ставиться у відповідність число $P(A)$), при цьому виконуються умови (аксіоми):

P1. $P(A)\ge 0$;

P2. $P(\Omega)=1$;

P3. $P(A\bigcup B)=P(A)+P(B)$ для довільних $A\in\mathfrak{A}$, $B\in\mathfrak{A}$ таких, що $A\bigcap B=\varnothing$ (аксіома скінченної адитивності);

P4. $P\left(\bigcup\limits_{i=1}^{\infty}A_i\right)=\sum\limits_{i=1}^{\infty}P(A_i)$ для довільних $A_i\in\mathfrak{A}$, $i=1,2,...$ таких, що $A_i\bigcap A_j=\varnothing$ при $i\ne j$ і $\bigcup\limits_{i=1}^{\infty}A_i\in\mathfrak{A}$ (аксіома зліченної адитивності).

Відзначимо, що із P4 випливає P3, але не навпаки. Пізніше ми будемо використовувати саме таку групу аксіом. В означенні ми вважали, що $\mathfrak{A}$ – алгебра подій. Але ймовірність можна визначити і на сігма-алгебрі. Наведемо без доведення наступну, відому в теорії міри, теорему Каратеодорі.

Теорема 1.1 (про продовження ймовірності). Нехай $P(\cdot)$ – визначена на алгебрі $\mathfrak{A}$ і виконуються аксіоми P1 – P4. Тоді існує єдина функція $Q(\cdot)$, яка визначена на мінімальній сігма–алгебрі, що породжена $\mathfrak{A}$, при цьому виконуються аксіоми P1 – P4 і для довільної події $A\in\mathfrak{A}$ $Q(A)=P(A)$.

На основі цієї теореми, ми можемо вважати, що в аксіоматичному означенні ймовірності $\mathfrak{A}$ є σ-алгеброю подій. Аксіоми P1 і P4 означають, що функція $P(A)$ є мірою, для якої виконується додаткова умова P2. Така міра називається ймовірнісною. Трійка $\{\Omega,\mathfrak{A},P\}$, де Ω – простір елементарних подій, $\mathfrak{A}-$ $\sigma-$алгебра подій, P – ймовірнісна міра, називається *ймовірнісним простором*.

1.2.5. Наслідки із аксіом.

Н1. Ймовірність неможливої події дорівнює нулю.

Дійсно, $\varnothing=\varnothing\bigcup\varnothing$, а за третьою аксіомою $P(\varnothing)=P(\varnothing)+P(\varnothing)$. Отже, $P(\varnothing)=0$.

Н2. $P(A)+P(\overline{A})=1$.

Оскільки $A+\overline{A}=\Omega$ і $A\bigcap\overline{A}=\varnothing$, то за аксіомами P2 і P3

$$1 = P(\Omega) = P\left(A + \overline{A}\right) = P(A) + P\left(\overline{A}\right).$$

Для ймовірностей двох протилежних подій використовують позначення: $P(A) = p$, $P(\overline{A}) = q$.

Н3. Якщо $A_1, ..., A_n$ – попарно несумісні $\left(A_i \bigcap A_j = \varnothing, \quad i \neq j\right)$, то

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

Це випливає із аксіоми Р3.

Н4. Якщо $A_1, ..., A_n$ – утворюють повну групу $\left(A_1 + \cdots + A_n = \Omega, A_i \bigcap A_j = \varnothing, i \neq j\right)$, то

$$P(A_1) + \cdots + P(A_n) = 1.$$

Випливає із попереднього наслідку і аксіоми Р2.

Н5. Якщо $A \subset B$, то $P(B \setminus A) = P(B) - P(A)$ і $P(A) \le P(B)$.

Подію B можна представити у вигляді $B = A + B \bigcap \overline{A}$, де події A і $B \bigcap \overline{A} = B \setminus A$ несумісні. Тому за аксіомою Р3 $P(B) = P(A) + P\left(B \bigcap \overline{A}\right)$ або $P(B \setminus A) = P(B) - P(A)$. За аксіомою Р1 $P(B \setminus A) \ge 0$, тому $P(B) \ge P(A)$.

Н6. Для довільної події A $P(A) \le 1$.

Випливає із того, що $A \subset \Omega$ і попереднього наслідку.

Н7. Нехай *A* і *B* – довільні події. Тоді

$$P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B).$$

Цю рівність називають теоремою додавання ймовірностей.

Подію $A \bigcup B$ можна представити у вигляді двох несумісних подій A і $B \setminus (A \bigcap B)$: $A \bigcup B = A \bigcup \left(B \setminus (A \bigcap B)\right)$. Крім того, $A \bigcap B \subset B$. Тому за п'ятим наслідком і аксіомою Р3 $P(A \bigcup B) = P(A) + P(B \setminus A \bigcap B) = P(A) + P(B) - P(A \bigcap B)$.

Н8. $P(A \bigcup B) \le P(A) + P(B)$, $P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$, $P\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} P(A_i)$.

Перша нерівність випливає із сьомого наслідку, а друга – із першої за методом математичної індукції. Для доведення третьої нерівності введемо такі події: $A_0 = \varnothing$, $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, …, $B_n = A_n \setminus \left(A_1 \bigcup A_2 \bigcup \cdots \bigcup A_{n-1}\right)$, … . Тоді $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Визначені події B_n попарно несумісні і $B_n \subset A_n$, тоді за аксіомою Р4 і наслідком 5 $P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) \le \sum_{n=1}^{\infty} P(A_n)$.

Вправи.

16. Нехай $P(A) \ge 0,8$; $P(B) \ge 0,8$. Довести, що $P(A \bigcap B) \ge 0,6$.

17. Довести рівності: $P(A \bigcup B \bigcup C) =$

$$= P(A) + P(B) + P(C) - P(A \bigcap B) - P(A \bigcap C) - P(B \bigcap C) + P(A \bigcap B \bigcap C),$$

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i_1 < i_2 \le n} P(A_{i_1} \bigcap A_{i_2}) + \sum_{1 \le i_1 < i_2 < i_3 \le n} P(A_{i_1} \bigcap A_{i_2} \bigcap A_{i_3}) -$$

$$- \cdots + (-1)^{n-1} P(A_1 \bigcap ... \bigcap A_n).$$

1.2.6. Аксіома неперервності. Наступна теорема показує, що можна сформулювати еквівалентну систему аксіом, змінюючи четверту аксіому.

Нехай A_n – зростаюча послідовність подій, тобто така, що $A_1 \subset A_2 \subset \ldots \subset A_n \subset A_{n+1} \subset \ldots$, тоді покладемо $\lim\limits_{n\to\infty} A_n = \bigcup\limits_{n=1}^{\infty} A_n$. Якщо ж A_n – спадна послідовність подій, тобто така, що $A_1 \supset A_2 \supset \ldots \supset A_n \supset A_{n+1} \supset \ldots$, тоді покладемо $\lim\limits_{n\to\infty} A_n = \bigcap\limits_{n=1}^{\infty} A_n$.

Теорема 1.2. Якщо має місце аксіома скінченої адитивності, то наступні твердження еквівалентні:

1) якщо $A_i \in \mathfrak{A}$, $i = 1, 2, \ldots$ такі, що $A_i \cap A_j = \varnothing$ при $i \neq j$ і $\bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{A}$, то $P\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty} P(A_i)$;

2) якщо $A_1, \ldots, A_n, \ldots$ – зростаюча послідовність подій ($A_n \subset A_{n+1}$), $A_i \in \mathfrak{A}$, $i = 1, 2, \ldots$, $\lim\limits_{n\to\infty} A_n = \bigcup\limits_{n=1}^{\infty} A_n \in \mathfrak{A}$, то $P\left(\lim\limits_{n\to\infty} A_n\right) = P\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \lim\limits_{n\to\infty} P(A_n)$;

3) якщо $A_1, \ldots, A_n, \ldots$ – спадна послідовність подій ($A_n \supset A_{n+1}$), $A_i \in \mathfrak{A}$, $i = 1, 2, \ldots$, $\lim\limits_{n\to\infty} A_n = \bigcap\limits_{n=1}^{\infty} A_n \in \mathfrak{A}$, то $P\left(\lim\limits_{n\to\infty} A_n\right) = P\left(\bigcap\limits_{n=1}^{\infty} A_n\right) = \lim\limits_{n\to\infty} P(A_n)$;

4) якщо $A_1, \ldots, A_n, \ldots$ – спадна послідовність подій ($A_n \supset A_{n+1}$), $A_i \in \mathfrak{A}$, $i = 1, 2, \ldots$ і $\lim\limits_{n\to\infty} A_n = \bigcap\limits_{n=1}^{\infty} A_n = \varnothing$, то $\lim\limits_{n\to\infty} P(A_n) = 0$.

Доведення. Проведемо доведення за такою схемою: $1) \Rightarrow 2) \Rightarrow 3) \Rightarrow 4) \Rightarrow 1)$, із якої буде випливати еквівалентність умов 1) – 4).

$1) \Rightarrow 2)$. Нехай $A_1, \ldots, A_n, \ldots$ – зростаюча послідовність подій ($A_n \subset A_{n+1}$) із $\mathfrak{A}$. Покладемо $A_0 = \varnothing$, $B_n = A_n \setminus A_{n-1}$, $n = 1, 2, \ldots$. Тоді $B_n \in \mathfrak{A}$ і події B_n будуть попарно несумісними $\left(B_n \cap B_m = (A_n \setminus A_{n-1}) \cap (A_m \setminus A_{m-1}) = \varnothing,\ n \neq m\right)$. Тоді $\bigcup\limits_{k=1}^{\infty} A_k = \bigcup\limits_{k=1}^{\infty} B_k$, $A_n = \bigcup\limits_{k=1}^{n} B_k$ і за третім наслідком $P(A_n) = P\left(\bigcup\limits_{k=1}^{n} B_k\right) = \sum\limits_{k=1}^{n} P(B_k)$. Події B_n попарно несумісні, тому із виконання умови 1) випливає, що

$$P\left(\lim_{n\to\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n\to\infty} \sum_{k=1}^{n} P(B_k) = \lim_{n\to\infty} P\left(\bigcup_{k=1}^{n} B_k\right) = \lim_{n\to\infty} P(A_n).$$

$2) \Rightarrow 3)$. Нехай $A_1, \ldots, A_n, \ldots$ – спадна послідовність подій ($A_n \supset A_{n+1}$) із $\mathfrak{A}$, тоді $\overline{A_n}$ – зростаюча. Застосовуючи умову 2) до $\overline{A_n}$, одержимо

$$P\left(\lim_{n\to\infty} A_n\right) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = 1 - P\left(\overline{\bigcap_{n=1}^{\infty} A_n}\right) = 1 - P\left(\bigcup_{n=1}^{\infty} \overline{A_n}\right) =$$

$$= 1 - \lim_{n\to\infty} P\left(\overline{A_n}\right) = \lim_{n\to\infty}\left(1 - P\left(\overline{A_n}\right)\right) = \lim_{n\to\infty} P(A_n).$$

$3) \Rightarrow 4)$. Оскільки виконуються умови 3) і $\bigcap\limits_{n=1}^{\infty} A_n = \varnothing$, то $0 = \lim\limits_{n\to\infty} P(A_n)$.

$4)\Rightarrow 1)$. Нехай $A_i \in \mathfrak{A}$, $i=1,2,...$ такі, що $A_i \bigcap A_j = \varnothing$ при $i \neq j$ і $\bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{A}$. Тоді $\bigcup\limits_{n=1}^{\infty} A_n = \left(\bigcup\limits_{k=1}^{n} A_k\right) \bigcup C_n$, де $C_n = \bigcup\limits_{k=n+1}^{\infty} A_k$. Послідовність C_n – спадна і $\bigcap\limits_{n=1}^{\infty} C_n = \varnothing$, тому за умовою 4) $\lim\limits_{n\to\infty} P\left(C_n\right) = 0$. За властивістю скінченної адитивності

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\left(\bigcup_{k=1}^{n} A_k\right) \bigcup C_n\right) = P\left(\bigcup_{k=1}^{n} A_k\right) + P\left(C_n\right) = \sum_{k=1}^{n} P\left(A_k\right) + P\left(C_n\right).$$

Переходячи до границі в останній рівності при $n \to \infty$ і, враховуючи, що ліва частина не залежить від n, а $\lim\limits_{n\to\infty} P\left(C_n\right) = 0$, одержимо

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

Теорема доведена.

Кожне із тверджень 2), 3), 4) називають аксіомою неперервності ймовірності. Твердження 2) і 3) можна сформулювати в такому вигляді: якщо A_n – монотонна послідовність подій, то $P\left(\lim\limits_{n\to\infty} A_n\right) = \lim\limits_{n\to\infty} P\left(A_n\right)$. Властивість, що виражається такою рівністю, називають властивістю неперервності ймовірнісної міри.

1.3. Умовні ймовірності. Незалежність подій

1.3.1. Умовні ймовірності. Теорема множення. Розглядаючи ймовірності випадкових подій, ми не мали ніякої інформації про ймовірнісний зв’язок між різними випадковими подіями. Очевидно, він існує. Якщо при обчисленні ймовірності події A не накладається додаткових умов, крім тих, якими визначається випадковий експеримент, то ймовірність $P(A)$ називають безумовною. Але в ряді випадків необхідно обчислити ймовірності подій при додатковій умові, що настала деяка подій B, яка має додатню ймовірність. Ймовірність події A, обчислена за припущенням, що подія B уже настала, називається *умовною ймовірністю* події A при умові B і позначається $P\left(A/B\right)$ або $P_B\left(A\right)$.

Розглянемо приклад знаходження умовної ймовірності в класичній моделі. Позначимо n_A, n_B, n_{AB} кількість елементарних подій, що спричиняють відповідно події A, B, $A \bigcap B$. Тоді

$$P(A) = \frac{n_A}{n}, \quad P(B) = \frac{n_B}{n}, \quad P(A \bigcap B) = \frac{n_{AB}}{n}.$$

Якщо подія B уже настала, то змінюються умови експерименту і у новому (умовному) експерименті число можливих наслідків буде рівне n_B – числу елементарних подій, що спричиняють подію B, а подію A будуть спричиняти тільки ті елементарні події, які спричиняють $A \bigcap B$. Тому

$$P\left(A/B\right) = \frac{n_{AB}}{n_B} = \frac{n_{AB}/n}{n_B/n} = \frac{P(A \bigcap B)}{P(B)}.$$

У загальному випадку умовна ймовірність вводиться за допомогою наступного означення.

Означення. *Умовною ймовірністю* події A при умові, що подія B настала з $P(B)>0$, називається число

$$P\left(A/B\right)=\frac{P\left(A\bigcap B\right)}{P\left(B\right)}.$$

Аналогічно, при виконанні умови $P(A)>0$,

$$P\left(B/A\right)=\frac{P\left(A\bigcap B\right)}{P\left(A\right)}.$$

Введена таким чином умовна ймовірність має всі властивості ймовірності. Для перевірки цього досить показати, що виконуються аксіоми ймовірності. Із означення і властивостей ймовірності випливає:

1) $0\le P\left(A/B\right)\le 1$,

2) $P\left(\Omega/B\right)=1$,

3) $P\left(B/B\right)=1$,

4) якщо $A\supset B$, то $P\left(A/B\right)=1$,

5) якщо $A_1,...,A_n,...$ – послідовність попарно несумісних подій ($A_i\bigcap A_j=\varnothing$ при $i\ne j$), то $P\left(\bigcup\limits_{n=1}^{\infty}A_n\Big/B\right)=\sum\limits_{n=1}^{\infty}P\left(A_n/B\right)$.

6) $P(A/B)+P(\overline{A}/B)=1$.

Властивості 2) і 3) випливають із 4), тому доведемо 4) і 5).

Якщо $A\supset B$, то $A\bigcap B=B$, тому $P\left(A/B\right)=\frac{P\left(A\bigcap B\right)}{P\left(B\right)}=\frac{P\left(B\right)}{P\left(B\right)}=1$.

$$P\left(\bigcup_{n=1}^{\infty}A_n\Big/B\right)=\frac{P\left(\left(\bigcup\limits_{n=1}^{\infty}A_n\right)\bigcap B\right)}{P(B)}=\frac{P\left(\bigcup\limits_{n=1}^{\infty}\left(A_n\bigcap B\right)\right)}{P(B)}=$$

$$=\frac{\sum\limits_{n=1}^{\infty}P\left(A_n\bigcap B\right)}{P(B)}=\sum_{n=1}^{\infty}\frac{P\left(A_n\bigcap B\right)}{P(B)}=\sum_{n=1}^{\infty}P\left(A_n/B\right).$$

Нехай $\mathfrak{A}_B$ – сігма-алгебра подій $A\bigcap B$, де $A\in\mathfrak{A}$, а $B\in\mathfrak{A}$ – фіксована, тоді $\left\{B,\mathfrak{A}_B,P\left(\cdot/B\right)\right\}$ – буде ймовірнісним простором.

Із означення умовної ймовірності одержуємо твердження, яке називається теоремою множення ймовірностей.

Теорема 1. Якщо $P(A)>0$ і $P(B)>0$, то

$$P\left(A\bigcap B\right)=P\left(A\right)P\left(B/A\right)=P\left(B\right)P\left(A/B\right).$$

Теорема 2. Якщо $A_1,...,A_n$ – такі випадкові події, що $P\left(A_1\bigcap...\bigcap A_{n-1}\right)>0$, тоді

$$P\left(\bigcap_{k=1}^{n}A_k\right)=P\left(A_1\right)P\left(A_2/A_1\right)P\left(A_3/A_1\bigcap A_2\right)\cdots P\left(A_n/A_1\bigcap...\bigcap A_{n-1}\right)$$

(загальна теорема множення ймовірностей).

Доведення. Оскільки $\bigcap\limits_{k=1}^{n-1}A_k\subset\bigcap\limits_{k=1}^{n-m}A_k$ для довільного $m=1,...,n-1$, то

$0<P\left(\bigcap\limits_{k=1}^{n-1}A_k\right)\le P\left(\bigcap\limits_{k=1}^{n-m}A_k\right)$, а це означає, що всі умовні ймовірності в теоремі визначені.

Застосовуючи теорему 1, одержимо

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P\left(\left(A_1 \cap \ldots \cap A_{n-1}\right) \cap A_n\right) = P\left(A_1 \cap \ldots \cap A_{n-1}\right) P\left(A_n / A_1 \cap \ldots \cap A_{n-1}\right) =$$
$$= P\left(A_1 \cap \ldots \cap A_{n-2}\right) P\left(A_{n-1} / A_1 \cap \ldots \cap A_{n-2}\right) P\left(A_n / A_1 \cap \ldots \cap A_{n-1}\right) = \ldots =$$
$$= P\left(A_1\right) P\left(A_2 / A_1\right) P\left(A_3 / A_1 \cap A_2\right) \cdots P\left(A_n / A_1 \cap \ldots \cap A_{n-1}\right).$$

Теорема доведена.

Приклад 1. В урні є 10 куль: 3 білих і 7 чорних. Із урни випадково вибирають дві кулі. Знайти ймовірність того, що вибрані кулі будуть білі.

Розв'язування. Експеримент, що розглядається в даній задачі, полягає у виборі двох куль із десяти. Нехай A – подія, яка полягає у тому, що перша вибрана куля буде білою, а B – друга вибрана куля буде білою. Тоді $A \cap B$ – обидві вибрані кулі будуть білі. За класичним означенням $P(A) = \frac{3}{10}$. Якщо A настала, то в урні залишилось дев'ять куль, із яких дві білі, тому $P(B / A) = \frac{2}{9}$. Отже, за теоремою множення $P\left(A \cap B\right) = P\left(A\right) P\left(B / A\right) = \frac{3}{10}\frac{2}{9} = \frac{1}{15}$.

Вправи.

18. Монету підкидають три рази. Описати простір елементарних подій; описати події: A – двічі випав герб, B – принаймні один раз випав герб; обчислити $P(B)$, $P(A \cap B)$ і $P(A / B)$.
19. Підкидаються два гральні кубики. Знайти ймовірність того, що випаде принаймні одна шістка, якщо випала сума очок рівна 8.
20. Дано: $P(A / B) = 0,7$, $P(A / \overline{B}) = 0,3$, $P(B / A) = 0,6$. Знайти $P(A)$.
21. Студент підготував 20 із 25 питань програми. Знайти ймовірність того, що студент знає відповіді на випадково запропоновані йому три питання.
22. Події A і B – несумісні і $P(B) > 0$. Обчислити $P(A / B)$.

1.3.2. Незалежність подій. Розглянемо тепер одне із найважливіших понять теорії ймовірностей – поняття незалежності випадкових подій. Саме це поняття виділило теорію ймовірностей із теорії міри і теорії функцій в самостійну дисципліну.

Означення 1. Події A і B називаються *незалежними*, якщо

$$P\left(A \cap B\right) = P\left(A\right) P\left(B\right).$$

Теорема 1.3. Для того, щоб події A і B були незалежними, необхідно і досить, щоб виконувалась одна із умов: $P\left(A / B\right) = P\left(A\right)$ (якщо $P(B) > 0$); або $P\left(B / A\right) = P\left(B\right)$ (якщо $P(A) > 0$).

Доведення. Якщо події A і B незалежні, то $P\left(A \cap B\right) = P\left(A\right) P\left(B\right)$, а за теоремою множення $P\left(A \cap B\right) = P\left(A\right) P\left(B / A\right) = P\left(B\right) P\left(A / B\right)$, тому $P\left(B\right) P\left(A / B\right) = P\left(A\right) P\left(B\right)$ і $P\left(A\right) P\left(B / A\right) = P\left(A\right) P\left(B\right)$. Тобто $P\left(A / B\right) = P\left(A\right)$ і $P\left(B / A\right) = P\left(B\right)$. І навпаки, нехай, наприклад, $P\left(A / B\right) = P\left(A\right)$. Тоді за теоремою множення $P\left(A \cap B\right) = P\left(B\right) P\left(A / B\right) = = P\left(A\right) P\left(B\right)$. Отже, події незалежні. Теорема доведена.

Таким чином, незалежність подій означає, що настання однієї із подій не змінює ймовірності іншої. На практиці висновок про незалежність подій роблять на основі незалежності експериментів, із якими ці події пов'язані.

Приклад 2. Двічі підкидають монету. Нехай подія A – при першому підкиданні випадає герб, B – при другому підкиданні випадає герб. Обчислити $P(A)$, $P(B)$, $P(A\bigcap B)$, $P(B/A)$. Показати, що події A і B незалежні.

Розв'язування. Простір елементарних подій $\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4\}$ складається із чотирьох рівноможливих елементарних подій: ω_1 – ГГ, ω_2 – ЦЦ, ω_3 – ГЦ, ω_4 – ЦГ. Тоді

$$A=\{\omega_1,\omega_3\},\ B=\{\omega_1,\omega_4\},\ A\bigcap B=\{\omega_1\},\text{ а } P(A)=\frac{2}{4}=0,5,\ P(B)=\frac{2}{4}=0,5,$$

$P(A\bigcap B)=\frac{1}{4}=0,25$. Події A і B незалежні, бо $P(A\bigcap B)=0,25=0,5\cdot 0,5=P(A)P(B)$.

Умовну ймовірність знайдемо за означенням: $P\left(B/A\right)=\dfrac{P(A\bigcap B)}{P(A)}=\dfrac{0,25}{0,5}=0,5$.

Висновок про незалежність подій A і B можемо зробити і на основі попередньої теореми: $P\left(B/A\right)=P\left(B\right)=0,5$, тому події A і B незалежні.

Приклад 3. В кожному із двох ящиків міститься по 10 куль: в першому – 8 білих і 2 чорні, в другому – 7 білих і 3 чорні. З кожного ящика вибирають по одній кулі. Знайти ймовірність того, що ці кулі будуть білі.

Розв'язування. Нехай подія A – куля, що вибрана із першого ящика, є білою, B – куля, що вибрана із другого ящика, є білою, тоді $A\bigcap B$ – обидві кулі є білими. Так як $P(A)=\frac{8}{10}=0,8$, $P(B)=\frac{7}{10}=0,7$, а події A і B – незалежні, то $P(A\bigcap B)=P(A)P(B)=$ $=0,8\cdot 0,7=0,56$.

Теорема 1.4. Якщо події A і B незалежні, то незалежними будуть і події $\overline{A}$ і B, A і $\overline{B}$, $\overline{A}$ і $\overline{B}$.

Доведення. Так як $\overline{A}\bigcap B=B\setminus(A\bigcap B)$ і $A\bigcap B\subset B$, то із незалежності подій ($P\left(A\bigcap B\right)=P\left(A\right)P\left(B\right)$) одержуємо

$$P\left(\overline{A}\bigcap B\right)=P\left(B\setminus(A\bigcap B)\right)=P\left(B\right)-P\left(A\bigcap B\right)=P\left(B\right)-P\left(B\right)P\left(A\right)=$$
$$=P\left(B\right)\left(1-P\left(A\right)\right)=P\left(B\right)P\left(\overline{A}\right).$$

Отже $\overline{A}$ і B незалежні. Аналогічно,

$$P\left(A\bigcap \overline{B}\right)=P\left(A\setminus(A\bigcap B)\right)=P\left(A\right)-P\left(A\bigcap B\right)=P\left(A\right)-P\left(B\right)P\left(A\right)=$$
$$=P\left(A\right)\left(1-P\left(B\right)\right)=P\left(A\right)P\left(\overline{B}\right),$$

що означає незалежність подій A і $\overline{B}$. І, нарешті,

$$P\left(\overline{A}\bigcap\overline{B}\right)=1-P\left(\overline{\overline{A}\bigcap\overline{B}}\right)=1-P\left(A\bigcup B\right)=1-P\left(A\right)-P\left(B\right)+P\left(A\bigcap B\right)=$$
$$=1-P\left(A\right)-P\left(B\right)+P\left(A\right)P\left(B\right)=\left(1-P\left(A\right)\right)\left(1-P\left(B\right)\right)=P\left(\overline{A}\right)P\left(\overline{B}\right).$$

Тобто, $\overline{A}$ і $\overline{B}$ незалежні. Теорема доведена.

Приклад 4. Для сигналізації про аварію встановлено два незалежно працюючих сигналізатори. Ймовірність того, що при аварії перший сигналізатор спрацює, дорівнює 0,7, а другий – 0,8. Знайти ймовірності таких подій: A – при аварії спрацює тільки один сигналізатор; B – при аварії сигнал про аварію буде поданий.

Розв'язування. Нехай A_i – подія, яка полягає в тому, що при аварії i – й ($i=1;2$) сигналізатор спрацює. Подія A настає, якщо спрацьовує тільки перший (подія $A_1\bigcap\overline{A_2}$), або тільки другий сигналізатор (подія $\overline{A_1}\bigcap A_2$), тому $A=A_1\bigcap\overline{A_2}\bigcup\overline{A_1}\bigcap A_2$. За умовою

$P(A_1)=0,7$, $P(A_2)=0,8$, події $A_1\bigcap\overline{A_2}$ і $\overline{A_1}\bigcap A_2$ несумісні, а події A_1 і A_2 незалежні (сигналізатори працюють незалежно), тоді $P(A)=P(A_1\bigcap\overline{A_2})+P(\overline{A_1}\bigcap A_2)=$
$=P(A_1)P(\overline{A_2})+P(\overline{A_1})P(A_2)=0,7\cdot 0,2+0,3\cdot 0,8=0,38$.

Для знаходження ймовірності події B вигідніше перейти до протилежної події. Подія $\overline{B}$ настає, якщо не спрацює жодний сигналізатор, тобто, $\overline{B}=\overline{A_1}\bigcap\overline{A_2}$. Тоді $P(\overline{B})=P(\overline{A_1})P(\overline{A_2})=0,3\cdot 0,2=0,06$, а $P(B)=1-P(\overline{B})=1-0,06=0,94$.

Даний приклад показує, що уже елементарні поняття теорії ймовірностей мають практичне застосування – пояснюється підвищення надійності при дублюванні.

Означення 2. Події $A_1, A_2, ..., A_n$ називаються *попарно незалежними*, якщо кожні дві – незалежні.

Означення 3. Події $A_1, A_2, ..., A_n$ називаються *незалежними* або *незалежними в сукупності*, якщо для будь-якого k $\left(1\le k\le n\right)$ і для будь-яких $i_1,...,i_k$ $\left(1\le i_1<i_2<...<i_k\le n\right)$

$$P\left(A_{i_1}\bigcap A_{i_2}\bigcap ...\bigcap A_{i_k}\right)=P\left(A_{i_1}\right)P\left(A_{i_2}\right)\cdots P\left(A_{i_k}\right).$$

Із незалежності в сукупності подій випливає їх попарна незалежність, але не навпаки. Нехай, наприклад, $\Omega=\left\{\omega_1,\omega_2,\omega_3,\omega_4\right\}$ і всі елементарні події рівноможливі. Тоді події $A=\left\{\omega_1,\omega_2\right\}$, $B=\left\{\omega_1,\omega_3\right\}$, $C=\left\{\omega_1,\omega_4\right\}$ попарно незалежні, але не є незалежними в сукупності. Дійсно, $A\bigcap B=A\bigcap C=B\bigcap C=A\bigcap B\bigcap C=\left\{\omega_1\right\}$, тому

$$P(A)=P(B)=P(C)=\frac{1}{2},$$

$$P\left(A\bigcap B\right)=P\left(A\right)P\left(B\right)=\frac{1}{4},\ P\left(A\bigcap C\right)=P\left(A\right)P\left(C\right)=\frac{1}{4},\ P\left(B\bigcap C\right)=P\left(B\right)P\left(C\right)=\frac{1}{4},$$

але
$$P(A\bigcap B\bigcap C)=\frac{1}{4}\ne\left(\frac{1}{2}\right)^3=P\left(A\right)P\left(B\right)P(C).$$

Відзначимо також, що із того, що для деяких подій A, B, C

$$P(A\bigcap B\bigcap C)=P\left(A\right)P\left(B\right)P(C),$$

ще не випливає попарна незалежність цих подій. Дійсно, нехай $\Omega=\left\{(i,j):i,j=1,2,...,6\right\}$ і всі пари рівноможливі. Тоді для $A=\left\{(i,j)\in\Omega: j=1,2,5\right\}$, $B=\left\{(i,j)\in\Omega: j=4,5,6\right\}$, $C=\left\{(i,j): i+j=9\right\}$ маємо $P(A)=P(B)=\frac{1}{2}$, $P(C)=\frac{1}{9}$, $P(A\bigcap B)=\frac{1}{6}\ne\frac{1}{4}=P(A)P(B)$, $P(A\bigcap C)=\frac{1}{36}\ne\frac{1}{18}=P(A)P(C)$, $P(B\bigcap C)=\frac{1}{12}\ne\frac{1}{18}=P(B)P(C)$, але $P(A\bigcap B\bigcap C)=\frac{1}{36}=P(A)P(B)P(C)$.

Вправи.

23. Події $A_1, A_2, ..., A_n$ – незалежні. Довести, що $P\left(\bigcup_{k=1}^{n}A_k\right)=1-\prod_{k=1}^{n}P(\overline{A_k})$.

24. Ймовірність того, що при аварії сигналізатор спрацює, дорівнює 0,7. Скільки необхідно встановити незалежно працюючих сигналізаторів, щоб із ймовірністю, яка не менше 0,9, при аварії сигнал про аварію був поданий.

25. Перший стрілець влучає в ціль з ймовірністю 0,6, другий – 0,5, а третій – 0,4. Стрільці зробили залп по цілі. Відомо, що є два влучення. Що більш ймовірно: попав третій стрілець у ціль, чи ні?
26. Розрив електричного кола настає внаслідок виходу із ладу елемента К або двох елементів K_1 і K_2, які виходять з ладу незалежно один від одного відповідно з ймовірностями 0,3; 0,2 і 0,2. Знайти ймовірність розриву електричного кола.
27. Довести, що коли події A і B незалежні і $A \subset B$, то $P(A) = 0$ або $P(B) = 1$.
28. Довести, що якщо $P(A) > 0$ і $P(A/\overline{B}) = P(A/B)$, то події A і B – незалежні.

1.3.3. Формула повної ймовірності. Формули Байєса. Нехай подія A може настати із однією із подій $H_1,...,H_n$, що утворюють повну групу і $P(H_i) > 0$, $i = 1,...,n$. Це означає, що події $H_1,...,H_n$ попарно несумісні і $\bigcup_{i=1}^{n} H_i = \Omega$. Тоді $A = A \bigcap \Omega = A \bigcap \left(H_1 \bigcup \cdots \bigcup H_n \right) = \bigcup_{i=1}^{n} A \bigcap H_i$. Із попарної несумісності подій $H_1,...,H_n$ випливає, що події $A \bigcap H_1, \cdots, A \bigcap H_n$ також попарно несумісні, тому $P\left(A\right) = \sum_{k=1}^{n} P\left(A \bigcap H_k\right) = \sum_{k=1}^{n} P\left(H_k\right) P\left(A/H_k\right)$. Тобто,

$$P\left(A\right) = \sum_{k=1}^{n} P\left(H_k\right) P\left(A/H_k\right).$$

Одержана формула називається *формулою повної ймовірності*. Відзначимо, що формула буде правильною і для зліченної кількості подій.

Приклад 5. Робітник обслуговує три верстати, на яких обробляються однотипні деталі. Ймовірність браку для першого верстата дорівнює 0,02, для другого – 0,03, для третього – 0,04. Оброблені деталі складаються в один ящик. Продуктивність першого верстата в три рази більша, ніж продуктивність третього, а другого – в два рази більша, ніж продуктивність третього. Знайти ймовірність того, що випадково взята із ящика деталь буде бракована.

Розв'язування. Нехай подія A – випадково взята із ящика деталь, буде бракована. Тоді ця подія може настати із однією із подій H_i – взята деталь оброблена на i – му верстаті ($i = 1,2,3$). Події H_1, H_2, H_3 утворюють повну групу. Застосуємо формулу повної ймовірності

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + P(H_3)P(A/H_3).$$

Ймовірності $P(A/H_i)$ задані в умові задачі;

$$P(A/H_1) = 0,02;\ P(A/H_2) = 0,03;\ P(A/H_3) = 0,04.$$

Для ймовірностей подій H_1, H_2, H_3 із умови задачі одержуємо систему:

$$P(H_1) = 3P(H_3);\ P(H_2) = 2P(H_3);\ P(H_1) + P(H_2) + P(H_3) = 1.$$

Звідки, $P(H_1) = \frac{1}{2}$, $P(H_2) = \frac{1}{3}$, $P(H_3) = \frac{1}{6}$. Підставляючи ймовірності $P(H_i)$ і $P(A/H_i)$ в формулу повної ймовірності, одержуємо:

$$P(A) = 0,02 \cdot \frac{1}{2} + 0,03 \cdot \frac{1}{3} + 0,04 \cdot \frac{1}{6} = \frac{0,08}{3} \approx 0,027.$$

Нехай виконуються, вказані на початку пункту, умови і $P(A) > 0$. Будемо вважати, що подія A настала. Необхідно знайти $P\left(H_i / A\right)$. За теоремою множення $P\left(A \cap H_i\right) = P\left(H_i\right) P\left(A / H_i\right) = P\left(A\right) P\left(H_i / A\right)$, тому $P\left(H_i / A\right) = \dfrac{P\left(H_i\right) P\left(A / H_i\right)}{P\left(A\right)}$, а за формулою повної ймовірності

$$P\left(H_i / A\right) = \frac{P\left(H_i\right) P\left(A / H_i\right)}{\sum_{k=1}^{n} P\left(H_k\right) P\left(A / H_k\right)} \quad \left(i = 1,..,n\right).$$

Одержані формули називаються *формулами Байєса*.

В статистичних застосуваннях події $H_1,...,H_n$ називаються гіпотезами, а $P(H_i)$ – апріорною ймовірністю гіпотези H_i. Умовна ймовірність $P(H_i / A)$ називається апостеріорною ймовірністю гіпотези H_i після настання події A.

Приклад 6. Вся продукція перевіряється двома контролерами. Перший контролер перевіряє 55% всієї продукції, а другий – 45%. Ймовірність того, що перший контролер пропустить нестандартний виріб, дорівнює 0,01, а другий – 0,02. Випадково взятий виріб, що визнаний при перевірці стандартним, виявився нестандартним. Знайти ймовірність того, що цей виріб перевірявся другим контролером.

Розв'язування. Нехай A – подія, яка полягає в тому, що випадково взятий виріб є нестандартним, але при перевірці визнаний стандартним, H_1 – виріб перевірявся першим контролером, H_2 – виріб перевірявся другим контролером. Нам необхідно знайти $P(H_2 / A)$. Використаємо формулу Байєса

$$P\left(H_2 / A\right) = \frac{P\left(H_2\right) P\left(A / H_2\right)}{P(H_1)P(A / H_1) + P(H_2)P(A / H_2)}.$$

За умовою задачі $P(A / H_1) = 0,01$, $P(A / H_2) = 0,02$. Так як перший контролер перевіряє 55% всієї продукції, а другий – 45%, то $P(H_1) = 0,55$, а $P(H_2) = 0,45$. Тому $P(H_2 / A) = \dfrac{0,45 \cdot 0,02}{0.55 \cdot 0,01 + 0,45 \cdot 0,02} = \dfrac{0,09}{0,0055 + 0,009} = \dfrac{18}{29}$.

Вправи.

29. Серед N екзаменаційних білетів є n "щасливих". Студенти підходять за білетами один за одним. У кого більша ймовірність взяти "щасливий" білет: у того, хто підійшов першим, чи у того, хто підійшов другим?
30. В кожній із двох урн міститься 6 білих і 4 чорних кулі. Із першої урни випадково вибираються дві кулі і перекладаються в другу. Знайти ймовірність того, що випадково вибрана після цього куля із другої урни буде біла.
31. Із 20 стрілків 4 попадають в мішень з ймовірністю 0,9, 10 – з ймовірністю 0,8, 6 – з ймовірністю 0,6. Знайти ймовірність того, що випадково взятий стрілок промахнеться.
32. Лічильник реєструє частинки трьох типів: A, B, C. Ймовірності появи цих частинок відповідно дорівнюють 0,2; 0,5; 0,3. Частинки кожного з цих типів лічильник реєструє відповідно з ймовірностями 0,8; 0,2; 0,4. Лічильник зареєстрував частинку. Знайти ймовірність того, що це була частинка типу A.
33. Три верстати виготовляють відповідно 25%, 35% і 40% всіх виробів. В їх продукції брак становить відповідно 2%, 3% і 5%. Випадково вибраний із всієї продукції виріб виявився бракованим. Яка ймовірність того, що він виготовлений першим верстатом?

Розділ 2. Схема Бернуллі. Формула Бернуллі

2.1. Послідовність незалежних випробувань. Формула Бернуллі

Нехай послідовно проводяться випадкові експерименти (випробування). Будемо вважати, що будь–яка подія, яка відноситься до фіксованого випробування, буде незалежною від будь–якої події, що відноситься до інших випробувань. Тоді говорять, що задано послідовність *незалежних випробувань*. Нехай проводиться послідовність незалежних випадкових експериментів, в кожному із яких може настати деяка подія A. Якщо ймовірність настання події A $P\left(A\right) = p$ є однаковою в кожному експерименті, то таку послідовність випробувань називають *схемою Бернуллі*.

Нехай проводяться серії із n випробувань. Позначимо $P_n\left(k\right)$ – ймовірність того, що в n випробуваннях подія A настане точно k разів. Результати серії із n випробувань можна подати у вигляді $\omega = (\omega_1, ..., \omega_n)$, де ω_k – подія, яка настає в k-му експерименті, $\omega_k = A$ або $\overline{A}$. Тоді $P(\omega_k) = P(A) = p$, якщо в k-му експерименті настала подія A, і $P(\omega_k) = P(\overline{A}) = 1 - p = q$, якщо $\omega_k = \overline{A}$. Тому, враховуючи незалежність, $P(\omega) = P(\omega_1) \cdot ... \cdot P(\omega_n)$, а для тих $\omega = (\omega_1, ..., \omega_n)$, у яких A входить рівно k разів, а $\overline{A}$ n–k разів $P(\omega) = p^k q^{n-k}$. Але k випробувань, у яких настає подія A, можна вибрати C_n^k способами і для кожного з них ймовірність рівна $p^k q^{n-k}$. Тоді за аксіомою адитивності,

$$P_n\left(k\right) = C_n^k p^k \left(1 - p\right)^{n-k}. \tag{1}$$

Одержана формула називається *формулою Бернуллі*.

Оскільки, $(p+q)^n = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = 1$, то $\sum_{k=0}^{n} P_n(k) = 1$.

Нехай n – деяке фіксоване число, k – змінне, $k = 0, 1, ..., n$. Дослідимо поведінку $P_n\left(k\right)$ при зміні k. Розглянемо відношення

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

Якщо $\left(n-k\right)p \geq \left(k+1\right)q$ або $k \leq np - q$, то $P_n(k+1) \geq P_n(k)$, тобто, ймовірності $P_n(k)$ – зростають. Аналогічно,

$$\frac{P_n\left(k\right)}{P_n\left(k-1\right)} = \frac{n-k+1}{k} \frac{p}{q} \leq 1$$

при $k \geq \left(n+1\right)p$. Тому при $k \geq \left(n+1\right)p$ ймовірності $P_n(k)$ – спадають. Якщо $np - q$ є цілим числом, то максимальне значення ймовірності $P_n\left(k\right)$ досягається при двох значеннях k, а саме: $k_0 = np - q$ і $k_0' = np - q + 1 = np + p$. Якщо ж $np - q$ не є цілим числом, то максимальне значення ймовірності $P_n\left(k\right)$ досягається тільки при одному значенні $k_0 = [np + p]$. Число k_0, при якому $P_n\left(k\right)$ є максимальною, називається найімовірнішим числом настання події. При цьому справедлива нерівність

$$np - q \leq k_0 \leq np + p. \tag{2}$$

Приклад 1. Ймовірність виготовлення на автоматичному верстаті стандартної деталі дорівнює 0,9. Знайти ймовірність того, що із 6 випадково взятих деталей 4 будуть стандартні. Знайти найімовірніше число стандартних деталей із 6 випадково взятих деталей.

Розв'язування. Використаємо формулу Бернуллі: $P_n(k) = C_n^k p^k (1-p)^{n-k}$. В нашому випадку $n = 6$, $p = 0,9$, $q = 1 - p = 0,1$, тому $P_6(4) = C_6^4 (0,9)^4 (0,1)^2 \approx$ $\approx 0,0984$. Для знаходження найімовірнішого числа стандартних деталей із 6 випадково взятих деталей застосуємо (2): $6 \cdot 0,9 - 0,1 \le k_0 \le 6 \cdot 0,9 + 0,9$. Тобто, $k_0 = 6$.

Застосування формули Бернуллі при великих n вимагає великої обчислювальної роботи, при цьому виникають похибки заокруглення, тому результат одержуємо наближений. Ми розглянемо тепер деякі наближені формули, що значно спрощують обчислення ймовірностей в схемі Бернуллі.

Вправи.

1. Ймовірність настання події A в кожному із 8 незалежних випробувань дорівнює 0,6. Знайти: a) найімовірніше число настання події – k_0 та $P_8(k_0)$; b) ймовірність того, що подія настане принаймні один раз.
2. Що ймовірніше: виграти у рівносильного супротивника 3 партії із 5, чи 4 партії із 8?
3. Подія B настає, коли подія A настане не менше двох раз. Знайти ймовірність настання події B, якщо буде проведено 6 незалежних випробувань, в кожному із яких $P(A) = 0,4$.

2.2. Граничні теореми в схемі Бернуллі

2.2.1. Теорема Пуассона. Якщо ймовірність настання події близька до нуля або одиниці, то для знаходження $P_n(k)$ в цьому випадку застосовують наближену формулу, що одержується із теореми Пуассона.

Теорема. Якщо в серії із n незалежних випробувань ймовірність настання події A є сталою в серії і рівною p_n і при $n \to \infty$ $p_n \to 0$ так, що $np_n \to \lambda$, $0 < \lambda < \infty$, то при довільному сталому k, $k = 0,1,2,...,$

$$\lim_{n\to\infty} P_n(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Доведення. Нехай $np_n = \lambda_n$, тоді

$$P_n(k) = \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} = \frac{1}{k!} n(n-1)\cdots(n-k+1)\left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} =$$

$$= \frac{\lambda_n^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda_n}{n}\right)^n\left(1-\frac{\lambda_n}{n}\right)^{-k}.$$

Так як k – стале, то при $n \to \infty$ кожен із множників $1-\frac{1}{n}$, $1-\frac{2}{n}$, $1-\frac{3}{n}$,…, $1-\frac{k-1}{n}$, $\left(1-\frac{\lambda_n}{n}\right)^{-k}$ прямує до одиниці, а $\left(1-\frac{\lambda_n}{n}\right)^n = \left(\left(1-\frac{\lambda_n}{n}\right)^{-\frac{n}{\lambda_n}}\right)^{-\lambda_n} \to e^{-\lambda}$, тому при $n \to \infty$

$P_n(k) \to e^{-\lambda}\frac{\lambda^k}{k!}$.

Теорема доведена.

Теорема Пуассона використовується при обчисленні ймовірностей рідких подій, коли число випробувань досить велике, а ймовірність настання події в кожному випробуванні досить мала. Таким чином, при великих n і малих p можемо використовувати наближену формулу

$$P_n(k) \approx e^{-\lambda} \frac{\lambda^k}{k!}, \tag{3}$$

де $\lambda = np$.

Приклад 2. Прилад містить 2000 однаково надійних елементів. Ймовірність відмови кожного із них дорівнює 0,0005. Яка ймовірність відмови приладу, якщо вона настає при відмові більше, ніж двох елементів?

Розв'язування. Нехай A – відмова більше, ніж двох елементів, A_i – відмова i елементів, тоді $\overline{A} = A_0 + A_1 + A_2$. Так як $n = 2000$ – досить велике, а $p = 0,0005$ – мале, то для знаходження ймовірностей подій A_i можемо використати наближену формулу (3), в якій $\lambda = np = 1$. Тому

$$P(\overline{A}) = P(A_0) + P(A_1) + P(A_2) = P_{2000}(0) + P_{2000}(1) + P_{2000}(2) \approx$$
$$= e^{-1}\frac{1}{0!} + e^{-1}\frac{1}{1!} + e^{-1}\frac{1}{2!} = \frac{2,5}{e},$$

а $P(A) = 1 - P(\overline{A}) = 1 - \frac{2,5}{e} \approx 0,08$.

Вправи.

4. Ймовірність подзвонити абоненту на телефонну станцію протягом хвилини дорівнює 0,005. Телефонна станція обслуговує 600 абонентів. Знайти ймовірність того, що протягом хвилини на станцію подзвонить принаймні один абонент.
5. Ймовірність виходу із ладу приладу за час випробування $p = 0,002$. Знайти ймовірність того, що із 500 приладів вийдуть з ладу за час випробування: а) рівно 5 приладів; б) менше 5 приладів.

2.2.2. Локальна гранична теорема Муавра-Лапласа. При великих n обчислення ймовірностей $P_n(k)$ досить громіздке, тому необхідно мати асимптотичні формули, за якими ці ймовірності можна знаходити наближено. Одну із таких наближених формул ми уже розглянули. Але, якщо ймовірність не є близькою до нуля або одиниці, то теорема Пуассона дає велику похибку. В цьому випадку вигідно застосовувати наступну теорему.

Теорема. Якщо ймовірність настання події A в кожному з n незалежних випробувань є сталою і дорівнює p, відмінною від 0 і 1, то рівномірно по k, для яких $x = \frac{k - np}{\sqrt{npq}}$ міститься в сталих межах $(-\infty < a \le x \le b < +\infty)$, має місце співвідношення

$$\frac{\sqrt{npq}P_n(k)}{\varphi(x)} \to 1, \quad n \to \infty, \tag{4}$$

де $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – функція Гауса.

Доведення. Використаємо формулу Стірлінга: $n! = \sqrt{2\pi n}\, n^n e^{-n} e^{\theta}$, $0 < \theta < \frac{1}{12n}$. Тоді із формули Бернуллі одержуємо:

$$\sqrt{npq}P_n(k)=\sqrt{npq}\,\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}=$$

$$=\sqrt{npq}\,\frac{\sqrt{2\pi n}n^n e^{-n}e^{\theta_1}}{\sqrt{2\pi k}k^k e^{-k}e^{\theta_2}\sqrt{2\pi(n-k)}\,(n-k)^{n-k}\,e^{-(n-k)}e^{\theta_3}}\,p^k(1-p)^{n-k}=$$

$$=\sqrt{\frac{n^2pq}{2\pi k(n-k)}}\left(\frac{n}{k}p\right)^k\left(\frac{n}{n-k}q\right)^{n-k}e^{\theta}, \qquad (5)$$

де $\theta=\theta_1-\theta_2-\theta_3$.

Із визначення x випливає, що

$$k=np+x\sqrt{npq},\quad n-k=nq-x\sqrt{npq}, \qquad (6)$$

тому при $n\to\infty$ $k\to\infty,\ n-k\to\infty$. Оскільки $a\le x\le b$, то $k\ge np+a\sqrt{npq}$, $n-k\ge nq-b\sqrt{npq}$, тому

$$|\theta|\le|\theta_1|+|\theta_2|+|\theta_3|\le\frac{1}{12n}+\frac{1}{12k}+\frac{1}{12(n-k)}\le\frac{1}{12n}+\frac{1}{12\left(np+a\sqrt{npq}\right)}+\frac{1}{12\left(nq-b\sqrt{npq}\right)}\to 0,$$

а

$$e^{\theta}\to 1 \qquad (7)$$

при $n\to\infty$ рівномірно по k, для яких $a\le x\le b$.

Враховуючи (6), для першого множника в (5) одержимо:

$$\sqrt{\frac{n^2pq}{2\pi k(n-k)}}=\sqrt{\frac{n^2pq}{2\pi\left(np+x\sqrt{npq}\right)\left(nq-x\sqrt{npq}\right)}}=\frac{1}{\sqrt{2\pi}}\sqrt{\frac{1}{\left(1+x\sqrt{\frac{q}{np}}\right)\left(1-x\sqrt{\frac{p}{nq}}\right)}}\to\frac{1}{\sqrt{2\pi}} \qquad (8)$$

при $n\to\infty$ рівномірно по k, для яких $a\le x\le b$.

Позначимо $A_n=\left(\frac{n}{k}p\right)^k\left(\frac{n}{n-k}q\right)^{n-k}$, тоді із (6)

$$\ln A_n=k\ln\frac{np}{k}+(n-k)\ln\frac{nq}{n-k}=$$

$$=-\left(np+x\sqrt{npq}\right)\ln\left(1+x\sqrt{\frac{q}{np}}\right)-\left(nq-x\sqrt{npq}\right)\ln\left(1-x\sqrt{\frac{p}{nq}}\right).$$

Використовуючи розклад функції $\ln(1+u)$ за формулою Тейлора і враховуючи, що величини $x\sqrt{\frac{q}{np}}$ і $x\sqrt{\frac{p}{nq}}$ прямують до нуля при $n\to\infty$ рівномірно по k, для яких $a\le x\le b$, то при $n\to\infty$

$$\ln\left(1+x\sqrt{\frac{q}{np}}\right)=x\sqrt{\frac{q}{np}}-\frac{x^2}{2}\frac{q}{np}+O\left(n^{-\frac{3}{2}}\right),$$

$$\ln\left(1-x\sqrt{\frac{p}{nq}}\right)=-x\sqrt{\frac{p}{nq}}-\frac{x^2}{2}\frac{p}{nq}+O\left(n^{-\frac{3}{2}}\right),$$

де оцінки залишкових членів рівномірні по x в будь-якому скінченному проміжку. Тому

$$\ln A_n=-\left(np+x\sqrt{npq}\right)\left(x\sqrt{\frac{q}{np}}-\frac{x^2}{2}\frac{q}{np}+O\left(n^{-\frac{3}{2}}\right)\right)-\left(nq-x\sqrt{npq}\right)\left(-x\sqrt{\frac{p}{nq}}-\frac{x^2}{2}\frac{p}{nq}+O\left(n^{-\frac{3}{2}}\right)\right)=$$

$$=-x\sqrt{npq}+\frac{x^2}{2}q+O\left(\frac{1}{\sqrt{n}}\right)-x^2q+\frac{x^3}{2}\frac{q^{\frac{3}{2}}}{\sqrt{np}}+O\left(\frac{1}{n}\right)+x\sqrt{npq}+\frac{x^2}{2}p+O\left(\frac{1}{\sqrt{n}}\right)-x^2p-\frac{x^3}{2}\frac{p^{\frac{3}{2}}}{\sqrt{nq}}+O\left(\frac{1}{n}\right)=$$

$$= -\frac{x^2}{2} + O\left(\frac{1}{\sqrt{n}}\right),$$

а це означає, що

$$A_n : e^{-\frac{x^2}{2}} \to 1 \qquad (9)$$

при $n \to \infty$ рівномірно по *k*, для яких $a \le x \le b$.

Із (5), (7), (8), (9) випливає (4). Теорема доведена.

Практичне використання доведеної теореми основане на наближеній формулі

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi(x), \qquad (10)$$

де $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – функція Гауса, а $x = \frac{k - np}{\sqrt{npq}}$. Значення функції Гауса можна знайти за таблицями (таблиця 1), які містяться у більшості підручників з теорії ймовірностей.

Приклад 3. Ймовірність настання події в кожному із 100 незалежних випробувань дорівнює 0,8. Знайти ймовірність того, що подія настане 90 разів.

Розв'язування. Використаємо локальну теорему Муавра-Лапласа в якій покладемо: n = 100, p = 0,8, q = 0,2, k = 90. Тоді $x = \frac{90 - 100 \cdot 0,8}{\sqrt{100 \cdot 0,8 \cdot 0,2}} = 2,5$, а за таблицею 1

$\varphi(2,5) = 0,0175$. Отже, із (10) одержуємо $P_{100}(90) \approx \frac{1}{\sqrt{100 \cdot 0,8 \cdot 0,2}} 0,0175 \approx 0,0044$.

Вправи.

6. Гральний кубик підкидають 120 разів. Знайти найімовірніше число випадання 5 очок та його ймовірність.
7. В коло вписано квадрат. Оцінити ймовірність того, що із 100 випадково кинутих точок в круг рівно 75 попадуть в квадрат.

2.2.3. Інтегральна теорема Муавра-Лапласа. При розв'язуванні задач часто виникає необхідність знаходити ймовірності того, що в n випробуваннях Бернуллі подія настане від k_1 до k_2 раз. За локальною граничною теоремою це робити незручно, а вигідно застосовувати наступну теорему.

Теорема. Нехай μ_n – число появ події в n незалежних випробуваннях в кожному із яких ймовірність її настання є сталою і дорівнює p ($0 < p < 1$), то рівномірно відносно a і b $\left(-\infty \le a \le b \le +\infty\right)$ справедливе співвідношення

$$\lim_{n\to\infty} P\left\{ a \le \frac{\mu_n - np}{\sqrt{npq}} < b \right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx. \qquad (11)$$

Доведення. Нехай $x_k = \frac{k - np}{\sqrt{npq}}$, $P_n(a,b) = P\left\{ a \le \frac{\mu_n - np}{\sqrt{npq}} < b \right\}$. Тоді

$$P_n(a,b) = \sum_{k: a \le x_k < b} P_n(k). \qquad (12)$$

Розглянемо функцію

$$\Pi_n(x) = \begin{cases} 0, \quad при \quad x < x_0 = -\dfrac{np}{\sqrt{npq}}, \\ \sqrt{npq}P_n(k), \quad при \quad x_k \le x < x_{k+1}, \quad k = 0,1,...,n, \\ 0, \quad при \quad x \ge x_{n+1} = x_n + \dfrac{1}{\sqrt{npq}}. \end{cases}$$

Оскільки функція $\Pi_n(x)$ є сталою на кожному із проміжків $[x_k, x_{k+1})$, то $\int\limits_{x_k}^{x_{k+1}} \Pi_n(x)dx = \sqrt{npq}P_n(k)(x_{k+1} - x_k) = P_n(k)$. Отже, із (12)

$$P_n(a,b) = \sum_{k=\underline{m}}^{\overline{m}-1} P_n(k) = \int\limits_{x_{\underline{m}}}^{x_{\overline{m}}} \Pi_n(x)dx, \tag{13}$$

де $\underline{m}$ і $\overline{m}$ – числа, які визначаються із нерівностей

$$a \le x_{\underline{m}} < a + \frac{1}{\sqrt{npq}}, \quad b \le x_{\overline{m}} < b + \frac{1}{\sqrt{npq}}. \tag{14}$$

Представимо (13) у вигляді

$$P_n(a,b) = \int\limits_a^b \Pi_n(x)dx + \rho_n, \tag{15}$$

де

$$\rho_n = \int\limits_b^{x_{\overline{m}}} \Pi_n(x)dx - \int\limits_a^{x_{\underline{m}}} \Pi_n(x)dx.$$

Із визначення функції $\Pi_n(x)$ випливає, що вона досягає найбільшого значення на проміжку $-\dfrac{1}{\sqrt{npq}} \le x_{k_0} = \dfrac{k_0 - np}{\sqrt{npq}} \le x < x_{k_0+1} = \dfrac{k_0 + 1 - np}{\sqrt{npq}} \le \dfrac{2}{\sqrt{npq}}$, де $k_0 = \left[p\left(n+1\right)\right]$ – найімовірніше число настання події. Оскільки проміжок зміни x скінченний, то за локальною граничною теоремою $\Pi_n(x) : \varphi(x_{k_0}) \to 1$ при $n \to \infty$. Це означає, що при великих n

$$\max \Pi_n(x) \le 2\frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x_{k_0}^2}{2}\right) \le \frac{2}{\sqrt{2\pi}}.$$

Тому із (14) одержуємо, що

$$|\rho_n| \le \int\limits_b^{x_{\overline{m}}} \Pi_n(x)dx + \int\limits_a^{x_{\underline{m}}} \Pi_n(x)dx \le \sqrt{\frac{2}{\pi}}(x_{\overline{m}} - b + x_{\underline{m}} - a) \le 2\sqrt{\frac{2}{\pi npq}},$$

тобто,

$$\rho_n \to 0 \tag{16}$$

при $n \to \infty$.

Нехай

$$\max(|a|,|b|) \le C, \tag{17}$$

де C – деяка стала. Тоді за локальною граничною теореммою при $a \le x_k < b$

$$\Pi_n(x_k) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x_k^2}{2}\right)\left(1 + \alpha_n\left(x_k\right)\right),$$

де $\alpha_n\left(x_k\right) \to 0$ при $n \to \infty$ рівномірно відносно x_k.

Нехай $x \in [x_k, x_{k+1})$. Тоді

$$\Pi_n(x)=\Pi_n(x_k)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x_k^2}{2}}\left(1+\alpha_n\left(x_k\right)\right)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}e^{\frac{x^2-x_k^2}{2}}\left(1+\alpha_n\left(x_k\right)\right).$$

Оскільки $\frac{\left|x^2-x_k^2\right|}{2}=\frac{\left(x-x_k\right)\left|x+x_k\right|}{2}\le\left(x-x_k\right)C\le\frac{1}{\sqrt{npq}}C\to 0$ при $n\to\infty$, то $\beta_n(x)=e^{\frac{x^2-x_k^2}{2}}(1+\alpha_n(x_k))-1\to 0$ при $n\to\infty$ рівномірно відносно $x\in[a,b]$ і рівномірно відносно a і b , а $\Pi_n(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\left(1+\beta_n\left(x\right)\right)$ при $n\to\infty$ рівномірно збігається на проміжку $[a,b]$ до $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Отже, із (15) і (16) для випадку $\max(|a|,|b|)\le C$ випливає твердження (11).

Знімемо тепер обмеження (17).

Із рівності

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}dx=1$$

випливає, що для довільного $\varepsilon>0$ знайдеться таке C, що

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-C}e^{-\frac{x^2}{2}}dx+\frac{1}{\sqrt{2\pi}}\int_{C}^{+\infty}e^{-\frac{x^2}{2}}dx<\frac{\varepsilon}{8}. \tag{18}$$

Зафіксуємо таке число C. За доведеним одержимо, що знайдеться таке n_1, що для всіх $n>n_1$ $\left|P_n\left(-C,C\right)-\frac{1}{\sqrt{2\pi}}\int_{-C}^{C}e^{-\frac{x^2}{2}}dx\right|<\frac{\varepsilon}{8}$, тому $P_n\left(-C,C\right)>\frac{1}{\sqrt{2\pi}}\int_{-C}^{C}e^{-\frac{x^2}{2}}dx-\frac{\varepsilon}{8}>1-\frac{\varepsilon}{4}$, а

$$P_n\left(-\infty,-C\right)+P_n\left(C,\infty\right)<\frac{\varepsilon}{4} \tag{19}$$

для $n>n_1$.

Розглянемо довільний проміжок $[a,b]$ і позначимо $[A,B]=[a,b]\bigcap[-C,C]$. Так як $-C\le A\le B\le C$, то, за доведеним, знайдеться таке n_2, що для всіх $n>n_2$

$$\left|P_n\left(A,B\right)-\frac{1}{\sqrt{2\pi}}\int_{A}^{B}e^{-\frac{x^2}{2}}dx\right|<\frac{\varepsilon}{2}. \tag{20}$$

Із нерівності

$$\left|P_n\left(a,b\right)-\frac{1}{\sqrt{2\pi}}\int_{a}^{b}e^{-\frac{x^2}{2}}dx\right|\le\left|P_n\left(A,B\right)-\frac{1}{\sqrt{2\pi}}\int_{A}^{B}e^{-\frac{x^2}{2}}dx\right|+$$

$$+P_n\left(-\infty,-C\right)+P_n\left(C,\infty\right)+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-C}e^{-\frac{x^2}{2}}dx+\frac{1}{\sqrt{2\pi}}\int_{C}^{+\infty}e^{-\frac{x^2}{2}}dx$$

одержуємо, врахувавши (18) – (20), для $n>\max(n_1,n_2)$

$$\left|P_n\left(a,b\right)-\frac{1}{\sqrt{2\pi}}\int_{a}^{b}e^{-\frac{x^2}{2}}dx\right|<\varepsilon$$

для всіх a і b. Теорема доведена.

Із теореми випливає, що при великих n має місце наближена рівність

$$P\left\{a\le\frac{\mu_n-np}{\sqrt{npq}}<b\right\}\approx\frac{1}{\sqrt{2\pi}}\int_{a}^{b}e^{-\frac{x^2}{2}}dx. \tag{21}$$

Нехай $\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$ – функція Лапласа, значення якої можна знайти за таблицею 2. Таблиця складена для значень $x\in[0,5]$, а при $x>5$ покладають $\Phi_0(x)=0,5$, крім того враховуємо, що $\Phi_0(-x)=-\Phi_0(x)$. Використовуючи рівність

$$\frac{1}{\sqrt{2\pi}}\int\limits_a^b e^{-\frac{x^2}{2}}dx=\Phi_0(b)-\Phi_0(a),$$

(21) можна записати у вигляді

$$P\left\{a\le\frac{\mu_n-np}{\sqrt{npq}}<b\right\}\approx\Phi_0(b)-\Phi_0(a). \qquad (22)$$

Позначимо через $P_n(k_1,k_2)$ – ймовірність того, що в n випробуваннях Бернуллі подія настане від k_1 до k_2 раз. Тоді із (21) і (22) одержуємо

$$P_n\left(k_1,k_2\right)=P\left\{k_1\le\mu_n\le k_2\right\}=P\left\{\frac{k_1-np}{\sqrt{npq}}\le\frac{\mu_n-np}{\sqrt{npq}}\le\frac{k_2-np}{\sqrt{npq}}\right\}\approx\frac{1}{\sqrt{2\pi}}\int\limits_{x_1}^{x_2}e^{-\frac{x^2}{2}}dx$$

або

$$P_n(k_1,k_2)\approx\Phi_0(x_2)-\Phi_0(x_1), \qquad (23)$$

де $x_i=\frac{k_i-np}{\sqrt{npq}}\quad(i=1,2)$.

Приклад 4. Ймовірність настання події в кожному із 100 незалежних випробувань дорівнює 0,8. Знайти ймовірність того, що подія настане не менше 80 і не більше 90 разів.

Розв'язування. Використаємо інтегральну теорему Муавра–Лапласа, в якій покладемо n = 100, p = 0,8, q = 0,2, k_1 = 80, k_2 = 90. Тоді $x_1=\frac{80-100\cdot0,8}{\sqrt{100\cdot0,8\cdot0,2}}=0$, $x_2=\frac{90-100\cdot0,8}{\sqrt{100\cdot0,8\cdot0,2}}=2,5$. За таблицею 2 $\Phi_0(0)$ = 0, $\Phi_0(2,5)$ = 0,4938. Отже, $P_{100}(80,90)\approx\Phi_0(2,5)-\Phi_0(0)=0,4938$.

Розглянемо ймовірність $P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}$, де μ_n – число настання події A в n експериментах, $\frac{\mu_n}{n}$ – відносна частота події, $\varepsilon>0$ – довільне число, а $p=P\left(A\right)$. Таким чином, ми розглядаємо ймовірність того, що відносна частота події відрізняється від ймовірності події за модулем менше, ніж на ε. Тоді

$$P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}=P\left\{-n\varepsilon<\mu_n-np<n\varepsilon\right\}=P\left\{-\varepsilon\sqrt{\frac{n}{pq}}<\frac{\mu_n-np}{\sqrt{npq}}<\varepsilon\sqrt{\frac{n}{pq}}\right\}. \qquad (24)$$

При $n\to\infty$ за інтегральною теоремою Муавра-Лапласа права частина рівності (24) прямує до $\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}e^{-\frac{x^2}{2}}dx=2\Phi_0(+\infty)=1$, тому

$$\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}=1.$$

Цю рівність називають теоремою Бернуллі, яка є найпростішою формою закону великих чисел. Дана теорема стверджує, що при великих n, з ймовірністю як завгодно близькою до 1, виконується наближена рівність $\frac{\mu_n}{n}\approx p$.

Покладаючи в рівності (22) $a=-\varepsilon\sqrt{\frac{n}{pq}}$, $b=\varepsilon\sqrt{\frac{n}{pq}}$, враховуючи непарність функції $\Phi_0(x)$, із рівності (24) одержимо

$$\beta=P\left\{\left|\frac{\mu_n}{n}-p\right|\le\varepsilon\right\}\approx 2\Phi_0\left(\varepsilon\sqrt{\frac{n}{pq}}\right). \tag{25}$$

Приклад 5. Ймовірність настання події в кожному із 10000 незалежних випробувань p = 0,75. Знайти ймовірність того, що відносна частота настання події відхилиться від її ймовірності за абсолютною величиною не більше, ніж на 0,01.

Розв'язування. Використаємо формулу (25). За умовою n = 10000, p = 0,75, q = 0,25, $\varepsilon=0,01$. Тоді $\beta\approx 2\Phi_0\left(0,01\sqrt{\frac{10000}{0,75\cdot 0,25}}\right)=2\Phi_0(2,31)$. За таблицею 2 знаходимо $\Phi_0(2{,}31)$ = 0,4893. Отже, $\beta\approx 2\Phi_0(2,31)\approx 0,979$.

Приклад 6. Ймовірність настання події в кожному із незалежних випробувань p = 0,9. Скільки необхідно провести випробувань, щоб з ймовірністю 0,95 можна було сподіватись, що відносна частота настання події відхилиться від її ймовірності за абсолютною величиною не більше, ніж на 0,01?

Розв'язування. Використаємо формулу (25):

$$\beta\approx 2\Phi_0\left(\varepsilon\sqrt{\frac{n}{pq}}\right).$$

За умовою β = 0,95, p = 0,9, q = 0,1, $\varepsilon=0,01$. Для знаходження n одержуємо рівняння $0,95=2\Phi_0\left(0,01\sqrt{\frac{n}{0,9\cdot 0,1}}\right)$ або $\Phi_0\left(\frac{\sqrt{n}}{30}\right)=0,475$. За таблицею 2 одержуємо $\frac{\sqrt{n}}{30}=1,96$. Тому $\sqrt{n}=30\cdot 1,96=58,8$, а n = 3458.

За допомогою формули (25) можна оцінити ймовірність по відносній частоті. Нехай проведена деяка серія із n експериментів, при цьому подія A настала k разів. Необхідно знайти межі, в яких із заданою ймовірністю β міститься невідома ймовірність $p=P(A)$. Для цього із рівності

$$\beta\approx 2\Phi_0\left(\varepsilon\sqrt{\frac{n}{pq}}\right)$$

необхідно знайти ε. Позначимо

$$t_\beta=\varepsilon\sqrt{\frac{n}{pq}}, \tag{26}$$

тоді за таблицею 2 із рівняння $\beta\approx 2\Phi_0\left(t_\beta\right)$ знаходимо t_β. Тому із (26) одержуємо

$$\varepsilon=t_\beta\sqrt{\frac{pq}{n}}.$$

Отже, із (25) випливає, що з ймовірністю β виконується нерівність

$$\left|\frac{k}{n}-p\right|<t_\beta\sqrt{\frac{pq}{n}}. \tag{27}$$

Якщо врахуємо, що $q=1-p$, і розв'яжемо нерівність (27) відносно p, то одержимо шукані межі для ймовірності p. Результат досить громіздкий. Для його спрощення використаємо, що $pq\le\frac{1}{4}$, тоді із (27) випливає нерівність $\left|\frac{k}{n}-p\right|<t_\beta\frac{1}{2\sqrt{n}}$, звідки одержуємо більш прості межі для p:

$$\frac{k}{n} - \frac{t_\beta}{2\sqrt{n}} < p < \frac{k}{n} + \frac{t_\beta}{2\sqrt{n}}.$$

Вправи.

8. Серед виготовлених заводом виробів 80% стандартних. Знайти ймовірність того, що серед 900 виробів стандартних буде не менше 700.
9. Підприємство випускає 75% першосортних виробів. Знайти ймовірність того, що із 300 вибраних виробів першосортними виявляться: а) рівно 200; б) від 219 до 234.
10. Скільки пострілів необхідно здійснити, щоб з надійністю $\beta = 0,995$ відхилення відносної частоти влучень від ймовірності $p = \frac{3}{8}$ за модулем не перевищувало 0,01?
11. Перевіряється 500 виробів, серед яких 90% стандартних. Знайти з ймовірністю 0,95 межі, в яких міститься число стандартних виробів серед перевірених.
12. Проведено $n = 2500$ незалежних дослідів, у яких подія A настала $k = 1250$ разів. Знайти межі для ймовірності $p = P(A)$, які можна гарантувати з ймовірністю $\beta = 0,9549$.

Розділ 3. Випадкові величини та їх розподіли

3.1. Випадкові величини. Функція розподілу випадкової величини

Розглядаючи випадкові експерименти, дотепер нас цікавили їх можливі наслідки. Але із можливими наслідками експерименту можна пов'язати певну величину, яка в залежності від наслідку експерименту приймає деякі числові значення. Наприклад, при киданні грального кубика нас цікавить число очок, які є на грані. При контролі якості продукції нас може цікавити число бракованих виробів серед випадково взятих n виробів. При проведенні пострілів по мішені кожній точці попадання ставимо у відповідність величину, яка дорівнює віддалі від цієї точки до центра мішені. Тобто, з кожним експериментом ми пов'язуємо величину, яка в залежності від випадку приймає деякі числові значення. Таку величину називають *випадковою*. Якщо в експерименті спостерігається певна випадкова величина ξ, то при виборі будь-якого ω із Ω повинно бути задане значення цієї величини $\xi(\omega)$. Отже, випадкову величину можна розглядати, як функцію визначену на просторі елементарних подій Ω. Ця функція повинна задовольняти деяким додатковим умовам. Для цього ми спочатку визначимо поняття борелєвої $\sigma-$алгебри.

Розглянемо на множині дійсних чисел R клас всіх напівінтервалів $[a,b)$, де $-\infty \le a < b \le \infty$. Тоді мінімальна $\sigma-$алгебра, яка містить такий клас множин, називається борелєвою. Будемо її позначати σ_R. Будь-який елемент борелєвої $\sigma-$алгебри називають борелєвою множиною. Зокрема, борелєвими є інтервали, відрізки, одноточкові множини, всі відкриті та замкнені множини. Аналогічно, мінімальна $\sigma-$алгебра, яка породжена класом всіх напіввідкритих паралелепіпедів із R^n, називають борелєвою $\sigma-$алгеброю підмножин із R^n. Функція $g: R^k \to R^n$ називається борелєвою, якщо прообраз будь-якої борелєвої множини в R^n є борелєвою множиною в R^k. Неперервні функції є борелєвими.

Означення 1. Нехай задано ймовірнісний простір $\{\Omega, \mathfrak{A}, P\}$. *Випадковою величиною* називається будь-яка дійсна функція $\xi = \xi(\omega)$, що відображає простір елементарних подій Ω в множину дійсних чисел R, для якої прообраз будь-якої борелєвої множини $B \in \sigma_R$ є множина із $\sigma-$алгебри $\mathfrak{A}$: $\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathfrak{A}$.

В теорії функцій такі функції називають вимірними відносно $\sigma-$алгебри $\mathfrak{A}$. Із властивостей вимірних функцій випливає, що в означенні достатньо було вимагати, щоб для довільного дійсного x прообраз півінтервалу $(-\infty, x)$ належав до $\mathfrak{A}$.

Сформулюємо деякі висновки, що випливають із властивостей вимірних функцій. Звідси, зокрема, одержуємо, що множини $\{\omega : \xi(\omega) \le x\}$, $\{\omega : \xi(\omega) \ge x\}$, $\{\omega : \xi(\omega) > x\}$, $\{\omega : a \le \xi(\omega) < b\}$, $\{\omega : a < \xi(\omega) < b\}$, $\{\omega : \xi(\omega) = x\}$ є випадковими подіями.

Якщо випадкові величини $\xi(\omega)$ і $\eta(\omega)$ визначені на одному і тому ж ймовірнісному просторі, то $\{\omega : \xi(\omega) < \eta(\omega)\}$, $\{\omega : \xi(\omega) \le \eta(\omega)\}$, $\{\omega : \xi(\omega) = \eta(\omega)\}$ є також випадковими подіями.

Ми будемо також розглядати і події, що пов'язані із нескінченними послідовностями випадкових величин. Нехай $\xi_n(\omega)$ - послідовність випадкових величин, що задані на ймовірнісному просторі $\{\Omega, \mathfrak{A}, P\}$, і $\xi(\omega)$ - випадкова величина на $\{\Omega, \mathfrak{A}, P\}$. Тоді $\{\omega : \lim\limits_{n\to\infty} \xi_n(\omega) \text{ існує}\} \in \mathfrak{A}$, $\{\omega : \lim\limits_{n\to\infty} \xi_n(\omega) = \xi(\omega)\} \in \mathfrak{A}$.

Із наведених фактів випливають наступні твердження.

Нехай $f:R^m \to R$ борелєва функція, а $\xi_1(\omega),\ldots,\xi_m(\omega)$ - випадкові величини на $\{\Omega,\mathfrak{A},P\}$, тоді $f(\xi_1(\omega),\ldots,\xi_m(\omega))$ є випадковою величиною.

Нехай $\xi_n(\omega)$ – послідовність випадкових величин, що задані на ймовірнісному просторі $\{\Omega,\mathfrak{A},P\}$, тоді $\lim\limits_{n\to\infty}\xi_n(\omega)$, $\underline{\lim}\,\xi_n(\omega)$, $\overline{\lim}\,\xi_n(\omega)$, $\sup\limits_n \xi_n(\omega)$, $\inf\limits_n \xi_n(\omega)$ є випадковими величинами (можливо невласними – це такі функції $\xi(\omega)$, які можуть приймати значення $\pm\infty$).

В позначеннях випадкової величини залежність від елементарної події часто не буде вказуватись: замість $\xi(\omega)$ будемо писати ξ.

Нехай задана випадкова величина ξ. Визначимо на сігма-алгебрі σ_R ймовірність $P_\xi(\cdot)$ за допомогою рівності $P_\xi(B)=P\{\omega:\xi(\omega)\in B\}=P\{\xi^{-1}(B)\}$. Можна показати, що вона задовольняє всім аксіомам ймовірності. Тому можемо розглядати новий ймовірнісний простір $\{R,\sigma_R,P_\xi\}$, що породжений випадковою величиною ξ.

Означення 2. Ймовірність $P_\xi(B)$ називається *розподілом випадкової величини* ξ.

Означення 3. Нехай $B=(-\infty,x),\,x\in R$. Тоді функція $F_\xi(x)=P_\xi((-\infty,x))$ називається *функцією розподілу випадкової величини* ξ. Тобто, функція, яка для кожного дійсного x визначається рівністю

$$F_\xi(x)=P\{\omega:\xi(\omega)<x\},$$

називається *функцією розподілу випадкової величини* ξ.

Функція розподілу випадкової величини повністю визначає її розподіл.

Там, де це не викличе непорозуміння, замість $F_\xi(x)$ писатимемо $F(x)$. Отже,

$$F(x)=P\{\xi<x\}.$$

Як приклад, розглянемо *рівномірний розподіл на відрізку* $[a,b]$. На рідрізок $[a,b]$ випадково кидається точка, причому вважаємо, що всі положення точки рівноможливі (ймовірність попадання точки на деяку частину відрізка пропорційна мірі цієї частини). Простір елементарних подій $\Omega=[a,b]$, $\mathfrak{A}$ є σ-алгебра борелєвих підмножин даного відрізка, $P(.)$ – така міра, що $P\{[\alpha,\beta]\}=\dfrac{\beta-\alpha}{b-a}$, якщо $[\alpha,\beta]\subset[a,b]$. Визначимо випадкову величину $\xi(\omega)=\omega$, якщо $\omega\in[a,b]$, тобто ξ - координата вибраної точки. Тоді для довільного x

$$\{\omega:\xi(\omega)<x\}=\begin{cases}\varnothing, & x\le a;\\ [a,x), & a<x\le b;\\ \Omega, & x>b.\end{cases}$$

Тому $\{\omega:\xi(\omega)<x\}\in\mathfrak{A}$ і $\xi(\omega)$ – випадкова величина, а її функція розподілу має вигляд:

$$F_\xi(x)=\begin{cases}0, & x\le a;\\ \dfrac{x-a}{b-a}, & a<x\le b;\\ 1, & x>b.\end{cases}$$

Відзначимо, що для різних випадкових величин їх функції розподілу можуть співпадати. Наприклад, нехай $\Omega=\{\omega_1,\omega_2\}$, $P(\{\omega_1\})=p_1$, $P(\{\omega_2\})=p_2$, $p_1+p_2=1$, $\xi(\omega)$ – випадкова величина, що визначена на $\Omega:\xi(\omega)=\begin{cases}-1, \omega=\omega_1;\\ 1, \omega=\omega_2.\end{cases}$ Тоді

$$\{\omega : \xi(\omega) < x\} = \begin{cases} \varnothing, & x \le -1; \\ \{\omega_1\}, & -1 < x \le 1; \\ \Omega, & x > 1. \end{cases}$$

Із визначення функції розподілу одержуємо:

$$F_\xi(x) = \begin{cases} 0, & x \le -1; \\ p_1, & -1 < x \le 1; \\ 1, & x > 1. \end{cases}$$

Розглянемо нову випадкову величину $\eta(\omega) = -\xi(\omega) = \begin{cases} -1, \omega = \omega_2; \\ 1, \omega = \omega_1. \end{cases}$ Тоді, очевидно,

$$F_\eta(x) = \begin{cases} 0, & x \le -1; \\ p_2, & -1 < x \le 1; \\ 1, & x > 1. \end{cases}$$

Якщо $p_1 = p_2 = 0,5$, то $F_\xi(x) = F_\eta(x)$. Отже, різні випадкові величини можуть мати однакові функції розподілу.

Вправи.

1. В круг із центром в початку координат і радіусом R випадково кидають точку. Нехай ξ - віддаль від даної точки до початку координат. Знайти функцію розподілу випадкової величини ξ.
2. Нехай $\{\Omega, \mathfrak{A}, P\}$ - довільний ймовірнісний простір, A - деяка випадкова подія. Розглянемо випадкову величину $I_A(\omega) = \begin{cases} 1, & \omega \in A; \\ 0, & \omega \notin A, \end{cases}$ яку називають індикатором випадкової події A. Знайти функцію розподілу даної випадкової величини.

3.2. Властивості функції розподілу

1. $0 \le F(x) \le 1$, бо $F(x) = P\{\xi < x\}$, а ймовірність міститься в межах від 0 до 1.

2. $F(x)$ - неспадна: для довільних $x_1 < x_2$, $F(x_1) \le F(x_2)$.

Дійсно, $\{\xi < x_2\} = \{\xi < x_1\} + \{x_1 \le \xi < x_2\}$ і події в правій частині несумісні, тому $F(x_2) = P\{\xi < x_2\} = P\{\xi < x_1\} + P\{x_1 \le \xi < x_2\} = F(x_1) + P\{x_1 \le \xi < x_2\}$. Оскільки $P\{x_1 \le \xi < x_2\} \ge 0$, то із попередньої рівності випливає, що $F(x_1) \le F(x_2)$. Одночасно одержуємо такий

Наслідок. $P\{x_1 \le \xi < x_2\} = F(x_2) - F(x_1)$.

3. Функція розподілу неперервна зліва, тобто, $F(x-0) = F(x)$.

Для доведення розглянемо таку зростаючу послідовність x_n, що $x_n \to x$. Тоді послідовність подій $A_n = \{\xi < x_n\}$ є зростаючою і $\bigcup_{n=1}^{\infty} A_n = \{\xi < x\}$. На основі аксіоми неперервності $F(x) = P\{\xi < x\} = \lim_{n\to\infty} P\{A_n\} = \lim_{n\to\infty} P\{\xi < x_n\} = \lim_{n\to\infty} F(x_n)$, а це означає, при врахуванні другої властивості, неперервність зліва.

4. $\lim_{x\to+\infty} F(x) = F(+\infty) = 1$, $\lim_{x\to-\infty} F(x) = F(-\infty) = 0$.

Розглянемо дві числові послідовності x_n і y_n: x_n – зростаюча і $x_n \to +\infty$, а y_n – спадна і $y_n \to -\infty$. Тому послідовність подій $A_n = \{\xi < x_n\}$ є зростаюча, а $B_n = \{\xi < y_n\}$ – спадна послідовність подій, при цьому, $\bigcup_{n=1}^{\infty} A_n = \Omega$, а $\bigcap_{n=1}^{\infty} B_n = \varnothing$. За аксіомою неперервності $1 = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} F(x_n)$ і $0 = P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} P(B_n) = \lim_{n\to\infty} F(y_n)$. Враховуючи монотонність функції розподілу, це і доводить дану властивість.

5. $P\{\xi = x\} = F(x+0) - F(x)$.

Розглянемо подію $A_n = \left\{x \le \xi < x + \frac{1}{n}\right\}$, тоді $P(A_n) = F\left(x + \frac{1}{n}\right) - F(x)$ і $\bigcap_{n=1}^{\infty} A_n = \{\xi = x\}$.

Тому $P\{\xi = x\} = P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty}\left(F\left(x + \frac{1}{n}\right) - F(x)\right) = F(x+0) - F(x)$.

Наслідок. $F(x)$ є неперервною в точці x тоді і тільки тоді, коли $P\{\xi = x\} = 0$.

6. $P\{\xi \le x\} = F(x+0)$.

Оскільки події $\{\xi < x\}$ і $\{\xi = x\}$ – несумісні і $\{\xi < x\} \cup \{\xi = x\} = \{\xi \le x\}$, тому $P\{\xi < x\} + P\{\xi = x\} = P\{\xi \le x\}$. Звідси, із врахуванням властивості 5, випливає справедливість даної властивості.

7. Множина точок розриву функції розподілу не більш як зліченна.

Точок розриву, у яких величина стрибка не менше 1, не більше 1;…; точок розриву, у яких величина стрибка не менше $\frac{1}{n}$, не більше n;…. Об'єднання не більше як зліченної сукупності не більше як зліченних множин є не більше як зліченна множина.

Вправи.

3. Задано функцію розподілу випадкової величини $F_\xi(x) = \begin{cases} 0, & x \le -1; \\ p, & -1 < x \le 1; \\ 1, & x > 1. \end{cases}$ Знайти ймовірність того, що випадкова величина приймає значення із проміжку $(0;2)$.
4. Довести, що величина ξ може набувати значення x_0 з додатною ймовірністю, якщо її функція розподілу має розрив у цій точці.
5. Якщо деяка функція $F(x)$ задовольняє властивостям 1 – 4, то існує випадкова величина, для якої дана функція буде функцією розподілу. Довести це.
6. Випадкова величина ξ задана функцією розподілу $F(x) = \begin{cases} 0, & x \le 0; \\ x^2, & 0 < x \le 1; \\ 1, & x > 1. \end{cases}$ Знайти ймовірність того, що випадкова величина в чотирьох випробуваннях три рази прийме значення із проміжку $\left(\frac{1}{4}, \frac{1}{2}\right)$.

3.3. Дискретні випадкові величини

Означення. Випадкова величина ξ називається *дискретною***,** якщо множина її значень скінченна або зліченна.

Можливі значення дискретної випадкової величини характеризують її ще не повністю, а для повної ймовірнісної характеристики випадкової величини необхідно знати і ймовірності, з якими ці значення приймаються.

Якщо $\xi(\omega)$ – дискретна випадкова величина, що приймає значення $x_1,...,x_n,...$, то для кожного *n* визначена ймовірність

$$p_n = P\{\omega : \xi(\omega) = x_n\} . \qquad (1)$$

Набір ймовірностей (1) називають *розподілом* випадкової величини ξ.

Відповідність між можливими значеннями випадкової величини та ймовірностями, з якими ці значення приймаються, називається *законом розподілу* (*рядом розподілу*) дискретної випадкової величини.

Оскільки події $\{\xi = x_n\}$ утворюють повну групу, то $p_n \geq 0$ і

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

Закон розподілу дискретної випадкової величини зручно подавати у вигляді таблиці:

Значення ξ	x_1	x_2	$\cdots$	x_n	$\cdots$
Ймовірності	p_1	p_2	$\cdots$	p_n	$\cdots$

За законом розподілу дискретної випадкової величини можна знайти її функцію розподілу. Якщо $x \leq x_1$, то подія $\{\xi < x\} = \varnothing$, тому $F(x) = 0$. Якщо $x_k < x \leq x_{k+1}$, то $\{\xi < x\} = \{\xi = x_1\} + \{\xi = x_2\} + \cdots + \{\xi = x_k\}$. Тому $F(x) = p_1 + p_2 + ... + p_k$. Отже,

$$F(x) = \sum_{i:x_i < x} p_i . \qquad (2)$$

Функція розподілу дискретної випадкової величини має скінченні розриви у кожній точці x_k і величина розриву, за властивістю 5, дорівнює $F(x_k + 0) - F(x_k) = p_k$. Між сусідніми точками розриву функція розподілу є сталою.

Розглянемо приклади важливих дискретних розподілів.

Біномний розподіл. Нехай проводиться n незалежних випробувань, в кожному із яких ймовірність настання події *A* дорівнює *p*. Біномним розподілом називається розподіл випадкової величини ξ, яка дорівнює числу настання події в n випробуваннях Бернуллі. Можливими значеннями випадкової величини ξ є числа *k* = 1, 2, … , *n*, а відповідні ймовірності знаходяться за формулою Бернуллі

$$p_k = P_n(k) = C_n^k p^k (1-p)^{n-k} .$$

Як ми уже відзначали, $\sum_{k=0}^{n} p_k = 1$.

Розподіл Пуассона. Випадкова величина ξ має розподіл Пуассона із параметром λ>0, якщо вона приймає значення 0, 1, … , *n*, …, а відповідні ймовірності знаходяться за формулою:

$$p_n = P(\xi = n) = e^{-\lambda} \frac{\lambda^n}{n!} .$$

Очевидно, $\sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} = 1$.

Розподіл Пуассона відіграє важливу роль в теорії випадкових процесів, теорії масового обслуговування, теорії надійності. Розподіл Пуассона є граничним для біномного розподілу.

Геометричний розподіл. Випадкова величина ξ має геометричний розподіл, якщо вона дорівнює числу експериментів до першої появи події в схемі Бернуллі (без експерименту, в якому подія настала).

Нехай $P(A)=p$, тоді множина $\Omega=\left\{A,\overline{A}A,...,\overline{A}....\overline{A}A,...\right\}$ може розглядатись як простір елементарних подій. Враховуючи незалежність подій, одержуємо: $P\Big(\underbrace{\overline{A}...\overline{A}}_{n}A\Big)=q^n p$.
Отже,

$$p_n=P(\xi=n)=q^n p \quad (n=0,1,2,...)$$

і

$$\sum_{n=0}^{\infty}q^n p=\frac{p}{1-q}=1.$$

Нехай $n\geq 0, \quad m\geq 0$, тоді $P\{\xi{=}n{+}m/\xi{\geq}n\}=P\{\xi{=}m\}$. Дійсно,

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

Наведена властивість геометричного розподілу називається властивістю відсутності післядії. Відзначимо, що серед всіх дискретних розподілів цю властивість має тільки геометричний розподіл.

Вправи.

7. Знайти функцію розподілу дискретної випадкової величини, що задана законом розподілу

Значення ξ	-1	0	2
Ймовірності	0,2	0,3	0,5

8. Нехай μ – число випадань герба при 4-х киданнях монети. Знайти закон розподілу випадкової величини μ та її функцію розподілу.
9. Стрілець, що має п’ять патронів, стріляє по мішені до першого попадання. Знайти закон розподілу числа зроблених пострілів, якщо ймовірність попадання в мішень при одному пострілі дорівнює *p*.
10. В партії із 8 деталей є 6 стандартних. Випадково вибирається три деталі. Знайти закон розподілу та функцію розподілу числа стандартних деталей із вибраних.

3.4. Абсолютно неперервні розподіли. Щільність розподілу випадкової величини

Означення. Нехай випадкова величина ξ має функцію розподілу $F(x)$. Якщо існує така інтегровна борелєва функція $p(x)$, що для всіх дійсних *x* виконується рівність

$$\int_{-\infty}^{x}p(u)du=F(x), \tag{3}$$

то розподіл називається *абсолютно неперервним*, а $p(x)$ називається *щільністю розподілу* випадкової величини.

Із означення випливає, що $F(x)$ - неперервна. Якщо x є точкою неперервності $p(x)$, тоді $F(x)$ диференційовна в точці *x* і $F'(x)=p(x)$. Так як $F(x)$ неспадна, то

$$p(x)\geq 0. \tag{4}$$

Із (3) і властивості 4 функції розподілу випливає, що

$$\int_{-\infty}^{+\infty}p(x)dx=1, \tag{5}$$

врахувавши рівність $P\{x_1\leq\xi<x_2\}=F(x_2)-F(x_1)$, одержуємо:

$$P\{x_1\leq\xi<x_2\}=\int_{x_1}^{x_2}p(x)dx. \tag{6}$$

Будь-яка функція $p(x)$, для якої мають місце властивості (4) і (5), є щільністю розподілу деякої випадкової величини.

Графік функції $y = p(x)$ називається *кривою розподілу*. Із (6) випливає, що ймовірність попадання випадкової величини у проміжок $[x_1, x_2)$ дорівнює площі фігури, що обмежена віссю абсцис, графіком кривої розподілу і прямими $x = x_1$, $x = x_2$.

Якщо щільність розподілу неперервна, то із рівності (6) при малих Δx

$$P\{x \le \xi < x + \Delta x\} = p(x)\Delta x + o(\Delta x).$$

Одержана рівність і наближена рівність $p(x) \approx \dfrac{P\{x \le \xi < x + \Delta x\}}{\Delta x}$, що із неї одержується, пояснюють ймовірнісний зміст щільності розподілу.

Відзначимо без доведення таке твердження.

Якщо випадкова величина ξ має щільність розподілу $p(x)$, то для довільної борелєвої множини B має місце рівність

$$P\{\omega : \xi(\omega) \in B\} = \int_B p(x)dx.$$

Ми розглянули два типи розподілів: дискретний і абсолютно неперервний. Але ними не вичерпуються всі типи розподілів. Сформулюємо теорему А. Лебега.

Теорема. Будь-яка функція розподілу $F(x)$ однозначно може бути представлена у вигляді суми

$$F(x) = p_1 F_Д(x) + p_2 F_{АН}(x) + p_3 F_C(x),$$

де $p_i \ge 0$, i=1, 2, 3, $p_1 + p_2 + p_3 = 1$, $F_Д(x)$ – функція розподілу дискретної випадкової величини, $F_{АН}(x)$ - абсолютно неперервна функція розподілу, $F_C(x)$ - сингулярна функція розподілу, що є неперервною функцією, множина точок росту якої має лебегову міру нуль.

Наведемо приклади найбільш важливих неперервних розподілів.

Рівномірний розподіл на відрізку $[a,b]$. На відрізок $[a,b]$ випадково кидається точка, причому вважаємо, що всі положення точки рівноможливі. Визначимо випадкову величину $\xi(\omega) = \omega$, якщо $\omega \in [a,b]$, тобто ξ – координата вибраної точки. В прикладі 1 ми показали, що функція розподілу випадкової величини $\xi(\omega)$ має вигляд:

$$F_\xi(x) = \begin{cases} 0, & x \le a; \\ \dfrac{x-a}{b-a}, & a < x \le b; \\ 1, & x > b. \end{cases}$$

Тому щільність розподілу випадкової величини $\xi(\omega)$

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & x \in (a,b); \\ 0, & x \notin (a,b). \end{cases}$$

Значення випадкової величини, що має рівномірний розподіл на відрізку $[0,1]$, одержане внаслідок експерименту, називають *випадковим числом*. На основі випадкових чисел із рівномірного розподілу можна моделювати випадкові величини із будь-яким наперед заданим розподілом.

Нормальний розподіл. Випадкова величина ξ має нормальний розподіл із параметрами a і σ^2, якщо її щільність розподілу

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

Позначаємо: $\xi \sim N\left(a,\sigma^2\right)$. Якщо $a=0$, а $\sigma=1$, то $p\left(x\right)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$ – функція Гауса, тоді розподіл називається стандартним.

Функція розподілу має вигляд:

$$F\left(x\right)=\frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{x}\exp\left(-\frac{\left(t-a\right)^2}{2\sigma^2}\right)dt.$$

Використовуючи інтеграл Пуассона, одержимо:

$$\int\limits_{-\infty}^{\infty}p\left(x\right)dx=\frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{\infty}\exp\left(-\frac{\left(x-a\right)^2}{2\sigma^2}\right)dx=\left|\frac{x-a}{\sigma}=t\right|=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}\exp\left(-\frac{t^2}{2}\right)dt=1.$$

Нехай $\Phi_0\left(x\right)=\frac{1}{\sqrt{2\pi}}\int\limits_{0}^{x}e^{-\frac{t^2}{2}}dt$ – функція Лапласа, $\Phi\left(x\right)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$ – функція розподілу стандартного нормального закону, тоді

$$P\left\{x_1\le\xi<x_2\right\}=\frac{1}{\sqrt{2\pi}\sigma}\int\limits_{x_1}^{x_2}\exp\left(-\frac{\left(x-a\right)^2}{2\sigma^2}\right)dx=\left|\frac{x-a}{\sigma}=t\right|=\frac{1}{\sqrt{2\pi}}\int\limits_{\frac{x_1-a}{\sigma}}^{\frac{x_2-a}{\sigma}}\exp\left(-\frac{t^2}{2}\right)dt=$$

$$=\Phi\left(\frac{x_2-a}{\sigma}\right)-\Phi\left(\frac{x_1-a}{\sigma}\right)=\Phi_0\left(\frac{x_2-a}{\sigma}\right)-\Phi_0\left(\frac{x_1-a}{\sigma}\right).$$

Тобто,

$$P\left\{x_1\le\xi<x_2\right\}=\Phi_0\left(\frac{x_2-a}{\sigma}\right)-\Phi_0\left(\frac{x_1-a}{\sigma}\right).$$

Як частинний випадок,

$$P\left\{\left|\xi-a\right|<\varepsilon\right\}=2\Phi_0\left(\frac{\varepsilon}{\sigma}\right).$$

Якщо у попередній формулі покласти $\varepsilon=3\sigma$, то $P\left\{\left|\xi-a\right|<3\sigma\right\}=2\Phi_0\left(3\right)\approx 0,997$. Тому нормально розподілена випадкова величина з ймовірністю досить близькою до одиниці приймає значення, що відрізняються від параметра *a* не більше, ніж на 3σ.

Якщо випадкова величина $\xi\sim N\left(a,\sigma^2\right)$, то $\eta=\frac{\xi-a}{\sigma}\sim N\left(0,1\right)$, тобто буде мати стандартний нормальний розподіл:

$$F_\eta\left(x\right)=P\left\{\eta<x\right\}=P\left\{\frac{\xi-a}{\sigma}<x\right\}=P\left\{\xi<a+x\sigma\right\}=\frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{x\sigma+a}\exp\left(-\frac{\left(t-a\right)^2}{2\sigma^2}\right)dt=\left|\frac{t-a}{\sigma}=z\right|=$$

$$=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}\exp\left(-\frac{z^2}{2}\right)dz=\Phi\left(x\right).$$

Показниковий розподіл. Випадкова величина ξ має показниковий розподіл із параметром $\lambda>0$, якщо

$$p\left(x\right)=\begin{cases}0, & x\le 0;\\ \lambda\exp\left(-\lambda x\right), & x>0.\end{cases}$$

Очевидно,

$$\int\limits_{-\infty}^{\infty} p(x)dx = \int\limits_{0}^{+\infty} \lambda \exp(-\lambda x)dx = -\exp(-\lambda x)\Big|_0^{+\infty} = 1.$$

Знайдемо функцію розподілу. Нехай $x \le 0$, тоді $F(x) = P\{\xi < x\} = \int\limits_{-\infty}^{x} p(t)dt = \int\limits_{-\infty}^{x} 0dt = 0$.

Якщо ж $x > 0$, то $F(x) = \int\limits_{-\infty}^{x} p(t)dt = \int\limits_{-\infty}^{0} p(t)dt + \int\limits_{0}^{x} p(t)dt = 1 - \exp(-\lambda x)$. Отже,

$$F(x) = \begin{cases} 1 - \exp(-\lambda x), & x > 0; \\ 0, \quad x \le 0. \end{cases}$$

Відзначимо, що показниковий розподіл має властивість відсутності післядії: при $0 < t < s$

$$P\{\xi < s / \xi > t\} = \frac{P\{t < \xi < s\}}{P\{\xi > t\}} =$$

$$= \frac{F(s) - F(t)}{1 - F(t)} = \frac{1 - e^{-\lambda s} - (1 - e^{-\lambda t})}{1 - (1 - e^{-\lambda t})} = 1 - e^{-\lambda(s-t)} = F(s-t) = P\{\xi < s - t\}.$$

Тобто, при $0 < t < s$

$$P\{\xi < s / \xi > t\} = P\{\xi < s - t\}.$$

Серед всіх неперервних розподілів цю властивість має тільки показниковий розподіл.

Гамма-розподіл. Розглянемо гамма-функцію $\Gamma(\alpha) = \int\limits_{0}^{\infty} x^{\alpha-1}e^{-x}dx, \quad \alpha > 0$. Зробимо в інтегралі заміну $x = \beta y$, де $\beta > 0$. Тоді

$$\Gamma(\alpha) = \int\limits_{0}^{\infty} (\beta y)^{\alpha-1} e^{-\beta y}\beta dy = \beta^{\alpha} \int\limits_{0}^{\infty} y^{\alpha-1}e^{-\beta y}dy$$

і

$$\frac{\beta^{\alpha}}{\Gamma(\alpha)} \int\limits_{0}^{\infty} x^{\alpha-1}e^{-\beta x}dx = 1.$$

Тому для функції

$$f(x, \alpha, \beta) = \begin{cases} 0, \quad x \le 0; \\ \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x}, \quad x > 0 \end{cases}$$

виконуються умови (4) і (5). Отже, вона є щільністю розподілу деякої випадкової величини. Розподіл із такою щільністю називається гамма-розподілом із параметрами α, β.

Вправи.

11. Випадкова величина ξ має показниковий розподіл із параметром $\lambda = \frac{1}{3}$.

    Знайти $P\{\xi > 3\}$, $P\{\xi < 6/\xi > 3\}$.

12. Функція $p(x)=\begin{cases} a\cdot\cos x, & |x|\le\dfrac{\pi}{2}; \\ 0, & |x|>\dfrac{\pi}{2} \end{cases}$ є щільністю розподілу випадкової величини ξ.

Знайти a і функцію розподілу ξ.

13. Показати, що показниковий розподіл є частинним випадком гамма-розподілу.

3.5. Функції від випадкової величини

Нехай $\varphi(x)$ – борелєва функція, ξ – випадкова величина, тоді $\varphi(\xi)=\eta$ також випадкова величина. Тоді

$$F_\eta(x)=P\{\eta<x\}=P\{\varphi(\xi)<x\}=P\{\xi\in\varphi^{-1}(-\infty,x)\}.$$

Якщо $\varphi(x)$ – монотонно зростаюча, то

$$F_\eta(x)=P\{\varphi(\xi)<x\}=P\{\xi<\varphi^{-1}(x)\}=F_\xi\left(\varphi^{-1}(x)\right),$$

а щільність розподілу

$$f_\eta(x)=f_\xi\left(\varphi^{-1}(x)\right)\frac{1}{\varphi'(x)}.$$

Якщо ж $\varphi(x)$ – монотонно спадна, то

$$F_\eta(x)=P\{\varphi(\xi)<x\}=P\{\xi>\varphi^{-1}(x)\}=1-F_\xi\left(\varphi^{-1}(x)+0\right),$$

а

$$f_\eta(x)=-f_\xi\left(\varphi^{-1}(x)\right)\frac{1}{\varphi'(x)}=f_\xi\left(\varphi^{-1}(x)\right)\frac{1}{|\varphi'(x)|}.$$

Для довільної монотонної функції щільність розподілу

$$f_\eta(x)=f_\xi\left(\varphi^{-1}(x)\right)\frac{1}{|\varphi'(x)|}.$$

Нехай $F_\xi(x)$ – строго зростаюча функція розподілу випадкової величини ξ. Розглянемо випадкову величину $\eta=F_\xi(\xi)$. Тоді при $x\in[0,1]$

$$F_\eta(x)=P\{F_\xi(\xi)<x\}=P\{\xi<F_\xi^{-1}(x)\}=F_\xi\left(F_\xi^{-1}(x)\right)=x,$$

при $x<0$ $F_\eta(x)=0$, а при $x>1$ $F_\eta(x)=1$. Тобто, випадкова величина η має рівномірний розподіл на відрізку [0,1]:

$$F_\eta(x)=\begin{cases} 0, & x<0; \\ x, & 0\le x\le 1; \\ 1, & x>1. \end{cases}$$

Нехай $F(x)$ – строго зростаюча функція розподілу. І нехай ξ – випадкова величина, що має рівномірний розподіл на відрізку [0,1]. Розглянемо випадкову величину $\eta=F^{-1}(\xi)$. Тоді

$$F_\eta(x)=P\{F^{-1}(\xi)<x\}=P\{\xi<F(x)\}=F(x).$$

Тобто, випадкова величина η має задану функцію розподілу $F(x)$.

Приклад. Випадкова величина ξ розподілена за стандартним нормальним законом. Знайти щільність розподілу випадкової величини $\eta=\xi^2$.

Розв'язування. Знайдемо спочатку функцію розподілу η. Оскільки випадкова величина η невід'ємна, то при $x \le 0$ $F_\eta(x) = P(\eta < x) = 0$. Нехай $x > 0$, тоді

$$F_\eta(x) = P(\xi^2 < x) = P(|\xi| < \sqrt{x}) = P(-\sqrt{x} < \xi < \sqrt{x}) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{t^2}{2}} dt = \frac{2}{\sqrt{2\pi}} \int_{0}^{\sqrt{x}} e^{-\frac{t^2}{2}} dt .$$

Тому $p_\eta(x) = F_\eta'(x) = \dfrac{2}{\sqrt{2\pi}} e^{-\frac{x}{2}} \dfrac{1}{2\sqrt{x}} = \dfrac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}$. Отже, $p_\eta(x) = \begin{cases} 0, & x \le 0; \\ \dfrac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}, & x > 0. \end{cases}$

Відзначимо, що одержаний розподіл є частинним випадком гамма-розподілу при $\alpha = \dfrac{1}{2}$, $\beta = \dfrac{1}{2}$.

Вправи.

14. Випадкова величина ξ має показниковий розподіл з параметром $\lambda = \frac{1}{2}$. Знайти щільність розподілу $\eta = \sqrt{\xi}$.
15. Випадкова величина ξ має показниковий розподіл з параметром λ. Знайти щільність розподілу випадкової величини $\eta = e^{\xi}$.
16. Випадкова величина ξ має функцію розподілу $F_\xi(x)$. Знайти функцію розподілу випадкової величини $\eta = e^{-\xi}$.
17. Випадкова величина ξ має рівномірний розподіл на [0,1]. Знайти функцію розподілу випадкової величини $\eta = \ln\dfrac{1}{\xi}$.

3.6. Багатовимірні випадкові величини

3.6.1. Випадкові вектори, їх розподіли. Нехай на ймовірнісному просторі $\{\Omega, \mathfrak{A}, P\}$ задані випадкові величини $\xi_1, ..., \xi_m$. Вектор $\xi = (\xi_1, ..., \xi_m)$ - називається m-вимірним *випадковим вектором* або *m-вимірною випадковою величиною*.

При цьому, відображення $\Omega \to R^m$ розглядається як вимірне відображення $\{\Omega, \mathfrak{A}\}$ на простір $\{R^m, \sigma_{R^m}\}$, де σ_{R^m} є σ-алгебра борелєвих множин в R^m. Тому для будь-якої борелєвої множини $B \subset R^m$ можна визначити ймовірність

$$P_\xi(B) = P\{\omega : (\xi_1(\omega), ..., \xi_m(\omega)) \in B\}.$$

Ймовірність $P_\xi(B)$ називають *розподілом випадкового вектора* ξ**.** У частинному випадку, якщо $B = (-\infty, x_1) \times \cdots \times (-\infty, x_m)$, одержимо функцію, яку називають функцією розподілу випадкового вектора ξ:

$$F_\xi(x_1, ..., x_m) = P\{\xi_1 < x_1, ..., \xi_m < x_m\}.$$

Подія $\{\xi_1 < x_1, ..., \xi_m < x_m\}$ розглядається як сумісне настання подій $\{\xi_1 < x_1\}, \dots, \{\xi_m < x_m\}$.

Очевидно,

$$0 \le F_\xi(x_1, ..., x_m) \le 1.$$

Оскільки $\{\xi_m < -\infty\} = \varnothing$, а $\{\xi_m < +\infty\} = \Omega$, то

$$\lim_{x_m \to -\infty} F_{(\xi_1,...,\xi_m)}(x_1,...,x_m) = 0,$$
$$\lim_{x_m \to +\infty} F_{(\xi_1,...,\xi_m)}(x_1,...,x_m) = F_{(\xi_1,...,\xi_{m-1})}(x_1,...,x_{m-1}).$$

Із останньої властивості випливає, що за функцією розподілу вектора $(\xi_1,...,\xi_m)$ можна знайти функцію розподілу вектора $(\xi_{i_1},...,\xi_{i_k})$ для довільного k=1,…,m, зокрема, можна знайти функцію розподілу $F_{\xi_k}(x_k)$ кожної компоненти вектора.

Відзначимо також, що функція розподілу по кожній із змінних неспадна і неперервна зліва.

Розглянемо двовимірний випадковий вектор (ξ_1,ξ_2). Тоді для функції розподілу $F_\xi(x_1,x_2)$ будуть справедливі рівності:

$$F_\xi(-\infty,x_2) = F_\xi(x_1,-\infty) = 0,$$
$$F_\xi(x_1,+\infty) = F_{\xi_1}(x_1),\ F_\xi(+\infty,x_2) = F_{\xi_2}(x_2),$$
$$P\{a_1 \le \xi_1 < b_1, a_2 \le \xi_2 < b_2\} = F_\xi(b_1,b_2) - F_\xi(a_1,b_2) - F_\xi(b_1,a_2) + F_\xi(a_1,a_2).$$

Якщо існує функція $p(x_1,...,x_m)$ така, що

$$\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} p(u_1,...,u_m) du_1 \cdots du_m = F_\xi(x_1,...,x_m),$$

то функцію $p(x) = p(x_1,...,x_m)$ називають *щільністю розподілу випадкового вектора* ξ, а розподіл випадкового вектора називають *абсолютно неперервним*.

Із даного означення випливає:

1) $p(x_1,...,x_m) \ge 0$;

2) якщо $(x_1,...,x_m)$ – точка неперервності p, то $p(x_1,...,x_m) = \dfrac{\partial^m F}{\partial x_1 \cdots \partial x_m}$;

3) $\int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} p(x_1,...,x_m) dx_1 \cdots dx_m = 1$;

4) для довільної борелєвої множини

$$P\{(\xi_1,...,\xi_m) \in B\} = \int (m) \int_B p(x_1,...,x_m) dx_1 \cdots dx_m.$$

Знаючи щільність розподілу випадкового вектора, можна знайти щільності розподілу його компонент. Розглянемо, як частинний випадок, двовимірну випадкову величину (ξ,η) зі щільністю $p(x,y)$. Тоді

$$F_\xi(x) = \lim_{y \to +\infty} F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} p(u,v) du dv,$$

звідки

$$p_\xi(x) = \frac{dF_\xi(x)}{dx} = \int_{-\infty}^{+\infty} p(x,y) dy,$$

а

$$p_\eta(y) = \int_{-\infty}^{+\infty} p(x,y) dx.$$

Як приклад розглянемо *багатовимірний нормальний розподіл*. Випадкова величина $\xi = (\xi_1,...,\xi_m)$ називається *нормально розподіленою*, якщо її щільність розподілу

$$p_\xi(x)=\frac{1}{(2\pi)^{\frac{m}{2}}}\frac{1}{|A|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-a)'A(x-a)\right),$$

де $x=\begin{pmatrix}x_1\\ \cdots\\ x_m\end{pmatrix}$; $a=\begin{pmatrix}a_1\\ \cdots\\ a_m\end{pmatrix}$, $(a_i\in R)$, A – невироджена квадратна матриця, додатньо-визначена, розміром $m\times m$, $|A|$ – детермінант матриці A.

У випадку двовимірного нормального розподілу щільність розподілу має вигляд:
$$p(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-a_1}{\sigma_1}\right)^2-2\rho\frac{x_1-a_1}{\sigma_1}\frac{x_2-a_2}{\sigma_2}+\left(\frac{x_2-a_2}{\sigma_2}\right)^2\right)\right),\quad |\rho|<1.$$

Якщо $\rho=0$, то $p(x_1,x_2)=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2}\left(\frac{x_1-a_1}{\sigma_1}\right)^2\right)\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{1}{2}\left(\frac{x_2-a_2}{\sigma_2}\right)^2\right)$, де

$\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2}\left(\frac{x_1-a_1}{\sigma_1}\right)^2\right)$ – щільність розподілу випадкової величини $\xi_1\sim N(a_1,\sigma_1^2)$, а

$\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{1}{2}\left(\frac{x_2-a_2}{\sigma_2}\right)^2\right)$ – щільність розподілу випадкової величини $\xi_2\sim N(a_2,\sigma_2^2)$.

За щільністю розподілу $p(x_1,x_2)$ випадкового вектора (ξ_1,ξ_2) знайдемо щільність розподілу випадкової величини ξ_1. Тоді

$$p_{\xi_1}(x_1)=\int_{-\infty}^{\infty}p(x_1,x_2)dx_2=$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-a_1}{\sigma_1}\right)^2-2\rho\frac{x_1-a_1}{\sigma_1}\frac{x_2-a_2}{\sigma_2}+\left(\frac{x_2-a_2}{\sigma_2}\right)^2\right)\right)dx_2=$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}\left(\frac{x_1-a_1}{\sigma_1}\right)^2\right)\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x_2-a_2}{\sigma_2}-\rho\frac{x_1-a_1}{\sigma_1}\right)^2\right)dx_2\,.$$

Після виконання в інтегралі підстановки $\frac{1}{\sqrt{1-\rho^2}}\left(\frac{x_2-a_2}{\sigma_2}-\rho\frac{x_1-a_1}{\sigma_1}\right)=t$, $\frac{dx_2}{\sigma_2\sqrt{1-\rho^2}}=dt$, одержимо:

$$p_{\xi_1}(x_1)=\frac{1}{2\pi\sigma_1}\exp\left(-\frac{1}{2}\left(\frac{x_1-a_1}{\sigma_1}\right)^2\right)\int_{-\infty}^{\infty}\exp\left(-\frac{t^2}{2}\right)dt=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2}\left(\frac{x_1-a_1}{\sigma_1}\right)^2\right).$$

Це означає, що величина $\xi_1\sim N(a_1,\sigma_1^2)$, аналогічно величина $\xi_2\sim N(a_2,\sigma_2^2)$. Тобто, якщо випадковий вектор (ξ_1,ξ_2) нормально розподілений, то кожна із випадкових величин ξ_1 і ξ_2 має нормальний розподіл.

Випадковий вектор називається дискретним, якщо множина його значень скінченна або зліченна. Розглянемо для спрощення двовимірну випадкову величину (ξ,η). Нехай $\xi=x_i$, $i=1,...,l$, $\eta=y_j$, $j=1,...,r$, тоді кожній парі (x_i,y_j) можна поставити у відповідність ймовірність $p_{ij} = P\{\xi = x_i; \eta = y_j\}$. Тут $\{\xi = x_i; \eta = y_j\} = \{\xi = x_i\} \bigcap \{\eta = y_j\}$. Таку відповідність називають законом розподілу. Його можна подати у вигляді такої таблиці:

$\xi \setminus \eta$	y_1	.....	y_r	
x_1	p_{11}	.....	p_{1r}	p_{x_1}
…	…	.....	…..	…..
x_l	p_{l1}	.....	p_{lr}	p_{x_l}
	p_{y_1}	.....	p_{y_r}	

Якщо $H_j = \{\eta = y_j\}$, то $H_1, ..., H_r$ утворюють повну групу і $\bigcup_{j=1}^{r} \{\eta = y_j\} = \Omega$, тому $\{\xi = x_i\} = \bigcup_{j=1}^{r} \{\xi = x_i; \eta = y_j\}$. Отже, за аксіомою адитивності $p_{x_i} = P\{\xi = x_i\} = \sum_{j=1}^{r} p_{ij}$. Так само $p_{y_j} = P\{\eta = y_j\} = \sum_{i=1}^{l} p_{ij}$. При цьому, $\sum_{i,j} p_{ij} = \sum_i p_{x_i} = \sum_j p_{y_j} = 1$.

Вправи.

18. Випадковий вектор (ξ, η) має рівномірний розподіл в області $D \subset R^2$, якщо його щільність розподілу має вигляд $p(x,y) = \begin{cases} c, & (x,y) \in D; \\ 0, & (x,y) \notin D. \end{cases}$ Знайти сталу *c*.
19. Випадковий вектор (ξ, η) має щільність розподілу $p(x,y) = \begin{cases} 8xy(1-x^2), & (x,y) \in D; \\ 0, & (x,y) \notin D, \end{cases}$ де $D = \{(x,y) : 0 \le x \le 1, 0 \le y \le 1\}$. Знайти $p_\xi(x)$ і $p_\eta(y)$.
20. Знайти закони розподілу дискретних випадкових величин ξ і η, якщо випадковий вектор (ξ, η) заданий законом розподілу

$\xi \setminus \eta$	-2	0	2	4
1	0,1	0,2		
2		0,1	0,2	
3		0.1	0,2	0,1

3.6.2. Незалежність випадкових величин. Випадкові величини $\xi_1, ..., \xi_m$ називаються *незалежними*, якщо

$$F_{(\xi_1,...,\xi_m)}(x_1, ..., x_m) = F_{\xi_1}(x_1) \cdot ... \cdot F_{\xi_m}(x_m).$$

Теорема. Нехай випадковий вектор $(\xi_1, ..., \xi_m)$ має щільність розподілу $p(x_1, ..., x_m)$. Для того, щоб випадкові величини $\xi_1, ..., \xi_m$ були незалежними необхідно і досить, щоб

$$p(x_1, ..., x_m) = p_{\xi_1}(x_1) \cdot ... \cdot p_{\xi_m}(x_m),$$

де $p_{\xi_i}(x_i)$ - щільність ξ_i.

Доведення. Якщо випадкові величини незалежні, то
$F_{(\xi_1,\dots,\xi_m)}\left(x_1,\dots,x_m\right)=F_{\xi_1}(x_1)\cdot\ldots\cdot F_{\xi_m}(x_m)=$

$$=\int_{-\infty}^{x_1} p_{\xi_1}(u_1)du_1\cdot\ldots\cdot\int_{-\infty}^{x_m} p_{\xi_m}(u_m)du_m=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_m} p_{\xi_1}(u_1)\cdot\ldots\cdot p_{\xi_m}(u_m)du_1\cdot\ldots\cdot du_m .$$

Тому $p(x_1,\dots x_m)=p_{\xi_1}(x_1)\cdot\ldots\cdot p_{\xi_m}(x_m)$.

Нехай навпаки $p\left(x_1,\dots,x_m\right)=p_{\xi_1}(x_1)\cdot\ldots\cdot p_{\xi_m}(x_m)$. Тоді

$$F\left(x_1,\dots,x_m\right)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_m} p\left(u_1,\dots,u_m\right)du_1\cdot\ldots\cdot du_m=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_m} p_{\xi_1}(u_1)\cdot\ldots\cdot p_{\xi_m}(u_m)du_1\cdot\ldots\cdot du_m=$$

$$=\int_{-\infty}^{x_1} p_{\xi_1}(u_1)du_1\cdot\ldots\cdot\int_{-\infty}^{x_m} p_{\xi_m}(u_m)du_m=F_{\xi_1}(x_1)\cdot\ldots\cdot F_{\xi_m}(x_m).$$

Тобто, випадкові величини незалежні.

Відзначимо, що у випадку двовимірного нормального розподілу умова $\rho=0$ є неохідною і достатньою умовою незалежності компонент вектора.

Якщо випадкова величина (ξ,η) дискретна, то із означення випливає, що умова незалежності випадкових величин еквівалентна умові

$$p_{ij}=P\left\{\xi=x_i;\eta=y_j\right\}=P\left\{\xi=x_i\right\}P\left\{\eta=y_j\right\}=p_{x_i}p_{y_j} .$$

Сформулюємо без доведення наступні властивості.

Якщо $\xi_1,\dots,\xi_m$ – незалежні випадкові величини, то для довільних борелєвих функцій $g_1(x),\dots,g_m(x)$ випадкові величини $g_1(\xi_1),\dots,g_m(\xi_m)$ також незалежні.

Якщо випадкові величини $\xi_1,\dots,\xi_m,\eta_1,\dots,\eta_n$ – незалежні, а $g(x_1,\dots,x_m)$ і $h(x_1,\dots,x_n)$ борелєві функції, то випадкові величини $g(\xi_1,\dots,\xi_m)$ і $h(\eta_1,\dots,\eta_n)$ – незалежні.

Вправи.

21. Задано функцію розподілу випадкового вектора (ξ,η) $F(x,y)=\begin{cases}0, & при\ x<0\ або\ y<0,\\ 1-2^{-x}-2^{-y}+2^{-x-y}, & при\ x\geq 0\ і\ y\geq 0.\end{cases}$ Чи будуть величини ξ і η незалежними?
22. Випадковий вектор (ξ,η) рівномірно розподілений в крузі $x^2+y^2\leq R^2$. Знайти $p_\xi(x)$ і $p_\eta(y)$. Чи будуть величини ξ і η незалежними?
23. Задано щільність розподілу випадкового вектора (ξ,η) $p(x,y)=xe^{-x-xy},\ x\geq 0,\ y\geq 0$. Чи будуть величини ξ і η незалежними?
24. Випадковий вектор (ξ,η) має щільність розподілу $p(x,y)=\begin{cases}8xy(1-x^2), & (x,y)\in D;\\ 0, & (x,y)\notin D,\end{cases}$ де $D=\{(x,y):0\leq x\leq 1,\ 0\leq y\leq 1\}$. Чи будуть величини ξ і η незалежними?
25. Задано розподіл випадкового вектора (ξ,η) (вправа 21). Чи будуть величини ξ і η незалежними?

3.6.3. Розподіл деяких функцій від декількох випадкових величин. Розглянемо спочатку приклад. Нехай дискретна випадкова величина ξ має розподіл $P\{\xi=-1\}=\dfrac{1}{4}$,

$P\{\xi=1\}=\frac{3}{4}$, а випадкова величина η має рівномірний на відрізку [0,1] розподіл. Необхідно знайти розподіл суми $\zeta=\xi+\eta$.

Для цього подію $\{\zeta<x\}$ зобразимо у вигляді:

$$\{\zeta<x\}=\{\xi+\eta<x,\xi=-1\}\cup\{\xi+\eta<x,\xi=1\},$$

Тоді

$$F_\zeta(x)=P\{\zeta<x\}=P\{\xi+\eta<x,\xi=-1\}+P\{\xi+\eta<x,\xi=1\}=$$
$$=P\{\xi=-1\}P\{\xi+\eta<x/\xi=-1\}+P\{\xi=1\}P\{\xi+\eta<x/\xi=1\}=$$
$$=\frac{1}{4}P\{-1+\eta<x\}+\frac{3}{4}P\{1+\eta<x\}=\frac{1}{4}F_\eta(x+1)+\frac{3}{4}F_\eta(x-1),$$

$F_\eta(x)=\begin{cases}0, & x\le 0;\\ x, & 0<x\le 1;\\ 1, & x>1,\end{cases}$ тоді $F_\eta(x+1)=\begin{cases}0, & x\le -1;\\ x+1, & -1<x\le 0;\\ 1, & x>0,\end{cases}$ а $F_\eta(x-1)=\begin{cases}0, & x\le 1;\\ x-1, & 1<x\le 2;\\ 1, & x>2.\end{cases}$

Тому при $x\le -1$ $F_\zeta(x)=0$, при $-1<x\le 0$ $F_\zeta(x)=\frac{1}{4}(x+1)$, при $0<x\le 1$ $F_\zeta(x)=\frac{1}{4}$, при $1<x\le 2$ $F_\zeta(x)=\frac{1}{4}+\frac{3}{4}(x-1)$ і при $x>2$ $F_\zeta(x)=1$. Тобто,

$$F_\zeta(x)=\begin{cases}0, & x\le -1;\\ \frac{1}{4}(x+1), & -1<x\le 0;\\ \frac{1}{4}, & 0<x\le 1;\\ \frac{1}{4}(3x-2), & 1<x\le 2;\\ 1, & x>2.\end{cases}$$

Нехай випадковий вектор $(\xi_1,...,\xi_m)$ має щільність розподілу $p(x_1,...,x_m)$. Розглянемо випадкову величину $\eta=\varphi(\xi_1,...,\xi_m)$. Функція розподілу η буде рівна

$$F_\eta(x)=P\{\eta<x\}=\int_{(x_1,...,x_m):}\cdots\int_{\varphi(x_1,...,x_m)<x} p(x_1,...,x_m)\,dx_1\cdots dx_m.$$

Розглянемо *розподіл суми* випадкових величин. Нехай $p(x_1,x_2)$ - щільність розподілу вектора (ξ_1,ξ_2), а $\eta=\xi_1+\xi_2$. Тоді функція розподілу суми буде рівна

$$F_\eta(x)=\iint_{(x_1,x_2):\,x_1+x_2<x} p(x_1,x_2)dx_1dx_2.$$

Область інтегрування можна подати у вигляді $\{(x_1,x_2):-\infty<x_1<+\infty;-\infty<x_2<x-x_1\}$ або $\{(x_1,x_2):-\infty<x_2<+\infty;-\infty<x_1<x-x_2\}$. У відповідності до цього розглянемо два варіанти знаходження функції розподілу:

$$F_\eta(x)=\int_{-\infty}^{+\infty}dx_1\int_{-\infty}^{x-x_1}p(x_1,x_2)dx_2=\left|\begin{matrix}x_1+x_2=u;\\ dx_2=du\end{matrix}\right|=\int_{-\infty}^{+\infty}dx_1\int_{-\infty}^{x}p(x_1,u-x_1)du=$$
$$=\int_{-\infty}^{x}du\int_{-\infty}^{\infty}p(x_1,u-x_1)dx_1,$$

$$F_\eta(x)=\int_{-\infty}^{+\infty}dx_2\int_{-\infty}^{x-x_2}p(x_1,x_2)dx_1=\begin{vmatrix}x_1+x_2=u;\\ dx_1=du\end{vmatrix}=\int_{-\infty}^{+\infty}dx_2\int_{-\infty}^{x}p(u-x_2,x_2)du=$$

$$=\int_{-\infty}^{x}du\int_{-\infty}^{\infty}p(u-x_2,x_2)dx_2\,.$$

Із одержаних рівностей і означення щільності розподілу випливає, що щільність розподілу суми випадкових величин знаходиться із рівностей

$$p_\eta(x)=\int_{-\infty}^{\infty}p(x_1,x-x_1)dx_1$$

або

$$p_\eta(x)=\int_{-\infty}^{\infty}p(x-x_2,x_2)dx_2\,.$$

Нехай ξ_1, ξ_2 – незалежні. Це означає, що $p(x_1,x_2)=p_{\xi_1}(x_1)p_{\xi_2}(x_2)$. Тому

$$p_{\xi_1+\xi_2}(x)=\int_{-\infty}^{+\infty}p_{\xi_1}(x_1)p_{\xi_2}(x-x_1)dx_1=\int_{-\infty}^{+\infty}p_{\xi_2}(x_2)p_{\xi_1}(x-x_2)dx_2\,.$$

Отже, щільність розподілу суми випадкових величин є *згорткою* функцій $p_{\xi_1}(x_1)$ і $p_{\xi_2}(x_2)$.

Із вигляду функції розподілу суми $\eta=\xi_1+\xi_2$ $F_\eta(x)=\int_{-\infty}^{+\infty}dx_1\int_{-\infty}^{x-x_1}p(x_1,x_2)dx_2$, для незалежних випадкових величин одержуємо:

$$F_\eta(x)=\int_{-\infty}^{+\infty}dx_1\int_{-\infty}^{x-x_1}p_{\xi_1}(x_1)p_{\xi_2}(x_2)dx_2=\int_{-\infty}^{+\infty}p_{\xi_1}(x_1)dx_1\int_{-\infty}^{x-x_1}p_{\xi_2}(x_2)dx_2=$$

$$=\int_{-\infty}^{+\infty}F_{\xi_2}(x-x_1)dF_{\xi_1}(x_1)\,,$$

$$F_\eta(x)=\int_{-\infty}^{+\infty}dx_2\int_{-\infty}^{x-x_2}p_{\xi_1}(x_1)p_{\xi_2}(x_2)dx_1=\int_{-\infty}^{+\infty}p_{\xi_2}(x_2)dx_2\int_{-\infty}^{x-x_2}p_{\xi_1}(x_1)dx_1=$$

$$=\int_{-\infty}^{+\infty}F_{\xi_1}(x-x_2)dF_{\xi_2}(x_2)\,.$$

Тобто, функція розподілу суми

$$F_\eta(x)=\int_{-\infty}^{+\infty}F_{\xi_2}(x-x_1)dF_{\xi_1}(x_1)=\int_{-\infty}^{+\infty}F_{\xi_1}(x-x_2)dF_{\xi_2}(x_2)$$

є *згорткою* або *композицією* функції розподілу $F_{\xi_1}(x)$ і $F_{\xi_2}(x)$.

Розглянемо випадок двовимірного нормального розподілу зі щільністю розподілу $p(x_1,x_2)=\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\dfrac{1}{2(1-\rho^2)}\left(\left(\dfrac{x_1-a_1}{\sigma_1}\right)^2-2\rho\dfrac{x_1-a_1}{\sigma_1}\dfrac{x_2-a_2}{\sigma_2}+\left(\dfrac{x_2-a_2}{\sigma_2}\right)^2\right)\right),\quad |\rho|<1.$

Знайдемо щільність розподілу суми

$$p_{\xi_1+\xi_2}(x)=\int\limits_{-\infty}^{+\infty} p(z,x-z)dz=$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int\limits_{-\infty}^{+\infty}\exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left(\left(\frac{z-a_1}{\sigma_1}\right)^2-2\rho\frac{z-a_1}{\sigma_1}\frac{x-z-a_2}{\sigma_2}+\left(\frac{x-z-a_2}{\sigma_2}\right)^2\right)\right)dz\,.$$

Для спрощення записів позначимо $v=x-a_1-a_2$ і зробимо в інтегралі заміну $z-a_1=u$, тоді

$$p_{\xi_1+\xi_2}\left(x\right)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int\limits_{-\infty}^{+\infty}\exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left(\frac{u^2}{\sigma_1^2}-2\rho\frac{u}{\sigma_1}\frac{v-u}{\sigma_2}+\frac{\left(v-u\right)^2}{\sigma_2^2}\right)\right)du\,.$$

Показник у підінтегральній функції перетворимо до вигляду

$$\frac{u^2}{\sigma_1^2}-2\rho\frac{u}{\sigma_1}\frac{v-u}{\sigma_2}+\frac{\left(v-u\right)^2}{\sigma_2^2}=\frac{u^2}{\sigma_1^2}+2\rho\frac{u^2}{\sigma_1\sigma_2}+\frac{u^2}{\sigma_2^2}-\frac{2uv}{\sigma_2^2}-2\rho\frac{uv}{\sigma_1\sigma_2}+\frac{v^2}{\sigma_2^2}=$$

$$=u^2\,\frac{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}{\sigma_1^2\sigma_2^2}-2uv\frac{\sigma_1+\rho\sigma_2}{\sigma_1\sigma_2^2}+\frac{v^2}{\sigma_2^2}=$$

$$=\left(u\frac{\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}{\sigma_1\sigma_2}-v\frac{\sigma_1+\rho\sigma_2}{\sigma_1\sigma_2^2}\frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}\right)^2-v^2\,\frac{\left(\sigma_1+\rho\sigma_2\right)^2}{\sigma_2^2\left(\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2\right)}+$$

$$+\frac{v^2}{\sigma_2^2}=\left(u\frac{\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}{\sigma_1\sigma_2}-v\frac{\sigma_1+\rho\sigma_2}{\sigma_2\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}\right)^2+\,v^2\,\frac{1-\rho^2}{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}$$

і виконаємо заміну $t=\frac{1}{\sqrt{1-\rho^2}}\left(u\frac{\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}{\sigma_1\sigma_2}-v\frac{\sigma_1+\rho\sigma_2}{\sigma_2\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}\right)$,

$dt=\frac{1}{\sqrt{1-\rho^2}}\frac{\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}{\sigma_1\sigma_2}du$. Врахувавши рівність $\int\limits_{-\infty}^{+\infty}\exp\left(-\frac{t^2}{2}\right)dt=\sqrt{2\pi}$, одержимо:

$$p_{\xi_1+\xi_2}\left(x\right)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}\frac{v^2}{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}\right)\times$$

$$\times\int\limits_{-\infty}^{\infty}\exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left(u\frac{\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}{\sigma_1\sigma_2}-v\frac{\sigma_1+\rho\sigma_2}{\sigma_2\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}\right)^2\right)du=$$

$$=\frac{1}{2\pi\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}\exp\left(-\frac{1}{2}\frac{v^2}{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}\right)\int\limits_{-\infty}^{\infty}\exp\left(-\frac{t^2}{2}\right)dt=$$

$$=\frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}}\exp\left(-\frac{1}{2}\frac{\left(x-a_1-a_2\right)^2}{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}\right).$$

Отже, якщо випадковий вектор $\left(\xi_1,\xi_2\right)$ – нормально розподілений, то сума $\xi_1+\xi_2$ – нормально розподілена з параметрами a_1+a_2 і $\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2$.

Якщо ξ_1 і ξ_2 - незалежні, тобто $\rho=0$, і $\xi_1\sim N\left(a_1,\sigma_1^2\right)$, $\xi_2\sim N\left(a_2,\sigma_2^2\right)$, то $\xi_1+\xi_2\sim N\left(a_1+a_2,\sigma_1^2+\sigma_2^2\right)$. Справедливе і обернене твердження (теорема Крамера). Якщо

ξ_1 і ξ_2 – незалежні і сума $\xi_1+\xi_2$ – нормально розподілена, то і кожен доданок ξ_1 і ξ_2 є нормально розподіленими випадковими величинами.

Розглянемо тепер *розподіл частки* випадкових величин. Нехай (ξ_1,ξ_2) – випадковий вектор із щільністю $p(x_1,x_2)$, а $\eta=\dfrac{\xi_1}{\xi_2}$, тоді $F_\eta(x)=\iint\limits_{(x_1,x_2):\frac{x_1}{x_2}<x} p(x_1,x_2)\,dx_1dx_2$.

Оскільки $\left\{(x_1,x_2):\dfrac{x_1}{x_2}<x\right\}=\{(x_1,x_2):x_2>0,x_1<xx_2\}\cup\{(x_1,x_2):x_2<0,x_1>xx_2\}$, то

$$F_\eta(x)=\iint\limits_{(x_1,x_2):\frac{x_1}{x_2}<x} p(x_1,x_2)\,dx_1dx_2=\int\limits_0^{+\infty}dx_2\int\limits_{-\infty}^{xx_2}p(x_1,x_2)dx_1+\int\limits_{-\infty}^{0}dx_2\int\limits_{xx_2}^{\infty}p(x_1,x_2)dx_1.$$

Диференціюючи дану рівність, одержимо щільність розподілу частки

$$p_\eta(x)=\int\limits_0^{+\infty}x_2p(xx_2,x_2)\,dx_2-\int\limits_{-\infty}^{0}x_2p(xx_2,x_2)dx_2=\int\limits_{-\infty}^{+\infty}|x_2|\,p(xx_2,x_2)\,dx_2.$$

Знайдемо ще і *розподіл добутку* випадкових величин. Нехай (ξ_1,ξ_2) – випадковий вектор із щільністю $p(x_1,x_2)$, а $\eta=\xi_1\xi_2$, тоді $F_\eta(x)=\iint\limits_{(x_1,x_2):\,x_1x_2<x} p(x_1,x_2)\,dx_1dx_2$.

Із рівності $\{(x_1,x_2):x_1x_2<x\}=\left\{(x_1,x_2):x_2>0,x_1<\dfrac{x}{x_2}\right\}\cup\left\{(x_1,x_2):x_2<0,x_1>\dfrac{x}{x_2}\right\}$ маємо

$$F_\eta(x)=\int\limits_0^{+\infty}dx_2\int\limits_{-\infty}^{x/x_2}p(x_1,x_2)dx_1+\int\limits_{-\infty}^{0}dx_2\int\limits_{x/x_2}^{+\infty}p(x_1,x_2)dx_1.$$

Диференціюючи дану рівність, одержимо щільність розподілу добутку

$$p_\eta(x)=\int\limits_0^{+\infty}p(x/x_2,x_2)\frac{1}{x_2}dx_2-\int\limits_{-\infty}^{0}p(x/x_2,x_2)\frac{1}{x_2}dx_2=\int\limits_{-\infty}^{+\infty}\frac{1}{|x_2|}p(x/x_2,x_2)\,dx_2.$$

Відзначимо, що при знаходженні щільності розподілу функції від випадкових величин, часто буває простіше знайти спочатку її функцію розподілу.

Вправи.

26. Нехай ξ_1 і ξ_2 – незалежні випадкові величини, що мають гамма-розподіл із параметрами (α_1,β) і (α_2,β) відповідно. Довести, що випадкова величина $\xi_1+\xi_2$ має гамма-розподіл із параметрами $(\alpha_1+\alpha_2,\beta)$.
27. Знайти розподіл $\xi+\eta$, якщо задано закон розподілу дискретного випадкового вектора (ξ,η). Таблиця вправи 20.

28. Задано щільність розподілу випадкового вектора (ξ,η): $p(x,y)=\begin{cases} c(x+y), & (x,y)\in D; \\ 0, & (x,y)\notin D, \end{cases}$ де $D=\{(x,y):0\le y\le x\le 1\}$. Знайти сталу c, щільність розподілу $\xi+\eta$.
29. Знайти закон розподілу $\xi+\eta$, якщо дискретні випадкові величини ξ і η незалежні і задано їх закони розподілу:

ξ	-1	1	2	3
p	0,2	0,4	0,3	0,1

η	-2	0	1
p	0,3	0,5	0,2

3.6.4. Деякі розподіли функцій від нормально розподілених випадкових величин. Нехай випадкові величини $\xi_1,...,\xi_n$ – незалежні, $\xi_i \sim N(0,1)$. Тоді випадкова величина $\chi^2(n)=\xi_1^2+\dots+\xi_n^2$ має *розподіл Пірсона (або розподіл* χ^2*)* із n ступенями вільності.

Нехай $\xi_0,\xi_1,...,\xi_n$ – незалежні нормально розподілені з параметрами $(0,1)$ випадкові величини. Тоді величина $\mathrm{T}=\dfrac{\xi_0}{\sqrt{\frac{1}{n}\left(\xi_1^2+\dots+\xi_n^2\right)}}=\dfrac{\xi_0}{\sqrt{\frac{1}{n}\chi^2(n)}}$ має *розподіл Стьюдента* (або t-розподіл) із n ступенями вільності.

Нехай $\xi_1,...,\xi_n$, $\eta_1,...,\eta_m$ – незалежні нормально розподілені з параметрами $(0,1)$ випадкові величини, $\chi^2(n)=\xi_1^2+\dots+\xi_n^2$, $\chi^2(m)=\eta_1^2+\dots+\eta_m^2$. Тоді величина $F=\dfrac{\frac{1}{n}\chi^2(n)}{\frac{1}{m}\chi^2(m)}$ має *розподіл Фішера* із (n,m) ступенями вільності.

Нормальний розподіл, розподіли Пірсона, Стьюдента, Фішера мають дуже широке використання у математичній статистиці.

3.7. Умовні розподіли

Нехай (ξ,η) - випадковий двовимірний вектор, ξ і η – дискретні випадкові величини, ξ приймає значення x_i, η приймає значення y_j. Кожній парі таких чисел можна поставити у відповідність ймовірність $P(x_i,y_j)=P\{\xi=x_i,\eta=y_j\}$. Тоді за теоремою множення

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

Позначимо $P_{\eta/\xi}(y_j/x_i)=P\{\eta=y_j/\xi=x_i\}$. Нехай $P_\xi(x_i)=\sum_j P(x_i,y_j)>0$, тоді

$$P_{\eta/\xi}(y_j/x_i)=\frac{P(x_i,y_j)}{P_\xi(x_i)}.$$

Розподіл ймовірностей $P_{\eta/\xi}\left(y_j/x_i\right)$ називається *умовним розподілом* випадкової величини η відносно випадкової величини ξ (при умові, що випадкова величина ξ прийняла фіксоване значення x_i).

Оскільки $P_\xi(x_i)=\sum\limits_j P(x_i,y_j)$, то очевидно,

$$\sum_j P_{\eta/\xi}\left(y_j/x_i\right)=\sum_j\frac{P\left(x_i,y_j\right)}{P_\xi\left(x_i\right)}=\frac{1}{P_\xi\left(x_i\right)}\sum_j P\left(x_i,y_j\right)=1.$$

Нехай розподіл вектора (ξ,η) – абсолютно неперервний із щільністю сумісного розподілу $p\left(x,y\right)$. Розглянемо подію $\{x\le\xi<x+h\}$, і нехай $P\{x\le\xi<x+h\}=\int\limits_x^{x+h}\int\limits_{-\infty}^{+\infty}p(x,y)dxdy$ додатна. Тоді умовна ймовірність події $\{\eta<y\}$ при умові $\{x\le\xi<x+h\}$ дорівнює

$$P\{\eta<y/x\le\xi<x+h\}=\frac{P\{x\le\xi<x+h,\eta<y\}}{P\{x\le\xi<x+h\}}=\frac{\int\limits_x^{x+h}\int\limits_{-\infty}^{y}p(u,v)dudv}{\int\limits_x^{x+h}\int\limits_{-\infty}^{+\infty}p(u,v)dudv}.$$

Це функція розподілу, що відповідає умовному розподілу η при умові $\{x\le\xi<x+h\}$. Нехай $h\to 0$. Тоді

$$\lim_{h\to0}P\{\eta<y/x\le\xi<x+h\}=\frac{\int\limits_{-\infty}^{y}p(x,v)dv}{\int\limits_{-\infty}^{+\infty}p(x,v)dv}=\frac{\int\limits_{-\infty}^{y}p(x,v)dv}{p_\xi(x)}.$$

При фіксованому *x* ця границя, як функція від *y*, є функцією розподілу

$$F_{\eta/\xi}(y/x)=\frac{\int\limits_{-\infty}^{y}p(x,v)dv}{p_\xi(x)}$$

і називається *умовною функцією розподілу* величини η при умові $\xi=x$.

Якщо $p(x,y)$ – неперервна по y, то одержану функцію розподілу можна продиференціювати по y, тоді одержимо відповідну *умовну щільність розподілу* η

$$p_{\eta/\xi}(y/x)=\frac{p(x,y)}{p_\xi(x)}$$

при умові $\xi=x$.

Очевидно, $\int\limits_{-\infty}^{+\infty}p_{\eta/\xi}(y/x)dy=1$.

Приклад. Випадкова величина $\left(\xi,\eta\right)$ має нормальний розподіл. Довести, що умовний розподіл є також нормальним.

Розв'язування. Так як щільність двовимірного нормального розподілу

$$p\left(x_1,x_2\right)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left(\left(\frac{x_1-a_1}{\sigma_1}\right)^2-2\rho\frac{x_1-a_1}{\sigma_1}\frac{x_2-a_2}{\sigma_2}+\left(\frac{x_2-a_2}{\sigma_2}\right)^2\right)\right),$$

а

$$p_\xi(x_1)=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2}\left(\frac{x_1-a_1}{\sigma_1}\right)^2\right),$$

то, ввівши тимчасові позначення $u=\frac{x_1-a_1}{\sigma_1}$, $v=\frac{x_2-a_2}{\sigma_2}$, одержимо:

$$p_{\eta/\xi}(x_2/x_1)=\frac{p(x_1,x_2)}{p_\xi(x_1)}=\frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}\right)}{\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{u^2}{2}\right)}=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(u^2-2\rho uv+v^2\right)+\frac{1}{2}u^2\right)=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\rho^2u^2-2\rho uv+v^2\right)\right)=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{(v-\rho u)^2}{2(1-\rho^2)}\right)$$

або

$$p_{\eta/\xi}(x_2/x_1)=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x_2-a_2}{\sigma_2}-\rho\frac{x_1-a_1}{\sigma_1}\right)^2\right).$$

Отже, умовна щільність розподілу

$$p_{\eta/\xi}(x_2/x_1)=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{\left(x_2-a_2-\rho\frac{\sigma_2}{\sigma_1}(x_1-a_1)\right)^2}{2\sigma_2^2(1-\rho^2)}\right)$$

співпадає із щільністю нормального закону із параметрами $a_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-a_1)$ і $\sigma_2^2\left(1-\rho^2\right)$.

Вправи.

30.Знайти умовний розподіл $P_{\eta/\xi}\left(y_j/x_i\right)$, якщо задано закон розподілу дискретного випадкового вектора (ξ,η). Таблиця вправи 20.

31.Задано щільність розподілу випадкового вектора (ξ,η) $p(x,y)=\begin{cases} xe^{-x-xy}, & (x,y)\in D; \\ 0, & (x,y)\notin D, \end{cases}$ якщо $D=\{(x,y): x\geq 0,\ y\geq 0\}$. Знайти умовну щільність розподілу η при умові $\xi=x$: $p_{\eta/\xi}(y/x)$.

Розділ 4. Числові характеристики випадкових величин

4.1. Математичне сподівання

4.1.1. Означення математичного сподівання. Приклади. Нехай випадкова величина ξ – дискретна із можливими значеннями x_i і відповідними ймовірностями $p_i = P(\xi = x_i)$. Сума добутків можливих значень випадкової величини на ймовірності, з якими ці значення приймаються, називається *математичним сподіванням* дискретної випадкової величини (якщо множина значень випадкової величини зліченна, то за умови збіжності ряду $\sum_i |x_i| p_i$) і позначається $M\xi$ або $M(\xi)$:

$$M\xi = \sum_i x_i p_i . \tag{1}$$

Нехай випадкова величина ξ – неперервна із щільністю $p_\xi(x)$, тоді

$$M\xi = \int_{-\infty}^{+\infty} x p_\xi(x) dx , \tag{2}$$

якщо інтеграл $\int_{-\infty}^{+\infty} |x| p_\xi(x) dx$ – збіжний.

Часто математичне сподівання називають середнім значенням випадкової величини. Для математичного сподівання використовують і таке позначення: $E\xi$.

Якщо відома функція розподілу $F_\xi(x)$, то обидва ці випадки можна об’єднати і подати математичне сподівання інтегралом Стільтьєса

$$M\xi = \int_{-\infty}^{+\infty} x dF_\xi(x) . \tag{3}$$

Знову ж, за умови абсолютної збіжності інтегралу.

Відзначимо, що математичне сподівання визначають і за допомогою інтеграла Лебега

$$M\xi = \int_\Omega \xi(\omega) P(d\omega) .$$

Приклади.

1. Математичне сподівання випадкової величини, що має біномний розподіл. Нехай ξ дорівнює числу появ події в n незалежних експериментах. Тоді ξ приймає значення $k = 0,1,...,n$ із ймовірностями $p_k = P_n(k) = C_n^k p^k q^{n-k}$. Тому

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

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

$$= np \sum_{l=0}^{n-1} P_{n-1}(l) = np .$$

Отже, для біномного розподілу $M\xi = np$.

2. Математичне сподівання випадкової величини, що має розподіл Пуассона. Випадкова величина ξ має розподіл Пуассона, якщо вона приймає значення $k = 0, 1, 2, \ldots$ з ймовірностями $p_k = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \lambda > 0$. Тоді

$$M\xi = \sum_{k=0}^{\infty} k p_k = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda .$$

3. Математичне сподівання випадкової величини, що має рівномірний на відрізку $[a,b]$ розподіл. Щільність розподілу випадкової величини ξ має вигляд:

$$p_\xi(x) = \begin{cases} 0, & x \notin [a,b], \\ \dfrac{1}{b-a}, & x \in [a,b]. \end{cases}$$

Тому за формулою (2)

$$M\xi = \int_{-\infty}^{+\infty} x p_\xi(x) dx = \int_{-\infty}^{a} x \cdot 0 dx + \int_{a}^{b} x \frac{1}{b-a} dx + \int_{b}^{+\infty} x \cdot 0 dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{a+b}{2} .$$

4. Математичне сподівання випадкової величини, що має нормальний розподіл. Випадкова величина ξ має нормальний розподіл, якщо її щільність має вигляд

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-a)^2}{2\sigma^2} \right).$$

Із (2) одержуємо:

$$M\xi = \int_{-\infty}^{+\infty} x p_\xi(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x \exp\left(-\frac{(x-a)^2}{2\sigma^2} \right) dx = \left| \frac{x-a}{\sigma} = t \right| =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (t\sigma + a) \exp\left(-\frac{t^2}{2} \right) dt = \frac{1}{\sqrt{2\pi}} \left(\sigma \int_{-\infty}^{+\infty} t \exp\left(-\frac{t^2}{2} \right) dt + a \int_{-\infty}^{+\infty} \exp\left(-\frac{t^2}{2} \right) dt \right) = a,$$

$$\text{бо} \int_{-\infty}^{+\infty} t \exp\left(-\frac{t^2}{2} \right) dt = 0, \quad \int_{-\infty}^{+\infty} \exp\left(-\frac{t^2}{2} \right) dt = \sqrt{2\pi} .$$

Отже, параметр a, що входить в щільність нормального розподілу, є $M\xi$.

5. Математичне сподівання випадкової величини, що має розподіл Коші. Щільність розподілу Коші має вигляд:

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

Тому із (2) $M\xi = \int_{-\infty}^{+\infty} x \frac{1}{\pi} \frac{1}{1+x^2} dx$. Але такий невласний інтеграл є розбіжним, тому для випадкової величини, що має розподіл Коші, математичне сподівання не існує.

4.1.2. Властивості математичного сподівання (випадок дискретно розподілених випадкових величин).

1. $M\xi = \sum\limits_{\omega \in \Omega} \xi(\omega) P(\omega)$.

Позначимо $H_i = \{\omega : \xi(\omega) = x_i\}$, тоді $H_1 + H_2 + \dots = \Omega$ і $H_i \bigcap H_j = \varnothing$, $i \neq j$, тому

$$\sum\limits_{\omega \in \Omega} \xi(\omega) P(\omega) = \sum\limits_i \sum\limits_{\omega \in H_i} \xi(\omega) P(\omega) = \sum\limits_i x_i \sum\limits_{\omega \in H_i} P(\omega) = \sum\limits_i x_i P(H_i) = \sum\limits_i x_i p_i = M\xi \cdot$$

2. Нехай $I_A(\omega) = \begin{cases} 0, \omega \notin A, \\ 1, \omega \in A \end{cases}$ – індикатор випадкової події A. Тоді $MI_A = P(A)$, тобто математичне сподівання індикатора випадкової події дорівнює ймовірності подій.

Дійсно, $MI_A = 0 \cdot P(\overline{A}) + 1 \cdot P(A) = P(A)$.

3. $MC = C$, де $C = const$.

4. $M(C\xi) = CM\xi$.

Справедливість властивості випливає із означення: $M(C\xi) = \sum\limits_i (Cx_i) p_i =$

$= C\sum\limits_i x_i p_i = CM\xi$.

5. $M(\xi + \eta) = M\xi + M\eta$.

За першою властивістю

$$M(\xi + \eta) = \sum\limits_{\omega \in \Omega} (\xi(\omega) + \eta(\omega)) P(\omega) = \sum\limits_{\omega \in \Omega} \xi(\omega) P(\omega) + \sum\limits_{\omega \in \Omega} \eta(\omega) P(\omega) =$$

$= M\xi + M\eta$.

Застосуємо дану властивість для знаходження математичного сподівання біномного розподілу. Нехай ξ_i – число появ події в i-ому експерименті, тоді ξ_i приймає два можливих значення: 0 з ймовірністю q і 1 з ймовірністю p. Тоді $M\xi_i = 0 \cdot q + 1 \cdot p = p$. Число появ події в n експериментах $\xi = \xi_1 + \dots + \xi_n$, тому $M\xi = M\xi_1 + \dots + M\xi_n = np$.

6. Якщо $\xi \le \eta$, то $M\xi \le M\eta$. $|M\xi| \le M|\xi|$. Якщо $\xi \ge 0$ і $M\xi = 0$, то $P\{\xi = 0\} = 1$.

За першою властивістю $M(\eta - \xi) = \sum\limits_{\omega \in \Omega} (\eta(\omega) - \xi(\omega)) P(\omega) \ge 0$, бо $\eta(\omega) - \xi(\omega) \ge 0$. А за властивостями 4 і 5 $\mathrm{M}(\eta - \xi) = \mathrm{M}\eta - \mathrm{M}\xi \ge 0$, тому $\mathrm{M}\eta \ge \mathrm{M}\xi$. Друга нерівність випливає із властивості 1. Якщо $\xi \ge 0$ і $\mathrm{M}\xi = 0$, то $\sum\limits_{\omega \in \Omega} \xi(\omega) P(\omega) = 0$, а це можливо, коли для будь-якого ω $\xi(\omega) P(\omega) = 0$. Із умови $P(\omega) > 0$ випливає $\xi(\omega) = 0$. Тоді $P\{\omega : \xi(\omega) = 0\} = 1$.

7. Нехай $\eta = g(\xi)$, тоді $M\eta = Mg(\xi) = \sum\limits_i g(x_i) p_i$.

За властивістю 1 $Mg(\xi) = \sum\limits_{\omega \in \Omega} g(\xi(\omega)) P(\omega)$. Аналогічно доведенню властивості 1, введемо події $H_i = \{\omega : \xi(\omega) = x_i\}$, тоді

$$\mathrm{M}g(\xi) = \sum\limits_i \sum\limits_{\omega \in H_i} g(\xi(\omega)) P(\omega) = \sum\limits_i \sum\limits_{\omega \in H_i} g(x_i) P(\omega) = \sum\limits_i g(x_i) \sum\limits_{\omega \in H_i} P(\omega) =$$

$$= \sum\limits_i g(x_i) P(H_i) = \sum\limits_i g(x_i) p_i .$$

8. Якщо ξ і η – незалежні, то $M(\xi\cdot\eta)=M\xi\cdot M\eta$.

Із рівностей $\xi=\sum_i x_i I_{\{\xi=x_i\}}$, $\eta=\sum_j y_j I_{\{\eta=y_j\}}$ одержуємо

$$\xi\cdot\eta=\sum_{i,j} x_i y_j I_{\{\xi=x_i\}} I_{\{\eta=y_j\}}=\sum_{i,j} x_i y_j I_{\{\xi=x_i,\eta=y_j\}}.$$

За властивостями 2 і 5, врахувавши незалежність випадкових величин, одержуємо

$$M(\xi\cdot\eta)=\sum_{i,j} x_i y_j P\{\xi=x_i,\eta=y_j\}=\sum_{i,j} x_i y_j P(x_i)P(y_j)=$$

$$=\sum_i x_i P(x_i)\sum_j y_j P(y_j)=M\xi\cdot M\eta.$$

9. Якщо функція $g(x)$ випукла вниз на R, то для будь-якої випадкової величини ξ

$$M(g(\xi))\ge g(M\xi).$$

Для доведення використаємо той факт, що для будь-якої випуклої функції $g(x)$ і будь-якої точки a знайдеться така стала C, що для всіх x $g(x)\ge g(a)+C(x-a)$. Покладемо в цій нерівності $x=\xi,\ \ a=M\xi$ і, взявши математичне сподівання обох частин нерівності $g(\xi)\ge g(M\xi)+C(\xi-M\xi)$, при врахуванні рівності $M(\xi-M\xi)=M\xi-M\xi=0$, одержимо $M(g(\xi))\ge g(M\xi)$.

10. *Нерівність Ляпунова*. Якщо $0<\alpha\le\beta$, де $\alpha,\beta\in R$, то

$$\left(M|\xi|^\alpha\right)^{\frac{1}{\alpha}}\le\left(M|\xi|^\beta\right)^{\frac{1}{\beta}}.$$

Для доведення розглянемо функцію $g(x)=x^{\frac{\beta}{\alpha}},\ x>0$. Тоді $g'(x)=\frac{\beta}{\alpha}x^{\frac{\beta}{\alpha}-1}$, а $g''(x)=\frac{\beta}{\alpha}\left(\frac{\beta}{\alpha}-1\right)x^{\frac{\beta}{\alpha}-2}\ge 0$, отже $g(x)$ – випукла вниз. Тоді за властивістю 9 для будь-якої випадкової величини $\eta\ge 0$ $M(g(\eta))\ge g(M\eta)$ або $M\left(\eta^{\frac{\beta}{\alpha}}\right)\ge(M\eta)^{\frac{\beta}{\alpha}}$. Якщо підставимо в останню нерівність $\eta=|\xi|^\alpha$ і піднесемо обидві частини одержаної нерівності $M\left(|\xi|^\beta\right)\ge\left(M|\xi|^\alpha\right)^{\frac{\beta}{\alpha}}$ до степеня $\frac{1}{\beta}$, то одержимо необхідну нерівність.

11. *Нерівність Коші-Буняковського*. Для будь-яких випадкових величин ξ і η

$$\left(M(\xi\cdot\eta)\right)^2\le M\left(\xi^2\right)\cdot M\left(\eta^2\right).$$

Оскільки для довільного a $(a\xi+\eta)^2\ge 0$, то і $M(a\xi+\eta)^2\ge 0$. Підносячи до квадрату, одержимо квадратну нерівність $a^2M\left(\xi^2\right)+2aM(\xi\cdot\eta)+M\left(\eta^2\right)\ge 0$, що справедлива для довільного a, а це можливо, коли дискримінант $\left(2M(\xi\cdot\eta)\right)^2-4M\left(\xi^2\right)\cdot M\left(\eta^2\right)\le 0$.

4.1.3. Доведення деяких властивостей математичного сподівання для неперервних випадкових величин. Властивості 1, 2, 3 стосуються дискретних випадкових величин, а властивості 4 – 11 справедливі і для неперервних випадкових величин. Доведемо деякі із цих властивостей.

4. $\mathrm{M}\left(C\xi\right)=C\mathrm{M}\xi$.

При $C=0$ рівність правильна. Розглянемо випадок, коли $C>0$. Тоді $F_{C\xi}\left(x\right)=P\left\{C\xi<x\right\}=P\left\{\xi<\dfrac{x}{C}\right\}=F_{\xi}\left(\dfrac{x}{C}\right)$, а $p_{C\xi}\left(x\right)=p_{\xi}\left(\dfrac{x}{C}\right)\dfrac{1}{C}$. Тому

$$M\left(C\xi\right)=\int_{-\infty}^{\infty}xp_{C\xi}\left(x\right)dx=\int_{-\infty}^{\infty}xp_{\xi}\left(\frac{x}{C}\right)\frac{1}{C}dx=\left|\frac{x}{C}=t\right|=\int_{-\infty}^{\infty}Ctp_{\xi}\left(t\right)dt=CM\xi .$$

Аналогічно розглядається випадок $C<0$.

5. $\mathrm{M}\left(\xi+\eta\right)=\mathrm{M}\xi+\mathrm{M}\eta$.

Якщо $\left(\xi,\eta\right)$ має щільність розподілу $p\left(x_1,x_2\right)$, то щільність розподілу суми $\xi+\eta$ має вигляд $p_{\xi+\eta}\left(x\right)=\int_{-\infty}^{+\infty}p\left(x-x_2,x_2\right)dx_2=\int_{-\infty}^{+\infty}p\left(x_1,x-x_1\right)dx_1$. Тому за формулою (2)

$$M\left(\xi+\eta\right)=\int_{-\infty}^{+\infty}xp_{\xi+\eta}\left(x\right)dx=\int_{-\infty}^{+\infty}xdx\int_{-\infty}^{+\infty}p\left(x_1,x-x_1\right)dx_1 .$$

Змінимо порядок інтегрування і зробимо заміну $x-x_1=u,\quad dx=du$ у внутрішньому інтегралі, тоді $M\left(\xi+\eta\right)=\int_{-\infty}^{+\infty}dx_1\int_{-\infty}^{+\infty}xp\left(x_1,x-x_1\right)dx=$

$$=\int_{-\infty}^{+\infty}dx_1\int_{-\infty}^{+\infty}\left(u+x_1\right)p\left(x_1,u\right)du=\int_{-\infty}^{+\infty}x_1dx_1\int_{-\infty}^{+\infty}p\left(x_1,u\right)du+\int_{-\infty}^{+\infty}dx_1\int_{-\infty}^{+\infty}up\left(x_1,u\right)du=M\xi+M\eta,$$

бо $\int_{-\infty}^{+\infty}x_1dx_1\int_{-\infty}^{+\infty}p\left(x_1,u\right)du=\int_{-\infty}^{+\infty}x_1p_{\xi}(x_1)dx_1=M\xi$, а $\int_{-\infty}^{+\infty}dx_1\int_{-\infty}^{+\infty}up\left(x_1,u\right)du=$

$$=\int_{-\infty}^{+\infty}udu\int_{-\infty}^{+\infty}p\left(x_1,u\right)dx_1=\int_{-\infty}^{+\infty}up_{\eta}(u)du=M\eta .$$

7. Якщо $g(x)$ – борелєва функція і $\eta=g\left(\xi\right)$, тоді

$$M\eta=Mg\left(\xi\right)=\int_{-\infty}^{+\infty}g\left(x\right)p_{\xi}\left(x\right)dx=\int_{-\infty}^{+\infty}g\left(x\right)dF_{\xi}\left(x\right)$$

Нехай $g(x)$ – зростаюча. Тоді із пункту 3.5 $p_{\eta}\left(x\right)=p_{\xi}\left(g^{-1}\left(x\right)\right)\dfrac{1}{g'\left(x\right)}$.

В інтегралі $M\eta=\int_{-\infty}^{+\infty}xp_{\eta}(x)dx=\int_{-\infty}^{+\infty}xp_{\xi}\left(g^{-1}(x)\right)\dfrac{1}{g'(x)}dx$, зробимо заміну $g^{-1}(x)=y$, тоді $M\eta=\int_{-\infty}^{+\infty}g(y)p_{\xi}(y)dy$. Аналогічно розглядається випадок спадної функції. В загальному випадку ця властивість випливає із властивостей інтеграла Лебега.

8. Якщо ξ і η – незалежні, то $M\left(\xi\cdot\eta\right)=M\xi\cdot M\eta$.

Використаємо вигляд щільності розподілу добутку випадкових величин із пункту 3.6.3, із врахуванням незалежності, маємо $p_{\xi\eta}(x)=\int\limits_{-\infty}^{+\infty}\frac{1}{|u|}p_\xi(x/u)\,p_\eta(u)\,du$. Тоді

$$M(\xi\eta)=\int\limits_{-\infty}^{+\infty}xp_{\xi\eta}(x)\,dx=\int\limits_{-\infty}^{+\infty}xdx\int\limits_{-\infty}^{+\infty}\frac{1}{|u|}p_\xi(x/u)\,p_\eta(u)\,du\cdot$$

Змінимо порядок інтегрування, тоді

$$M(\xi\eta)=\int\limits_{-\infty}^{+\infty}\frac{1}{|u|}p_\eta(u)\,du\int\limits_{-\infty}^{+\infty}xp_\xi(x/u)\,dx=\int\limits_{0}^{+\infty}\frac{1}{u}p_\eta(u)\,du\int\limits_{-\infty}^{+\infty}xp_\xi(x/u)\,dx-\int\limits_{-\infty}^{0}\frac{1}{u}p_\eta(u)\,du\int\limits_{-\infty}^{+\infty}xp_\xi(x/u)\,dx\cdot$$

Після виконання заміни $x=uy,\quad dx=udy$ у внутрішньому інтегралі, одержимо:

$$M(\xi\eta)=\int\limits_{0}^{+\infty}up_\eta(u)\,du\int\limits_{-\infty}^{+\infty}yp_\xi(y)\,dy-\int\limits_{-\infty}^{0}up_\eta(u)\,du\int\limits_{+\infty}^{-\infty}yp_\xi(y)\,dy=\int\limits_{-\infty}^{+\infty}up_\eta(u)\,du\int\limits_{-\infty}^{+\infty}yp_\xi(y)\,dy=M\xi\cdot M\eta\,.$$

Властивість 6 випливає із зображення математичного сподівання інтегралом Лебега

$$M\xi=\int\limits_{\Omega}\xi(\omega)P(d\omega)$$

і властивостей інтеграла Лебега. Властивості 9 – 11 випливають із властивості 6 і властивостей інтеграла Лебега.

Вправи.

1. Випадкова величина ξ розподілена за законом Пуассона з параметром $\lambda>0$. Знайти $M\left(\frac{1}{\xi+1}\right)$.
2. Знайти $M\xi$ та $P\{|\xi-M\xi|\le 1\}$, якщо дискретна випадкова величина ξ задана законом розподілу

ξ	-2	-1	0	1
P	0,4	0,3	0,1	0,2

3. Випадкова величина ξ має гамма-розподіл: $p_\xi(x)=\begin{cases}0, & x\le 0;\\ \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, & x>0,\end{cases}$ $(\lambda>0,\quad \alpha>0)$. Знайти $M\xi$.

4.2. Дисперсія випадкової величини

Математичне сподівання квадрата відхилення випадкової величини від її математичного сподівання називається *дисперсією* (позначається $D\xi$):

$$D\xi=M\left(\xi-M\xi\right)^2. \tag{4}$$

Квадратний корінь із дисперсії називається *середнім квадратичним відхиленням* або *стандартним відхиленням*: $\sigma_\xi=\sqrt{D\xi}$.

Дисперсія характеризує розсіювання значень випадкової величини відносно її математичного сподівання.

Розглянемо деякі властивості дисперсії.

1. Дисперсія, що характеризує розсіювання відносно математичного сподівання, є найменшою характеристикою розсіювання відносно будь-якого центру C, тобто:

$$D\xi=\min_C M\left(\xi-C\right)^2.$$

Дійсно, за властивостями математичного сподівання

$$M\left(\xi - C\right)^2 = M\left(\xi - M\xi + M\xi - C\right)^2 =$$

$$= M\left(\left(\xi - M\xi\right)^2 + 2\left(\xi - M\xi\right)\left(M\xi - C\right) + \left(M\xi - C\right)^2\right) = D\xi + \left(M\xi - C\right)^2,$$

бо $\left(M\xi - C\right) = const$ і $M\left(\xi - M\xi\right) = 0$.

Для будь-якого C права частина одержаної рівності $D\xi + \left(M\xi - C\right)^2 \ge D\xi$. Мінімум $D\xi$ досягається, якщо $C = M\xi$.

В результаті доведення властивості 1 одержуємо таку формулу:

$$D\xi = M\left(\xi - C\right)^2 - \left(M\xi - C\right)^2,$$

що справедлива для будь-якого C. Зокрема, покладаючи в останній формулі $C = 0$, одержимо:

$$D\xi = M\xi^2 - \left(M\xi\right)^2. \tag{5}$$

2. $D\xi \ge 0$. $DC = 0$. Якщо $D\xi = 0$, то з ймовірністю 1 ξ - стала.

Перше очевидне, оскільки $D\xi = M\left(\xi - M\xi\right)^2 \ge 0$. Нехай $\xi = C$, тоді $M\xi = C$, а $M\xi^2 = C^2$, тому $D\xi = C^2 - C^2 = 0$. Якщо $D\xi = 0$, то із нерівності $\left(\xi - M\xi\right)^2 \ge 0$ за властивістю 6 математичного сподівання $P\{(\xi - M\xi)^2 = 0\} = 1$, тому і $P\{\xi - M\xi = 0\} = 1$ або $P\{\xi = M\xi\} = 1$.

Із цієї властивості і (5) одержуємо такі нерівності:

$$\left|M\xi\right| \le \sqrt{M\xi^2},$$

$$D\xi \le M\xi^2.$$

3. Для будь-яких a і b $D\left(a\xi + b\right) = a^2 D\xi$.

За означенням дисперсії $D(a\xi + b) = M(a\xi + b - M(a\xi + b))^2 =$

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

4. $D\left(\xi + \eta\right) = D\xi + D\eta + 2M\left(\left(\xi - M\xi\right)\left(\eta - M\eta\right)\right)$.

За властивостями математичного сподівання

$$D\left(\xi + \eta\right) = M\left(\xi + \eta - M\xi - M\eta\right)^2 = M\left(\left(\xi - M\xi\right) + \left(\eta - M\eta\right)\right)^2 =$$

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

Величина

$$M\left(\left(\xi - M\xi\right)\left(\eta - M\eta\right)\right) = \operatorname{cov}\left(\xi, \eta\right)$$

називається коваріацією випадкових величин ξ і η. Тому

$$D\left(\xi \pm \eta\right) = D\xi + D\eta \pm 2\operatorname{cov}\left(\xi, \eta\right).$$

Для незалежних випадкових величин ξ і η

$$\operatorname{cov}\left(\xi, \eta\right) = M\left(\left(\xi - M\xi\right)\left(\eta - M\eta\right)\right) = M\left(\xi - M\xi\right) M\left(\eta - M\eta\right) = 0.$$

Тому, якщо випадкові величини ξ і η – незалежні, то

$$D\left(\xi + \eta\right) = D\xi + D\eta.$$

Якщо випадкові величини $\xi_1, ..., \xi_n$ – незалежні і однаково розподілені, $\sigma^2 = D\xi_i$, то

$$D\left(\frac{\xi_1 + ... + \xi_n}{n}\right) = \frac{\sigma^2}{n}.$$

Із властивості 11 математичного сподівання випливає нерівність

$$\left|\operatorname{cov}\left(\xi, \eta\right)\right| \le \sqrt{D\xi \cdot D\eta}.$$

Із властивостей математичного сподівання випливає, що дисперсію можна обчислити за формулами

$$D\xi = \int_{-\infty}^{+\infty} \left(x - M\xi\right)^2 dF\left(x\right) = \int_{-\infty}^{+\infty} x^2 dF\left(x\right) - \left(\int_{-\infty}^{+\infty} x dF\left(x\right)\right)^2 .$$

Якщо ξ – дискретна і приймає значення x_k із ймовірностями p_k, то

$$D\xi = \sum_k \left(x_k - M\xi\right)^2 p_k = \sum_k x_k^2 p_k - \left(\sum_k x_k p_k\right)^2 ,$$

якщо величина ξ має щільність $p(x)$, то

$$D\xi = \int_{-\infty}^{+\infty} \left(x - M\xi\right)^2 p(x)dx = \int_{-\infty}^{+\infty} x^2 p(x)dx - \left(\int_{-\infty}^{+\infty} xp(x)dx\right)^2 .$$

Розглянемо приклади обчислення дисперсії.

1. Біномний розподіл. Нехай ξ дорівнює числу появ події в n незалежних експериментах. $\text{M}\xi = np$. Нехай ξ_i – число появ події в i-ому експерименті, тоді ξ_i приймає два можливих значення: 0 з ймовірністю q і 1 з ймовірністю p;

ξ_i	0	1
P	q	p

Тоді $M\xi_i = 0 \cdot q + 1 \cdot p = p$, $M\xi_i^2 = 0 \cdot q + 1 \cdot p = p$. За формулою (5) $D\xi_i = M\xi_i^2 - \left(M\xi_i\right)^2 = p - p^2 = pq$. Число появ події в n експериментах $\xi = \xi_1 + ... + \xi_n$ і випадкові величини $\xi_1, ..., \xi_n$ – незалежні. Тому

$$D\xi = D\xi_1 + \cdots + D\xi_n = npq.$$

2. Розподіл Пуассона. Випадкова величина ξ приймає значення $k = 0,1,2,...$ з ймовірностями $p_k = e^{-\lambda} \dfrac{\lambda^k}{k!}$, $\lambda > 0$. Тоді $M\xi = \lambda$. А

$$M\xi^2 = \sum_{k=0}^{\infty} k^2 p_k = \sum_{k=1}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{\left(k-1\right)!} = \lambda e^{-\lambda} \left(\sum_{k=1}^{\infty} \left(k-1\right) \frac{\lambda^{k-1}}{\left(k-1\right)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{\left(k-1\right)!}\right) =$$

$$= \lambda e^{-\lambda} \left(\sum_{k=2}^{\infty} \frac{\lambda^{k-1}}{\left(k-2\right)!} + e^{\lambda}\right) = \lambda e^{-\lambda} \left(\lambda e^{\lambda} + e^{\lambda}\right) = \lambda\left(1 + \lambda\right).$$

Тому

$$D\xi = M\xi^2 - \left(M\xi\right)^2 = \lambda\left(1 + \lambda\right) - \lambda^2 = \lambda .$$

3. Рівномірний розподіл на відрізку $[a,b]$. Щільність розподілу випадкової величини ξ має вигляд:

$$p_\xi\left(x\right) = \begin{cases} 0, & x \notin [a,b], \\ \dfrac{1}{b-a}, & x \in [a,b]. \end{cases}$$

Ми уже знайшли $M\xi = \dfrac{a+b}{2}$. Знайдемо

$$M\xi^2 = \int\limits_{-\infty}^{+\infty} x^2 p(x)\,dx = \int\limits_a^b x^2 \frac{1}{b-a}\,dx = \frac{1}{b-a}\frac{b^3-a^3}{3} = \frac{b^2+ab+a^2}{3}.$$

Тоді

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(a-b)^2}{12}.$$

4. Нормальний розподіл. Нехай випадкова величина ξ має нормальний розподіл з параметрами a і σ^2. Її щільність $p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$ і ми уже знайшли $M\xi = a$. За формулою для обчислення дисперсії

$$D\xi = M(\xi - M\xi)^2 = \int\limits_{-\infty}^{+\infty} (x-a)^2 \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)dx = \left|\frac{x-a}{\sigma} = t\right| =$$

$$= \frac{\sigma^2}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} t^2 \exp\left(-\frac{t^2}{2}\right)dt = \frac{2\sigma^2}{\sqrt{2\pi}}\int\limits_0^{+\infty} t^2 \exp\left(-\frac{t^2}{2}\right)dt.$$

Після виконання підстановки $\frac{t^2}{2} = z$, $t = \sqrt{2z}$, $dt = \frac{\sqrt{2}dz}{2\sqrt{z}}$, одержуємо:

$$D\xi = \frac{2\sigma^2}{\sqrt{2\pi}}\int\limits_0^{+\infty} 2ze^{-z}\frac{\sqrt{2}dz}{2\sqrt{z}} = \frac{2\sigma^2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right) = \sigma^2.$$

Отже, нормальний закон розподілу повністю визначається параметрами $a = M\xi$ і $\sigma^2 = D\xi$.

Вправи.

4. Нехай випадкова величина ξ набуває значень $0, \pm1, \pm2, ..., \pm n$ з ймовірностями $P\{\xi = i\} = \frac{1}{2n+1}$. Знайти $M\xi$ і $D\xi$.
5. Знайти дисперсію дискретної випадкової величини за законом розподілу із вправи 2.
6. Дискретна випадкова величина ξ приймає два можливих значення x_1 і x_2 $(x_1 < x_2)$. $P\{\xi = x_1\} = 0,2$. Знайти закон розподілу ξ, якщо $M\xi = 2,6$; $\sigma_\xi = 0,8$.
7. Тривалість безвідмовної роботи приладу є випадкова величина τ, що має показниковий розподіл із щільністю $p(x) = \begin{cases} 0, & x < 0; \\ \lambda e^{-\lambda x}, & x \ge 0, \end{cases}$ $\lambda > 0$. Знайти $M\tau$, $D\tau$, $P\{\tau < M\tau\}$.
8. Задано функцію розподілу випадкової величини $F_\xi(x) = \begin{cases} 0, & x \le 1; \\ \ln x, & 1 < x \le e; \\ 1, & x > e. \end{cases}$ Знайти $M\xi$, $D\xi$, $M(\ln\xi)$.

4.3. Нерівність Чебишова

1. Нехай функція $g(x)$ – невід'ємна і неспадна на множині значень випадкової величини ξ і $g(x)>0$ при $x>0$, існує $M\left|g(\xi)\right|$, тоді для будь-якого $\varepsilon>0$

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

Нехай $F(x)$ – функція розподілу ξ. Оскільки $g(x)$ – неспадна і невід'ємна, то за властивістю математичного сподівання функції від випадкової величини

$$M\left(g(\xi)\right)=\int_{-\infty}^{\infty} g(x)dF(x) \geq \int_{\varepsilon}^{\infty} g(x)dF(x) \geq g(\varepsilon)\int_{\varepsilon}^{\infty} dF(x)=g(\varepsilon)P\{\xi \geq \varepsilon\}.$$

Поділивши останню нерівність на $g(\varepsilon)$, дістанемо необхідну нерівність.

2. Розглянемо частинний випадок одержаної нерівності. Нехай випадкова величина ξ – невід'ємна, має скінченне математичне сподівання, тоді для будь-якого $\varepsilon>0$

$$P\{\xi \geq \varepsilon\} \leq \frac{M\xi}{\varepsilon}.$$

Цю нерівність називають ще нерівністю Маркова. Вона випливає із попередньої нерівності, якщо покладемо $g(x)=x$.

3. Якщо випадкова величина ξ має скінченне математичне сподівання, тоді для будь-якого $\varepsilon>0$

$$P\{|\xi| \geq \varepsilon\} \leq \frac{M|\xi|}{\varepsilon}$$

Випливає із застосування попередньої до випадкової величини $\eta=|\xi|$.

4.Якщо випадкова величина ξ має скінченну дисперсію, то для будь-якого $\varepsilon>0$

$$P\{|\xi - M\xi| \geq \varepsilon\} \leq \frac{D\xi}{\varepsilon^2}.$$

Застосуємо першу нерівність до випадкової величини $\eta=|\xi-M\xi|$ і $g(x)=x^2$:

$$P\{|\xi - M\xi| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2}M\left(\xi-M\xi\right)^2=\frac{D\xi}{\varepsilon^2}.$$

Якщо перейдемо до протилежної події, то одержимо нерівність

$$P\{|\xi - M\xi| < \varepsilon\} \geq 1-\frac{D\xi}{\varepsilon^2}.$$

Приклад. Нехай випадкова величина ξ приймає два можливих значення з ймовірностями $P\{\xi=-1\}=0,5$ і $P\{\xi=1\}=0,5$. Тоді $M\xi=0$, $D\xi=1$. За нерівністю Чебишова $p_1=P\{|\xi|\geq 1\}\leq 1$. Але безпосередньо $p_1=P\{\xi=-1\}+P\{\xi=1\}=1$. Тобто, нерівність Чебишова дає оцінку, яку покращити неможливо.

Вправа.

9. Використовуючи нерівність Чебишова, оцінити ймовірність того, що при 1000 киданнях монети число випадань герба μ буде лежати між 450 і 550.

4.4. Моменти випадкових величин. Деякі інші числові характеристики

Моментом k-го порядку $(k \ge 0)$ випадкової величини ξ (початковим моментом k-го порядку) називається величина $\alpha_k = M\xi^k$, якщо вона існує. В цьому випадку існує також величина $\beta_k = M\left(|\xi|^k\right)$, яка називається *абсолютним моментом k-го порядку*. Моменти випадкової величини $\xi - M\xi$ називаються *центральними моментами* k-го порядку: $\mu_k = M\left(\xi - M\xi\right)^k$. Величина $\nu_k = M\left(|\xi - M\xi|^k\right)$ називається *центральним абсолютним моментом* k-го порядку. Зокрема, $\mu_1 = 0$, $\mu_2 = D\xi$.

Використовуючи введені позначення, нерівність Ляпунова може бути записана у вигляді: при $0 < k < m$, $\beta_k^{\frac{1}{k}} \le \beta_m^{\frac{1}{m}}$.

Обчислимо центральні моменти випадкової величини ξ, що має нормальний розподіл з параметрами a і σ^2:

$$\mu_n = M\left(\xi - a\right)^n = \int_{-\infty}^{+\infty} \left(x-a\right)^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(x-a\right)^2}{2\sigma^2}\right) dx .$$

Після виконання підстановки $\frac{x-a}{\sigma} = z$, одержимо:

$$\mu_n = \frac{\sigma^n}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^n e^{-\frac{z^2}{2}} dz .$$

Якщо n – непарне, то $\mu_n = 0$. Нехай $n = 2k$. Тоді підстановка $z = \sqrt{2t}$, аналогічно до обчислення дисперсії, дає

$$\mu_{2k} = \frac{2^k \sigma^{2k}}{\sqrt{\pi}} \int_0^{+\infty} t^{k-\frac{1}{2}} e^{-t} dt = \frac{2^k \sigma^{2k}}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right) = \left(2k-1\right)!!\sigma^{2k} .$$

За допомогою центральних моментів вводяться такі характеристики, як *асиметрія* розподілу $A_s = \frac{\mu_3}{\sigma^3}$ і *ексцес* розподілу $e_s = \frac{\mu_4}{\sigma^4} - 3$. Для нормального розподілу ці характеристики рівні нулю. Легко показати, що асиметрія і ексцес інваріантні відносно лінійного перетворення випадкової величини: для будь-яких дійсних $a \ne 0$ і b випадкові величини ξ і $\eta = a\xi + b$ мають однакову асиметрію і однаковий ексцес.

Якщо функція розподілу $F(x)$ неперервна і строго монотонна, то для довільного $q \in (0,1)$ рівняння $F\left(x\right) = q$ має єдиний розв'язок x_q. Цей розв'язок x_q називають *q-квантиллю* або *квантиллю порядку q* розподілу $F(x)$. Але для довільної функції розподілу $F(x)$ рівняння $F\left(x\right) = q$ може мати багато розв'язків, тоді будь-який із них називають *q-квантиллю* розподілу $F(x)$, а може і не мати розв'язків. В загальному випадку *q-квантиллю* розподілу $F(x)$ називається таке число x_q, що $F(x_q) \le q$, а $F(x_q + 0) \ge q$. Інколи q-квантиль визначають за допомогою рівності $x_q = \max\left\{x : F(x) \le q\right\}$.

Якщо функція розподілу $F(x)$ випадкової величини ξ неперервна і строго монотонна і $q' > q$, то $P\left\{x_q < \xi < x_{q'}\right\} = q' - q$.

У випадку $q=\frac{1}{2}$, $x_{\frac{1}{2}}$ називають *медіаною* розподілу. Якщо $x_{\frac{1}{2}}$ – медіана розподілу, то $F(x_{\frac{1}{2}})\le\frac{1}{2}$ і $F(x_{\frac{1}{2}}+0)\ge\frac{1}{2}$.

Якщо випадкова величина дискретна, то те із її значень, яке вона приймає із найбільшою ймовірністю, називають *модою* розподілу, якщо ж випадкова величина неперервна, то модою називають такі значення x, при яких щільність розподілу має максимум.

Вправа.

10. Знайти асиметрію і ексцес для рівномірного розподілу на відрізку $[a,b]$.

4.5. Числові характеристики багатовимірних випадкових величин

Нехай $\xi=\left(\xi_1,...,\xi_m\right)$ – $m-$вимірна випадкова величина. Тоді її математичним сподіванням називається вектор $M\xi=\left(M\xi_1,...,M\xi_m\right)$, якщо $M\xi_1,...,M\xi_m$ існують.

Величини $M(\xi_1^{k_1}\cdot...\cdot\xi_m^{k_m})$, де $k_i\ge 0,\ \ i=1,...,m$, $k_1+...+k_m=k$, називають змішаними моментами порядку k випадкових величин $\xi_1,...,\xi_m$. Серед змішаних моментів важливу роль відіграють моменти другого порядку – коваріації випадкових величин ξ_i і ξ_j:

$$M\left(\xi_i-M\xi_i\right)\left(\xi_j-M\xi_j\right)=\operatorname{cov}\left(\xi_i,\xi_j\right).$$

Для введення характеристик випадкового вектора $\xi=\left(\xi_1,...,\xi_m\right)$ з функцією розподілу $F_{\xi_1,...,\xi_m}\left(x_1,...,x_m\right)$ використаємо рівність

$$Mg(\xi)=\int\limits_{-\infty}^{\infty}...\int\limits_{-\infty}^{\infty}g(x_1,...,x_m)dF_{\xi_1,...,\xi_m}\left(x_1,...,x_m\right),$$

де $g(x_1,...,x_m)$ – довільна борелєва функція, для якої принаймні одна частина рівності має зміст.

Зокрема,

$$M\xi_i=\int\limits_{-\infty}^{\infty}...\int\limits_{-\infty}^{\infty}x_i dF_{\xi_1,...,\xi_m}\left(x_1,...,x_m\right),$$

$$M(\xi_1^{k_1}\cdot...\cdot\xi_m^{k_m})=\int\limits_{-\infty}^{\infty}...\int\limits_{-\infty}^{\infty}x_1^{k_1}\cdot...\cdot x_m^{k_m}dF_{\xi_1,...,\xi_m}\left(x_1,...,x_m\right).$$

Для неперервного розподілу із щільністю $p\left(x_1,...,x_m\right)$

$$M(\xi_1^{k_1}\cdot...\cdot\xi_m^{k_m})=\int\limits_{-\infty}^{\infty}...\int\limits_{-\infty}^{\infty}x_1^{k_1}\cdot...\cdot x_m^{k_m}p\left(x_1,...,x_m\right)dx_1\cdot...\cdot dx_m,$$

а якщо врахуємо, що $\int\limits_{-\infty}^{\infty}...\int\limits_{-\infty}^{\infty}p\left(x_1,...,x_m\right)dx_1...dx_{i-1}dx_{i+1}...dx_m=p_{\xi_i}(x_i)$, то

$$M\xi_i^k = \int\limits_{-\infty}^{\infty} ... \int\limits_{-\infty}^{\infty} x_i^k p\left(x_1,...,x_m\right) dx_1...dx_m = \int\limits_{-\infty}^{\infty} x_i^k dx_i \int\limits_{-\infty}^{\infty} ... \int\limits_{-\infty}^{\infty} p\left(x_1,...,x_m\right) dx_1...dx_{i-1}dx_{i+1}...dx_m =$$

$$= \int\limits_{-\infty}^{\infty} x_i^k p_{\xi_i}\left(x_i\right) dx_i .$$

Якщо для величин ξ_i існують $M\xi_i^2$, $i = 1,...,m$, то матриця $B = \left\|b_{ij}\right\|_{i,j=\overline{1,m}}$, де $b_{ij} = \operatorname{cov}\left(\xi_i, \xi_j\right)$, називається *коваріаційною* (*кореляційною*) матрицею випадкових величин $\xi_1,...,\xi_m$:

$$B = \begin{pmatrix} b_{11} & ... & b_{1m} \\ ... & ... & ... \\ b_{m1} & ... & b_{mm} \end{pmatrix}.$$

Коваріаційна матриця додатньо визначена.

Введемо ще одну досить важливу характеристику

$$r_{ij} = \frac{\operatorname{cov}\left(\xi_i, \xi_j\right)}{\sqrt{D\xi_i \cdot D\xi_j}},$$

яку називають *коефіцієнтом кореляції* випадкових величин ξ_i і ξ_j. Враховуючи, що

$$\operatorname{cov}\left(\xi_i, \xi_j\right) = M\left(\xi_i - M\xi_i\right)\left(\xi_j - M\xi_j\right) = M\left(\xi_i\xi_j - \xi_i M\xi_j - \xi_j M\xi_i + M\xi_i M\xi_j\right) =$$
$$= M\left(\xi_i\xi_j\right) - M\xi_i M\xi_j,$$

то для коефіцієнта кореляції справедлива формула

$$r_{ij} = \frac{M\left(\xi_i\xi_j\right) - M\xi_i M\xi_j}{\sqrt{D\xi_i \cdot D\xi_j}}.$$

Розглянемо деякі властивості коефіцієнта кореляції. Для спрощення розглянемо двовимірний вектор $\left(\xi, \eta\right)$, а коефіцієнт кореляції позначимо $r = \dfrac{\operatorname{cov}\left(\xi, \eta\right)}{\sqrt{D\xi \cdot D\eta}}$.

Із визначення коваріації випливає, що для незалежних випадкових величин $\operatorname{cov}\left(\xi, \eta\right) = 0$, тому і $r = 0$. Якщо $r = 0$, то випадкові величини називаються *некорельованими.* Для нормального розподілу некорельованість означає незалежність.

$\left|r\right| \le 1$, якщо $r = \pm 1$, то з ймовірністю 1 між випадковими величинами ξ і η існує лінійний зв'язок. Навпаки, якщо ξ і η зв'язані лінійною залежністю, то $r = \pm 1$. Тому r можна розглядати як міру лінійної залежності між величинами ξ і η.

Для доведення цього введемо позначення: $\xi_1 = \dfrac{\xi - M\xi}{\sqrt{D\xi}}$; $\eta_1 = \dfrac{\eta - M\eta}{\sqrt{D\eta}}$. Тоді

$M\xi_1 = 0$, $M\eta_1 = 0$, $D\xi_1 = M\left(\dfrac{\xi - M\xi}{\sqrt{D\xi}}\right)^2 = 1$, $D\eta_1 = 1$, а

$$M\left(\xi_1\eta_1\right) = M\frac{\xi - M\xi}{\sqrt{D\xi}}\frac{\eta - M\eta}{\sqrt{D\eta}} = \operatorname{cov}\left(\xi_1, \eta_1\right) = r.$$

Із властивостей дисперсії одержуємо

$$D\left(\xi_1 \pm \eta_1\right) = D\xi_1 + D\eta_1 \pm 2\operatorname{cov}\left(\xi_1, \eta_1\right) = 2\left(1 \pm r\right) \ge 0.$$

Якщо $r \geq 0$, то із нерівності $1-r \geq 0$ випливає, що $r \leq 1$, а якщо $r \leq 0$, то із нерівності $1+r \geq 0$ одержимо $r \geq -1$. Отже, $|r| \leq 1$.

Нехай $\eta = a\xi + b$, тоді $\operatorname{cov}(\xi,\eta) = \mathrm{M}\left((\xi - \mathrm{M}\xi)(\eta - \mathrm{M}\eta)\right) =$
$= \mathrm{M}\left((\xi - \mathrm{M}\xi)\left(a\xi + b - (a\mathrm{M}\xi + b)\right)\right) = aM(\xi - \mathrm{M}\xi)^2 = aD\xi$, а $D\eta = a^2 D\xi$, тому

$$r = \frac{aD\xi}{\sqrt{D\xi \cdot a^2 D\xi}} = \frac{a}{|a|} = \pm 1.$$

Нехай $r = 1$. Тоді із рівності $D(\xi_1 - \eta_1) = 2(1-r) = 0$ випливає, що з ймовірністю 1 $\xi_1 - \eta_1 = C$, а із рівностей $M\xi_1 = 0,\quad M\eta_1 = 0$ одержуємо $0 = M\xi_1 - M\eta_1 = C$. Отже, $\xi_1 = \eta_1$ або $\dfrac{\xi - \mathrm{M}\xi}{\sigma_\xi} = \dfrac{\eta - \mathrm{M}\eta}{\sigma_\eta}$. Звідки, $\eta = \dfrac{\sigma_\eta}{\sigma_\xi}(\xi - M\xi) + M\eta$, тобто випадкові величини зв'язані лінійною залежністю.

Нехай $r = -1$. Тоді із рівності $D(\xi_1 + \eta_1) = 2(1+r) = 0$, аналогічно попередньому, одержуємо, що з ймовірністю 1 $\xi_1 + \eta_1 = 0$. Звідки, $\eta = -\dfrac{\sigma_\eta}{\sigma_\xi}(\xi - M\xi) + M\eta$, тобто знову випадкові величини зв'язані лінійною залежністю.

Вправи.

11. Щільність сумісного розподілу вектора (ξ,η) $p(x,y) = \begin{cases} 0, & x^2 + y^2 > R^2; \\ \dfrac{1}{\pi R^2}, & x^2 + y^2 \leq R^2. \end{cases}$

    Знайти коефіцієнт кореляції.
12. Знайти коефіцієнт кореляції дискретного випадкового вектора (ξ,η) за його розподілом

$\xi \setminus \eta$	-1	0	1
-1	0,1	0,15	0,2
1	0,15	0,25	0,15

4.6. Умовне математичне сподівання. Регресія

Нехай випадковий вектор (ξ,η) – дискретний із розподілом $P(x_i, y_j) = P\{\xi = x_i, \eta = y_j\}$ і нехай $P_{\eta/\xi}(y_j / x_i) = \dfrac{P(x_i, y_j)}{P_\xi(x_i)}$ – умовний розподіл η при умові $\xi = x_i$. Тоді величина

$$M(\eta / \xi = x_i) = M(\eta / x_i) = \sum_j y_j P_{\eta/\xi}(y_j / x_i)$$

називається *умовним математичним сподіванням* випадкової величини η при умові $\xi = x_i$.

Нехай випадковий вектор (ξ,η) – неперервний із щільністю $p(x,y)$ і нехай $p_{\eta/\xi}(y/x) = \dfrac{p(x,y)}{p_\xi(x)}$ – умовна щільність розподілу випадкової величини η при умові $\xi = x$, тоді величина

$$M\left(\eta/\xi=x\right)=M\left(\eta/x\right)=\int_{-\infty}^{+\infty} yp_{\eta/\xi}\left(y/x\right)dy$$

називається *умовним математичним сподіванням* випадкової величини η при умові $\xi=x$.

Якщо випадковий вектор (ξ,η) – дискретний із розподілом $P\left(x_i,y_j\right)$, то

$$M\left(\eta/x_i\right)=\sum_j y_j\frac{P\left(x_i,y_j\right)}{P_\xi\left(x_i\right)}=\frac{\sum_j y_jP\left(x_i,y_j\right)}{\sum_j P\left(x_i,y_j\right)},$$

а якщо випадковий вектор (ξ,η) – неперервний із щільністю $p(x,y)$, то

$$M\left(\eta/x\right)=\int_{-\infty}^{+\infty} y\frac{p(x,y)}{p_\xi(x)}dy=\frac{\int_{-\infty}^{+\infty} yp(x,y)dy}{\int_{-\infty}^{+\infty} p(x,y)dy}.$$

Якщо величини ξ і η – незалежні, то $M\left(\eta/x\right)=M\eta$. Аналогічно вводиться умовне математичне сподівання $M\left(\xi/\eta=y\right)$.

Ми можемо вважати $M\left(\eta/\xi=x_k\right)$ значеннями випадкової величини $M\left(\eta/\xi\right)$, що є функцією від ξ і рівною $M\left(\eta/\xi=x_k\right)$ при $\xi=x_k$. Випадкову величину $M\left(\eta/\xi\right)$ називають *умовним математичним сподіванням* випадкової величини η при заданому ξ. Від цієї величини можна знайти математичне сподівання

$$M\left(M\left(\eta/\xi\right)\right)=\sum_k P\{\xi=x_k\}M\left(\eta/\xi=x_k\right)=$$

$$=\sum_k P_\xi\left(x_k\right)\frac{\sum_j y_jP\left(x_k,y_j\right)}{P_\xi\left(x_k\right)}=\sum_k\sum_j y_jP\left(x_k,y_j\right)=M\eta.$$

Аналогічно у неперервному випадку

$$M\left(M\left(\eta/\xi\right)\right)=\int_{-\infty}^{+\infty} M\left(\eta/x\right)p_\xi(x)dx=\int_{-\infty}^{+\infty} p_\xi(x)dx\int_{-\infty}^{+\infty} y\frac{p(x,y)}{p_\xi(x)}dy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yp(x,y)dxdy=M\eta.$$

Отже, справедлива рівність

$$M\left(M\left(\eta/\xi\right)\right)=M\eta.$$

Очевидно, умовне математичне сподівання $M\left(\eta/x\right)$ є функцією від x. Цю функцію називають *регресією* η на ξ, а її графік називають *кривою регресії*.

Умовне математичне сподівання функції $g(\xi,\eta)$ при умові $\xi=x$ у випадку неперервного розподілу визначається рівністю

$$M\left(g(\xi,\eta)/\xi=x\right)=\int_{-\infty}^{+\infty} g(x,y)\frac{p(x,y)}{p_\xi(x)}dy=\frac{\int_{-\infty}^{+\infty} g(x,y)p(x,y)dy}{p_\xi(x)}.$$

Якщо цю рівність помножити на $p_\xi(x)$ і проінтегрувати по x, то одержимо:

$$\int_{-\infty}^{+\infty} M\left(g(\xi,\eta)/\xi=x\right)p_\xi(x)dx=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)p(x,y)dxdy=Mg(\xi,\eta).$$

Тобто, $M\left(M\left(g(\xi,\eta)/\xi\right)\right)=Mg(\xi,\eta)$.

Умовною дисперсією випадкової величини η при умові $\xi = x$ називається величина

$$D(\eta / x) = M\left(\left(\eta - M(\eta / x)\right)^2 / \xi = x\right).$$

У випадку неперервного розподілу

$$D(\eta / x) = \int_{-\infty}^{+\infty} \left(y - M(\eta / x)\right)^2 \frac{p(x,y)}{p_\xi(x)} dy .$$

Дисперсія характеризує на скільки при фіксованому x значення величини η розсіяні відносно лінії регресії.

Якщо повернемось до прикладу щільності умовного нормального розподілу із попереднього розділу

$$p_{\eta/\xi}(y/x) = \frac{1}{\sqrt{2\pi}\sigma_\eta\sqrt{1-\rho^2}} \exp\left(-\frac{\left(y - b - \rho\dfrac{\sigma_\eta}{\sigma_\xi}(x-a)\right)^2}{2\sigma_\eta^2\left(1-\rho^2\right)}\right),$$

то одержимо, що нормальна регресія є лінійною: $M(\eta / x) = b + \rho\dfrac{\sigma_\eta}{\sigma_\xi}(x-a)$. Умовна дисперсія дорівнює $D(\eta / x) = \sigma_\eta^2\left(1-\rho^2\right)$.

Лінія регресії має одну важливу властивість мінімальності. Серед всіх функцій $g(x)$ мінімум виразу $M\left(\left(\eta - g(\xi)\right)^2\right)$ досягається при $g(x) = M(\eta / x)$.

В рівності

$$M\left(\eta - g(\xi)\right)^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(y - g(x)\right)^2 p(x,y)\,dxdy = \int_{-\infty}^{+\infty} p_\xi(x)dx \int_{-\infty}^{+\infty} \left(y - g(x)\right)^2 p_{\eta/\xi}(y/x)\,dy$$

права частина буде мінімальна, якщо внутрішній інтеграл для довільного x буде мінімальним. Перетворимо його до вигляду

$$\int_{-\infty}^{+\infty} \left(y - g(x)\right)^2 p_{\eta/\xi}(y/x)\,dy = \int_{-\infty}^{+\infty} \left(y - M(\eta / x) + M(\eta / x) - g(x)\right)^2 p_{\eta/\xi}(y/x)\,dy =$$

$$= \int_{-\infty}^{+\infty} \left(y - M(\eta / x)\right)^2 p_{\eta/\xi}(y/x)\,dy + 2\left(M(\eta / x) - g(x)\right) \int_{-\infty}^{+\infty} \left(y - M(\eta / x)\right) p_{\eta/\xi}(y/x)\,dy +$$

$$+ \left(M(\eta / x) - g(x)\right)^2 \int_{-\infty}^{+\infty} p_{\eta/\xi}(y/x)\,dy =$$

$$= \int_{-\infty}^{+\infty} \left(y - M(\eta / x)\right)^2 p_{\eta/\xi}(y/x)\,dy + \left(M(\eta / x) - g(x)\right)^2 .$$

Отже, $M\left(\eta - g(\xi)\right)^2$ буде мінімальним, якщо $g(x) = M(\eta / x)$.

Розглянемо задачу про знаходження такої лінійної функції $g(x) = ax + b$, для якої $M\left(\eta - g(\xi)\right)^2$ буде мінімальним. Тоді її називають *лінійною середньоквадратичною регресією*. Для цього параметри a і b необхідно вибрати так, щоб $M\left(\eta - a\xi - b\right)^2$ було мінімальним (позначимо $M\xi = m_1$, $M\eta = m_2$):

$$M\left(\eta - a\xi - b\right)^2 = M\left(\eta - m_2 + m_2 - m_1 a - b - a\left(\xi - m_1\right)\right)^2 =$$

$$= M\left(\eta - m_2\right)^2 + \left(m_2 - m_1 a - b\right)^2 + a^2 M\left(\xi - m_1\right)^2 +$$

$$+2M\left(\left(\eta-m_2\right)\left(m_2-m_1a-b\right)\right)-2aM\left(\left(\xi-m_1\right)\left(\eta-m_2\right)\right)-2a\left(m_2-m_1a-b\right)M\left(\xi-m_1\right)=$$

$$=\sigma_\eta^2+a^2\sigma_\xi^2+\left(m_2-m_1a-b\right)^2-2a\sigma_\xi\sigma_\eta r=$$

$$=\left(a\sigma_\xi-\sigma_\eta r\right)^2+\sigma_\eta^2\left(1-r^2\right)+\left(m_2-m_1a-b\right)^2.$$

Щоб математичне сподівання було мінімальним, необхідно a і b вибрати так, щоб $\left(a\sigma_\xi-\sigma_\eta r\right)^2=0$ і $\left(m_2-m_1a-b\right)^2=0$. Одержимо: $a=\dfrac{\sigma_\eta}{\sigma_\xi}r,\quad b=m_2-m_1a$, $g\left(x\right)=r\dfrac{\sigma_\eta}{\sigma_\xi}x+m_2-m_1\dfrac{\sigma_\eta}{\sigma_\xi}r$. Отже, пряма середньоквадратичної регресії задається рівнянням

$$y=m_2+r\frac{\sigma_\eta}{\sigma_\xi}\left(x-m_1\right).$$

Коефіцієнт $\dfrac{\sigma_\eta}{\sigma_\xi}r$ називають коефіцієнтом середньоквадратичної регресії η на ξ, $\min M\left(\eta-a\xi-b\right)^2=\sigma_\eta^2\left(1-r^2\right)$ називається залишковою дисперсією.

Вираз $M\left(\eta-a\xi-b\right)^2$ можна подати у вигляді

$$M\left(\eta-a\xi-b\right)^2=M\left(\eta-M\left(\eta/\xi\right)\right)^2+$$

$$+2M\left(\left(\eta-M\left(\eta/\xi\right)\right)\left(M\left(\eta/\xi\right)-a\xi-b\right)\right)+M\left(M\left(\eta/\xi\right)-a\xi-b\right)^2.$$

Другий доданок рівний нулю, бо

$$M\left(\left(\eta-M\left(\eta/\xi\right)\right)\left(M\left(\eta/\xi\right)-a\xi-b\right)\right)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(y-M\left(\eta/x\right)\right)\left(M\left(\eta/x\right)-ax-b\right)p\left(x,y\right)dxdy=$$

$$=\int_{-\infty}^{+\infty}\left(M\left(\eta/x\right)-ax-b\right)p_\xi\left(x\right)dx\int_{-\infty}^{+\infty}\left(y-M\left(\eta/x\right)\right)p_{\eta/\xi}\left(y/x\right)dy,$$

а внутрішній інтеграл рівний нулю. Тому для будь-яких a і b

$$M\left(\eta-a\xi-b\right)^2=M\left(\eta-M\left(\eta/\xi\right)\right)^2+M\left(M\left(\eta/\xi\right)-a\xi-b\right)^2.$$

Перший доданок правої частини не залежить від a і b, тому другий доданок буде мати мінімум при тих самих a і b, що і ліва частина. Це означає, що пряма середньоквадратичної регресії може розглядатись як пряма, яка дає найкраще середньоквадратичне наближення до кривої регресії.

Покладемо в рівності

$$M\left(\eta-a\xi-b\right)^2=M\left(\eta-M\left(\eta/\xi\right)\right)^2+M\left(M\left(\eta/\xi\right)-a\xi-b\right)^2,$$

що справедлива для будь-яких a і b, $a=0$, $b=m_2$, тоді одержимо:

$$D\eta=M\left(\eta-m_2\right)^2=M\left(\eta-M\left(\eta/\xi\right)\right)^2+M\left(M\left(\eta/\xi\right)-M\eta\right)^2.$$

Величину $\theta_{\eta/\xi}$, що визначається рівністю

$$\theta_{\eta/\xi}^2=\frac{M\left(M\left(\eta/\xi\right)-M\eta\right)^2}{D\eta},$$

називають кореляційним відношенням η на ξ.

Із рівності $\theta_{\eta/\xi}^2=1-\dfrac{M\left(\eta-M\left(\eta/\xi\right)\right)^2}{D\eta}$ випливають такі властивості кореляційного відношення: 1) $0\le\theta_{\eta/\xi}\le1$; 2) $\theta_{\eta/\xi}=0$ тоді і тільки тоді, коли $M\left(\eta/x\right)$ не залежить від x;

3) $\theta_{\eta/\xi}=1$ тоді і тільки тоді, коли розподіл вектора зосереджений на лінії регресії $y=M\left(\eta/x\right)$.

Якщо в рівності

$$M\left(\eta-a\xi-b\right)^2=M\left(\eta-M\left(\eta/\xi\right)\right)^2+M\left(M\left(\eta/\xi\right)-a\xi-b\right)^2,$$

що справедлива для будь-яких a і b, виберемо a і b, як коефіцієнти прямої середньо-квадратичної регресії і врахуємо, що в цьому випадку $M\left(\eta-a\xi-b\right)^2=\sigma_\eta^2\left(1-r^2\right)$, то одержимо $\sigma_\eta^2\left(1-r^2\right)=\sigma_\eta^2\left(1-\theta_{\eta/\xi}^2\right)+M\left(M\left(\eta/\xi\right)-a\xi-b\right)^2$. Звідки

$$\theta_{\eta/\xi}^2=r^2+\frac{1}{\sigma_\eta^2}M\left(M\left(\eta/\xi\right)-a\xi-b\right)^2.$$

Тому $\theta_{\eta/\xi}^2\geq r^2$.

Якщо регресія η на ξ лінійна, то $\theta_{\eta/\xi}^2=r^2$. Тобто, $\theta_{\eta/\xi}$ можна розглядати, як характеристику того, наскільки залежність між ξ і η близька до лінії регресії.

Вправи.

13. Знайти умовне математичне сподівання $M\left(\eta/\xi=x_i\right)$ за розподілом із вправи 12.
14. Задано щільність розподілу випадкового вектора (ξ,η) $p(x,y)=xe^{-x-xy}$, $x\geq 0,\ y\geq 0$. Знайти умовне математичне сподівання $M\left(\eta/x\right)$.
15. Знайти кореляційне відношення η на ξ за розподілом із вправи 12.

Розділ 5. Закон великих чисел

5.1. Деякі типи збіжностей послідовностей випадкових величин

5.1.1. Основні означення. Нехай задана послідовність випадкових величин $\xi_1,...,\xi_n,\ldots$, що визначені на одному і тому ж ймовірнісному просторі.

Послідовність ξ_n називається *збіжною за ймовірністю* до випадкової величини ξ, якщо для будь-якого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\xi_n - \xi\right| \geq \varepsilon\right\} = 0 \tag{1}$$

або $\lim\limits_{n\to\infty} P\left\{\left|\xi_n - \xi\right| < \varepsilon\right\} = 1$. Позначається $\xi_n \xrightarrow{p} \xi$.

Послідовність ξ_n називається збіжною до випадкової величини ξ *з ймовірністю 1* (*майже напевне*), якщо

$$P\left\{\omega : \xi_n(\omega) \to \xi(\omega)\right\} = 1 \tag{2}$$

або $P\left\{\omega : \xi_n(\omega) \not\to \xi(\omega)\right\} = 0$. Позначається $\xi_n \xrightarrow{м.н} \xi$ або $\xi_n \to \xi$ при $n \to \infty$ з ймовірністю 1.

Послідовність ξ_n називається збіжною до випадкової величини ξ *в середньому порядку* $r > 0$, якщо

$$\lim_{n\to\infty} M\left(\left|\xi_n - \xi\right|^r\right) = 0. \tag{3}$$

При $r = 1$ збіжність називається збіжністю в середньому, а при $r = 2$ – збіжністю в середньоквадратичному.

Нехай маємо послідовність функцій розподілу $F_n(x)$. Дана послідовність називається *збіжною в основному* до функції $F(x)$, якщо ця збіжність має місце в кожній точці неперервності функції $F(x)$. Якщо $F_n(x)$ - функція розподілу ξ_n, а $F(x)$ – функція розподілу випадкової величини ξ, то в цьому випадку кажуть, що послідовність випадкових величин ξ_n збігається *за розподілом* до випадкової величини ξ $\left(\xi_n \xrightarrow{d} \xi\right)$.

5.1.2. Властивості збіжності за ймовірністю.

Теорема 1. Нехай $\xi_n \xrightarrow{p} \xi$ і $f(x)$ – неперервна на R функція, тоді

$$f(\xi_n) \xrightarrow{p} f(\xi). \tag{4}$$

Доведення. Із неперервності на R функції $f(x)$ випливає, що для довільного $c > 0$ і для довільного $\varepsilon > 0$ існує таке $\delta > 0$, що як тільки $|x_1| \leq c$ і $|x_1 - x_2| < \delta$, то $\left|f(x_1) - f(x_2)\right| < \varepsilon$. Покладемо $x_1 = \xi$, $x_2 = \xi_n$ і розглянемо події $\left\{|\xi| \leq c\right\}$, $\left\{|\xi - \xi_n| < \delta\right\}$. Із одночасного настання цих двох подій випливає настання події $\left\{\left|f(\xi) - f(\xi_n)\right| < \varepsilon\right\}$, тобто $\left\{|\xi| \leq c\right\} \cap \left\{|\xi - \xi_n| < \delta\right\} \subset \left\{\left|f(\xi) - f(\xi_n)\right| < \varepsilon\right\}$, а для протилежних події будемо мати $\left\{\left|f(\xi) - f(\xi_n)\right| \geq \varepsilon\right\} \subset \left\{|\xi| > c\right\} \cup \left\{|\xi - \xi_n| \geq \delta\right\}$. Тому

$$P\left\{\left|f(\xi_n) - f(\xi)\right| \geq \varepsilon\right\} \leq P\left\{|\xi| > c\right\} + P\left\{|\xi - \xi_n| \geq \delta\right\}.$$

Із того, що $F(-c) \to 0$, а $F(c) \to 1$ при $c \to +\infty$ випливає, що для будь-якого $\varepsilon_1 > 0$ число c можна вибрати так, щоб $P\left\{|\xi| > c\right\} = F(-c) + 1 - F(c) < \dfrac{\varepsilon_1}{2}$. Зафіксуємо таке число

c. Розглянемо другий доданок: $P\{|\xi-\xi_n|\geq\delta\}$. За умовою $\xi_n \xrightarrow{p} \xi$, тому для заданого довільного $\varepsilon_1>0$ існує таке n_{ε_1}, що для всіх $n>n_{\varepsilon_1}$ $P\{|\xi-\xi_n|\geq\delta\}<\frac{\varepsilon_1}{2}$. Отже, для довільного $\varepsilon_1>0$ існує таке n_{ε_1}, що для всіх $n>n_{\varepsilon_1}$ $P\{|f(\xi_n)-f(\xi)|\geq\varepsilon\}<\varepsilon_1$, а це означає, що має місце (4). Теорема доведена.

Теорема 2. Нехай задано m-вимірну послідовність $\left(\xi_1^{(n)},...,\xi_m^{(n)}\right)$, при цьому $\xi_k^{(n)} \xrightarrow{p} \xi_k$ при $n\to\infty$ $(k=1,...,m)$, $f(x_1,...,x_m)$ – неперервна на R^m функція, тоді $f\left(\xi_1^{(n)},...,\xi_m^{(n)}\right) \xrightarrow{p} f(\xi_1,...,\xi_m)$.

Доведення аналогічне доведенню теореми 1.

Теорема 3. Нехай випадкові величини ξ_n мають функцію розподілу $F_n(x)$, а випадкова величина ξ – $F(x)$. Якщо $\xi_n \xrightarrow{p} \xi$, то послідовність $F_n(x)$ збігається в основному до $F(x)$.

Доведення. Нехай $A_n=\{\omega:|\xi_n(\omega)-\xi(\omega)|<\varepsilon\}=\{\omega:\xi-\varepsilon<\xi_n<\xi+\varepsilon\}$, де $\varepsilon>0$ - довільне. Оскільки $\xi_n \xrightarrow{p} \xi$, то

$$P\left(\overline{A_n}\right)=P\{|\xi_n-\xi|\geq\varepsilon\}\xrightarrow[n\to\infty]{}0. \tag{5}$$

Нехай x – довільна точка неперервності $F(x)$. Якщо настає подія A_n, то настає подія $\{\xi-\varepsilon<\xi_n\}$ і $\{\xi_n<x\}\cap A_n\subset\{\xi_n<x\}\cap\{\xi-\varepsilon<\xi_n\}=\{\xi-\varepsilon<\xi_n<x\}\subset\{\xi-\varepsilon<x\}$, тоді $\{\xi_n<x\}=(\{\xi_n<x\}\cap A_n)\cup(\{\xi_n<x\}\cap\overline{A_n})\subset\{\xi<x+\varepsilon\}\cup\overline{A_n}$. Тому

$$F_n(x)=P\{\xi_n<x\}\leq P\{\xi<x+\varepsilon\}+P\left\{\overline{A_n}\right\}.$$

Із даної нерівності і (5) випливає, що

$$\overline{\lim_{n\to\infty}}F_n(x)\leq F(x+\varepsilon). \tag{6}$$

Аналогічно, $A_n\subset\{\xi_n<\xi+\varepsilon\}$, тоді $\{\xi+\varepsilon<x\}\cap A_n\subset\{\xi_n<\xi+\varepsilon\}\cap\{\xi+\varepsilon<x\}=$
$=\{\xi_n<\xi+\varepsilon<x\}\subset\{\xi_n<x\}$, а $\{\xi+\varepsilon<x\}=(\{\xi+\varepsilon<x\}\cap A_n)\cup(\{\xi+\varepsilon<x\}\cap\overline{A_n})\subset$
$\subset\{\xi_n<x\}\cup\overline{A_n}$. Тому

$$P\{\xi<x-\varepsilon\}\leq P\{\xi_n<x\}+P\left\{\overline{A_n}\right\}=F_n(x)+P(\overline{A_n}).$$

Звідки, врахувавши (5), одержимо:

$$F(x-\varepsilon)\leq\underline{\lim_{n\to\infty}}F_n(x). \tag{7}$$

Із (6) і (7) випливає нерівність

$$F(x-\varepsilon)\leq\underline{\lim_{n\to\infty}}F_n(x)\leq\overline{\lim_{n\to\infty}}F_n(x)\leq F(x+\varepsilon).$$

У останній нерівності перейдемо до границі, коли $\varepsilon\to0$. За рахунок того, що x – точка неперервності $F(x)$, верхня і нижня границі співпадають, тому існує $\lim\limits_{n\to\infty}F_n(x)=F(x)$. Що і треба було довести.

Теорема 4. Якщо послідовність функцій розподілу $F_n(x)$ збігається в основному до $F(x)$, де $F(x)=\begin{cases}0, x\leq c,\\ 1, x>c,\end{cases}$ $c=const$, то $\xi_n \xrightarrow{p} c$.

Доведення. Нам треба показати, що

$$P\left\{\left|\xi_n - c\right| < \varepsilon\right\} \xrightarrow[n\to\infty]{} 1. \tag{8}$$

Але $P\left\{\left|\xi_n - c\right| < \varepsilon\right\} = P\left\{c - \varepsilon < \xi_n < c + \varepsilon\right\} = F_n\left(c+\varepsilon\right) - F_n\left(c-\varepsilon\right)$. Виберемо $\varepsilon > 0$ таким, щоб точки $c \pm \varepsilon$ були точками неперервності $F\left(x\right)$. Тоді $F_n\left(c+\varepsilon\right) \to 1$, а $F_n\left(c-\varepsilon\right) \to 0$ при $n \to \infty$. Тому із попередньої рівності випливає (8).

5.1.3. Збіжність з ймовірністю 1. За означенням $\xi_n \xrightarrow{м.н} \xi$, якщо

$$P\left\{\omega : \xi_n\left(\omega\right) \to \xi\left(\omega\right)\right\} = 1. \tag{9}$$

Тоді збіжність $\xi_n(\omega) \to \xi(\omega)$ при $n \to \infty$ означає, що для будь-якого $\varepsilon > 0$ існує $N > 0$ таке, що для всіх $n \geq N$ виконується нерівність $\left|\xi_n\left(\omega\right) - \xi\left(\omega\right)\right| < \varepsilon$. Для вияснення змісту даного речення, введемо деякі події. Виберемо $\varepsilon = \frac{1}{k}$ і позначимо $A_{nk} = \left\{\omega : \left|\xi_n\left(\omega\right) - \xi\left(\omega\right)\right| < \frac{1}{k}\right\}$. Але починаючи із N настають всі події A_{nk}, тому настає подія $\bigcap_{n=N}^{\infty} A_{nk} = B_{Nk}$. Так як існує $N > 0$ таке, що настає подія B_{Nk}, то настає і подія $\bigcup_{N=1}^{\infty} B_{Nk} = \bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty} A_{nk} = B_k$. І нарешті, для довільного $k > 0$ настає подія B_k, тому настає подія $\bigcap_{k=1}^{\infty} B_k = \bigcap_{k=1}^{\infty}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty} A_{nk}$. Таким чином, справедлива така рівність

$$\left\{\omega : \xi_n\left(\omega\right) \to \xi\left(\omega\right)\right\} = \bigcap_{k=1}^{\infty}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left\{\omega : \left|\xi_n\left(\omega\right) - \xi\left(\omega\right)\right| < \frac{1}{k}\right\} = \bigcap_{k=1}^{\infty} B_k . \tag{10}$$

Враховуючи (10), із (9) випливає, що збіжність $\xi_n \xrightarrow{м.н} \xi$ еквівалентна умові $P\left\{\bigcap_{k=1}^{\infty} B_k\right\} = 1$, а це еквівалентно тому, що

$$P\left(B_k\right) = P\left\{\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left\{\omega : \left|\xi_n\left(\omega\right) - \xi\left(\omega\right)\right| < \frac{1}{k}\right\}\right\} = 1, \tag{11}$$

бо якщо існує k_0 таке, що $P\left(B_{k_0}\right) < 1$, то із того, що $\bigcap_{k=1}^{\infty} B_k \subset B_{k_0}$ одержали б, що $P\left\{\bigcap_{k=1}^{\infty} B_k\right\} < 1$. Оскільки подія $B_k = \bigcup_{N=1}^{\infty} B_{Nk}$, а події $B_{Nk} = \bigcap_{n=N}^{\infty} A_{nk}$ утворюють зростаючу послідовність подій, тому за аксіомою неперервності із (11) одержимо $1 = P\left\{B_k\right\} = \lim_{N\to\infty} P\left\{\bigcap_{n=N}^{\infty}\left\{\omega : \left|\xi_n\left(\omega\right) - \xi\left(\omega\right)\right| < \frac{1}{k}\right\}\right\}$. Тому ймовірність протилежної події має прямувати до 0, тобто, $\lim_{N\to\infty} P\left\{\bigcup_{n=N}^{\infty}\left\{\omega : \left|\xi_n\left(\omega\right) - \xi\left(\omega\right)\right| \geq \frac{1}{k}\right\}\right\} = 0$. Отже, ми довели наступне твердження.

Теорема 5. Збіжність з ймовірністю 1 еквівалентна виконанню однієї із умов: для будь-якого $\varepsilon > 0$

$$\lim_{N\to\infty} P\left\{\bigcap_{n=N}^{\infty}\left\{\omega : \left|\xi_n\left(\omega\right) - \xi\left(\omega\right)\right| < \varepsilon\right\}\right\} = 1$$

або

$$\lim_{N\to\infty} P\left\{\bigcup_{n=N}^{\infty}\left\{\omega:\left|\xi_n(\omega)-\xi(\omega)\right|\ge\varepsilon\right\}\right\}=0.$$

Друга умова рівносильна умові:

$$\lim_{N\to\infty} P\left\{\omega:\sup_{n\ge N}\left|\xi_n(\omega)-\xi(\omega)\right|\ge\varepsilon\right\}=0,$$

а це означає, що $\sup_{k\ge n}\left|\xi_k(\omega)-\xi(\omega)\right|\xrightarrow{p}0$ при $n\to\infty$.

Наслідок 1. Із збіжності з ймовірністю 1 випливає збіжність за ймовірністю.

Із того, що для всіх $n\ge N$ $\left\{\left|\xi_n(\omega)-\xi(\omega)\right|\ge\varepsilon\right\}\subset\bigcup_{n=N}^{\infty}\left\{\omega:\left|\xi_n(\omega)-\xi(\omega)\right|\ge\varepsilon\right\}$ випливає, що

$$P\left\{\left|\xi_n(\omega)-\xi(\omega)\right|\ge\varepsilon\right\}\le P\left(\bigcup_{n=N}^{\infty}\left\{\omega:\left|\xi_n(\omega)-\xi(\omega)\right|\ge\varepsilon\right\}\right). \tag{12}$$

Якщо має місце збіжність з ймовірністю 1, то за теоремою 5 права частина в (12) прямує до нуля при $N\to\infty$, тому із нерівності (12) випливає, що $P\left\{\left|\xi_n(\omega)-\xi(\omega)\right|\ge\varepsilon\right\}\to 0$ при $n\to\infty$, а це означає збіжність за ймовірністю і справедливість наслідку.

Наслідок 2. Якщо для будь-якого $\varepsilon>0$ числовий ряд $\sum_{n=1}^{\infty}P\left\{\left|\xi_n-\xi\right|\ge\varepsilon\right\}$ є збіжним, то ξ_n збігається до випадкової величини ξ з ймовірністю 1.

Розглянемо нерівність

$$P\left(\bigcup_{n=N}^{\infty}\left\{\left|\xi_n-\xi\right|\ge\varepsilon\right\}\right)\le\sum_{n=N}^{\infty}P\left\{\left|\xi_n-\xi\right|\ge\varepsilon\right\}, \tag{13}$$

у якій права частина є залишком ряду $\sum_{n=1}^{\infty}P\left\{\left|\xi_n-\xi\right|\ge\varepsilon\right\}$. Якщо ряд збіжний, то його залишок прямує до нуля. Тому, за нерівністю (13) і теоремою 5, одержуємо справедливість наслідку 2.

Теорема 6. Якщо $\xi_n\xrightarrow{p}\xi$, то із послідовності ξ_n можна виділити підпослідовність, яка буде збігатися до випадкової величини ξ з ймовірністю 1.

Доведення. Дано: $P\left\{\left|\xi_n-\xi\right|\ge\varepsilon\right\}\xrightarrow[n\to\infty]{}0$. Це означає, що для довільного $\varepsilon_1>0$ існує таке n_{ε_1}, що для всіх $n>n_{\varepsilon_1}$ виконується нерівність $P\left\{\left|\xi_n-\xi\right|\ge\varepsilon\right\}<\varepsilon_1$. Покладемо $\varepsilon_1=\frac{1}{k^2}$, k – натуральне. Тоді існує n_k, для якого $P\left\{\left|\xi_{n_k}-\xi\right|\ge\varepsilon\right\}<\frac{1}{k^2}$. Ряд $\sum_{k=1}^{\infty}\frac{1}{k^2}$ збіжний, тому збіжний і ряд $\sum_{k=1}^{\infty}P\left\{\left|\xi_{n_k}-\xi\right|\ge\varepsilon\right\}$, а із попереднього наслідку випливає, що підпослідовність $\xi_{n_k}\to\xi$ з ймовірністю 1.

5.1.4. Лема Бореля-Кантеллі. Розглянемо довільну послідовність подій $A_1,\dots,A_n,\dots$. Позначимо через C подію, яка полягає у тому, що у послідовності подій $A_1,\dots,A_n,\dots$ настає нескінченна кількість подій. Тоді $C=\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k$.

Лема. 1) Якщо числовий ряд $\sum_{n=1}^{\infty} P(A_n)$ – збіжний, то $P(C)=0$, а якщо події $A_1,...,A_n,...$ – незалежні і $P(C)=0$, то ряд $\sum_{n=1}^{\infty} P(A_n)$ – збіжний.

2) Якщо події $A_1,...,A_n,...$ – незалежні, і ряд $\sum_{n=1}^{\infty} P(A_n)$ – розбіжний, то $P(C)=1$.

Доведення. Послідовність $\bigcup_{k=n}^{\infty} A_k$ – спадна, тому за аксіомою неперервності $P(C)=\lim_{n\to\infty} P\left\{\bigcup_{k=n}^{\infty} A_k\right\}$. В нерівності $P\left\{\bigcup_{k=n}^{\infty} A_k\right\} \le \sum_{k=n}^{\infty} P(A_k)$ права частина є залишок збіжного ряду, тому $\sum_{k=n}^{\infty} P(A_k) \to 0$ при $n\to\infty$, отже, $P(C)=\lim_{n\to\infty} P\left\{\bigcup_{k=n}^{\infty} A_k\right\}=0$. Розглянемо тепер подію $\overline{C}=\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\overline{A_k}$, що є сумою зростаючої послідовності подій $\bigcap_{k=n}^{\infty}\overline{A_k}$. За аксіомою неперервності

$$P\left(\overline{C}\right)=\lim_{n\to\infty} P\left\{\bigcap_{k=n}^{\infty}\overline{A_k}\right\}. \tag{14}$$

Використовуючи незалежність подій і нерівність $1+x\le e^x$, одержимо

$$P\left\{\bigcap_{k=n}^{\infty}\overline{A_k}\right\}=\prod_{k=n}^{\infty} P\left(\overline{A_k}\right)=\prod_{k=n}^{\infty}\left(1-P(A_k)\right)\le\prod_{k=n}^{\infty}\exp\left(-P(A_k)\right)=\exp\left(-\sum_{k=n}^{\infty} P(A_k)\right)\le 1. \tag{15}$$

Якщо $P(C)=0$, то $P\left(\overline{C}\right)=1$, тому із (14) випливає, що $P\left\{\bigcap_{k=n}^{\infty}\overline{A_k}\right\}\xrightarrow[n\to\infty]{}1$. Переходячи до границі в нерівності (15) при $n\to\infty$, одержуємо $\exp\left(-\sum_{k=n}^{\infty} P(A_k)\right)\to 1$, а $\sum_{k=n}^{\infty} P(A_k)\xrightarrow[n\to\infty]{}0$, звідки випливає, що ряд $\sum_{n=1}^{\infty} P(A_n)$ – збіжний.

Якщо ряд $\sum_{n=1}^{\infty} P(A_n)$ – розбіжний, то $\sum_{k=n}^{\infty} P(A_k)\xrightarrow[n\to\infty]{}+\infty$. В цьому випадку права частина нерівності $0\le P\left\{\bigcap_{k=n}^{\infty}\overline{A_k}\right\}\le\exp\left(-\sum_{k=n}^{\infty} P(A_k)\right)$, що одержана із (15), прямуватиме до нуля, тому і $P\left\{\bigcap_{k=n}^{\infty}\overline{A_k}\right\}\xrightarrow[n\to\infty]{}0$. Звідки випливає, що $P\left(\overline{C}\right)=0$, а $P(C)=1$. Лема доведена.

Нехай $A_n^{\varepsilon}=\left\{\left|\xi_n-\xi\right|\ge\varepsilon\right\}$, $C=\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k^{\varepsilon}$ – подія, яка полягає в тому, що у послідовності $\left\{A_n^{\varepsilon}\right\}$ настає нескінченна кількість подій. Порівнюючи із (11), одержуємо, що умова $\xi_n\xrightarrow{м.н.}\xi$ еквівалентна умові $P(C)=0$. За доведеною лемою для цього достатньо збіжності ряду $\sum_{n=1}^{\infty} P\left(A_n^{\varepsilon}\right)$. Таким чином, із леми Бореля-Кантеллі ми одержали результат, що еквівалентний наслідку 2.

5.1.5. Властивості збіжності у середньому.

Теорема 7. Із збіжності у середньому порядку $r>0$ випливає збіжність за ймовірністю ($\xi_n \xrightarrow{r} \xi \Rightarrow \xi_n \xrightarrow{p} \xi$).

Доведення. Використаємо нерівність Чебишова

$$P\{|\xi_n-\xi|\geq\varepsilon\}=P\{|\xi_n-\xi|^r\geq\varepsilon^r\}\leq\frac{M\left(|\xi_n-\xi|^r\right)}{\varepsilon^r},$$

де $\varepsilon>0$ – довільне, але фіксоване. Якщо $\xi_n \xrightarrow{r} \xi$, то чисельник правої частини нерівності прямує до 0, тому і права частина теж прямує до 0, а звідси, в свою чергу, випливає, що і ліва частина прямує до 0, а це рівносильне умові $\xi_n \xrightarrow{P} \xi$.

Теорема 8. Якщо існують $M\xi_n^2$, $M\xi^2$ і $\xi_n\to\xi$ в середньоквадратичному, то $\xi_n\to\xi$ в середньому, $M\xi_n\to M\xi$ і $M\xi_n^2\to M\xi^2$.

Доведення. Із нерівності Ляпунова для моментів

$$M|\xi_n-\xi|\leq(M|\xi_n-\xi|^2)^{\frac{1}{2}}$$

випливає, що із збіжності в середньоквадратичному випливає збіжність в середньому. Оскільки $|M\xi_n-M\xi|\leq M|\xi_n-\xi|$, то із збіжності в середньому випливає, що $M\xi_n\to M\xi$.

За нерівністю $\left|M\xi_n^2-M\xi^2\right|=|M(\xi_n-\xi)(\xi_n+\xi)|\leq(M(\xi_n-\xi)^2)^{\frac{1}{2}}(M(\xi_n+\xi)^2)^{\frac{1}{2}}\leq$

$\leq(M(\xi_n-\xi)^2)^{\frac{1}{2}}(2(M\xi_n^2+M\xi^2))^{\frac{1}{2}}$ із умов теореми випливає, що із збіжності в середньоквадратичному випливає третє твердження теореми.

Із наведених результатів випливає, що зв’язок між різними видами збіжності можна подати такою схемою:

$$\begin{matrix}\xi_n \xrightarrow{м.н.} \xi \Rightarrow \\ \xi_n \xrightarrow{r} \xi \Rightarrow\end{matrix}\xi_n \xrightarrow{p} \xi \Rightarrow \xi_n \xrightarrow{d} \xi.$$

Наведемо деякі приклади, які показують, що в загальному випадку зворотне місця не має.

1. Нехай простір елементарних подій $\Omega=[0,1]$. $F=B([0,1])$ - борелева σ-алгебра підмножин відрізка $[0,1]$ і нехай $\xi_n^i=I_{A_n^i}(\omega)$, $i=1,...,n$ – індикатори подій $A_n^i=\left\{\omega\in\left[\frac{i-1}{n},\frac{i}{n}\right]\right\}$, тоді $P\left(A_n^i\right)=\frac{1}{n}$. Розглянемо послідовність випадкових величин $\xi_1^1,\xi_2^1,\xi_2^2,\xi_3^1,\xi_3^2,\xi_3^3,...$ заданих на проміжку $[0,1]$ і випадкову величину $\xi(\omega)=0$, якщо $\omega\in[0,1]$. Тоді $P\{|\xi_n^i-\xi|\geq\varepsilon\}=\frac{1}{n}\xrightarrow[n\to\infty]{}0$. Тому побудована послідовність збігається за ймовірністю. Але $P\{\omega:\xi_n^i\to\xi\}=P\{\omega:\xi_n^i=0\}=P\left\{\omega\notin\left[\frac{i-1}{n},\frac{i}{n}\right]\right\}=1-\frac{1}{n}\neq1$, тому $\xi_n^i\not\to\xi=0$ з ймовірністю 1. Тобто, із збіжності $\xi_n \xrightarrow{P} \xi$ не випливає збіжність $\xi_n \xrightarrow{м.н.} \xi$.

Дослідимо збіжність побудованої послідовності в середньому: $M\left(|\xi_n^i-\xi|^r\right)=\int\limits_{\frac{i-1}{n}}^{\frac{i}{n}}1dx=\frac{1}{n}\xrightarrow[n\to\infty]{}0$. Одержимо, що $\xi_n \xrightarrow{r} \xi \not\Rightarrow \xi_n \xrightarrow{м.н.} \xi$.

2. Розглянемо послідовність $\xi_n(\omega) = \begin{cases} e^n, & 0 \le \omega \le \dfrac{1}{n}, \\ 0, & \dfrac{1}{n} < \omega \le 1. \end{cases}$. Тоді $\xi_n(\omega) \to \xi(\omega) = 0$, $\omega \in (0,1]$ і $P\{\omega : \xi_n(\omega) \to \xi(\omega)\} = P\{\omega \in (0,1]\} = 1$. Отже, $\xi_n \xrightarrow{м.н.} \xi$, але $M\left(\left|\xi_n - \xi\right|^r\right) = \int_0^{\frac{1}{n}} e^{nr} dx = \dfrac{e^{nr}}{n} \xrightarrow[n\to\infty]{} \infty$, тобто, $\xi_n \not\xrightarrow{r} \xi$.

Вправа.

1. Нехай $\{\xi_n, n \ge 1\}$ – послідовність незалежних випадкових величин, що мають рівномірний на $(0,1)$ розподіл. Довести, що $\left(e^n \prod_{k=1}^{n} \xi_k\right)^{\frac{1}{n}} \xrightarrow[n\to\infty]{} 1$ за ймовірністю.

5.2. Закон великих чисел. Посилений закон великих чисел

5.2.1. Закон великих чисел. Нехай маємо послідовність випадкових величин $\xi_1, ..., \xi_n, ...$, розглянемо нову послідовність випадкових величин $\eta_n = \dfrac{1}{n}(\xi_1 + \cdots + \xi_n)$. Кажуть, що для послідовності $\xi_1, ..., \xi_n, ...$ має місце *закон великих чисел*, якщо існує числова послідовність a_n така, що $\eta_n - a_n \xrightarrow[n\to\infty]{p} 0$.

Якщо ця збіжність має місце з ймовірністю 1, то кажуть, що має місце *посилений закон великих чисел*.

Теорема 1 (Чебишова). Якщо випадкові величини $\xi_1, ..., \xi_n, ...$ – незалежні, мають скінченні математичні сподівання $M\xi_i = a_i$ і дисперсії $D\xi_i$, що рівномірно обмежені сталою C $(D\xi_i \le C, i = 1, 2, ...)$, то $\dfrac{\xi_1 + ... + \xi_n}{n} - \dfrac{a_1 + ... + a_n}{n} \xrightarrow[n\to\infty]{p} 0$ $\left(\dfrac{1}{n}\sum_{k=1}^{n}(\xi_k - a_k) \xrightarrow[n\to\infty]{p} 0\right)$.

Доведення. Використаємо нерівність Чебишова у вигляді $P\{|\xi - M\xi| \ge \varepsilon\} \le \dfrac{D\xi}{\varepsilon^2}$ і застосуємо її до випадкової величини $\eta_n = \dfrac{1}{n}(\xi_1 + \cdots + \xi_n)$. Врахуємо, що

$$M\eta_n = \frac{1}{n} M(\xi_1 + \cdots + \xi_n) = \frac{1}{n}(a_1 + \cdots + a_n),$$

$$D\eta_n = \frac{1}{n^2} D(\xi_1 + \cdots + \xi_n) = \frac{1}{n^2}(D\xi_1 + \cdots + D\xi_n) \le \frac{nC}{n^2} = \frac{C}{n}.$$

Тоді $P\{|\eta_n - M\eta_n| \ge \varepsilon\} \le \dfrac{D\eta_n}{\varepsilon^2} \le \dfrac{C}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0$. А це означає, що $\eta_n - M\eta_n \xrightarrow{p} 0$ при $n \to \infty$. Теорема доведена.

Теорема 2 (Бернуллі). Нехай μ_n - число появ події в n випробуваннях Бернуллі, p – ймовірність настання події в кожному випробуванні. Тоді $\dfrac{\mu_n}{n} \xrightarrow[n\to\infty]{p} p$ (відносна частота події збігається за ймовірністю до ймовірності появи події).

Доведення. Нехай ξ_i – число появ події в i - му випробуванні, тоді $\mu_n = \xi_1 + \cdots + \xi_n$, $M\xi_i = p,\ D\xi_i = pq \le \frac{1}{4}$, $M\mu_n = np$, $M\left(\frac{\mu_n}{n}\right) = p$. Отже, виконуються умови теореми 1, тому $\frac{\mu_n}{n} \xrightarrow[n\to\infty]{p} p$.

Аналізуючи доведення теореми 1, бачимо, що можна не вимагати незалежності подій і послабити її умови щодо дисперсій.

Теорема 3 (Маркова). Нехай випадкові величини $\xi_1,...,\xi_n,...$ мають скінченні математичні сподівання $M\xi_i = a_i$. Тоді, якщо $\frac{1}{n^2} D\sum_{k=1}^{n}\xi_k \xrightarrow[n\to\infty]{} 0$, то має місце закон великих чисел: $\frac{1}{n}\sum_{k=1}^{n}\left(\xi_k - a_k\right) \xrightarrow{p} 0$ при $n \to \infty$.

5.2.2. Необхідна і достатня умова для закону великих чисел.

Теорема 4. Нехай випадкові величини $\xi_1,...,\xi_n,...$ – незалежні, мають скінченні математичні сподівання $M\xi_i$. Для того, щоб

$$\eta_n = \frac{1}{n}\sum_{i=1}^{n}\left(\xi_i - M\xi_i\right) \xrightarrow[n\to\infty]{p} 0, \qquad (16)$$

необхідно і достатньо виконання умови

$$M\frac{\eta_n^2}{1+\eta_n^2} \xrightarrow[n\to\infty]{} 0 \qquad (17)$$

Доведення. Використаємо, що функція $y = \frac{x^2}{1+x^2}$ при $x > 0$ є зростаючою. Тоді для будь-якого $\varepsilon > 0$

$$P\left\{\left|\eta_n\right| \ge \varepsilon\right\} = \int_{-\infty}^{-\varepsilon} dF_{\eta_n}(x) + \int_{\varepsilon}^{\infty} dF_{\eta_n}(x) = \int_{|x|\ge\varepsilon} dF_{\eta_n}(x) \le$$

$$\le \frac{1+\varepsilon^2}{\varepsilon^2}\int_{|x|\ge\varepsilon}\frac{x^2}{1+x^2}dF_{\eta_n}(x) \le \frac{1+\varepsilon^2}{\varepsilon^2}\int_{-\infty}^{\infty}\frac{x^2}{1+x^2}dF_{\eta_n}(x) = \frac{1+\varepsilon^2}{\varepsilon^2}M\frac{\eta_n^2}{1+\eta_n^2}.$$

Якщо виконується умова (17), то із одержаної нерівності випливає, що $\lim_{n\to\infty} P\left\{\left|\eta_n\right| \ge \varepsilon\right\} = 0$, а це означає, що умова (16) виконується.

Для доведення необхідності врахуємо, що при $|x| \le \varepsilon$ $\frac{x^2}{1+x^2} \le \varepsilon^2$, а $\frac{x^2}{1+x^2} \le 1$ для всіх x, тоді

$$M\frac{\eta_n^2}{1+\eta_n^2}=\int_{-\infty}^{\infty}\frac{x^2}{1+x^2}dF_{\eta_n}(x)=\int_{-\varepsilon}^{\varepsilon}\frac{x^2}{1+x^2}dF_{\eta_n}(x)+\int_{|x|\geq\varepsilon}\frac{x^2}{1+x^2}dF_{\eta_n}(x)\leq$$

$$\leq\varepsilon^2\int_{-\varepsilon}^{\varepsilon}dF_{\eta_n}(x)+\int_{|x|\geq\varepsilon}dF_{\eta_n}(x)\leq\varepsilon^2+P\{|\eta_n|\geq\varepsilon\}.$$

Умова (16) еквівалентна умові $P\{|\eta_n|\geq\varepsilon\}\xrightarrow[n\to\infty]{}0$, тому із одержаної нерівності випливає, що для будь-якого $\varepsilon>0$ $\overline{\lim_{n\to\infty}}M\frac{\eta_n^2}{1+\eta_n^2}\leq\varepsilon^2$, а це означає, що верхня границя може бути рівна лише 0, отже, і сама границя рівна 0. Тобто, із (16) випливає (17). Теорема доведена.

5.2.3. Нерівність Колмогорова. Нехай незалежні випадкові величини $\xi_1,...,\xi_n$ мають скінченні математичні сподівання $M\xi_i$ і дисперсії $D\xi_i$. Позначимо $S_n=\sum_{i=1}^{n}(\xi_i-M\xi_i)$. Тоді для будь-якого $a>0$ справедлива нерівність

$$P\left\{\sup_{1\leq k\leq n}|S_k|\geq a\right\}\leq\frac{1}{a^2}\sum_{k=1}^{n}D\xi_k\,, \tag{18}$$

яку називають нерівністю Колмогорова.

Доведення. Позначимо $\xi_i-M\xi_i=\xi_i'$, тоді $M\xi_i'=0,\ D\xi_i'=D\xi_i$. Введемо випадкові події $A=\left\{\sup_{1\leq k\leq n}|S_k|\geq a\right\},\quad A_k=\{\ |S_1|<a,...,|S_{k-1}|<a,|S_k|\geq a\ \}$, тоді $A_k\cap A_j=\varnothing,\quad k\neq j$ і $A=\bigcup_{k=1}^{n}A_k$. Нехай $\chi_k=I_{A_k}$, тоді $A=\left\{\sum_{k=1}^{n}\chi_k=1\right\}$ і $I_A=\sum_{k=1}^{n}\chi_k$. Для всіх елементарних подій справедлива нерівність $S_n^2\geq S_n^2I_A=\sum_{k=1}^{n}\chi_kS_n^2$. Оскільки $MS_n=0$ і $S_n=\xi_1'+...+\xi_k'+\xi_{k+1}'+...+\xi_n'=S_k+(S_n-S_k)$, то

$$DS_n=MS_n^2\geq M\sum_{k=1}^{n}\chi_kS_n^2=M\sum_{k=1}^{n}\chi_k\left(S_k+(S_n-S_k)\right)^2=$$

$$=M\left(\sum_{k=1}^{n}\chi_kS_k^2+2\sum_{k=1}^{n}\chi_kS_k(S_n-S_k)+\sum_{k=1}^{n}\chi_k(S_n-S_k)^2\right)=$$

$$=M\left(\sum_{k=1}^{n}\chi_kS_k^2\right)+2M\left(\sum_{k=1}^{n}\chi_kS_k(S_n-S_k)\right)+M\left(\sum_{k=1}^{n}\chi_k(S_n-S_k)^2\right). \tag{19}$$

Випадкові величини χ_k і S_k визначаються величинами $\xi_1,...,\xi_k$, а $S_n-S_k=\xi_{k+1}'+...+\xi_n'$. За умовою випадкові величини $\xi_1,...,\xi_n$ - незалежні і $M\xi_i'=0$, тому величини χ_kS_k і S_n-S_k – незалежні і $M\left(\sum_{k=1}^{n}\chi_kS_k(S_n-S_k)\right)=\sum_{k=1}^{n}M(\chi_kS_k)M(\xi_{k+1}'+\cdots+\xi_n')=0$. Якщо врахуємо нерівність $M\left(\sum_{k=1}^{n}\chi_k(S_n-S_k)^2\right)\geq0$, то із (19) одержуємо

$$DS_n \geq M\left(\sum_{k=1}^{n} \chi_k S_k^2\right). \qquad (20)$$

Для всіх елементарних подій справедлива нерівність $\chi_k S_k^2 \geq \chi_k a^2$, тоді із (20) одержимо: $DS_n \geq M\left(\sum_{k=1}^{n} \chi_k S_k^2\right) \geq a^2 M\left(\sum_{k=1}^{n} \chi_k\right) = a^2 P\left\{\sup_{1\leq k\leq n} |S_k| \geq a\right\}$. Якщо ми врахуємо, що $MS_n^2 = DS_n = \sum_{k=1}^{n} D\xi_k$ і поділимо попередню нерівність на a^2, то одержимо нерівність Колмогорова (18).

Оскільки, $\left\{\sup_{k\geq 1} |S_k| \geq a\right\} = \bigcup_{n=1}^{\infty} \left\{\sup_{1\leq k\leq n} |S_k| \geq a\right\}$ і $P\left\{\sup_{k\geq 1} |S_k| \geq a\right\} = \lim_{n\to\infty} P\left\{\sup_{1\leq k\leq n} |S_k| \geq a\right\}$, то перехід до границі у нерівності Колмогорова дає таку нерівність

$$P\left\{\sup_{k\geq 1} |S_k| \geq a\right\} \leq \frac{1}{a^2} \sum_{k=1}^{\infty} D\xi_k .$$

5.2.4. Посилений закон великих чисел.

Теорема Колмогорова. Нехай випадкові величини $\xi_1, ..., \xi_n, ...$ – незалежні, мають скінченні математичні сподівання $M\xi_i$ і дисперсії $D\xi_i$. Якщо ряд $\sum_{k=1}^{\infty} \frac{D\xi_k}{k^2}$ – збіжний, то має місце посилений закон великих чисел $\left(\frac{1}{n} \sum_{k=1}^{n} (\xi_k - M\xi_k) \xrightarrow[n\to\infty]{м.н.} 0\right)$.

Доведення. Позначимо $\xi_i - M\xi_i = \xi_i'$, $S_n = \sum_{i=1}^{n} \xi_i'$, $\eta_n = \sup\limits_{1 \leq k \leq 2^n} |S_k|$. Тоді, для довільного натурального m існує n таке, що $2^{n-1} \leq m < 2^n$ і

$$\left|\frac{1}{m} \sum_{k=1}^{m} \xi_k'\right| = \frac{|S_m|}{m} \leq \frac{\eta_n}{2^{n-1}} = 2^{1-n} \eta_n .$$

Тому для доведення теореми досить довести, що $2^{1-n} \eta_n \xrightarrow{м.н.} 0$. За властивістю збіжності з ймовірністю 1 (наслідок 2), останнє буде виконуватись, якщо для будь-якого $\varepsilon > 0$ ряд

$$\sum_{n=1}^{\infty} P\left\{2^{1-n} \eta_n \geq \varepsilon\right\} = \sum_{n=1}^{\infty} P\left\{\sup_{1 \leq k \leq 2^n} |S_k| \geq \varepsilon \cdot 2^{n-1}\right\} \qquad (21)$$

буде збіжним. Застосувавши нерівність Колмогорова

$$P\left\{\sup_{1 \leq k \leq 2^n} |S_k| \geq \varepsilon \cdot 2^{n-1}\right\} \leq \frac{1}{\varepsilon^2 4^{n-1}} \sum_{k=1}^{2^n} D\xi_k ,$$

із (21) одержимо

$$\sum_{n=1}^{\infty} P\left\{\sup_{1 \leq k \leq 2^n} |S_k| \geq \varepsilon \cdot 2^{n-1}\right\} \leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 4^{n-1}} \sum_{k=1}^{2^n} D\xi_k . \qquad (22)$$

Змінимо у правій частині нерівності (22) порядок додавання, тоді $(n \geq \log_2 k \Leftrightarrow 2^n \geq k)$

$$\sum_{n=1}^{\infty} P\left\{\sup_{1 \leq k \leq 2^n} |S_k| \geq \varepsilon \cdot 2^{n-1}\right\} \leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 4^{n-1}} \sum_{k=1}^{2^n} D\xi_k = \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \sum_{n \geq \log_2 k} \frac{D\xi_k}{4^n} =$$

$$=\frac{4}{\varepsilon^2}\sum_{k=1}^{\infty} D\xi_k \sum_{n\ge\log_2 k}\frac{1}{4^n}\le\frac{4^2}{3\varepsilon^2}\sum_{k=1}^{\infty}\frac{D\xi_k}{k^2}.$$

За умовою, ряд $\sum_{k=1}^{\infty}\frac{D\xi_k}{k^2}$ – збіжний, тому збіжний і ряд (21), а звідси, як відзначено вище, випливає справедливість теореми.

Теорема Бореля. Якщо μ_n – число появ події A в n випробуваннях Бернуллі і $P(A)=p$, то $\frac{\mu_n}{n}\xrightarrow{\text{м.н.}}p$ (відносна частота настання події збігається до p з ймовірністю 1).

Доведення. Як і при доведенні теореми Бернуллі, нехай ξ_i – число появ події в i-му випробуванні, тоді $\mu_n=\xi_1+\dots+\xi_n$, $M\xi_i=p,\;\; D\xi_i=pq$, тому ряд $\sum_{k=1}^{\infty}D\xi_k\frac{1}{k^2}=pq\sum_{k=1}^{\infty}\frac{1}{k^2}$ – збіжний, а звідси, за теоремою Колмогорова, випливає, що $\frac{\mu_n}{n}\xrightarrow{\text{м.н.}}p$.

Теорема Бореля є підтвердженням статистичного підходу до означення ймовірності як границі відносної частоти.

Зауваження. Нехай випадкові величини $\xi_1,...,\xi_n,...$ – незалежні і однаково розподілені. А.М. Колмогоров показав, що в цьому випадку умова існування скінченного математичного сподівання є необхідною і достатньою для справедливості посиленого закону великих чисел.

Вправи.

2. Задано послідовність незалежних випадкових величин $\xi_1,...,\xi_n,...$, що мають рівномірний розподіл на відрізку *а*) $[0,n]$; *б*) $[0,\sqrt{n}]$; *в*) $[0,\frac{1}{\sqrt{n}}]$. Чи можна застосувати до даної послідовності теорему Чебишова про закон великих чисел?
3. Перевірити, чи можна застосувати теорему Чебишова про закон великих чисел до послідовності незалежних випадкових величин $\xi_1,...,\xi_n,...$, якщо

ξ_n	$-n\alpha$	0	$n\alpha$
p	$\frac{1}{2n^2}$	$1-\frac{1}{n^2}$	$\frac{1}{2n^2}$

$\alpha>0$.

Розділ 6. Характеристичні функції

6.1. Характеристична функція випадкової величини, її властивості

Якщо випадкові величини ξ_1 і ξ_2 – дійсні, то випадкова величина $\xi=\xi_1+i\xi_2$ буде комплексно значною випадковою величиною. Визначимо її математичне сподівання за допомогою рівності $M\xi=M\xi_1+iM\xi_2$.

Означення. *Характеристичною функцією* випадкової величини ξ називають функцію

$$f(t)=Me^{it\xi},\quad t\in R. \tag{1}$$

Математичне сподівання завжди існує, бо e^{itx} неперервна і $\left|e^{itx}\right|=1$. Використовуючи формулу Ейлера $e^{it\xi}=\cos t\xi+i\sin t\xi$, одержимо $f(t)=M(\cos t\xi)+iM(\sin t\xi)$.

Якщо випадкова величина ξ приймає значення x_k з ймовірностями p_k, то

$$f(t)=\sum_k e^{itx_k}p_k\ . \qquad (2)$$

Якщо ж випадкова величина ξ – неперервна і має щільність розподілу $p(x)$, то

$$f(t)=\int_{-\infty}^{+\infty}e^{itx}p(x)\,dx\,. \qquad (3)$$

У загальному випадку

$$f(t)=\int_{-\infty}^{+\infty}e^{it\xi}dF(x)\,. \qquad (4)$$

Розглянемо деякі приклади знаходження характеристичних функцій.

Розподіл Пуассона. Випадкова величина ξ має розподіл $p_k=e^{-\lambda}\dfrac{\lambda^k}{k!},\quad k=0,1,...$

Тому за формулою (2) $f(t)=\sum_{k=0}^{\infty}e^{itk}e^{-\lambda}\dfrac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\dfrac{\left(\lambda e^{it}\right)^k}{k!}=e^{-\lambda\left(1-e^{it}\right)}$.

Рівномірний розподіл. Випадкова величина має щільність $p(x)=\begin{cases}\dfrac{1}{b-a},\ x\in(a,b);\\ 0,\quad x\notin(a,b).\end{cases}$

За формулою (3) $f(t)=Me^{it\xi}=\int_{-\infty}^{+\infty}e^{itx}p(x)\,dx=\int_a^b e^{itx}\dfrac{1}{b-a}dx=\dfrac{1}{b-a}\dfrac{e^{itx}}{it}\Big|_a^b=\dfrac{e^{itb}-e^{ita}}{(b-a)it}$,

якщо ж випадкова величина ξ має рівномірний розподіл на $[-a,a]$, то

$$f(t)=\frac{e^{ita}-e^{-ita}}{2ait}=\frac{\sin at}{at}\,.$$

Нормальний розподіл. Нехай випадкова величина $\xi\sim N(0,1)$. Це означає, що $p(x)=\dfrac{1}{\sqrt{2\pi}}\exp\left(-\dfrac{x^2}{2}\right)$. За формулою (3)

$$f(t)=\int_{-\infty}^{+\infty}\exp(itx)\,p(x)\,dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(itx-\frac{x^2}{2}\right)dx\,.$$

Продиференціюємо дану функцію по параметру t $f'(t)=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{itx-\frac{x^2}{2}}ix\,dx$ (це можливо, бо інтеграл в правій частині рівномірно збіжний відносно $t\in R$), а далі інтегруємо частинами, тоді одержимо

$$f'(t)=-\frac{i}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{itx}d\left(e^{-\frac{x^2}{2}}\right)=\frac{i}{\sqrt{2\pi}}\left(-e^{itx-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty}+it\int_{-\infty}^{+\infty}e^{itx-\frac{x^2}{2}}dx\right)=-tf(t)\,.$$

Розв’язуючи рівняння $f'(t)=-tf(t)$ при початковій умові $f(0)=1$, одержимо $f(t)=e^{-\frac{t^2}{2}}$.

Розглянемо тепер деякі властивості характеристичних функцій.

1. Для довільного $t\in R$ $\left|f(t)\right|\le 1,\ f(0)=1$.

За формулою (4) для довільного $t\in R$ $\left|f(t)\right|\le\int_{-\infty}^{+\infty}\left|e^{itx}\right|dF(x)=1$, а $f(0)=\int_{-\infty}^{+\infty}e^{i0x}dF(x)=1$.

2. $f(-t)=\overline{f(t)}$.

$$f(t)=\int_{-\infty}^{\infty}(\cos tx+i\sin tx)dF(x), \quad f(-t)=\int_{-\infty}^{\infty}(\cos tx-i\sin tx)dF(x)=\overline{f(t)}.$$

3. Якщо випадкова величина ξ має характеристичну функцію $f_\xi(t)$, то випадкова величина $\eta=a\xi+b$ буде мати характеристичну функцію

$$f_\eta(t)=e^{itb}f_\xi(at).$$

Із (1) $f_\eta(t)=M\exp(it(a\xi+b))=\exp(itb)M\exp(ita\xi)=\exp(itb)f_\xi(at)$.

Наприклад, якщо випадкова величина $\xi\sim N(0,1)$, а випадкова величина $\eta=a+\sigma\xi$, то $\xi\sim N(0,1)$, Тоді $f_\eta(t)=\exp\left(ita-\frac{\sigma^2t^2}{2}\right)$.

4. Якщо випадкові величини ξ_1,ξ_2 – незалежні, мають характеристичні функції $f_{\xi_1}(t)$ і $f_{\xi_2}(t)$, то характеристична функція $f_{\xi_1+\xi_2}(t)=f_{\xi_1}(t)f_{\xi_2}(t)$.

Оскільки випадкові величини ξ_1,ξ_2 – незалежні, то незалежними будуть і величини $e^{it\xi_1}$ і $e^{it\xi_2}$, тому $f_{\xi_1+\xi_2}(t)=Me^{it(\xi_1+\xi_2)}=Me^{it\xi_1}e^{it\xi_2}=Me^{it\xi_1}Me^{it\xi_2}=f_{\xi_1}(t)f_{\xi_2}(t)$.

Із властивості (4) випливає: якщо випадкові величини $\xi_1,...,\xi_k$ – незалежні, то характеристична функція їх суми дорівнює добутку характеристичних функцій доданків.

Як приклад застосування цієї властивості знайдемо характеристичну функцію *біномного розподілу*. Нехай ξ – число появ події в n експериментах Бернуллі, тоді $\xi=\xi_1+\cdots+\xi_k$, де

ξ_k	0	1
p	q	p

За формулою (2) $f_{\xi_k}(t)=Me^{it\xi}=q+e^{it}p$, а за властивістю 4 $f_\xi(t)=\left(f_{\xi_k}(t)\right)^n=\left(q+pe^{it}\right)^n$.

Для доведення наступних властивостей необхідна наступна лема.

Лема. Для будь-якого дійсного α і цілого $n\ge 0$ справедливі нерівності

$$\left|e^{i\alpha}-\sum_{k=0}^{n}\frac{(i\alpha)^k}{k!}\right|\le\frac{|\alpha|^{n+1}}{(n+1)!}, \tag{Н1}$$

$$\left|e^{i\alpha}-\sum_{k=0}^{n}\frac{(i\alpha)^k}{k!}\right|\le\frac{2|\alpha|^{n}}{n!}. \tag{Н2}$$

Доведення. Доведемо спочатку нерівність (Н1). Із рівності $\int_0^\alpha e^{it}dt=\left.\frac{e^{it}}{i}\right|_0^\alpha=\frac{1}{i}\left(e^{i\alpha}-1\right)$ випливає, що $\left|e^{i\alpha}-1\right|=\left|\int_0^\alpha e^{it}dt\right|\le\int_0^{|\alpha|}dt=|\alpha|$. Тобто при $n=0$ нерівність (Н1) правильна.

Із рівності $\int_0^\alpha\left(e^{it}-1\right)dt=\frac{1}{i}\left(e^{i\alpha}-1-i\alpha\right)$, при врахуванні попередньої нерівності, одержуємо $\left|e^{i\alpha}-1-i\alpha\right|=\left|\int_0^\alpha\left(e^{it}-1\right)dt\right|\le\int_0^{|\alpha|}\left|e^{it}-1\right|dt\le\int_0^{|\alpha|}tdt=\frac{\alpha^2}{2}$. Тобто, нерівність правильна і при $n=1$.

Припустимо, що нерівність вірна при n, і доведемо, що вона вірна і при $n+1$. Із рівності

$$\int_0^\alpha\left(e^{it}-\sum_{k=0}^{n}\frac{(it)^k}{k!}\right)dt=\frac{1}{i}\left(e^{i\alpha}-\sum_{k=0}^{n+1}\frac{(i\alpha)^k}{k!}\right)$$

випливає, що

$$\left|e^{i\alpha}-\sum_{k=0}^{n+1}\frac{(i\alpha)^k}{k!}\right|\le\int_0^{|\alpha|}\left|e^{it}-\sum_{k=0}^{n}\frac{(it)^k}{k!}\right|dt\le\frac{1}{(n+1)!}\int_0^{|\alpha|}t^{n+1}dt=\frac{|\alpha|^{n+2}}{(n+2)!}.$$

Це означає, що нерівність (Н1) правильна при $n+1$. Нерівність (Н2) випливає із (Н1):

$$\left|e^{i\alpha}-\sum_{k=0}^{n}\frac{(i\alpha)^k}{k!}\right|=\left|\left(e^{i\alpha}-\sum_{k=0}^{n-1}\frac{(i\alpha)^k}{k!}\right)-\frac{(i\alpha)^n}{n!}\right|\le\left|e^{i\alpha}-\sum_{k=0}^{n-1}\frac{(i\alpha)^k}{k!}\right|+\frac{|\alpha|^n}{n!}\le\frac{2|\alpha|^n}{n!}.$$

Лема доведена.

5. Характеристична функція рівномірно неперервна на R.

Розглянемо різницю $f(t+\Delta t)-f(t)=\int\limits_{-\infty}^{+\infty}e^{i(t+\Delta t)x}dF(x)-\int\limits_{-\infty}^{+\infty}e^{itx}dF(x)=\int\limits_{-\infty}^{+\infty}e^{itx}\left(e^{i\Delta tx}-1\right)dF(x)$.

Тоді $\left|f(t+\Delta t)-f(t)\right|\le\int\limits_{-\infty}^{\infty}\left|e^{ix\Delta t}-1\right|dF(x)=\int\limits_{-A}^{A}\left|e^{ix\Delta t}-1\right|dF(x)+\int\limits_{|x|>A}\left|e^{ix\Delta t}-1\right|dF(x)$. За нерівностями (Н1) і (Н2) при $n=0$ $\left|e^{ix\Delta t}-1\right|\le|x\Delta t|$ і $\left|e^{ix\Delta t}-1\right|\le 2$, тому

$$\left|f(t+\Delta t)-f(t)\right|\le|\Delta t|\int\limits_{-A}^{A}|x|dF(x)+2\int\limits_{|x|>A}dF(x)\le$$

$$\le|\Delta t|A\big(F(A)-F(-A)\big)+2\big(1-F(A)+F(-A)\big)\le|\Delta t|A+2\big(1-F(A)+F(-A)\big)$$

Із того, що $F(x)\to 1$ при $x\to+\infty$ і $F(x)\to 0$ при $x\to-\infty$ випливає, що для довільного $\varepsilon>0$ число A можна вибрати настільки великим, щоб виконувалась нерівність $1-F(A)+F(-A)<\frac{\varepsilon}{4}$. Зафіксуємо таке A і покладемо $\frac{\varepsilon}{2A}=\delta(\varepsilon)$. Тоді при $|\Delta t|<\delta(\varepsilon)$

$$\left|f(t+\Delta t)-f(t)\right|\le|\Delta t|A+2\big(1-F(A)+F(-A)\big)<\varepsilon,$$

а це означає, що характеристична функція рівномірно неперервна на R.

6. Нехай для випадкової величини ξ існує математичне сподівання $M|\xi|^n$, тоді характеристична функція має в кожній точці похідні $f^{(k)}(t)$ для $k\le n$ і має місце рівність

$$f^{(k)}(0)=i^kM\left(\xi^k\right),\quad k=1,2,...,n, \tag{5}$$

крім того,

$$f(t)=\sum_{k=0}^{n}\frac{(it)^k}{k!}M\left(\xi^k\right)+R_n(t), \tag{6}$$

де $R_n(t)=o\left(t^n\right)$ при $t\to 0$.

Оскільки $\int\limits_{-\infty}^{+\infty}|x|dF(x)$ – збіжний, то інтеграл $i\int\limits_{-\infty}^{+\infty}xe^{itx}dF(x)$, що одержується формальним диференціюванням характеристичної функції $f(t)=\int\limits_{-\infty}^{+\infty}e^{itx}dF(x)$, є рівномірно збіжним при $t\in R$. Тому диференціювання під знаком інтеграла законне і

$$f'(t)=i\int\limits_{-\infty}^{\infty}xe^{itx}dF(x),\ f'(0)=iM\xi.$$

Припустимо, що існує $f^{(k-1)}(t)=i^k\int\limits_{-\infty}^{\infty}x^{k-1}e^{itx}dF(x)$. Із збіжності інтеграла $\int\limits_{-\infty}^{+\infty}|x|^n dF(x)$ випливає, що інтеграл $i^k\int\limits_{-\infty}^{+\infty}x^ke^{itx}dF(x)$ є рівномірно збіжним при $t\in R$, тому попередню рівність можна диференціювати по t і буде справедливою рівність $f^{(k)}(t)=i^k\int\limits_{-\infty}^{\infty}x^k e^{itx}dF(x)$, при цьому, $f^{(k)}(0)=i^k M(\xi^k)$.

У зв'язку із тим, що існують похідні до n-го порядку і $f^{(k)}(0)=i^k M(\xi^k)$, то ми можемо розкласти $f(t)$ за формулою Тейлора:

$$f(t)=\sum_{k=0}^{n}\frac{(it)^k}{k!}M(\xi^k)+R_n(t).$$

Оцінимо залишковий член

$$|R_n(t)|=\left|f(t)-\sum_{k=0}^{n}\frac{(it)^k}{k!}M(\xi^k)\right|=\left|\int\limits_{-\infty}^{+\infty}\left(e^{itx}-\sum_{k=0}^{n}\frac{(itx)^k}{k!}\right)dF(x)\right|\le$$

$$\le\int\limits_{-A}^{A}\left|e^{itx}-\sum_{k=0}^{n}\frac{(itx)^k}{k!}\right|dF(x)+\int\limits_{|x|>A}\left|e^{itx}-\sum_{k=0}^{n}\frac{(itx)^k}{k!}\right|dF(x).$$

Для оцінки першого інтеграла правої частини застосуємо нерівність (Н1) $\left|e^{itx}-\sum_{k=0}^{n}\frac{(itx)^k}{k!}\right|\le\frac{|tx|^{n+1}}{(n+1)!}$, а другого – (Н2) $\left|e^{itx}-\sum_{k=0}^{n}\frac{(itx)^k}{k!}\right|\le\frac{2|tx|^n}{n!}$. Тоді

$$R_n(t)\le\frac{|t|^{n+1}}{(n+1)!}\int\limits_{-A}^{A}|x|^{n+1}dF(x)+\frac{2|t|^n}{n!}\int\limits_{|x|>A}|x|^n dF(x)\le\frac{|t|^{n+1}}{(n+1)!}AM|\xi|^n+\frac{2|t|^n}{n!}\int\limits_{|x|>A}|x|^n dF(x).$$

Нехай $\varepsilon>0$. Виберемо і зафіксуємо число A таким, щоб $\frac{2}{n!}\int\limits_{|x|>A}|x|^n dF(x)<\frac{\varepsilon}{2}$. Це можна зробити, бо $M|\xi|^n=\int\limits_{-\infty}^{+\infty}|x|^n dF(x)$ існує, тому залишок цього інтегралу $\int\limits_{|x|>A}|x|^n dF(x)\to 0$ при $A\to+\infty$. Покладемо $\frac{\varepsilon(n+1)!}{2AM|\xi|^n}=\delta(\varepsilon)$, тоді при $|t|<\delta(\varepsilon)$

$\frac{|t|A}{(n+1)!}M|\xi|^n<\frac{\varepsilon}{2}$. Отже, $|R_n(t)|\le|t|^n\varepsilon$. А це означає, що $R_n(t)=o(t^n)$ при $t\to 0$.

7. Характеристична функція додатно визначена: для будь-якого натурального n, довільних дійсних $t_1,...,t_n$ і довільних комплексних $c_1,...,c_n$ $\sum_{k=1}^{n}\sum_{l=1}^{n}f(t_k-t_l)c_k\overline{c_l}\ge 0$.

Дійсно,

$$\sum_{k=1}^{n}\sum_{l=1}^{n}Me^{i(t_k-t_l)\xi}c_k\overline{c_l}=M\left(\sum_{k=1}^{n}c_k e^{it_k\xi}\right)\left(\overline{\sum_{l=1}^{n}c_l e^{it_l\xi}}\right)=M\left|\sum_{k=1}^{n}c_k e^{it_k\xi}\right|^2\ge 0.$$

Має місце більш загальне твердження, яке ми наведемо без доведення.

Теорема (Бохнера-Хінчина). Для того, щоб функція $f(t)$ була характеристичною функцією деякого розподілу, необхідно і достатньо, щоб вона була неперервна, додатно визначена і $f(0)=1$.

Означення. Характеристичною функцією випадкового вектора $\xi=(\xi_1,...,\xi_m)$ називається функція векторного аргументу $t=(t_1,...,t_m)$ $f_\xi(t)=Me^{i(t,\xi)}=M\exp\left(i\sum_{k=1}^{m}t_k\xi_k\right)$.

Характеристична функція випадкового вектора має властивості аналогічні до властивостей одновимірної випадкової величини.

Якщо значення випадкових величин не довільні, то замість характеристичних функцій зручніше застосовувати деякі інші. Якщо величина ξ має лише невід'ємні значення, то розглядають її перетворення Лапласа $g(\lambda) = Me^{-\lambda\xi} = \int_0^{+\infty} e^{-\lambda x} dF(x)$.

Нехай випадкова величина ξ приймає лише цілі невід'ємні значення з ймовірностями $p_k = P\{\xi = k\}$, $k = 0,1,2,\ldots$. *Твірною функцією* (*генератрисою*) випадкової величини ξ називається функція $\varphi(\lambda) = M\lambda^{\xi} = \sum_{k=0}^{\infty} p_k \lambda^k$, що визначена при комплексних λ, для яких $|\lambda| \le 1$.

Твірна функція суми незалежних випадкових величин дорівнює добуткові твірних функцій доданків. Те саме стосується і перетворення Лапласа. Важливо відзначити, що характеристична функція, твірна функція, а також перетворення Лапласа однозначно визначають розподіл випадкової величини (для характеристичних функцій це ми пізніше доведемо).

Вправи.

16. Нехай ξ набуває значень 1 і –1 з ймовірностями ½ кожне. Знайти характеристичну функцію ξ.
17. Довести, що функція $\cos^2 t$ є характеристичною функцією, і знайти відповідний розподіл ймовірностей.
18. Випадкова величина ξ має гамма-розподіл: $p_\xi(x) = \begin{cases} 0, & x \le 0; \\ \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0, \end{cases}$ $(\lambda > 0, \quad \alpha > 0)$. Знайти характеристичну функцію.
19. Використовуючи вигляд характеристичної функції гауссової (нормально розподіленої) випадкової величини ξ $f(t) = \exp\left\{ita - \frac{1}{2}bt^2\right\}$, де $a = M\xi$, $b = D\xi$, довести, що сума будь-якої кількості гауссових випадкових величин є також гауссова величина.
20. Довести, що характеристична функція $f_\xi(t)$ дійсна тоді і тільки тоді, коли розподіл ξ симетричний (ξ і $-\xi$ мають однаковий розподіл).
21. Довести, що характеристична функція $f_\xi(t)$ дійсна тоді і тільки тоді, коли відповідна функція розподілу $F_\xi(x)$ задовольняє рівність $F_\xi(x) = 1 - F_\xi(-x-0)$.
22. Використовуючи властивості характеристичних функцій показати, що функції $\sin t$, $|\cos t|$ не можуть бути характеристичними.

6.2. Формули обернення для характеристичних функцій

Якщо задана функція розподілу $F(x)$, тоді характеристична функція $f(t)$ виражається через $F(x)$ однозначно за формулою $f(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x)$. Поставимо обернену задачу: знайти функцію розподілу $F(x)$, якщо відомо характеристичну функцію $f(t)$.

Теорема 1. Нехай $F(x)$ – функція розподілу, а $f(t)$ – характеристична функція випадкової величини ξ. Для довільних x_1 і x_2, які є точками неперервності $F(x)$, справедлива рівність

$$F(x_2)-F(x_1)=\lim_{c\to+\infty}\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-itx_1}-e^{-itx_2}}{it}f(t)\,dt. \quad (7)$$

Доведення. Розглянемо інтеграл $I_c=\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-itx_1}-e^{-itx_2}}{it}f(t)\,dt$ і підставимо $f(t)=\int_{-\infty}^{+\infty}e^{itx}dF(x)$, тоді $I_c=\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-itx_1}-e^{-itx_2}}{it}dt\int_{-\infty}^{+\infty}e^{itz}dF(z)$ і, змінивши порядок інтегрування, одержимо $I_c=\frac{1}{2\pi}\int_{-\infty}^{+\infty}dF(z)\int_{-c}^{c}\frac{\exp(it(z-x_1))-\exp(it(z-x_2))}{it}dt$.

Оскільки інтеграл $\int_{-c}^{c}\frac{\exp(it(z-x_1))-\exp(it(z-x_2))}{it}dt=$

$$=\int_{-c}^{0}\frac{\exp(it(z-x_1))-\exp(it(z-x_2))}{it}dt+\int_{0}^{c}\frac{\exp(it(z-x_1))-\exp(it(z-x_2))}{it}dt,$$

а $\int_{-c}^{0}\frac{\exp(it(z-x_1))-\exp(it(z-x_2))}{it}dt=\left|\begin{matrix}t=-\tau\\ dt=-d\tau\end{matrix}\right|=\int_{c}^{0}\frac{\exp(-i\tau(z-x_1))-\exp(-i\tau(z-x_2))}{i\tau}d\tau$, то

$$I_c=\frac{1}{2\pi}\int_{-\infty}^{+\infty}dF(z)\int_{0}^{c}\frac{\exp(it(z-x_1))-\exp(it(z-x_2))-\exp(-it(z-x_1))+\exp(-it(z-x_2))}{it}dt=$$

$$=\frac{1}{\pi}\int_{-\infty}^{+\infty}dF(z)\int_{0}^{c}\frac{\sin t(z-x_1)-\sin t(z-x_2)}{t}dt.$$ Отже,

$$I_c=\int_{-\infty}^{+\infty}\psi(c,z,x_1,x_2)\,dF(z),$$

де

$$\psi(c,z,x_1,x_2)=\frac{1}{\pi}\int_{0}^{c}\left(\frac{\sin t(z-x_1)}{t}-\frac{\sin t(z-x_2)}{t}\right)dt. \quad (8)$$

Врахуємо, що $\int_{0}^{+\infty}\frac{\sin\alpha t}{t}dt=0$ при $\alpha=0$, $\int_{0}^{+\infty}\frac{\sin\alpha t}{t}dt=\frac{\pi}{2}$ при $\alpha>0$, $\int_{0}^{+\infty}\frac{\sin\alpha t}{t}dt=-\frac{\pi}{2}$ при $\alpha<0$ і $\int_{0}^{+\infty}\frac{\sin\alpha t}{t}dt$ є рівномірно збіжним в кожному із проміжків $\alpha\le-\delta$ і $\alpha\ge\delta$ для будь-якого $\delta>0$ і при $|\alpha|\le\delta$ для всіх досить великих c

$$\frac{1}{\pi}\left|\int_{0}^{c}\frac{\sin\alpha t}{t}dt\right|<1. \quad (9)$$

Нехай $x_1<x_2$ і виберемо $\delta>0$ так, щоб $x_1+\delta<x_2-\delta$. Тоді

$$I_c = \underbrace{\int_{-\infty}^{x_1-\delta} \psi dF(z)}_{I_1} + \underbrace{\int_{x_1-\delta}^{x_1+\delta} \psi dF(z)}_{I_2} + \underbrace{\int_{x_1+\delta}^{x_2-\delta} \psi dF(z)}_{I_3} + \underbrace{\int_{x_2-\delta}^{x_2+\delta} \psi dF(z)}_{I_4} + \underbrace{\int_{x_2+\delta}^{+\infty} \psi dF(z)}_{I_5} =$$

$$= I_1 + I_2 + I_3 + I_4 + I_5 . \tag{10}$$

Розглянемо окремо кожен інтеграл.

Якщо $-\infty < z < x_1 - \delta < x_2 - \delta$, то $z - x_1 < -\delta,\ z - x_2 < -\delta$, тому при $c \to +\infty$ інтеграли $\int_0^c \frac{\sin t(z - x_i)}{t} dt$ $(i = 1,2)$ рівномірно збігаються до $-\frac{\pi}{2}$. Отже, із (8) одержуємо

$$I_1 \xrightarrow[c\to+\infty]{} 0 . \tag{11}$$

Оскільки в I_5 $z > x_2 + \delta$, то $z - x_1 > \delta$ і $z - x_2 > \delta$, тому при $c \to +\infty$ інтеграли $\int_0^c \frac{\sin t(z - x_i)}{t} dt$ $(i = 1,2)$ рівномірно збігаються до $\frac{\pi}{2}$. Тому

$$I_5 \xrightarrow[c\to+\infty]{} 0 . \tag{12}$$

Із того, що в I_3 $z - x_1 > \delta$, а $z - x_2 < -\delta$ випливає, що рівномірно $\int_0^c \frac{\sin t(z - x_1)}{t} dt \xrightarrow[c\to+\infty]{} \frac{\pi}{2}$, а $\int_0^c \frac{\sin t(z - x_2)}{t} dt \xrightarrow[c\to+\infty]{} -\frac{\pi}{2}$ також рівномірно, тому із (8) випливає, що $\psi(c, z, x_1, x_2) \xrightarrow[c\to+\infty]{} 1$ рівномірно по z, а це означає, що

$$I_3 = \int_{x_1+\delta}^{x_2-\delta} \psi(c, z, x_1, x_2) dF(z) \xrightarrow[c\to+\infty]{} F(x_2 - \delta) - F(x_1 + \delta). \tag{13}$$

Для I_2 $x_1 - \delta \le z \le x_1 + \delta$ або $|z - x_1| \le \delta$, тому із (9) і (8) для великих c одержуємо $|\psi(c, z, x_1, x_2)| \le 2$, а це означає, що для всіх досить великих c

$$|I_2| = \left| \int_{x_1-\delta}^{x_1+\delta} \psi(c, z, x_1, x_2) dF(z) \right| \le 2\big(F(x_1 + \delta) - F(x_1 - \delta)\big). \tag{14}$$

Аналогічно,

$$|I_4| \le 2\big(F(x_2 + \delta) - F(x_2 - \delta)\big). \tag{15}$$

Отже, із (10) – (15) одержуємо

$$\lim_{c\to+\infty} I_c = F(x_2 - \delta) - F(x_1 + \delta) + R(\delta, x_1, x_2), \tag{16}$$

де $|R(\delta, x_1, x_2)| \le 2\big(F(x_1 + \delta) - F(x_1 - \delta)\big) + 2\big(F(x_2 + \delta) - F(x_2 - \delta)\big)$.

Враховуючи, що x_1, x_2 – точки неперервності $F(x)$ і інтеграл I_c не залежить від δ, то перейшовши в (16) до границі, коли $\delta \to 0$, і враховуючи, що $R(\delta, x_1, x_2) \to 0$ при $\delta \to 0$, одержуємо $\lim_{c\to+\infty} I_c = F(x_2) - F(x_1)$.

Наслідок. Функція розподілу випадкової величини однозначно визначається своєю характеристичною функцією.

Доведення. Покладемо в (7) $x_2 = x$ і перейдемо до границі, коли $x_1 \to -\infty$, тоді $F(x_1) \to 0$, а із (7) одержимо $F(x) = \lim_{x_1\to-\infty} \lim_{c\to+\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{\exp(-itx_1) - \exp(-itx)}{it} f(t) dt$.

Теорема 2. Нехай характеристична функція $f(t)$ – абсолютно інтегровна на R, тобто $\int_{-\infty}^{+\infty}|f(t)|dt$ – збіжний. Тоді відповідна їй функція розподілу буде неперервною в кожній точці $x \in R$; існує щільність розподілу

$$p(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-itx}f(t)\,dt; \tag{17}$$

$p(x)$ – неперервна.

Доведення. Із абсолютної інтегровності характеристичної функції $f(t)$ і теореми 1 випливає, що

$$F(x_2)-F(x_1)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{\exp(-itx_1)-\exp(-itx_2)}{it}f(t)\,dt. \tag{18}$$

Нехай $x \in R$, покладемо $x_2=x+\delta,\ \ x_1=x-\delta$ так, щоб x_1,x_2 були точками неперервності функції $F(x)$. Тоді

$$F(x+\delta)-F(x-\delta)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{\exp(-it(x-\delta))-\exp(-it(x+\delta))}{it}f(t)\,dt=$$

$$=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp(-itx)\frac{\exp(it\delta)-\exp(-it\delta)}{it}f(t)\,dt=\frac{1}{\pi}\int_{-\infty}^{+\infty}\exp(-itx)\frac{\sin t\delta}{t}f(t)\,dt.$$

Звідки, $F(x+\delta)-F(x-\delta)\le\frac{1}{\pi}\int_{-\infty}^{+\infty}\left|\frac{\sin t\delta}{t}\right||f(t)|\,dt\le\frac{\delta}{\pi}\int_{-\infty}^{+\infty}|f(t)|dt\xrightarrow[\delta\to 0]{}0$, а це означає, що функція розподілу є неперервною в кожній точці $x \in R$.

Із збіжності інтеграла $\int_{-\infty}^{+\infty}|f(t)|dt$ випливає, що інтеграл в рівності (17) існує. Тоді із (18) одержимо:

$$\rho=\frac{F(x+h)-F(x)}{h}-p(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{1-e^{-ith}-ith}{ith}e^{-itx}f(t)\,dt.$$

Тоді

$$|\rho|\le\frac{1}{2\pi}\int_{-\infty}^{+\infty}\left|\frac{e^{-ith}-1+ith}{ith}\right||f(t)|\,dt=\frac{1}{2\pi}\int_{|t|\le A}\left|\frac{e^{-ith}-1+ith}{ith}\right||f(t)|\,dt+\frac{1}{2\pi}\int_{|t|>A}\left|\frac{e^{-ith}-1+ith}{ith}\right||f(t)|\,dt.$$

Із нерівностей (Н1) і (Н2) $\left|e^{-ith}-1+ith\right|\le\frac{1}{2}(th)^2$ і $\left|e^{-ith}-1+ith\right|\le 2|th|$, застосувавши їх відповідно до першого і другого інтегралів правої частини попередньої нерівності, маємо

$$|\rho|\le\frac{1}{2\pi}\int_{|t|\le A}\frac{|th|}{2}|f(t)|\,dt+\frac{1}{\pi}\int_{|t|>A}|f(t)|dt\le|h|\frac{A}{4\pi}\int_{-\infty}^{+\infty}|f(t)|dt+\frac{1}{\pi}\int_{|t|>A}|f(t)|dt. \tag{19}$$

Нехай задане довільне $\varepsilon>0$. Із збіжності інтеграла $\int_{-\infty}^{+\infty}|f(t)|dt$ випливає, що існує A таке, що буде виконуватись нерівність

$$\frac{1}{\pi}\int_{|t|>A}|f(t)|dt<\frac{\varepsilon}{2}. \tag{20}$$

Зафіксуємо таке A. Тоді, якщо покладемо $\delta=\varepsilon\dfrac{2\pi}{A}\left(\int\limits_{-\infty}^{+\infty}|f(t)|dt\right)^{-1}$, то при $|h|<\delta$ із (19) і (20) одержуємо $|\rho|<\varepsilon$. А це означає, що існує $F'(x)=\lim\limits_{h\to0}\dfrac{F(x+h)-F(x)}{h}=p(x)$. Тобто, $p(x)$ є щільністю розподілу і справедлива рівність (17).

Доведемо, що $p(x)$ – неперервна. Оскільки $p(x+h)-p(x)=$

$$=\frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}\left(e^{-it(x+h)}-e^{-itx}\right)f(t)\,dt=\frac{1}{2\pi}\int\limits_{|t|\le A}e^{-itx}\left(e^{-ith}-1\right)f(t)\,dt+\frac{1}{2\pi}\int\limits_{|t|>A}e^{-itx}\left(e^{-ith}-1\right)f(t)\,dt,$$

то із нерівності (Н1)

$$|p(x+h)-p(x)|\le\frac{1}{2\pi}\int\limits_{|t|\le A}\left|e^{-ith}-1\right||f(t)|\,dt+\frac{1}{2\pi}\int\limits_{|t|>A}\left|e^{-ith}-1\right||f(t)|\,dt\le$$

$$\le\frac{|h|}{2\pi}\int\limits_{|t|\le A}|t||f(t)|\,dt+\frac{1}{\pi}\int\limits_{|t|>A}|f(t)|\,dt\le\frac{A|h|}{2\pi}\int\limits_{-\infty}^{+\infty}|f(t)|\,dt+\frac{1}{\pi}\int\limits_{|t|>A}|f(t)|\,dt. \quad (21)$$

При $A\to+\infty$ другий доданок у нерівності (21) прямує до 0. Тому для довільного $\varepsilon>0$ існує таке A, що $\dfrac{1}{\pi}\int\limits_{|t|>A}|f(t)|\,dt<\dfrac{\varepsilon}{2}$. Зафіксуємо таке A. Якщо покладемо $\delta=\dfrac{\varepsilon\pi}{A}\left(\int\limits_{-\infty}^{+\infty}|f(t)|dt\right)^{-1}$, то при $|h|<\delta$, перший доданок в (21) буде менший за $\dfrac{\varepsilon}{2}$, а тоді і $|p(x+h)-p(x)|<\varepsilon$. Це означає, що в точці x функція $p(x)$ – неперервна.

Вправи.

23. Нехай $f(t)$ – характеристична функція розподілу $F(x)$. Довести, що для точок неперервності $F(x)$ має місце рівність $F(x)=\lim\limits_{a\to0}\dfrac{1}{2\pi}\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{+\infty}e^{-\frac{1}{2}z^2a^2}f(z)e^{-izy}\,dz\,dy$.

24. Випадкова величина ξ має щільність розподілу $p(x)=\begin{cases}0, & |x|>a;\\ \dfrac{1}{a}\left(1-\dfrac{|x|}{a}\right), & |x|\le a.\end{cases}$

 Довести, що характеристична функція ξ рівна $f(t)=2\dfrac{1-\cos az}{a^2z^2}$.

25. Випадкова величина ξ має щільність розподілу $p(x)=\dfrac{1}{\pi}\dfrac{1-\cos ax}{ax^2}$. Довести, що характеристична функція ξ рівна $f(t)=\begin{cases}0, & |t|>a;\\ 1-\dfrac{|t|}{a}, & |t|\le a.\end{cases}$

6.3. Збіжність в основному послідовності функцій розподілу. Теореми Хеллі. Слабка збіжність послідовності функцій розподілу

Будемо розглядати множину всіх функцій розподілу, а також множину узагальнених функцій розподілу $G(x)$, які неспадні на R, неперервні зліва і

$0 \le G(-\infty) \le G(+\infty) \le 1$. Узагальнена функція розподілу $G(x)$ буде функцією розподілу тоді і тільки тоді, коли $G(-\infty)=0$, $G(+\infty)=1$.

Означення. Нехай $F(x)$, $F_n(x), n=1,2,...$ – узагальнені функції розподілу. Послідовність $F_n(x)$ називається *збіжною в основному* до функції $F(x)$, якщо для будь-якої точки неперервності x функції $F(x)$ $F_n(x) \xrightarrow[n\to\infty]{} F(x)$.

Відзначимо, що якщо послідовність функцій розподілу $F_n(x)$ збігається в основному до функції $F(x)$, то функція $F(x)$ може і не бути функцією розподілу. Наприклад, якщо $F_n(x)=\begin{cases}0, & x \le n;\\ 1, & x > n,\end{cases}$ то $F(x)=0$ для всіх $x \in R$.

Множина $D \subset R$ називається всюди щільною в R, якщо для довільного $\varepsilon > 0$ і для довільного $y \in R$ існує $x \in D$ таке, що $|x-y| < \varepsilon$.

Теорема 1. Нехай $D \subset R$ – довільна всюди щільна в R множина, $F(x)$, $F_n(x), n=1,2,...$ – узагальнені функції розподілу. Якщо для довільного $x \in D$ $F_n(x) \xrightarrow[n\to\infty]{} F(x)$, то $F_n(x)$ збігається в основному до $F(x)$. Якщо для всіх $x \in D$ існує $\lim\limits_{n\to\infty} F_n(x)$, то для деякої узагальненої функції розподілу $F(x)$ послідовність $F_n(x)$ збігається в основному до $F(x)$.

Доведення. Нехай x будь-яка точка неперервності функції $F(x)$, x_1 і x_2 – будь-які точки із D $(x_1 < x < x_2)$. Для довільного $x \in D$ $F_n(x) \xrightarrow[n\to\infty]{} F(x)$, тому при переході в нерівності $F_n(x_1) \le F_n(x) \le F_n(x_2)$ до границі, коли $n \to \infty$, одержимо $F(x_1)=\lim\limits_{n\to\infty} F_n(x_1)=\varliminf\limits_{n\to\infty} F_n(x_1) \le \varliminf\limits_{n\to\infty} F_n(x) \le \varlimsup\limits_{n\to\infty} F_n(x) \le \varlimsup\limits_{n\to\infty} F_n(x_2)=\lim\limits_{n\to\infty} F_n(x_2)=F(x_2)$ або $F(x_1) \le \varliminf\limits_{n\to\infty} F_n(x) \le \varlimsup\limits_{n\to\infty} F_n(x) \le F(x_2)$. У цій нерівності перейдемо до границі, коли $x_1 \to x-0$, а $x_2 \to x+0$, тоді $F(x-0) \le \varliminf\limits_{n\to\infty} F_n(x) \le \varlimsup\limits_{n\to\infty} F_n(x) \le F(x+0)$. Оскільки x – точка неперервності $F(x)$, тобто $F(x-0)=F(x+0)=F(x)$, то із попередньої нерівності випливає, що існує $\lim\limits_{n\to\infty} F_n(x)=F(x)$.

Для доведення другої частини теореми для будь-якого $x \in R$ покладемо $F(x)=\sup\limits_{y\in D, y<x} G(x)$, де $G(x)=\lim\limits_{n\to\infty} F_n(x)$, $x \in D$. Оскільки $F_n(x)$ – неспадні, то і $G(x)$ є неспадною на D. Із визначення $F(x)$ випливає, що і $F(x)$ буде неспадною. Із монотонності $F(x)$ одержуємо $F(x-0)=\sup\limits_{y<x} F(x)=\sup\limits_{y\in D, y<x} G(x)=F(x)$. Отже $F(x)$ є узагальненою функцією розподілу. Нехай x – точка неперервності $F(x)$.Для довільних дійсних u і v таких, що $u<x<v$ існують $u' \in D$ і $v' \in D$ і $u<u'<x<v'<v$. Тоді $F(u) \le G(u')=\lim\limits_{n\to\infty} F_n(u') \le \varliminf\limits_{n\to\infty} F_n(x) \le \varlimsup\limits_{n\to\infty} F_n(x) \le \lim\limits_{n\to\infty} F(v')=G(v') \le F(v)$. Якщо в цих нерівностях перейдемо до границі, коли $u \to x$ і $v \to x$, то одержимо $\lim\limits_{n\to\infty} F_n(x)=F(x)$.

Теорема 2 (1-ша теорема Хеллі). Із будь-якої послідовності функцій розподілу $F_1(x),...,F_n(x),...$ можна вибрати підпослідовність $F_{nn}(x)$, яка збігається в основному до деякої узагальненої функції розподілу $F(x)$.

Доведення. Нехай $D=\{x_1,x_2,...\}$ – зліченна всюди щільна в R множина. Розглянемо послідовність $\{F_n(x_1)\}$. Оскільки вона обмежена, то вона містить збіжну підпослідовність $F_{11}(x_1),...,F_{1n}(x_1),...$. Її границю позначимо $G(x_1)$.

Розглянемо послідовність $F_{11}(x),...,F_{1n}(x),...$ і покладемо $x=x_2$. Дістанемо знову обмежену числову послідовність $F_{11}(x_2),...,F_{1n}(x_2),...$. Із неї знову можна виділити збіжну підпослідовність $F_{21}(x_2),...,F_{2n}(x_2),...$. Її границю позначимо $G(x_2)$. Продовжуючи цей процес, ми для кожної точки $x_n\in D$ одержимо підпослідовність $F_{n1}(x),...,F_{nn}(x),...$, яка в кожній із точок x_k, $k=1,...,n$, буде збігатися до $G(x_k)$. На основі цих підпослідовностей побудуємо нову підпослідовність $F_{11}(x),F_{22}(x),...,F_{nn}(x),...$. Так побудована підпослідовність буде збігатися в кожній точці множини D до функції $G(x)$.

Покладемо $F(x)=\sup\limits_{y\in D,y<x} G(x)$, тоді за теоремою 1 побудована підпослідовність буде збігатися в основному до функції $F(x)$.

Теорема 3 (2-га теорема Хеллі). Нехай послідовність функцій розподілу $F_n(x)$ збігається в основному до функції розподілу $F(x)$ ($F(-\infty)=0$, $F(+\infty)=1$). Тоді для довільної неперервної і обмеженої на R функції $f(x)$

$$\lim_{n\to\infty}\int_{-\infty}^{+\infty} f(x)\,dF_n(x)=\int_{-\infty}^{+\infty} f(x)\,dF(x). \tag{22}$$

Доведення. Розглянемо довільний скінченний проміжок $[a,b]$, де a і b – точки неперервності $F(x)$. Із неперервності функції $f(x)$ на відрізку $[a,b]$ випливає, що для довільного $\varepsilon>0$ ми можемо вибрати таке розбиття проміжку $[a,b]$ точками $a=x_0<x_1<\cdots<x_m=b$, де x_k – точки неперервності $F(x)$, що для всіх $x\in[a,b]$ виконується нерівність

$$\left|f(x)-f_\varepsilon(x)\right|<\varepsilon, \tag{23}$$

де функція $f_\varepsilon(x)=f(x_{k-1})$ при $x\in[x_{k-1},x_k)$.

Розглянемо нерівність

$$\left|\int_a^b f(x)\,dF_n(x)-\int_a^b f(x)\,dF(x)\right|\le\left|\int_a^b f(x)\,dF_n(x)-\int_a^b f_\varepsilon(x)\,dF_n(x)\right|+$$
$$+\left|\int_a^b f_\varepsilon(x)\,dF_n(x)-\int_a^b f_\varepsilon(x)\,dF(x)\right|+\left|\int_a^b f_\varepsilon(x)\,dF(x)-\int_a^b f(x)\,dF(x)\right|. \tag{24}$$

Із (23) і умови теореми одержуємо

$$\left|\int_a^b f(x)\,dF_n(x)-\int_a^b f_\varepsilon(x)\,dF_n(x)\right|\le\int_a^b\left|f(x)-f_\varepsilon(x)\right|dF_n(x)<\varepsilon\left(F_n(b)-F_n(a)\right)\le\varepsilon, \tag{25}$$

$$\left|\int_a^b f_\varepsilon(x)\,dF(x)-\int_a^b f(x)\,dF(x)\right|<\varepsilon\left(F(b)-F(a)\right)\le\varepsilon. \tag{26}$$

Другий доданок правої частини (24) зобразимо у вигляді

$$\left|\int_a^b f_\varepsilon(x)\,dF_n(x)-\int_a^b f_\varepsilon(x)\,dF(x)\right|=\left|\sum_{k=1}^m\left(\int_{x_{k-1}}^{x_k} f_\varepsilon(x)\,dF_n(x)-\int_{x_{k-1}}^{x_k} f_\varepsilon(x)\,dF(x)\right)\right|=$$

$$=\left|\sum_{k=1}^m f(x_{k-1})\left(\int_{x_{k-1}}^{x_k} dF_n(x)-\int_{x_{k-1}}^{x_k} dF(x)\right)\right|=\left|\sum_{k=1}^m f(x_{k-1})\Big(\big(F_n(x_k)-F(x_k)\big)-\big(F_n(x_{k-1})-F(x_{k-1})\big)\Big)\right|.$$

Функція $f(x)$ обмежена, тому існує M, що $|f(x)|\le M$. Тому будемо мати

$$\left|\int_a^b f_\varepsilon(x)\,dF_n(x)-\int_a^b f_\varepsilon(x)\,dF(x)\right|\le 2M\sum_{k=0}^m\left|F_n(x_k)-F(x_k)\right|.$$

Послідовність $F_n(x)$ збігається в основному до $F(x)$, а x_k - точки неперервності функції $F(x)$ і цих точок скінченна кількість, тому для заданого $\varepsilon>0$ існує n_ε, що для всіх $n\ge n_\varepsilon$ буде виконуватись нерівність $\left|F_n(x_k)-F(x_k)\right|<\varepsilon$ для всіх x_k $(k=0,1,...,m)$. Отже,

$$\left|\int_a^b f_\varepsilon(x)\,dF_n(x)-\int_a^b f_\varepsilon(x)\,dF(x)\right|<2Mm\varepsilon. \tag{27}$$

Таким чином, із (24) – (27) випливає, що для довільного $\varepsilon>0$ існує n_ε, що для всіх $n\ge n_\varepsilon$ буде виконуватись нерівність $\left|\int_a^b f(x)dF_n(x)-\int_a^b f(x)dF(x)\right|<\varepsilon(2Mm+2)$, а це означає, що для довільного скінченного проміжку $[a,b]$

$$\lim_{n\to\infty}\int_a^b f(x)dF_n(x)=\int_a^b f(x)dF(x). \tag{28}$$

Перейдемо до доведення (22). Для довільного $\varepsilon>0$ виберемо A і B так, щоб A і B були точками неперервності $F(x)$ і $F(A)+1-F(B)<\varepsilon$. Тоді існує n_0, що для всіх $n\ge n_0$ $F_n(A)+1-F_n(B)<2\varepsilon$. Зафіксуємо такі A і B.

$$\left|\int_{-\infty}^{+\infty} f(x)dF_n(x)-\int_{-\infty}^{+\infty} f(x)dF(x)\right|\le\left|\int_A^B f(x)dF_n(x)-\int_A^B f(x)dF(x)\right|+$$

$$+\left|\int_{-\infty}^{A} f(x)dF_n(x)\right|+\left|\int_{B}^{+\infty} f(x)dF_n(x)\right|+\left|\int_{-\infty}^{A} f(x)dF(x)\right|+\left|\int_{B}^{+\infty} f(x)dF(x)\right|\le$$

$$\le\left|\int_A^B f(x)dF_n(x)-\int_A^B f(x)dF(x)\right|+M\left(F_n(A)+1-F_n(B)+F(A)+1-F(B)\right)<$$

$$\le\left|\int_A^B f(x)dF_n(x)-\int_A^B f(x)dF(x)\right|+M3\varepsilon.$$

Із (28) випливає, що і перший доданок в правій частині вибором n може бути як завгодно малим, тому із умов теореми одержуємо (22).

Означення. Нехай $F_n(x)$, $n=1,2,...$, і $F(x)$ є функціями розподілу. Якщо для довільної неперервної і обмеженої на R функції $f(x)$ $\lim_{n\to\infty}\int_{-\infty}^{+\infty} f(x)dF_n(x)=\int_{-\infty}^{+\infty} f(x)dF(x)$, то кажуть, що послідовність $F_n(x)$ *слабко збігається* до $F(x)$ (позначають $F_n(x)\xrightarrow[n\to\infty]{W}F(x)$ або $F_n(x)\Rightarrow F(x)$).

Теорема 4. Для того, щоб послідовність функцій розподілу $F_n(x)$ слабко збігалась до функції розподілу $F(x)$, необхідно і достатньо, щоб $F_n(x)$ збігалась в основному до $F(x)$.

Доведення. Із теореми 3 випливає, що із збіжності в основному випливає слабка збіжність. Нехай має місце слабка збіжність. Тобто, для довільної неперервної і обмеженої на R функції $f(x)$ $\lim\limits_{n\to\infty}\int\limits_{-\infty}^{+\infty} f(x)dF_n(x) = \int\limits_{-\infty}^{+\infty} f(x)dF(x)$.

Виберемо функцію $f(u)$ у вигляді: $f(u) = \begin{cases} 1, & u \le x; \\ 1 - \dfrac{u-x}{\varepsilon}, & u \in (x, x+\varepsilon); \\ 0, & u \ge x+\varepsilon. \end{cases}$

Тоді

$$F_n(x) = \int\limits_{-\infty}^{x} dF_n(u) = \int\limits_{-\infty}^{x} f(u)dF_n(u) \le \int\limits_{-\infty}^{+\infty} f(u)dF_n(u)$$

і

$$\overline{\lim_{n\to\infty}} F_n(x) \le \lim_{n\to\infty}\int\limits_{-\infty}^{+\infty} f(u)dF_n(u) = \int\limits_{-\infty}^{+\infty} f(u)dF(u) = \int\limits_{-\infty}^{x+\varepsilon} f(u)dF(u) \le \int\limits_{-\infty}^{x+\varepsilon} dF(u) = F(x+\varepsilon).$$

А тепер виберемо $f(u) = \begin{cases} 1, & u \le x-\varepsilon; \\ \dfrac{x-u}{\varepsilon}, & u \in (x-\varepsilon, x); \\ 0, & u \ge x. \end{cases}$

Тоді

$$F_n(x) = \int\limits_{-\infty}^{x} dF_n(u) \ge \int\limits_{-\infty}^{x} f(u)dF_n(u) = \int\limits_{-\infty}^{+\infty} f(u)dF_n(u).$$

Звідки,

$$\underline{\lim_{n\to\infty}} F_n(x) \ge \lim_{n\to\infty}\int\limits_{-\infty}^{+\infty} f(x)dF_n(x) = \int\limits_{-\infty}^{+\infty} f(x)dF(x) \ge \int\limits_{-\infty}^{x-\varepsilon} f(u)dF(u) = \int\limits_{-\infty}^{x-\varepsilon} dF(u) = F(x-\varepsilon).$$

Отже, ми одержали таку нерівність:

$$F(x-\varepsilon) \le \underline{\lim_{n\to\infty}} F_n(x) \le \overline{\lim_{n\to\infty}} F_n(x) \le F(x+\varepsilon).$$

Якщо x – точка неперервності $F(x)$, то при переході до границі в останній нерівності, коли $\varepsilon \to 0$, одержимо: $\underline{\lim\limits_{n\to\infty}} F_n(x) = \overline{\lim\limits_{n\to\infty}} F_n(x) = F(x)$. Тобто існує $\lim\limits_{n\to\infty} F_n(x) = F(x)$, а це і означає, що $F_n(x)$ збігається до $F(x)$ в основному.

6.4. Неперервна відповідність між збіжністю функцій розподілу і характеристичних функцій

Розглянемо тепер одну із найбільш важливих властивостей характеристичних функцій, яка є основним засобом доведення теорем про слабку збіжність розподілів. Нехай $F_n(x)$, $F(x)$ – функції розподілу, а $f_n(t)$, $f(t)$ – відповідні характеристичні функції.

Теорема 1. Нехай послідовність функцій розподілу $F_n(x)$ слабко збігається до функції розподілу $F(x)$, тоді послідовність характеристичних функцій $f_n(t)\xrightarrow[n\to\infty]{}f(t)$, де $f(t)=\int\limits_{-\infty}^{+\infty}e^{itx}dF(x)$. При цьому ця збіжність є рівномірною в кожному скінченному проміжку зміни t.

Доведення. Для кожного t функції $\cos(tx)$ і $\sin(tx)$ – неперервні і обмежені при $x\in R$, $F_n(x)\xrightarrow[n\to\infty]{}F(x)$ слабко, а ця збіжність еквівалентна збіжності в основному, тому за другою теоремою Хеллі

$$f_n(t)=\int\limits_{-\infty}^{+\infty}\exp(itx)dF_n(x)=\int\limits_{-\infty}^{+\infty}\cos(tx)dF_n(x)+i\int\limits_{-\infty}^{+\infty}\sin(tx)dF_n(x)\xrightarrow[n\to\infty]{}$$

$$\xrightarrow[n\to\infty]{}\int\limits_{-\infty}^{+\infty}\cos(tx)dF(x)+i\int\limits_{-\infty}^{+\infty}\sin(tx)dF(x)=\int\limits_{-\infty}^{+\infty}\exp(itx)dF(x)=f(t),$$

тобто, послідовність характеристичних функцій $f_n(t)\xrightarrow[n\to\infty]{}f(t)$.

Доведення рівномірної збіжності є повторенням доведення теореми 3.

Теорема 2. Якщо послідовність характеристичних функцій $f_n(t)$ збігається до функції $f(t)$, що неперервна в точці $t=0$, тоді відповідна послідовність функцій розподілу $F_n(x)$ буде збігатися в основному до функції розподілу $F(x)$, для якої $f(t)$ буде відповідною характеристичною функцією, тобто $f(t)=\int\limits_{-\infty}^{+\infty}e^{itx}dF(x)$.

Доведення. За першою теоремою Хеллі із $F_n(x)$ можна виділити підпослідовність $F_{n_k}(x)$, яка слабко збігається до деякої неспадної функції $F(x)$. Покажемо, що $F(x)$ буде функцією розподілу. Для цього досить показати, що $F(+\infty)-F(-\infty)=1$. Для довільного $a>0$ покладемо $b=\dfrac{2}{a}$ і розглянемо інтеграл

$$\frac{1}{2a}\int\limits_{-a}^{a}\left(1-f_{n_k}(t)\right)dt=\frac{1}{2a}\int\limits_{-a}^{a}dt\int\limits_{-\infty}^{+\infty}\left(1-e^{itx}\right)dF_{n_k}(x)=\frac{1}{2a}\int\limits_{-\infty}^{+\infty}dF_{n_k}(x)\int\limits_{-a}^{a}\left(1-e^{itx}\right)dt=$$

$$=\int\limits_{-\infty}^{+\infty}\left(1-\frac{\sin ax}{ax}\right)dF_{n_k}(x)\geq\frac{1}{2}\int\limits_{|x|\geq b}dF_{n_k}(x)\geq\frac{1}{2}\left(F_{n_k}(-b)+1-F_{n_k}(b)\right). \qquad (29)$$

Ми використали рівність $\dfrac{1}{2a}\int\limits_{-a}^{a}e^{itx}dt=\dfrac{\sin ax}{ax}$, невід'ємність виразу $1-\dfrac{\sin ax}{ax}$ і нерівність $\left|\dfrac{\sin ax}{ax}\right|\leq\dfrac{1}{a|x|}\leq\dfrac{1}{2}$ при $|x|\geq b=\dfrac{2}{a}$. Нехай $-b$ і b – точки неперервності $F(x)$. Перейдемо в нерівності (29) до границі, коли $n_k\to\infty$, і врахуємо, що $\int\limits_{-a}^{a}f_{n_k}(t)dt\to\int\limits_{-a}^{a}f(t)dt$, тоді одержимо $\dfrac{1}{2a}\int\limits_{-a}^{a}\left(1-f(t)\right)dt\geq\dfrac{1}{2}\left(F(-b)+1-F(b)\right)$. Нехай тепер $a\to 0$. Оскільки функція $f(t)$ – неперервна в точці $t=0$ і $f(0)=1$, то $\dfrac{1}{2a}\int\limits_{-a}^{a}\left(1-f(t)\right)dt\to 0$, а це означає, що

$F(+\infty)-F(-\infty)=1$. Тобто $F(x)$ є функцією розподілу, а $f(t)$ буде відповідною характеристичною функцією, тобто $f(t)=\int\limits_{-\infty}^{+\infty} e^{itx}dF(x)$.

Щоб завершити доведення теореми, необхідно показати, що і вся послідовність $F_n(x)$ збігається в основному до $F(x)$. Припустимо, що це не виконується. Тоді існує підпослідовність $F_{m_k}(x)$, яка збігається в основному до $G(x)$ і $G(x)\neq F(x)$ принаймні в одній точці неперервності функції $G(x)$. Але відповідна послідовність характеристичних функцій $f_{m_k}(t)\to f(t)$, яка є характеристичною функцією для $G(x)$. А це означає, що $G(x)=F(x)$. Одержали протиріччя, яке доводить теорему. Тобто, $F_n(x)$ збігається в основному до функції розподілу $F(x)$.

Розділ 7. Класична центральна гранична теорема

7.1. Центральна гранична теорема (ЦГТ)

7.1.1. Постановка задачі. Умова Ліндеберга, її ймовірнісний зміст. До ЦГТ відносяться твердження, у яких іде мова про збіжність розподілів сум до нормального закону. Нехай маємо послідовність випадкових величин $\xi_1,...,\xi_n,...$. Якщо існують послідовності дійсних чисел a_n і $b_n > 0$ такі, що функція розподілу суми $\dfrac{\sum_{i=1}^{n}\xi_i - a_n}{b_n}$ збігається до функції розподілу стандартного нормального закону при $n \to \infty$, то кажуть, що для послідовності $\xi_1,...,\xi_n,...$ має місце ЦГТ. В цьому випадку говорять, що випадкова величина $\xi_1 + ... + \xi_n$ асимптотично нормальна з параметрами a_n і b_n^2. Асимптотична нормальність відіграє важливу роль в теорії і на практиці, вона дозволяє замінити складні розподіли сум добре вивченим нормальним розподілом. ЦГТ є підставою для широкого використання нормального закону в статистиці.

Нехай задана послідовність незалежних випадкових величин $\xi_1,...,\xi_n,...$ із скінченними математичними сподіваннями $M\xi_k = a_k$, скінченними дисперсіями $D\xi_k = \sigma_k^2$ і функціями розподілу $F_k(x) = P\{\xi_k < x\}$. Позначимо через $B_n^2 = \sigma_1^2 + \cdots + \sigma_n^2 = D(\xi_1 + \cdots + \xi_n)$ і нехай для довільного $\varepsilon > 0$

$$L_n(\varepsilon) = \frac{1}{B_n^2}\sum_{k=1}^{n}\int\limits_{|x-a_k| > \varepsilon B_n}(x-a_k)^2 dF_k(x).$$

Якщо для довільного $\varepsilon > 0$

$$\lim_{n\to\infty} L_n(\varepsilon) = 0, \tag{1}$$

то (1) називають *умовою Ліндеберга*.

Вияснимо ймовірнісний зміст умови Ліндеберга, для цього розглянемо події $A_k = \{\omega : |\xi_k - a_k| > \varepsilon B_n\}$, $k = 1,...,n$, і подію

$$A = \left\{\omega : \max_{1\le k\le n}|\xi_k - a_k| > \varepsilon B_n\right\}.$$

Тоді $A = A_1 + A_2 + \cdots + A_n$, а

$$P(A) \le \sum_{k=1}^{n} P(A_k) = \sum_{k=1}^{n}\int\limits_{|x-a_k| > \varepsilon B_n} dF_k(x).$$

Якщо врахуємо, що в області інтегрування $\dfrac{|x-a_k|}{\varepsilon B_n} \ge 1$, то одержимо

$$P(A) \le \sum_{k=1}^{n}\int\limits_{|x-a_k| > \varepsilon B_n}\frac{(x-a_k)^2}{\varepsilon^2 B_n^2} dF_k(x) \le \frac{1}{\varepsilon^2} L_n(\varepsilon). \tag{2}$$

Якщо виконується умова Ліндеберга, то із (2) випливає, що

$$\lim_{n\to\infty} P\left\{\omega : \max_{1\le k\le n}|\xi_k - a_k| > \varepsilon B_n\right\} = 0.$$

Тобто, якщо виконується умова Ліндеберга, то випадкові величини $\frac{\xi_k - a_k}{B_n}$ будуть рівномірно нескінченно малі.

7.1.2. Теорема Ліндеберга. Будемо використовувати позначення, що введені в попередньому пункті. Крім того, позначимо $S_n = \frac{1}{B_n}\sum_{k=1}^{n}(\xi_k - a_k)$, $\Phi_n(x) = P\{S_n < x\}$ – функція розподілу випадкової величини S_n, а $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$ – функція розподілу стандартного нормального закону.

Теорема 1. Нехай задана послідовність незалежних випадкових величин $\xi_1,...,\xi_n,...$ із скінченними математичними сподіваннями $M\xi_i = a_i$ і скінченними дисперсіями $D\xi_i = \sigma_i^2$. Якщо для довільного $\varepsilon > 0$ виконується умова Ліндеберга (1), то рівномірно відносно x

$$\lim_{n\to\infty}\Phi_n(x) = \Phi(x). \tag{3}$$

Доведення. Позначимо $\xi_{nk} = \frac{\xi_k - a_k}{B_n}$, $F_{nk}(x)$ – функцію розподілу ξ_{nk}. Тоді $S_n = \sum_{k=1}^{n}\xi_{nk}$, $M\xi_{nk} = 0$, $D\xi_{nk} = \frac{\sigma_k^2}{B_n^2}$, $\sum_{k=1}^{n}D\xi_{nk} = 1$, $F_{nk}(x) = P\left\{\frac{\xi_k - a_k}{B_n} < x\right\} = P\{\xi_k < xB_n + a_k\} = = F_k(xB_n + a_k)$.

В цих позначеннях $L_n(\varepsilon)$, після виконання в інтегралі заміни $\frac{x-a_k}{B_n} = y$, набуває вигляду:

$$L_n(\varepsilon) = \frac{1}{B_n^2}\sum_{k=1}^{n}\int\limits_{|x-a_k|>\varepsilon B_n}(x-a_k)^2 dF_k(x) = \sum_{k=1}^{n}\int\limits_{|y|>\varepsilon} y^2 dF_k(yB_n + a_k) =$$

$$= \sum_{k=1}^{n}\int\limits_{|x|>\varepsilon} x^2 dF_{nk}(x). \tag{4}$$

Нехай $f_{nk}(t) = M\exp(it\xi_{nk})$, тоді характеристична функція S_n

$$\varphi_n(t) = \prod_{k=1}^{n} f_{nk}(t).$$

Покажемо, що $f_{nk}(t)$ при $n\to\infty$ рівномірно відносно k $(1 \le k \le n)$ прямують до одиниці. Врахувавши, що $M\xi_{nk} = 0$, одержимо

$$f_{nk}(t) - 1 = \int_{-\infty}^{+\infty}\left(e^{itx} - 1 - itx\right)dF_{nk}(x).$$

Із нерівності (Н1) із попереднього розділу при $n = 1$ для довільного дійсного α

$$\left|\exp(i\alpha) - 1 - i\alpha\right| \le \frac{\alpha^2}{2},$$

тому із (4) для будь-якого додатного ε

$$\left|f_{nk}(t) - 1\right| \le \left|\int_{-\infty}^{\infty}\frac{t^2x^2}{2}dF_{nk}(x)\right| = \frac{t^2}{2}\left(\int\limits_{|x|\le\varepsilon} x^2 dF_{nk}(x) + \int\limits_{|x|>\varepsilon} x^2 dF_{nk}(x)\right) \le \frac{t^2}{2}\left(\varepsilon^2 + L_n(\varepsilon)\right), \tag{5}$$

а при $|t| \le T$

$$\max_{1\le k\le n}\left|f_{nk}(t)-1\right|\le\frac{T^2}{2}\left(\varepsilon^2+L_n(\varepsilon)\right). \qquad (6)$$

Отже, із умови Ліндеберга (1) випливає, що

$$\lim_{n\to\infty}\max_{1\le k\le n}\left|f_{nk}(t)-1\right|=0$$

і при всіх досить великих n і $|t|\le T$

$$\max_{1\le k\le n}\left|f_{nk}(t)-1\right|<\frac{1}{2}. \qquad (7)$$

Це означає, що $f_{nk}(t)$ не перетворюється в 0, тому при $|t|\le T$ $\varphi_n(t)$ можна логарифмувати:

$$\ln\varphi_n(t)=\sum_{k=1}^{n}\ln f_{nk}(t).$$

Щоб довести (3), на основі теореми про неперервну відповідність між збіжністю функцій розподілу і відповідних характеристичних функцій, нам треба довести, що

$$\varphi_n(t)\xrightarrow[n\to\infty]{}e^{-\frac{t^2}{2}},$$

а для цього достатньо довести, що

$$\ln\varphi_n(t)\xrightarrow[n\to\infty]{}-\frac{t^2}{2}. \qquad (8)$$

Використовуючи розклад $\ln(1+z)=z-\frac{z^2}{2}+\frac{z^3}{3}-\cdots$, при $|z|<\frac{1}{2}$ одержимо $\left|\ln(1+z)-z\right|\le\sum_{k=2}^{\infty}\frac{|z|^k}{k}\le\frac{1}{2}\sum_{k=2}^{\infty}|z|^k=\frac{|z|^2}{2(1-|z|)}\le|z|^2$. Звідки, при $|z|<\frac{1}{2}$

$$\ln(1+z)-z=\frac{\ln(1+z)-z}{z^2}z^2=\theta z^2, \qquad (9)$$

де $|\theta|=\left|\frac{\ln(1+z)-z}{z^2}\right|\le1$.

Із (7) і (9) випливає справедливість рівності

$$\ln\varphi_n(t)=\sum_{k=1}^{n}\ln\left(\left(f_{nk}(t)-1\right)+1\right)=\sum_{k=1}^{n}\left(f_{nk}(t)-1\right)+\sum_{k=1}^{n}\theta_k\left(f_{nk}(t)-1\right)^2.$$

Тоді

$$\left|\ln\varphi_n(t)+\frac{t^2}{2}\right|\le\left|\frac{t^2}{2}+\sum_{k=1}^{n}\left(f_{nk}(t)-1\right)\right|+\sum_{k=1}^{n}\left|f_{nk}(t)-1\right|^2=\alpha_n+\beta_n, \qquad (10)$$

де $\alpha_n=\left|\frac{t^2}{2}+\sum_{k=1}^{n}\left(f_{nk}(t)-1\right)\right|,\quad \beta_n=\sum_{k=1}^{n}\left|f_{nk}(t)-1\right|^2$.

Із (5) випливає, що $\left|f_{nk}(t)-1\right|\le\frac{t^2}{2}\int_{-\infty}^{\infty}x^2dF_{nk}(x)\le\frac{t^2}{2}\frac{\sigma_k^2}{B_n^2}$, а оскільки $\sum_{k=1}^{n}\frac{\sigma_k^2}{B_n^2}=1$, то звідси одержуємо наступну нерівність

$$\beta_n\le\max_{1\le k\le n}\left|f_{nk}(t)-1\right|\sum_{k=1}^{n}\left|f_{nk}(t)-1\right|\le\max_{1\le k\le n}\left|f_{nk}(t)-1\right|\sum_{k=1}^{n}\frac{t^2\sigma_k^2}{2B_n^2}\le\max_{1\le k\le n}\left|f_{nk}(t)-1\right|\frac{t^2}{2}.$$

Тому із (6) випливає, що при всіх досить великих n і $|t|\le T$

$$\beta_n \le \max_{1\le k\le n}\left|f_{nk}(t)-1\right|\frac{t^2}{2} \le \frac{T^4}{4}\left(\varepsilon^2 + L_n(\varepsilon)\right). \qquad (11)$$

Переходимо до оцінки α_n:

$$\alpha_n = \left|\frac{t^2}{2} + \sum_{k=1}^{n}\left(f_{nk}(t)-1\right)\right| = \left|\sum_{k=1}^{n}\left(\frac{\sigma_k^2 t^2}{2B_n^2} + f_{nk}(t)-1\right)\right| = \left|\sum_{k=1}^{n}\int_{-\infty}^{+\infty}\left(e^{itx}-1-itx-\frac{(itx)^2}{2}\right)dF_{nk}(x)\right| \le$$

$$\le \sum_{k=1}^{n}\int_{-\infty}^{+\infty}\left|e^{itx}-1-itx-\frac{(itx)^2}{2}\right|dF_{nk}(x) = \sum_{k=1}^{n}\int_{|x|\le\varepsilon}\left|e^{itx}-1-itx-\frac{(itx)^2}{2}\right|dF_{nk}(x) +$$

$$+\sum_{k=1}^{n}\int_{|x|>\varepsilon}\left|e^{itx}-1-itx-\frac{(itx)^2}{2}\right|dF_{nk}(x).$$

У першому інтегралі використаємо нерівність (Н1) $\left|e^{itx}-1-itx-\frac{(itx)^2}{2}\right| \le \frac{|tx|^3}{6}$, а у другому – (Н2) $\left|e^{itx}-1-itx-\frac{(itx)^2}{2}\right| \le 2\frac{|tx|^2}{2}$, тоді

$$\alpha_n \le \frac{|t|^3}{6}\sum_{k=1}^{n}\int_{|x|\le\varepsilon}|x|^3 dF_{nk}(x) + t^2\sum_{k=1}^{n}\int_{|x|>\varepsilon}x^2 dF_{nk}(x) \le \frac{\varepsilon|t|^3}{6}\sum_{k=1}^{n}\int_{|x|\le\varepsilon}x^2 dF_{nk}(x) + t^2 L_n(\varepsilon).$$

Врахувавши, що $\sum_{k=1}^{n}\int_{|x|\le\varepsilon}x^2 dF_{nk}(x) \le \sum_{k=1}^{n}D\xi_{nk} = 1$, в кожному скінченному проміжку $|t|\le T$

$$\alpha_n \le \frac{\varepsilon T^3}{6} + T^2 L_n(\varepsilon). \qquad (12)$$

Із (10), (11) і (12) одержуємо, що при всіх досить великих n і $|t|\le T$

$$\left|\ln\varphi_n(t) + \frac{t^2}{2}\right| \le \frac{T^4}{4}\left(\varepsilon^2 + L_n(\varepsilon)\right) + \frac{\varepsilon T^3}{6} + T^2 L_n(\varepsilon).$$

Звідки, із довільності ε і умови Ліндеберга (1) випливає, що рівномірно по t в кожному скінченному проміжку виконується (8), а це рівносильно (3). Теорема доведена.

7.1.3. ЦГТ для послідовності серій. Розглянемо послідовність серій випадкових величин

$$\xi_{n1}, \xi_{n2}, \ldots, \xi_{nm_n}, \quad n = 1, 2, \ldots \qquad (13)$$

Будемо вважати, що величини, які входять в одну серію, незалежні. Покладемо

$$S_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nm_n}.$$

Теорема 2. Нехай послідовність серій незалежних випадкових величин (13) така, що величини ξ_{nk} мають скінченні моменти другого порядку, $M\xi_{nk} = 0$, $D\xi_{nk} = \sigma_{nk}^2$, $\sum_{k=1}^{m_n}\sigma_{nk}^2 = 1$ і виконується умова Ліндеберга

$$L_n(\varepsilon) = \sum_{k=1}^{m_n}\int_{|x|>\varepsilon}x^2 dF_{nk}(x) \xrightarrow[n\to\infty]{} 0, \quad \forall\varepsilon > 0,$$

де $F_{nk}(x)$ – функція розподілу ξ_{nk}, тоді послідовність сум S_n асимптотично нормальна з параметрами 0 і 1.

Доведення практично не відрізняється від доведення теореми 1.

7.1.4. ЦГТ для однаково розподілених випадкових величин.

Теорема 3. Якщо випадкові величини $\xi_1,...,\xi_n,...$ – незалежні, однаково розподілені, мають скінченні математичні сподівання $M\xi_i = a$ і скінченні дисперсії $D\xi_i = \sigma^2$, то $\lim\limits_{n\to\infty} P\left\{\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(\xi_k - a) < x\right\} = \Phi(x)$.

Доведення. В цьому випадку $B_n^2 = n\sigma^2$, $B_n = \sigma\sqrt{n}$. Позначимо через $F(x)$ функцію розподілу ξ_k. Тоді

$$L_n(\varepsilon) = \frac{1}{n\sigma^2}\sum_{k=1}^{n}\int\limits_{|x-a|>\varepsilon\sigma\sqrt{n}}(x-a)^2 dF(x) = \frac{1}{\sigma^2}\int\limits_{|x-a|>\varepsilon\sigma\sqrt{n}}(x-a)^2 dF(x) =$$

$$= \frac{1}{\sigma^2}\left(\int\limits_{-\infty}^{a-\varepsilon\sigma\sqrt{n}}(x-a)^2\,dF(x) + \int\limits_{a+\varepsilon\sigma\sqrt{n}}^{+\infty}(x-a)^2\,dF(x)\right)\xrightarrow[n\to\infty]{}0,$$

бо із існування дисперсії $\sigma^2 = \int\limits_{-\infty}^{+\infty}(x-a)^2\,dF(x)$ випливає, що інтеграл збіжний, тому

$$\int\limits_{-\infty}^{a-\varepsilon\sigma\sqrt{n}}(x-a)^2\,dF(x)\xrightarrow[n\to\infty]{}0 \text{ і } \int\limits_{a+\varepsilon\sigma\sqrt{n}}^{+\infty}(x-a)^2\,dF(x)\xrightarrow[n\to\infty]{}0.$$

А це означає, що виконується умова Ліндеберга (1), тому виконується і (3).

Наслідок (Інтегральна теорема Муавра-Лапласа). Якщо μ_n є число появ події A в n незалежних випробуваннях, у кожному із яких ймовірність настання цієї події дорівнює p і $0 < p < 1$, то рівномірно відносно a і b $(-\infty \le a \le b \le \infty)$

$$\lim_{n\to\infty} P\left\{a \le \frac{\mu_n - np}{\sqrt{npq}} < b\right\} = \frac{1}{\sqrt{2\pi}}\int\limits_a^b e^{-\frac{x^2}{2}}dx. \tag{14}$$

Доведення. Позначимо через ξ_k число появ події в k-му експерименті, тоді число появ події в n експериментах $\mu_n = \xi_1 + \cdots + \xi_n$. Випадкові величини ξ_k – незалежні однаково розподілені за законом

ξ_k	0	1
P	q	p

$M\xi_k = p,\;\; D\xi_k = pq$. Тоді $S_n = \frac{\xi_1 + \cdots + \xi_n - np}{\sqrt{npq}} = \frac{\mu_n - np}{\sqrt{npq}}$ і виконуються умови теореми 3. Тому для довільного x

$$\lim_{n\to\infty} P\left\{\frac{\mu_n - np}{\sqrt{npq}} < x\right\} = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}dt. \tag{15}$$

Оскільки $P\left\{a \le \frac{\mu_n - np}{\sqrt{npq}} < b\right\} = P\left\{\frac{\mu_n - np}{\sqrt{npq}} < b\right\} - P\left\{\frac{\mu_n - np}{\sqrt{npq}} < a\right\}$, то із (15) одержуємо (14).

7.1.5. Теорема Ляпунова.

Теорема 4. Нехай $\xi_1,...,\xi_n,...$ – незалежні випадкові величини, що мають скінченні математичні сподівання $M\xi_k = a_k$, дисперсії $D\xi_k = \sigma_k^2$, для деякого $\delta > 0$ існують $M\left|\xi_k\right|^{2+\delta} < \infty$ і

$$\lim_{n\to\infty}\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n} M\left|\xi_k - a_k\right|^{2+\delta} = 0, \qquad (16)$$

тоді $\lim_{n\to\infty}\Phi_n\left(x\right) = \Phi\left(x\right)$.

Доведення. Для довільного $\varepsilon > 0$

$$L_n\left(\varepsilon\right) = \frac{1}{B_n^2}\sum_{k=1}^{n}\int\limits_{\left|x-a_k\right| > \varepsilon B_n}\left(x-a_k\right)^2 dF_k\left(x\right) \le \frac{1}{\varepsilon^\delta B^{2+\delta}}\sum_{k=1}^{n}\int\limits_{\left|x-a_k\right| > \varepsilon B_n}\left|x-a_k\right|^{2+\delta} dF_k\left(x\right) \le$$

$$\le \frac{1}{\varepsilon^\delta B^{2+\delta}}\sum_{k=1}^{n}\int_{-\infty}^{+\infty}\left|x-a_k\right|^{2+\delta} dF_k\left(x\right) = \frac{1}{\varepsilon^\delta B_n^{2+\delta}}\sum_{k=1}^{n} M\left|\xi_k - a_k\right|^{2+\delta}.$$

Отже, із умови (16) випливає виконання умови Ліндеберга (1), тому виконується і (3).

Вправи.

1. Випадкова величина ξ має розподіл Пуассона з параметром $\lambda > 0$. Показати, що $\eta = \dfrac{\xi - M\xi}{\sqrt{D\xi}}$ асимптотично нормальна при $\lambda \to \infty$.
2. Випадкова величина ξ має гамма-розподіл із щільністю $p\left(x\right) = \begin{cases} 0, & x \le 0; \\ \dfrac{b^a}{\Gamma\left(a\right)} x^{a-1} e^{-bx}, & x > 0 \end{cases}$ $(a > 0,\quad b > 0)$. Показати, що $\eta = \dfrac{\xi - M\xi}{\sqrt{D\xi}}$ асимптотично нормальна при $a \to \infty$.
3. Випадкова величина ξ має розподіл χ^2 із щільністю $p\left(x\right) = \begin{cases} 0, & x \le 0; \\ \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0. \end{cases}$ Показати, що $\eta = \dfrac{\xi - n}{\sqrt{2n}}$ асимптотично нормальна при $n \to \infty$.
4. Нехай μ_n є число появ події A в n незалежних випробуваннях, у кожному із яких ймовірність настання цієї події дорівнює p і $0 < p < 1$. Використовуючи ЦГТ, оцінити ймовірність того, що μ_n буде відрізнятися від p не більше, ніж на ε. Знайти n, при якому для $\varepsilon = 0,01$ ця ймовірність не перевищує 0,05.
5. Нехай ξ_k, $k \ge 1$ – незалежні випадкові величини, причому $P\{\xi_1 = \pm 1\} = \dfrac{1}{2}$, а при $k > 1$ $P\{\xi_k = \pm 1\} = \dfrac{1}{2}\left(1 - \dfrac{1}{k^2}\right)$ і $P\{\xi_k = \pm\sqrt{k}\} = \dfrac{1}{2k^2}$. Чи виконується для сум S_n ЦГТ?

7.2. Гратчасті розподіли. Локальна гранична теорема для гратчастих розподілів

Дискретна випадкова величина ξ має *гратчастий* (*арифметичний*) *розподіл*, якщо її можливі значення можна представити у вигляді $a + kh$, де $a \in R$, $h > 0$ – крок розподілу,

k може приймати будь-які цілі значення. Крок h називається *максимальним*, якщо ні для якого числа b і $h_1 > h$ можливі значення не можуть бути представлені у вигляді $b + kh_1$.

Теорема 1. Для того, щоб випадкова величина ξ мала гратчастий розподіл необхідно і досить, щоб існувало таке число $t_0 \neq 0$, що $\left|f\left(t_0\right)\right| = 1$.

Доведення. Нехай випадкова величина ξ має гратчастий розподіл і приймає значення $a + kh$ з ймовірностями p_k ,тоді характеристична функція випадкової величини ξ

$$f\left(t\right) = Me^{it\xi} = \sum_{k=-\infty}^{\infty} e^{it\left(a+kh\right)} p_k = e^{ita} \sum_k e^{itkh} p_k . \tag{17}$$

Виберемо $t_0 = \frac{2\pi}{h}$, тоді $f\left(\frac{2\pi}{h}\right) = e^{i\frac{2\pi}{h}a} \sum_k e^{i2\pi k} p_k = e^{i\frac{2\pi}{h}a}$. Тому $\left|f\left(\frac{2\pi}{h}\right)\right| = 1$.

Нехай існує $t_0 \neq 0$ таке, що $\left|f\left(t_0\right)\right| = 1$, тоді $f\left(t_0\right) = e^{i\varphi}$, де φ – фіксоване дійсне число. За означенням $f\left(t\right) = \int\limits_{-\infty}^{+\infty} e^{itx} dF\left(x\right)$. Покладемо $t = t_0$, тоді одержимо: $e^{i\varphi} = \int\limits_{-\infty}^{+\infty} e^{it_0x} dF\left(x\right)$ або $1 = \int\limits_{-\infty}^{+\infty} e^{i\left(t_0x-\varphi\right)} dF\left(x\right)$. Звідки одержуємо рівність

$$1 = \int\limits_{-\infty}^{+\infty} \cos\left(t_0x - \varphi\right) dF\left(x\right).$$

Ця рівність може виконуватись тільки тоді, коли в точках росту функції $F\left(x\right)$ виконується рівність $\cos\left(t_0x - \varphi\right) = 1$. Це означає, що можливі значення ξ мають вигляд $x = \frac{\varphi}{t_0} + \frac{2\pi}{t_0}k$, де $a = \frac{\varphi}{t_0}$, $h = \left|\frac{2\pi}{t_0}\right|$. Що і треба було довести.

Відзначимо, що із (17) випливає така властивість: якщо випадкова величина має гратчастий розподіл із кроком h, то модуль характеристичної функції є функція періодична із періодом $\frac{2\pi}{h}$, тобто $\left|f\left(t + \frac{2\pi}{h}\right)\right| = \left|f\left(t\right)\right|$.

Теорема 2. Для того, щоб крок h був максимальним необхідно і досить, щоб для довільного $t \in \left(0, \frac{2\pi}{h}\right)$ $\left|f\left(t\right)\right| < 1$, а $\left|f\left(\frac{2\pi}{h}\right)\right| = 1$.

Доведення. Якщо крок h - максимальний, то можливі значення не можуть бути представлені у вигляді $b + kh_1$, де $h_1 > h$. Нехай h - максимальний. Припустимо, що в деякій точці $t_0 \in \left(0, \frac{2\pi}{h}\right)$ $\left|f\left(t_0\right)\right| = 1$. По доведеному це означає, що можливі значення x можна представити у вигляді $x = \frac{\varphi}{t_0} + \frac{2\pi}{t_0}k$ із кроком $h_1 = \frac{2\pi}{t_0}$. Але $0 < t_0 < \frac{2\pi}{h}$, тому $h < \frac{2\pi}{t_0} = h_1$. Отже, одержали протиріччя, яке доводить, що $\left|f\left(t\right)\right| < 1$ при $t \in \left(0, \frac{2\pi}{h}\right)$.

Наслідок. Якщо h – максимальний крок, то для довільного $\varepsilon > 0$ і $\varepsilon < \frac{\pi}{h}$ існує $c_0 > 0$ таке, що для $t \in \left[\varepsilon, \frac{2\pi}{h} - \varepsilon\right]$ $\left|f\left(t\right)\right| \le e^{-c_0}$.

Доведення випливає із того, що $\left|f(t)\right|$ неперервна на відрізку $\left[\varepsilon, \frac{2\pi}{h}-\varepsilon\right]$, тому досягає на цьому відрізку свого найбільшого значення. Позначимо його c_1. Припустимо, що $c_1 = 1$, тоді знайдеться точка $t_0 \in \left[\varepsilon, \frac{2\pi}{h}-\varepsilon\right]$, що $\left|f(t_0)\right| = 1$, а за теоремою 2 це протирічить тому, що крок h – максимальний.

На основі наведених результатів можна довести важливу локальну граничну теорему для гратчастих розподілів, яка є узагальненням локальної граничної теореми Муавра-Лапласа.

Розглянемо послідовність незалежних однаково розподілених випадкових величин $\xi_1, ..., \xi_n, ...$, що мають гратчастий розподіл, тобто можливі значення ξ_i можна представити у вигляді $a + kh$. Позначимо через $S_n = \xi_1 + \cdots + \xi_n$. Тоді можливі значення S_n можна подати у вигляді $na + kh$. Позначимо через $P_n(k) = P\{S_n = na + kh\}$. Будемо вважати, що випадкові величини мають скінченні математичні сподівання $M\xi_i = m$, скінченні дисперсії $D\xi_i = \sigma^2$. Позначимо $B_n = \sqrt{n}\sigma,\ B_n^2 = DS_n,\ z_{nk} = \frac{na + kh - nm}{B_n}$.

Теорема 3 (Локальна гранична теорема для гратчастих розподілів). Нехай $\xi_1, ..., \xi_n, ...$ – послідовність незалежних однаково розподілених випадкових величин, що мають гратчастий розподіл, скінченні математичні сподівання і дисперсії. Для того, щоб рівномірно по k $(-\infty < k < +\infty)$ $\lim\limits_{n\to\infty}\left(\frac{B_n}{h}P_n(k) - \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z_{nk}^2}{2}\right)\right) = 0$, необхідно і досить, щоб крок h був максимальним.

Вправа.

6. Показати, що розподіли Пуассона і біномний є гратчастими.

7.3. Поняття про граничні закони відмінні від нормального. Нескінченно подільні закони, стійкі закони

У теорії граничних теорем виникає питання про те, які закони, крім нормального, можуть бути граничними для сум незалежних випадкових величин. Нормальним законом не вичерпується клас граничних законів. Із закону великих чисел випливає, що вироджений розподіл $E(x-a) = \begin{cases} 0, & x \le a, \\ 1, & x > a \end{cases}$ є граничним для середнього арифметичного незалежних однаково розподілених випадкових величин із скінченним математичним сподіванням. В схемі Бернуллі при виконанні певних умов граничним виступає розподіл Пуассона.

Нехай задано послідовність серій випадкових величин

$$\xi_{n1}, \xi_{n2}, ..., \xi_{nm_n}, \quad n = 1, 2, ...$$

Будемо вважати, що величини, які входять в одну серію, незалежні. Покладемо

$$S_n = \xi_{n1} + \xi_{n2} + ... + \xi_{nm_n}.$$

Виявляється, що при певних обмеженнях на випадкові величини клас граничних законів для сум S_n співпадає з класом нескінченно подільних законів. Якщо не накладати на випадкові величини ніяких обмежень, то граничним може бути будь-який розподіл. Дійсно, якщо ξ – довільна випадкова величина з функцією розподілу $F(x)$ і $\xi_{n1} = \xi$, а $\xi_{nk} = 0$ $(2 \le k \le m_n)$, то $S_n = \xi$ і розподіл S_n, а отже і граничний, співпадає з $F(x)$.

Функція розподілу $F(x)$ називається *нескінченно подільною*, якщо для довільного $n \ge 1$ її можна подати у вигляді n-кратної згортки деякої функції розподілу $F_n(x)$

$$F(x) = F_n^{*n}(x).$$

У термінах характеристичних функцій це означає, що відповідна характеристична функція $f(t)$ (яку також називають нескінченно подільною) для довільного $n \ge 1$ є n-им степенем деякої іншої характеристичної функції $f_n(t)$:

$$f(t) = \left(f_n(t)\right)^n.$$

Відповідна випадкова величина ξ також називається нескінченно подільною. В цьому випадку існують незалежні однаково розподілені випадкові величини $\xi_1, ..., \xi_n$ такі, що $\xi \overset{d}{=} \xi_1 + \cdots + \xi_n$ (тобто випадкова величина ξ і сума випадкових величин мають однакові розподіли).

Теорема 1. Характеристична функція нескінченно подільного розподілу ніде не перетворюється в 0, тобто, $f(t) \ne 0, \quad \forall t \in R$.

Доведення. Доведемо спочатку одну нерівність для характеристичних функцій. Нехай $f(t)$ - дійсна характеристична функція, а $F(x)$ - відповідна функція розподілу. Тоді із нерівності $1 - \cos 2tx = 2\sin^2 tx = 2\left(1 - \cos^2 tx\right) = 2\left(1 - \cos tx\right)\left(1 + \cos tx\right) \le 4\left(1 - \cos tx\right)$, одержуємо

$$1 - f(2t) = \int_{-\infty}^{+\infty} (1 - \cos 2tx) dF(x) \le 4 \int_{-\infty}^{+\infty} (1 - \cos tx) dF(x) = 4\left(1 - f(t)\right). \quad (18)$$

Із властивостей характеристичних функцій випливає, що для довільної характеристичної функції $f(t)$ справедлива рівність $f(-t) = \overline{f(t)}$, тому і $\left|f(t)\right|^2$ є характеристичною функцією, крім того, $\left|f(t)\right|^2$ – дійсна.

Нехай характеристична функція $f(t)$ є нескінченно подільною. Це означає, що для довільного $n \ge 1$ $f(t) = \left(f_n(t)\right)^n$, де $f_n(t)$ – деяка характеристична функція. Оскільки $f(t)$ неперервна функція і $f(0) = 1$, то існує *a*, що при $|t| \le a$ $f(t) \ne 0$, а отже і функція $f_n(t) \ne 0$. При досить великих *n* і при довільному $|t| \le a$ величину $f_n(t) = \sqrt[n]{f(t)}$ можемо зробити як завгодно близькою до одиниці. Із того, що $\left|f_n(t)\right|^2$ є дійсною характеристичною функцією із (18) випливає, що виконується нерівність

$$1 - \left|f_n(2t)\right|^2 \le 4\left(1 - \left|f_n(t)\right|^2\right),$$

із якої одержуємо

$$1 - \left|f_n(2t)\right| \le 1 - \left|f_n(2t)\right|^2 \le 4\left(1 - \left|f_n(t)\right|^2\right) \le 8\left(1 - \left|f_n(t)\right|\right). \quad (19)$$

Із (19) випливає, що при великих *n* і $|t| \le a$ $1 - \left|f_n(t)\right| < \varepsilon$, то в тому ж проміжку $1 - \left|f_n(2t)\right| < 8\varepsilon$. Звідки випливає, що $f_n(t) \ne 0$, а отже і $f(t) \ne 0$, при досить великих *n* і $|t| \le 2a$. Аналогічно одержуємо, що $f(t) \ne 0$ при досить великих *n* і $|t| \le 4a$ і так далі. Що і доводить теорему.

Теорема 2. Сума незалежних нескінченно подільних випадкових величин є величина нескінченно подільна.

Доведення. Нехай випадкові величини ξ_1, ξ_2 – нескінченно подільні і мають характеристичні функції $f_1(t), f_2(t)$ відповідно. Тоді для довільного $n \geq 1$ існують характерастичні функції $\varphi_{n1}(t)$, $\varphi_{n2}(t)$, що $f_1(t) = (\varphi_{n1}(t))^n$ і $f_2(t) = (\varphi_{n2}(t))^n$. Тому характеристична функція суми $\xi = \xi_1 + \xi_2$ зображається у вигляді $f(t) = f_1(t) f_2(t) = (\varphi_{n1}(t) \cdot \varphi_{n2}(t))^n$, тобто є нескінченно подільною.

Теорема 3. Якщо $F_k(x)$ при кожному k є нескінченно подільною і $F_k(x) \xrightarrow{w} F(x)$, то функція розподілу $F(x)$ є нескінченно подільною.

Доведення. Нехай $f_k(t)$ – характеристична функція розподілу $F_k(x)$, а $f(t)$ – характеристична функція розподілу $F(x)$, тоді рівномірно в кожному скінченному проміжку зміни t

$$\lim_{k \to \infty} f_k(t) = f(t). \tag{20}$$

За умовою $f_k(t)$ – нескінченно подільна характеристична функція, тому для довільного натурального n $f_k(t) = (f_{kn}(t))^n$, де $f_{kn}(t)$ – характеристична функція. Із рівності $f_{kn}(t) = \sqrt[n]{f_k(t)}$ і (20) випливає, що при кожному натуральному n існує границя $\lim_{k \to \infty} f_{kn}(t) = \sqrt[n]{f(t)} = \varphi_n(t)$. Із неперервності $f(t)$ випливає неперервність $\varphi_n(t)$, тому $\varphi_n(t)$ є характеристичною функцією при кожному натуральному n і $f(t) = (\varphi_n(t))^n$, а це означає, що $f(t)$ є нескінченно подільною.

Важливість класу нескінченно подільних законів випливає із наступної теореми, яку ми наведемо без доведення.

Теорема 4. Нехай $\xi_{n1}, ..., \xi_{nn}$ в кожній серії незалежні однаково розподілені випадкові величини, позначимо $S_n = \xi_{n1} + \cdots + \xi_{nn}$, $F_n(x)$ – функція розподілу S_n. Для того, щоб $F_n(x) \xrightarrow{w} F(x)$ при $n \to \infty$, необхідно і досить, щоб функція розподілу $F(x)$ була нескінченно подільною.

Справедливе і більш загальне твердження. Нехай для послідовності серій незалежних в кожній серії випадкових величин $\xi_{n1}, \xi_{n2}, ..., \xi_{nm_n}$, $n = 1, 2, ...$ виконується умова: для довільного $\varepsilon > 0$ $\lim_{n \to \infty} \sup_{1 \leq k \leq m_n} P\{|\xi_{nk}| > \varepsilon\} = 0$. Тоді граничні розподіли для $S_n = \xi_{n1} + \xi_{n2} + ... + \xi_{nm_n}$ при $n \to \infty$ можуть бути тільки нескінченно подільними.

Якщо $f(t)$ – нескінченно подільна характеристична функція, то її логарифм може бути єдиним способом поданий у вигляді

$$\ln f(t) = ita + \int_{-\infty}^{+\infty} \left(e^{itx} - 1 - \frac{itx}{1 + x^2} \right) \frac{1 + x^2}{x^2} dG(x),$$

де $a \in R$ і $G(x)$ – неспадна функція обмеженої варіації.

Функція розподілу $F(x)$ називається стійкою, якщо для довільних $a_1 > 0$, b_1, $a_2 > 0$, b_2 існують $a > 0$ і b такі, що $F(a_1 x + b_1) * F(a_2 x + b_2) = F(ax + b)$, де $*$ – позначає згортку. Характеристична функція $f(t)$ називається стійкою, якщо для довільних $a_1 > 0$, $a_2 > 0$, існують $a > 0$ і b такі, що $f(a_1 t) \cdot f(a_2 t) = e^{itb} f(at)$.

Очевидно, що стійкий розподіл є нескінченно подільним.

Нехай $\xi_1, \xi_2, \ldots$ – послідовність незалежних, однаково розподілених випадкових величин і $S_n = \xi_1 + \ldots + \xi_n$. Якщо існують a_n і $b_n > 0$ такі, що $P\left\{ \frac{S_n - a_n}{b_n} < x \right\} \xrightarrow[n \to \infty]{} F(x)$, то функція розподілу $F(x)$ є стійкою.

Вправи.

7. Узагальнимо поняття закону Пуассона, яке зустрічалося раніше. Випадкова величина ξ має розподіл Пуассона, якщо вона може приймати лише значення $a + kh$, де a і $h > 0$ дійсні сталі, а $k = 0, 1, 2, \ldots$ і $P\{\xi = a + kh\} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad (\lambda > 0)$. Довести, що розподіл Пуассона нескінченно подільний.
8. Довести, що нескінченно подільними є: а) нормальний закон; б) вироджений закон; в) показниковий розподіл.
9. Показати, що для всіх $a > 0$ характеристична функція $f(t) = \frac{1}{(1 + it)^a}$ є нескінченно подільною.

Розділ 8. Елементи теорії випадкових процесів

8.1. Ланцюги Маркова

8.1.1. Означення ланцюга Маркова. Ймовірності переходу. Ланцюги Маркова можна розглядати як важливе узагальнення схеми незалежних випробувань на випадок, коли випробування залежні.

Нехай проводиться послідовність випробувань і в s-му випробуванні можуть настати події $E_1^{(s)},...,E_n^{(s)}$. Ці події можна розглядати як стан деякої системи в момент часу s, при цьому ми будемо вважати, що стан системи в момент часу s залежить лише від стану системи у попередній момент часу $s-1$ і не залежить від інших станів системи, тобто ймовірність того, що в s-му випробуванні настане подія $E_k^{(s)}$ залежить тільки від того, яка із подій $E_i^{(s-1)}$ настає в ($s-1$)-му випробуванні і не залежить від результатів попередніх випробувань, тобто, $P\left\{E_k^{(s)} / E_{n_1}^{(1)},...,E_{n_{S-1}}^{(s-1)}\right\} = P\left\{E_k^{(s)} / E_{n_{S-1}}^{(s-1)}\right\}$. Такі системи називаються *марковськими системами* або *ланцюгами Маркова*. Множину всіх станів системи називають *фазовим простором* ланцюга Маркова.

Ми будемо вважати, що множина станів скінченна або зліченна, а випробування проводяться в дискретні моменти часу. Для того, щоб вказати в якому стані перебуває система, досить вказати лише номер цього стану.

Позначимо через $\xi(t)$ стан системи у момент часу t. При цьому $\xi(t)$ – стан системи в момент часу t – є випадковою величиною. Система знаходиться в стані $E_k^{(t)}$ означає, що $\xi(t)=k$, $k=0,1,....$, $t=0,1,...$. Таким чином, поведінка системи повністю визначається послідовністю випадкових величин $\xi(0),\xi(1),...,\xi(n),...$. Якщо

$$P\left\{\xi(t)=k / \xi(0)=i,...,\xi(t-1)=j\right\} = P\left\{\xi(t)=k / \xi(t-1)=j\right\},$$

тоді $\xi(t)$ називають *ланцюгом Маркова*. Для ланцюга Маркова при відомому стані $\xi(t-1)=j$ на наступному кроці система переходить в стан $\xi(t)=k$ з умовною ймовірністю $P\left\{\xi(t)=k / \xi(t-1)=j\right\}$. Якщо ця ймовірність

$$P\left\{\xi(t)=k / \xi(t-1)=j\right\} = p_{jk}$$

не залежить від t, то ланцюг Маркова називають *однорідним*. При цьому ймовірності p_{jk} називаються *ймовірностями переходу* із стану j в стан k і

$$\sum_k p_{jk} = 1 .$$

Надалі будемо розглядати лише однорідні ланцюги Маркова.

Позначимо $P\left\{\xi(0)=k\right\} = p_k$, $k=0,1,...$, – початковий розподіл ланцюга Маркова. Тоді поведінка системи в момент часу n повністю визначається початковим розподілом та ймовірностями переходу p_{jk}:

$$P\left\{\xi(0)=k_0,\xi(1)=k_1,...,\xi(n-1)=k_{n-1},\xi(n)=k_n\right\} =$$
$$= P\left\{\xi(0)=k_0,\xi(1)=k_1,...,\xi(n-1)=k_{n-1}\right\} p_{k_{n-1}k_n} = p_{k_0} p_{k_0k_1} p_{k_1k_2} \cdot ... \cdot p_{k_{n-1}k_n} .$$

Розглянемо матрицю $P = \left\| p_{jk} \right\|$. Для скінченної кількості станів

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ p_{k1} & p_{2k} & \cdots & p_{kk} \end{pmatrix}.$$

Матрицю P називають матрицею ймовірностей переходу ланцюга Маркова (за один крок).

Для матриці P характерно, що $p_{ij} \geq 0\,(i,j=1,...,k)$ та $\sum_j p_{ij} = 1\,(i=1,...,k)$. Матриці, для яких виконуються такі умови, називаються *стохастичними*. Будь-яка стохастична матриця може бути матрицею переходу і разом з початковим розподілом вона визначає деякий ланцюг Маркова.

Розглянемо приклад блукання по відрізку $[0,k]$, де k - натуральне. Якщо точка знаходиться в положенні i, то з ймовірностями $\frac{1}{2}$ вона може перейти в положення $i-1$ або $i+1$. При $i=0$ з ймовірностями $\frac{1}{2}$ вона може перейти в положення $i=1$ або залишитись на місці, аналогічно, при $i=k$ з ймовірностями $\frac{1}{2}$ вона може перейти в положення $i=k-1$ або залишитись на місці. Тоді матриця переходу буде мати вигляд

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & ... & 0 \\ 1/2 & 0 & 1/2 & ... & 0 \\ 0 & 1/2 & 0 & ... & 0 \\ . & . & . & . & . \\ 0 & ... & 0 & 1/2 & 1/2 \end{pmatrix}.$$

Розглянемо ймовірності $p_{ik}(n) = P\{\xi(n) = k / \xi(0) = i\}$. Такі ймовірності називаються ймовірностями переходу із стану i в стан k за n кроків. Тоді $\sum_j p_{ij}(n) = 1$. Покажемо, що

$$p_{ik}(n) = \sum_l p_{il}(m)\, p_{lk}(n-m), \quad m = 1,...,n-1. \tag{1}$$

Для обґрунтування цієї формули позначимо через A подію, яка полягає у тому, що система із стану i перейде в стан k за n кроків, тоді ця подія може настати із однією із подій H_l – система із стану i перейде в стан l за m кроків $m=1,...,n-1$. Тоді за формулою повної ймовірності $P(A) = \sum_l P(H_l)P(A/H_l)$, де $P(A) = p_{ik}(n)$, $P(H_l) = p_{il}(m)$, $P(A/H_l) = p_{lk}(n-m)$, одержуємо $p_{ik}(n) = \sum_l p_{il}(m)\, p_{lk}(n-m)$.

Позначимо через P_n матрицю ймовірностей переходу за n кроків, тобто $P_n = \left\| p_{ij}(n) \right\|$. Тоді із формули (1) випливає, що для довільного $m=1,...,n-1$ буде справедлива рівність

$$P_n = P_m P_{n-m}.$$

Зокрема, $P_n = P_1 P_{n-1} = P_1 P_1 P_{n-2} = \cdots = P_1^n$, де $P_1 = P$ – матриця ймовірностей переходу за один крок.

Позначимо $P\{\xi(n) = k\} = p_k(n)$ – розподіл ланцюга Маркова в момент часу n. Тоді

$$p_k(n) = \sum_i p_i\, p_{ik}(n),$$

тобто, розподіл ланцюга Маркова в момент часу n визначається початковим розподілом і ймовірностями переходу за n кроків. Дана рівність випливає з формули повної ймовірності.

8.1.2. Класифікація станів однорідного ланцюга Маркова. Кажуть, що стан j *досягається* із стану i, якщо існує $n \geq 1$, що $p_{ij}(n) > 0$ або існують $k_1,...,k_{n-1}$, що $p_{ik_1} > 0$, $p_{k_1k_2} > 0, \ldots,$ $p_{k_{n-1}j} > 0$.

Кажуть, що стан i та стан j *сполучаються*, якщо кожен із них досягається із іншого, тобто існують m і n такі, що $p_{ij}(m) > 0$ і $p_{ji}(n) > 0$.

Розглянемо довільний стан x і позначимо через $K(x)$ множину всіх станів, які сполучаються з x. Тоді множина всіх станів розбивається на класи без спільних елементів. Тобто для двох станів x і y класи $K(x)$ і $K(y)$ або співпадають або не перетинаються. Якщо $y \in K(x)$, то $K(x) = K(y)$ – класи будуть співпадати, якщо ж $y \notin K(x)$, то $K(x) \bigcap K(y) = \varnothing$.

Стан j називається *істотним*, якщо для довільного i, що досягається із j існує n, що $p_{ij}(n) > 0$. У протилежному випадку стан називається *неістотним*. Тобто стан j називається неістотним, якщо існує стан i, що досягається із j, але $p_{ij}(n) = 0$ для довільного n. Таким чином, неістотний стан характеризується тим, що із нього з додатньою ймовірністю можна попасти в деякий інший стан, але назад повернутися уже не можна. Якщо ж система попадає в один із станів певного класу істотних станів, то вона уже не може вийти за межі цього класу.

Розглянемо систему, яка складається із чотирьох станів E_1, E_2, E_3, E_4 із матрицею переходу

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

В даному прикладі стани E_1 і E_2 неістотні, а E_3 і E_4 є істотними станами, що сполучаються.

Якщо ланцюг Маркова складається лише з одного класу істотних станів, то такий ланцюг Маркова називається *незвідним*.

Введемо ймовірність $f_j(n)$ того, що система, яка вийшла із стану j рівно за n кроків вперше повернеться в стан j: $f_j(n) = P\{\xi(0) = j, \xi(1) \neq j, ..., \xi(n-1) \neq j, \xi(n) = j\}$. $f_j(n)$ не співпадає з $p_{jj}(n)$.

Ймовірність того, що система, яка вийшла із стану j, знову коли-небудь повернеться в нього, дорівнює $F_j = \sum_{n=1}^{\infty} f_j(n)$ – ймовірність повернення.

Якщо $F_j = 1$, то стан j називається *рекурентним*, якщо ж $F_j < 1$, то стан j називається *не рекурентним*.

Стан j називається *нульовим*, якщо $\lim\limits_{n\to\infty} p_{jj}(n) = 0$, а в протилежному випадку – *ненульовим*.

Стан j називається *періодичним* з періодом d, якщо повернення в цей стан можливе тільки через число кроків кратне d.

Наведемо деякі властивості.

Якщо стан i неістотний, то для довільного j $\lim\limits_{n\to\infty} p_{ji}(n)=0$.

Стан j рекурентний тоді і тільки тоді, коли $\sum\limits_{n=1}^{\infty} p_{jj}(n)=\infty$.

Для нерекурентнного стану j

$$F_j=\frac{\sum\limits_{n=1}^{\infty} p_{jj}(n)}{1+\sum\limits_{n=1}^{\infty} p_{jj}(n)}.$$

Нерекурентний стан завжди є нульовим.

Для доведення цього за формулою повної ймовірності маємо

$$p_{jj}(n)=f_j(1)\,p_{jj}(n-1)+f_j(2)\,p_{jj}(n-2)+\ldots+f_j(n-1)\,p_{jj}(1)+f_j(n)1.$$

Розглянемо твірні функції послідовностей $p_{jj}(n)$ і $f_j(n)$:

$$P_j(z)=\sum_{n=1}^{\infty} p_{jj}(n)z^n\,,\ F_j(z)=\sum_{n=1}^{\infty} f_j(n)z^n\,.$$

Помножимо рівність для $p_{jj}(n)$ на z^n і додамо одержані рівності по $n=1,2,\ldots$, тоді одержимо

$$P_j(z)=zf_j(1)\big(1+P_j(z)\big)+z^2 f_j(2)\big(1+P_j(z)\big)+\ldots=\big(1+P_j(z)\big)F_j(z).$$

Отже, $F_j(z)=\dfrac{P_j(z)}{1+P_j(z)}$, $P_j(z)=\dfrac{F_j(z)}{1-F_j(z)}$. Якщо $\sum\limits_{n=1}^{\infty} p_{jj}(n)=\infty$, то $P_j(z)\to\infty$ при $z\to 1$, тому $F_j(z)\to 1$. Отже, $F_j=1$. Справедливо і навпаки. Якщо ж $\sum\limits_{n=1}^{\infty} p_{jj}(n)<\infty$, то при $z=1$ одержуємо формулу для F_j.

8.1.3. Ергодична теорема Маркова. Нехай множина станів ланцюга Маркова скінченна.

Теорема. Нехай існує $s\ge 1$, для якого $p_{ij}(s)>0$, $i,j=1,\ldots,k$. Тоді для довільного $j=1,\ldots,k$ існують границі

$$\lim_{n\to\infty} p_{ij}(n)=p_j\,. \tag{2}$$

Граничні ймовірності p_j не залежать від початкового стану i і є єдиним розв'язком системи

$$\sum_{i=1}^{k} p_i p_{ij}=p_j\,,\ j=1,\ldots,k\,,\ \sum_{j=1}^{k} p_j=1\,. \tag{3}$$

Доведення. Із (1) одержимо $p_{ij}(n)=\sum\limits_{r=1}^{k} p_{ir}p_{rj}(n-1)$. Звідки, для всіх $i,j=1,\ldots,k$ і $n>1$ $p_{ij}(n)\ge \min\limits_{1\le r\le k} p_{rj}(n-1)$, бо $\sum\limits_{r=1}^{k} p_{ir}=1$. Тому виконується і нерівність

$$\min_{1\le i\le k} p_{ij}(n)\ge \min_{1\le r\le k} p_{rj}(n-1).$$

Аналогічно, для для всіх $i,j=1,\ldots,k$ і $n>1$ $p_{ij}(n)\le \max\limits_{1\le r\le k} p_{rj}(n-1)$,

$$\max_{1\le i\le k} p_{ij}(n) \le \max_{1\le r\le k} p_{rj}(n-1).$$

Отже, для всіх $i, j = 1,...,k$ і $n > 1$

$$\min_{1\le r\le k} p_{rj}(n-1) \le \min_{1\le i\le k} p_{ij}(n) \le p_{ij}(n) \le \max_{1\le i\le k} p_{ij}(n) \le \max_{1\le r\le k} p_{rj}(n-1). \tag{4}$$

Таким чином, при зростанні n відрізки $\left[\min_{1\le i\le k} p_{ij}(n), \max_{1\le i\le k} p_{ij}(n)\right]$ є вкладеними. Якщо ми доведемо, що при $n \to \infty$ довжина відрізка

$$\max_{1\le i\le k} p_{ij}(n) - \min_{1\le i\le k} p_{ij}(n) \to 0, \tag{5}$$

то із (4) випливатиме , що $p_{ij}(n)$ має границю і ця границя не залежить від i.

Будемо вважати, що $n > s$. Тоді із (1) при $m = s$

$$p_{ij}(n) - p_{lj}(n) = \sum_{r=1}^{k} p_{ir}(s) p_{rj}(n-s) - \sum_{r=1}^{k} p_{lr}(s) p_{rj}(n-s) = \sum_{r=1}^{k} \left(p_{ir}(s) - p_{lr}(s)\right) p_{rj}(n-s). \tag{6}$$

Позначимо через β_{irl} невід'ємні різниці $p_{ir}(s) - p_{lr}(s)$ і через $-\beta'_{irl}$ – від'ємні різниці $p_{ir}(s) - p_{lr}(s)$ у (6). У будь-якому випадку β_{irl} і β'_{irl} невід'ємні. Враховуючи, що $\sum_{r=1}^{k} p_{ir}(s) = 1$, $\sum_{r=1}^{k} p_{lr}(s) = 1$, одержимо: $\sum_{r=1}^{k}\left(p_{ir}(s) - p_{lr}(s)\right) = 0$, а це означає, що $\sum_r{}^{+}\beta_{irl} = \sum_r{}^{-}\beta'_{irl} = h_{il}$. Якщо для всіх $r = 1,...,k$ $p_{ir}(s) - p_{lr}(s) = 0$, то із (6) випливає, що для всіх $i, j = 1,...,k$ і $n > 1$ $p_{ij}(n) - p_{lj}(n) = 0$, а це означає, що $p_{ij}(n)$ не залежить від i і n, тому існують границі (2). Якщо ж принаймні при одному r $p_{ir}(s) - p_{lr}(s) \ne 0$, то із умови $p_{ij}(s) > 0$, $i, j = 1,...,k$, одержимо, що $h_{il} = \sum_r{}^{+}\left(p_{ir}(s) - p_{lr}(s)\right) < \sum_r{}^{+} p_{ir}(s) < 1$, а тому і $h = \max_{i,l} h_{il} < 1$. Тоді із рівності (6) одержуємо

$$\begin{aligned} p_{ij}(n) - p_{lj}(n) &= \sum_r{}^{+}\left(p_{ir}(s) - p_{lr}(s)\right) p_{rj}(n-s) + \sum_r{}^{-}\left(p_{ir}(s) - p_{lr}(s)\right) p_{rj}(n-s) = \\ &= \sum_r{}^{+}\beta_{irl} p_{rj}(n-s) - \sum_r{}^{-}\beta'_{irl} p_{rj}(n-s) \le \sum_r{}^{+}\beta_{irl} \max_{1\le r\le k} p_{rj}(n-s) - \sum_r{}^{-}\beta'_{irl} \min_{1\le r\le k} p_{rj}(n-s) = \\ &= h_{il}\left(\max_{1\le r\le k} p_{rj}(n-s) - \min_{1\le r\le k} p_{rj}(n-s)\right) \le h\left(\max_{1\le r\le k} p_{rj}(n-s) - \min_{1\le r\le k} p_{rj}(n-s)\right). \end{aligned}$$

Тому і $\max_{i,l}\left|p_{ij}(n) - p_{lj}(n)\right| \le h\left(\max_{1\le r\le k} p_{rj}(n-s) - \min_{1\le r\le k} p_{rj}(n-s)\right)$. Якщо ми врахуємо, що $\max_{i,l}\left|p_{ij}(n) - p_{lj}(n)\right| = \max_{1\le i\le k} p_{ij}(n) - \min_{1\le l\le k} p_{lj}(n)$, то одержимо:

$$\max_{1\le i\le k} p_{ij}(n) - \min_{1\le i\le k} p_{ij}(n) \le h\left(\max_{1\le i\le k} p_{ij}(n-s) - \min_{1\le i\le k} p_{ij}(n-s)\right). \tag{7}$$

До другого множника у правій частині можна знову застосувати цю нерівність. Застосувавши нерівність (7) $n_s = \left[\frac{n}{s}\right]$ раз (n_s – ціла частина $\frac{n}{s}$), одержимо:

$$\max_{1\le i\le k} p_{ij}(n) - \min_{1\le i\le k} p_{ij}(n) \le h^{n_s}\left(\max_{1\le i\le k} p_{ij}(n - n_s s) - \min_{1\le i\le k} p_{ij}(n - n_s s)\right) \le h^{n_s} \xrightarrow[n\to\infty]{} 0,$$

бо $n_s \xrightarrow[n\to\infty]{} \infty$, $0 < h < 1$. Отже, (5) виконується, тому виконується і (2).

Враховуючи рівності $p_{ij}(n+1) = \sum_{r=1}^{k} p_{ir}(n) p_{rj}$, $\sum_{j=1}^{k} p_{ij}(n) = 1$, і переходячи в них до границі при $n \to \infty$, одержимо (3). Теорема доведена.

Приклад. Задано матрицю ймовірностей переходу за один крок $P=\begin{pmatrix} 0 & 0,5 & 0,5 \\ 0,7 & 0,15 & 0,15 \\ 0,5 & 0 & 0,5 \end{pmatrix}$.
Знайти матрицю переходу за два кроки. Знайти граничні ймовірності $p_j=\lim\limits_{n\to\infty} p_{ij}(n)$.

Матриця переходу за два кроки є добуток матриць переходу за один крок:

$$P_2=P\cdot P=\begin{pmatrix} 0 & 0,5 & 0,5 \\ 0,7 & 0,15 & 0,15 \\ 0,5 & 0 & 0,5 \end{pmatrix}\begin{pmatrix} 0 & 0,5 & 0,5 \\ 0,7 & 0,15 & 0,15 \\ 0,5 & 0 & 0,5 \end{pmatrix}=\begin{pmatrix} 0,6 & 0,075 & 0,325 \\ 0,18 & 0,3725 & 0,4475 \\ 0,25 & 0,25 & 0,5 \end{pmatrix}.$$

Всі елементи матриці P_2 додатні, тому за теоремою існують граничні ймовірності. Для їх знаходження складемо систему (3):

$0,7p_2+0,5p_3=p_1$, $0,5p_1+0,15p_2=p_2$, $0,5p_1+0,15p_2+0,5p_3=p_3$, $p_1+p_2+p_3=1$.

Із другого рівняння $p_1=1,7p_2$, а із третього $p_3=2p_2$. Підставляючи їх у четверте рівняння, одержимо $p_2=\frac{1}{4,7}$. Тому $p_1=\frac{17}{47}$, $p_3=\frac{20}{47}$. Отже, граничні ймовірності рівні $p_1=\frac{17}{47}$, $p_2=\frac{10}{47}$, $p_3=\frac{20}{47}$. Відзначимо, що ми зовсім не використовували перше рівняння. В загальному випадку одне із рівнянь системи (3) є наслідком решти.

Вправи.

1. В урні міститься 5 куль, білі і чорні. Випробування полягає у тому, що із урни випадково вибирається одна куля і повертається в урну куля іншого кольору (замість білої – чорна і навпаки). Знайти матрицю ймовірностей переходу ланцюга Маркова за один крок, якщо його станами є кількість білих куль в урні. Знайти ймовірності переходу за два кроки.
2. Задано матрицю переходу системи із стану i в стан j ($i,j=1,2$) за один крок $P=\begin{pmatrix} 0,4 & 0,6 \\ 0,1 & 0,9 \end{pmatrix}$. Знайти матрицю переходу за два кроки. Знайти граничні ймовірності $p_j=\lim\limits_{n\to\infty} p_{ij}(n)$, $j=1,2$.

8.2. Випадкові процеси. Основні поняття. Деякі класи процесів

В багатьох реальних системах виникає потреба враховувати вплив випадкових факторів, що є функціями часу. Випадкові процеси (випадкові функції) – це випадкові величини, що залежать від параметрів (дискретних або неперервних). До випадкових процесів може бути віднесена велика кількість виробничих процесів, а також процесів, що зустрічаються в економіці, фінансах, страхуванні.

Нехай задано ймовірнісний простір $\{\Omega,\mathfrak{A},P\}$ і деяка множина T. Функція $\xi(t,\omega)$, де $\omega\in\Omega$, $t\in T$, називається *випадковою функцією* або *випадковим процесом*, якщо при кожному фіксованому $t\in T$ $\xi(t,\omega)$ є випадковою величиною, що визначена на ймовірнісному просторі $\{\Omega,\mathfrak{A},P\}$.

Якщо (X,σ_X) – вимірний простір і випадковий процес приймає значення із *X*, тобто, $\xi(t,\omega):T\times\Omega\to X$, то (X,σ_X) називають фазовим простором. Найбільш важливим є випадок X=R, тоді процес $\xi(t)=\xi(t,\omega)$ називають числовим або дійснозначним (одновимірним). Якщо X= R^m, то процес $\xi(t)$ називають векторним (багатовимірним).

Що стосується параметра t, то у випадку $T\subset R$, t частіше розглядається як час, а $\xi(t)$ називають *випадковим процесом*, якщо ж t пробігає лише цілочислові значення, то *випадковою послідовністю* або *часовим рядом*. Якщо значеннями t є точки деякого векторного простору, то $\xi(t)$ називають *випадковим полем*.

Якщо ми зафіксуємо $t\in T$, то $\xi(t,\omega)$ називається *перерізом* випадкового процесу в точці t. При фіксованому $\omega\in\Omega$ $\xi(t,\omega)$, що розглядається як функція часу, називається *реалізацією* випадкового процесу або *вибірковою функцією* (*траєкторією*, якщо t розглядається як час).

Основними характеристиками випадкового процесу є його скінченновимірні розподіли. Нехай $X=R$. Функція $F_t(x)=P\{\xi(t,\omega)<x\}$ є функцією розподілу випадкового процесу, але вона не дає повної ймовірнісної характеристики випадкового процесу, бо не враховує залежності між випадковими величинами $\xi(t)$ при різних t. У зв'язку із цим для випадкового процесу вводяться скінченновимірні функції розподілу випадкового процесу.

Для будь-якого натурального n і будь-яких $t_1,...,t_n$, де $t_i\in T$ $i=1,...,n$, функція $F_{t_1,...,t_n}(x_1,...,x_n)=P\{\xi(t_1)<x_1,...,\xi(t_n)<x_n\}$ називається *скінченновимірною функцією розподілу* випадкового процесу. На скінченновимірні функції розподілу накладають наступні умови узгодженості:

$$F_{t_1,...,t_n,t_{n+1},...,t_{n+m}}(x_1,...,x_n,+\infty,...,+\infty)=F_{t_1,...,t_n}(x_1,...,x_n),$$

$$F_{t_{i_1},...,t_{i_n}}(x_{i_1},...,x_{i_n})=F_{t_1,...,t_n}(x_1,...,x_n),$$

де $i_1,...,i_n$ – довільна перестановка індексів $1,...,n$.

Відзначимо, що за відомою теоремою Колмогорова для кожного набору скінченновимірних функцій розподілу, для яких виконуються умови узгодженості, існує випадкова функція із заданою множиною скінченновимірних функцій розподілу.

Для характеристики випадкового процесу вживаються також моментні функції. Величина $m_n(t_1,...,t_n)=M(\xi(t_1)\cdot...\cdot\xi(t_n))$ називається $n-$ною моментною функцією випадкового процесу. Найбільш важливими моментними функціями є середнє значення $m_1(t)=M\xi(t)$ та кореляційна функція випадкового процесу

$$R(t_1,t_2)=M\big((\xi(t_1)-m_1(t_1))(\xi(t_2)-m_1(t_2))\big)=m_2(t_1,t_2)-m_1(t_1)m_1(t_2).$$

Існує великий клас випадкових процесів, у якому використовуються лише ці дві функції.

Для випадкового процесу вводять сукупні характеристичні функції

$$f_{t_1,...,t_n}(z_1,...,z_n)=M\exp\big(i\big(z_1\xi(t_1)+z_2\xi(t_2)+\cdots+z_n\xi(t_n)\big)\big)=$$

$$=\int_{-\infty}^{+\infty}(n)\int_{-\infty}^{+\infty}\exp\big(i(z_1x_1+z_2x_2+\cdots+z_nx_n)\big)dF_{t_1,...,t_n}(x_1,...,x_n).$$

Розглянемо деякі типи випадкових процесів, всі скінченновимірні розподіли яких визначаються одновимірними розподілами значень $\xi(t)$ і двовимірними розподілами пар $(\xi(t_1),\xi(t_2))$.

1) Якщо для довільних $t_1 < t_2 < ... < t_n$ величини $\xi(t_1),....,\xi(t_n)$ – незалежні, то випадковий процес $\xi(t)$ називається *процесом із незалежними значеннями*.

2) Якщо для будь-яких $t_1 < t_2 < ... < t_n$ величини $\xi(t_1), \xi(t_2) - \xi(t_1),...., \xi(t_n) - \xi(t_{n-1})$ – незалежні, то випадковий процес називається *процесом із незалежними приростами*.

Покажемо, що процес із незалежними приростами повністю характеризується двовимірними скінченними розподілами. Для довільних $t_0 < t_1 < t_2 < ... < t_n$ розглянемо сумісну характеристичну функцію вектора $\xi(t_0), \xi(t_1),..., \xi(t_n)$

$$f_{t_0,t_1,...,t_n}(z_0, z_1,..., z_n) = M \exp\Big(i\big(z_0\xi(t_0) + z_1\xi(t_1) + \cdots + z_n\xi(t_n)\big)\Big).$$

Оскільки

$$\begin{aligned}
&z_0\xi(t_0) + z_1\xi(t_1) + \cdots + z_{n-1}\xi(t_{n-1}) + z_n\xi(t_n) = \\
&= z_n\big(\xi(t_n) - \xi(t_{n-1})\big) + (z_n + z_{n-1})\big(\xi(t_{n-1}) - \xi(t_{n-2})\big) + \cdots + \\
&+ (z_n + z_{n-1} + \cdots + z_1)\big(\xi(t_1) - \xi(t_0)\big) + (z_n + z_{n-1} + \cdots + z_0)\xi(t_0),
\end{aligned}$$

то

$$\begin{aligned}
f_{t_0,...,t_n}(z_0,..., z_n) = M \exp\Big(&i(z_n + z_{n-1} + \cdots + z_0)\xi(t_0) + i(z_n + z_{n-1} + \cdots + z_1)\big(\xi(t_1) - \xi(t_0)\big) + \\
&+ ... + i(z_n + z_{n-1})\big(\xi(t_{n-1}) - \xi(t_{n-2})\big) + iz_n\big(\xi(t_n) - \xi(t_{n-1})\big)\Big).
\end{aligned}$$

Ввівши позначення $\varphi_t(z) = M \exp(iz\xi(t))$ і $\psi_{t_1,t_2}(z_1, z_2) = M \exp\Big(iz\big(\xi(t_2) - \xi(t_1)\big)\Big)$, одержимо

$$f_{t_0,...,t_n}(z_0,..., z_n) = \varphi_{t_0}(z_0 + z_1 + \cdots + z_n) \cdot \psi_{t_0,t_1}(z_1 + \cdots + z_n) \cdot ... \cdot \psi_{t_{n-1},t_n}(z_n).$$

Тобто, сумісна характеристична функція виражається через функції φ і ψ, а це означає, що для визначення скінченновимірних розподілів процесу із незалежними приростами досить знати одновимірні розподіли процесу і розподіли приростів процесу, що визначаються двовимірними розподілами.

3) Процес $\xi(t)$ називається *гауссівським*, якщо всі багатовимірні розподіли випадкових векторів $\big(\xi(t_1),....,\xi(t_n)\big)$ є нормальними. Так як нормальний розподіл однозначно визначається своїми першим і другим моментами, то для задання гауссівського процесу $\xi(t)$ достатньо вказати значення функцій $M\xi(t) = m(t)$ і $M\xi(t)\xi(s) = B(t,s)$, де $B(t,s)$ має бути невід'ємно визначеним ядром.

4) Важливим класом процесів є марковські процеси. Розглянемо моменти часу $u < t < s$. Якщо поведінка процесу в момент часу s не залежить від u, а залежить лише від значень в момент часу t, то такий процес називаться *марковським*.

5) Випадковий процес $\xi(t)$ називається *стаціонарним*, якщо для будь-яких $t_1 < t_2 < ... < t_n$ і h розподіл випадкового вектора $\big(\xi(t_1 + h),....,\xi(t_n + h)\big)$ не залежить від h. Для багатьох задач, що пов'язані із такими процесами, досить знати лише значення перших двох моментів $M\xi(t) = m$ і $M\xi(t)\xi(t + s) = B(s)$, тобто моменти $M\xi(t)$ і $M\xi(t)\xi(t + s)$ не залежать від t. Такі процеси називають *стаціонарними в широкому розумінні*.

Вправи.

3. Нехай η – випадкова величина з функцією розподілу $F(x)$. Знайти всі скінченновимірні розподіли випадкового процесу $\xi(t)=\eta+t$, $t\in R$.
4. Нехай ξ і η – незалежні випадкові величини, що мають однаковий нормальний розподіл з $a=0$ і $\sigma^2=\frac{1}{2}$. Знайти всі скінченновимірні розподіли випадкового процесу $\xi(t)=\frac{\xi+\eta}{t}$, $t>0$.

8.3. Пуассонівський процес

Розглянемо одну досить загальну схему, за допомогою якої можна описувати велику кількість експериментів, в яких спостерігаються події, що можуть відбуватися чи не відбуватися в кожен момент часу.

Нехай подія A може наставати у випадкові моменти часу і нехай $\xi(t)$ – число появ події за проміжок часу t, при цьому виконуються наступні умови:

1) Число появ події за проміжки часу, що не перетинаються є незалежними випадковими величинами: при $0=t_0<t_1<t_2<...<t_n$ $\xi(0)=0,\xi(t_1)-\xi(t_0),...,\xi(t_n)-\xi(t_{n-1})$ незалежні. Тобто $\xi(t)$ є процесом із незалежними приростами.

2) Для будь-якого проміжку часу ймовірність настання події залежить тільки від довжини проміжку і не залежить від його положення: розподіл $\xi(t+h)-\xi(t)$ не залежить від t (процес є однорідним).

3) При $t\to 0$ $P\{\xi(t)=1\}=\lambda t+o(t)$, $\lambda>0$, $P\{\xi(t)>1\}=o(t)$.

Тоді процес, що задовольняє таким умовам, називається *процесом Пуассона*.

Введемо ймовірність $P_k(t)=P\{\xi(t)=k\}$. Очевидно $P_0(0)=1,$ $P_k(0)=0$ при $k\ge 1$.

Якщо виконуються умови 1) – 3), *то для будь-якого* $t\ge 0$

$$P_k(t)=e^{-\lambda t}\frac{(\lambda t)^k}{k!},\ k=0,1,2,....$$

Знайдемо спочатку $P_0(t)$. Тоді $P_0(t+\Delta t)=P_0(t)P_0(\Delta t)=P_0(t)\left(1-P\{\xi(\Delta t)\ge 1\}\right)=$ $=P_0(t)\left(1-\lambda\Delta t-o(\Delta t)\right)$ і $P_0(t+\Delta t)-P_0(t)=-\lambda\Delta tP_0(t)-o(\Delta t)$. Звідси випливає неперервність $P_0(t)$ і існування $P_0'(t)$. Тому після ділення останньої рівності на Δt і переходу до границі, коли $\Delta t\to 0$, одержуємо диференціальне рівняння $P_0'(t)=-\lambda P_0(t)$, розв'язок якого задовольняє умові $P_0(0)=1$. Тому

$$P_0(t)=e^{-\lambda t}.$$

При $n\ge 0$ $P_{n+1}(t+\Delta t)=P_{n+1}(t)P_0(\Delta t)+P_n(t)P_1(\Delta t)+P_{n-1}(t)P_2(\Delta t)+...+P_0(t)P_{n+1}(\Delta t)$.
Із умови 3) випливає, що $P_0(\Delta t)=1-\lambda\Delta t-o(\Delta t)$, $P_1(\Delta t)=\lambda\Delta t+o(\Delta t)$, $P_k(\Delta t)=o(\Delta t)$ при $k>1$. Тому $P_{n+1}(t+\Delta t)-P_{n+1}(t)=P_{n+1}(t)(-\lambda\Delta t)+P_n(t)(\lambda\Delta t)+o(\Delta t)$. Після ділення на Δt і переходу до границі, коли $\Delta t\to 0$, одержуємо диференціальне рівняння

$$P_{n+1}'(t)=-\lambda P_{n+1}(t)+\lambda P_n(t),$$

розв'язок якого задовольняє умові $P_k(0)=0$, $k\ge 1$. Виконавши заміну $P_n(t)=y_n(t)e^{-\lambda t}$, для $y_n(t)$ одержуємо диференціальне рівняння $y'_{n+1}(t)=\lambda y_n(t)$, розв'язок якого задовольняє умові $y_n(0)=0$, $n\ge 1$. Оскільки $y_0(t)=1$, то за індукцією для $n\ge 0$ одержуємо $y_n(t)=\frac{(\lambda t)^n}{n!}$. Отже, $P_n(t)=\frac{(\lambda t)^n}{n!}e^{-\lambda t}$.

Введемо випадкову величину τ - момент першого настання події. Нехай $\xi(0)=0$. Тоді $\{\tau>t\}=\{\xi(t)=0\}$ і $P\{\tau>t\}=P_0(t)=e^{-\lambda t}$. Отже, момент першого настання події має показниковий розподіл: $F_\tau(t)=1-e^{-\lambda t}$ при $t\ge 0$ і $F_\tau(t)=0$ при $t<0$.

Нехай $\nu(t)$ – процес Пуассона, для якого $\nu(0)=0$, і нехай $\xi_1,...,\xi_n,...$ – незалежні, однаково розподілені випадкові величини, які для будь-яких i і t не залежать від $\nu(t)$.

Розглянемо новий процес

$$\xi(t)=\sum_{k=1}^{\nu(t)}\xi_k$$

($\sum_{k=1}^{0}=0$). Такий процес називається *узагальненим процесом Пуассона*.

Цей процес кусково-сталий, його стрибки відбуваються в тих точках, де відбуваються стрибки $\nu(t)$, величина k-го стрибка дорівнює ξ_k. Розглянемо послідовність випадкових величин $\tau_1,\tau_2,...,\tau_k,...$, для якої $\nu(t)=0$ при $t<\tau_1$, $\nu(t)=k$ при $\sum_{i=1}^{k}\tau_i\le t<\sum_{i=1}^{k+1}\tau_i$ (τ_1 – момент 1-го настання події, τ_2 – час від моменту 1-го настання події до моменту 2-го настання події і т. д.). Тоді

$$\xi(t)=\begin{cases}0, & t<\tau_1;\\ \xi_1, & \tau_1\le t<\tau_1+\tau_2;\\ \ldots\ldots & \\ \xi_1+\cdots+\xi_k, & \sum_{i=1}^{k}\tau_i\le t<\sum_{i=1}^{k+1}\tau_i.\end{cases}$$

Покажемо, що *узагальнений процес Пуассона є однорідним процесом із незалежними приростами*.

Спочатку доведемо, що для будь-яких $t_1<t_2<...<t_n$ випадкові величини $\xi(t_1)$, $\xi(t_2)-\xi(t_1),\ldots$, $\xi(t_n)-\xi(t_{n-1})$ є незалежними. Для довільних $l\ge 0$, $m\ge 0$ і дійсних $x_1,\ldots,x_m$, x події $\{\xi_1<x_1,...,\xi_m<x_m,\nu(t_{n-1})=m\}$ і $\{\xi_{m+1}+...+\xi_{m+l}<x,\nu(t_n)-\nu(t_{n-1})=l\}$ є незалежними, бо випадкові величини $\xi_1,...,\xi_m$, $\nu(t_{n-1})$, $\xi_{m+1},...,\xi_{m+l}$, $\nu(t_n)-\nu(t_{n-1})$ незалежні. Це означає, що величини $\xi(t_{n-1})$ і $\xi(t_n)-\xi(t_{n-1})$ є незалежними. Звідси випливає, що $\xi(t)$ є процесом із незалежними приростами.

Процес $\xi(t)$ – однорідний, якщо $P\{\xi(t+s)-\xi(t)<x\}$ не залежать від t. За формулою повної ймовірності

$$P\left\{\xi(t+s)-\xi(t)<x\right\}=P\left\{\sum_{k=\nu(t)+1}^{\nu(t+s)}\xi_k<x\right\}=$$

$$=\sum_{\substack{l\ge 0\\ m\ge 0}}P\left\{\nu(t)=m,\nu(s+t)-\nu(t)=l\right\}P\left\{\sum_{k=\nu(t)+1}^{\nu(t+s)}\xi_k/\nu(t)=m,\nu(t+s)-\nu(s)=l\right\}=$$

$$=\sum_{\substack{l\ge 0\\ m\ge 0}}P\left\{\nu(t)=m\right\}P\left\{\nu(s+t)-\nu(t)=l\right\}P\left(\xi_{m+1}+...+\xi_{m+l}<x\right).$$

Позначимо $F_k(x)=P\{\xi_1+\dots+\xi_k<x\}$ для $k=1,2,...$, $F_0(x)=0$ при $x\le 0$ і $F_0(x)=1$ при $x>0$. Оскільки $\xi_1,...,\xi_n,...$ – однаково розподілені випадкові величини, то розподіли величин $\xi_{m+1}+\dots+\xi_{m+l}$ і $\xi_1+\dots+\xi_l$ будуть однаковими. Тому $P\left\{\xi_{m+1}+\dots+\xi_{m+l}<x\right\}=F_l(x)$.

Врахуємо, що $\sum_{m\ge 0}\exp(-\lambda t)\frac{(\lambda t)^m}{m!}=1$. Тоді

$$P\left\{\xi(t+s)-\xi(t)<x\right\}=$$

$$=\sum_{\substack{l\ge 0\\ m\ge 0}}F_l(x)P\left\{\nu(t)=m\right\}P\left\{\nu(s+t)-\nu(t)=l\right\}=\sum_{l\ge 0}F_l(x)e^{-\lambda s}\frac{(\lambda s)^l}{l!}\sum_{m\ge 0}e^{-\lambda t}\frac{(\lambda t)^m}{m!}=$$

$$=\sum_{l\ge 0}F_l(x)e^{-\lambda s}\frac{(\lambda s)^l}{l!}.$$

Отже, розподіл від t не залежить. Тому узагальнений процес Пуассона є однорідним.

Вправа.

5. Нехай $\xi(t)$ – узагальнений процес Пуассона. Знайти характеристичну функцію величини $\xi(t+s)-\xi(t)$, якщо $F(x)$ – функція розподілу ξ_n.

8.4. Вінерівський процес

Неперервний однорідний процес $w(t)$ з незалежними приростами такий, що $w(t)-w(s)$ при $t>s$ має нормальний розподіл, називається процесом броунівського руху. Якщо $w(t)-w(s)\sim N\left(a(t-s),\sigma^2(t-s)\right)$, то a називають *коефіцієнтом переносу*, σ^2 – *коефіцієнтом дифузії*.

Якщо ж $w(0)=0$ і $w(t)\sim N(0,t)$, то процес $w(t)$ називається вінерівським процесом.

Вінерівський процес є математичною моделлю явища руху мікроскопічних частинок в рідині або газі під дією молекул середовища, що знаходяться в тепловому хаотичному русі. Якщо середовище однорідне, то рух частинки із кожної точки в будь-який момент буде однаковим в ймовірнісному розумінні, тобто $w(t+s)-w(t)$ не залежить від t. Очевидно також, що траєкторія частинки є неперервною функцією t.

Розглянемо дискретну модель броунівського руху, в якій частинка змінює своє положення стрибками в фіксовані моменти часу. Позначимо через $\xi(t)$ положення частинки в момент часу $t\ge 0$, $\xi(0)=0$. Будемо вважати, що частинка змінює своє положення тільки в моменти кратні Δt, а величина стрибка дорівнює Δx. Якщо в момент часу t частинка знаходиться в точці x, то в момент $t+\Delta t$ з рівними ймовірностями вона може перейти в точки $x-\Delta x$ і $x+\Delta x$. Визначений таким чином процес $\xi(t)$ буде однорідним із незалежними приростами. Неперервну модель броунівського руху одержуємо в результаті граничного переходу при $\Delta x \to 0$ і $\Delta t \to 0$. Положення частинки в момент часу $t+s$

$$\xi(t+s) = \left(\xi(t+s)-\xi(s)\right)+\left(\xi(s)-\xi(0)\right),$$

де доданки в правій частині є незалежними випадковими величинами і розподіли $\xi(t+s)-\xi(s)$ і $\xi(t)-\xi(0)$ однакові.

Тому

$$D\left(\xi(t+s)\right) = D\left(\xi(t+s)-\xi(s)\right)+D\left(\xi(s)-\xi(0)\right) =$$
$$= D\left(\xi(t)-\xi(0)\right)+D\left(\xi(s)-\xi(0)\right) = D\left(\xi(t)\right)+D\left(\xi(s)\right).$$

Дисперсія задовольняє функціональному рівнянню $f(x+y)=f(x)+f(y)$, а такому рівнянню задовольняє єдина лінійна функція. Оскільки $D\left(\xi(0)\right)=0$, то

$$D\left(\xi(t)\right)=\sigma^2 t,$$

де стала σ^2 називається коефіцієнтом дифузії. Позначимо через ξ_k випадкову величину, яка приймає значення $\pm\Delta x$ з ймовірностями $\frac{1}{2}$. Тоді $M\xi_k=0$, $D\xi_k=M\xi_k^2=\Delta x^2$. За час t частинка зробить $\frac{t}{\Delta t}$ стрибків, тому $\xi(t)=\sum_{k=1}^{\frac{t}{\Delta t}}\xi_k$ має $M\xi(t)=0$ і дисперсію $D\xi(t)=\Delta x^2\frac{t}{\Delta t}=\frac{\Delta x^2}{\Delta t}t$. Оскільки $D\left(\xi(t)\right)=\sigma^2 t$, то $\frac{\Delta x^2}{\Delta t}=\sigma^2$. Тоді за центральною граничною теоремою величина $\frac{\xi(t)}{\sigma\sqrt{t}}$ при $\Delta t\to 0$ і $\Delta x\to 0$ так, що $\frac{\Delta x^2}{\Delta t}=\sigma^2$ залишається сталим, буде збігатися за розподілом до випадкової величини $w(t)$ із стандартним нормальним розподілом. Величину $w(t)$ будемо розглядати як неперервну модель броунівського руху.

Отже, процес броунівського руху є вінерівським процесом.

Позначимо через τ_x – момент 1-го досягнення процесом точки x, тобто це мінімальний момент часу t, для якого $\xi(t)=x$.

Розглянемо подію $\{\tau_x<t\}$. Це означає, що точка до моменту часу t побувала в положенні x. Тоді з однаковими ймовірностями вона може виявитись як праворуч цієї точки, так і ліворуч. Тому

$$P\{\tau_x<t\}=2P\{\xi(t)>x\}=2P\left\{\frac{\xi(t)}{\sigma\sqrt{t}}>\frac{x}{\sigma\sqrt{t}}\right\}=2\frac{1}{\sqrt{2\pi}}\int_{\frac{x}{\sigma\sqrt{t}}}^{\infty}\exp\left(-\frac{z^2}{2}\right)dz$$

є функцією розподілу випадкової величини τ_x. Ми використали, що $\frac{\xi(t)}{\sigma\sqrt{t}}\sim N(0,1)$.

Щільність розподілу випадкової величини τ_x має вигляд

$$p_{\tau_x}(t)=\frac{x}{\sqrt{2\pi t^3}\sigma}\exp\left(-\frac{x^2}{2\sigma^2 t}\right).$$

Розглянемо випадковий процес $\zeta(t)=\max\limits_{0\le s\le t}\xi(s)$ – максимальний досягнутий рівень за проміжок часу t. Тоді $\{\zeta(t)\ge x\}=\{\tau_x<t\}$. Отже,

$$P\left\{\max_{0\le s\le t}\xi(s)\ge x\right\}=P\{\tau_x<t\}=\frac{2}{\sqrt{2\pi}}\int_{\frac{x}{\sigma\sqrt{t}}}^{\infty}\exp\left(-\frac{z^2}{2}\right)dz .$$

8.5. Стаціонарні процеси

Серед випадкових процесів важливе місце займають процеси, ймовірнісні характеристики яких не змінюються із часом.

Процес $\xi(t)$ називається *стаціонарним*, якщо для будь-яких $t_1,...,t_n,t$, сумісний розподіл випадкового вектора $\{\xi(t_1+t),...,\xi(t_n+t)\}$ не залежить від t.

Для будь-якої функції $f(x_1,...,x_n)$ – неперервної і обмеженої в R^n $Mf(\xi(t_1+h),...,\xi(t_n+h))$ не залежить від h.

Зокрема, якщо для процесу $\xi(t)$ існують скінченні моменти другого порядку, то $M\xi(t)=m(t)=m$ і $D\xi(t)=\sigma^2$ не залежать від t, а $M\xi(t)\xi(s)$ залежить тільки від $t-s$ $(t\ge s)$.

В тих задачах, де використовуються тільки моменти першого і другого порядків, умову стаціонарності можна послабити, замінивши її умовами, що відносяться тільки до моментів перших двох порядків. При вивченні стаціонарних процесів частіше розглядають процеси $\xi(t)$ із значеннями в множині комплексних чисел, тобто при кожному фіксованому t $\xi(t)$ є комплекснозначна випадкова величина. Випадкова величина ξ буде комплекснозначною, якщо її можна представити у вигляді $\xi=\xi_1+i\xi_2$, де ξ_1 і ξ_2 – дійсні випадкові величини. Тоді $M\xi=M\xi_1+iM\xi_2$, $D\xi=M\left((\xi-M\xi)\overline{(\xi-M\xi)}\right)=M|\xi-M\xi|^2$, $\operatorname{cov}(\xi,\eta)=M\left((\xi-M\xi)\overline{(\eta-M\eta)}\right)$.

Випадковий процес $\xi(t)$ $(t\in R)$ називається *стаціонарним у широкому розумінні* (*слабо стаціонарним*), якщо $M|\xi(t)|^2<\infty$ і $M\xi(t)=m=const$, а

$$R(t,s)=M\left((\xi(t)-m)\overline{(\xi(s)-m)}\right)=R(t-s),$$

де $R(t)$ є неперервною функцією від t.

Функцію $R(t)$ називають *кореляційною функцією* процесу $\xi(t)$. Її можна подати у вигляді $R(t)=M\left((\xi(t)-m)\overline{(\xi(0)-m)}\right)$. Відзначимо, що $R(0)=D\xi(t)$.

Процес $\xi(t)$ називається *неперервним в середньоквадратичному*, якщо

$$M\left(|\xi(t+h)-\xi(t)|^2\right)\xrightarrow[h\to0]{}0 .$$

Якщо процес неперервний в середньоквадратичному, то його кореляційна функція є неперервною.

За нерівністю Коші-Буняковського

$$|R(t+\Delta t)-R(t)|=\left|M\left((\xi(t+\Delta t)-m)\overline{(\xi(0)-m)}\right)-M\left((\xi(t)-m)\overline{(\xi(0)-m)}\right)\right|=$$

$$=\left|M\left((\xi(t+\Delta t)-\xi(t))\overline{(\xi(0)-m)}\right)\right|\le\left(M|\xi(t+\Delta t)-\xi(t)|^2M|\xi(0)-m|^2\right)^{\frac12}\xrightarrow[\Delta t\to0]{}0 .$$

Отже, кореляційна функція є неперервною.

Розглянемо приклад стаціонарного процесу із дискретним спектром. Нехай випадкові величини η_k – комплекснозначні, мають $M\eta_k = 0$, $D\eta_k = c_k^2$, не корельовані, тобто $M\left(\eta_k \overline{\eta_l}\right) = 0$, $k \neq l$, $k, l = 1,...,n$. На основі цих випадкових величин побудуємо випадковий процес

$$\xi(t) = \sum_{k=1}^{n} \eta_k \exp(itu_k),$$

де $u_k \in R$. Покажемо, що такий процес є стаціонарним в широкому розумінні. Дійсно, $M\xi(t) = \sum_{k=1}^{n} M(\eta_k)\exp(itu_k) = 0$ – не залежить від t, а

$$R(t,s) = M\left(\xi(t)\overline{\xi(s)}\right) = M\left(\sum_{k=1}^{n} \eta_k \exp(itu_k)\right)\left(\overline{\sum_{l=1}^{n} \eta_l \exp(isu_l)}\right) =$$

$$= \sum_{k=1}^{n} M\left(\eta_k \overline{\eta_k}\right)\exp\left(i(t-s)u_k\right) = \sum_{k=1}^{n} c_k^2 \exp\left(i(t-s)u_k\right) = R(t-s).$$

Кореляційна функція процесу $R(t,s) = R(t-s)$ залежить тільки від $t-s$. Отже, процес є стаціонарним у широкому розумінні і його кореляційна функція задається формулою

$$R(t) = \sum_{k=1}^{n} c_k^2 \exp(itu_k).$$

Цю формулу називають *спектральним зображенням* кореляційної функції. Вона визначає спектр стаціонарного випадкового процесу $u_1,...,u_n$ – сукупність частот гармонійних коливань, складових процесу. Величину c_k^2 називають середнім значенням потужності гармонійної складової процесу з частотою u_k, а $\sum_{k=1}^{n} c_k^2 = R(0)$ називають потужністю стаціонарного процесу.

Функція

$$F(u) = \sum_{k: u_k < u} c_k^2$$

називається спектральною функцією стаціонарного процесу. За допомогою спектральної функції кореляційна функція процесу $\xi(t)$ може бути записана у вигляді

$$R(t) = \int_{-\infty}^{+\infty} \exp(itu)\, dF(u).$$

Виявляється, що поняття спектральної функції і зображення кореляційної функції за допомогою попередньої формули можна узагальнити на довільні випадкові процеси. Наведемо без доведення теорему.

Теорема (Хінчіна). Для того, щоб функція $R(t)$ була кореляційною функцією неперервного в середньоквадратичному стаціонарного в широкому розумінні випадкового процесу, необхідно і досить, щоб вона допускала зображення $R(t) = \int_{-\infty}^{+\infty} \exp(itu)\, dF(u)$.

Якщо функція $F(u)$ – абсолютно неперервна, тобто $F(u) = \int_{-\infty}^{u} f(x)\, dx$, то функцію $f(x)$ називають спектральною щільністю випадкового процесу. В цьому випадку

$$R(t) = \int_{-\infty}^{+\infty} \exp(itu)\, f(u)\, du.$$

Якщо $R(t)$ – абсолютно інтегровна, то існує спектральна щільність і справедлива рівність

$$f(u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itu) R(t) dt .$$

Вправи.

6. Нехай ξ – випадкова величина, що має нормальний розподіл з математичним сподіванням m і дисперсією σ^2, b – дійсне число. Знайти кореляційну функцію процесу $\xi(t) = \xi t + b$, $t > 0$.
7. Знайти спектральну функцію випадкового процесу $\xi(t)$, кореляційна функція якого дорівнює $R(t) = ce^{-\alpha|t|}$, $c > 0$, $\alpha > 0$, $t \in R$.

8.6. Марковські процеси

Нехай система може знаходитись в одному із станів множини X (фазового простору) і переходить із одного стану в другий у будь-які моменти часу. Нехай випадковий процес $\xi(t)$ приймає значення із множини X, де $X = \{x_1, x_2, ...\}$ – скінченна або зліченна.

Процес $\xi(t)$ називається *марковським*, якщо для довільних $0 \le t_1 < t_2 < ... < t_n$ $P\{\xi(t_n) = j / \xi(t_1) = k, ..., \xi(t_{n-1}) = i\} = P\{\xi(t_n) = j / \xi(t_{n-1}) = i\}$, тобто значення процесу в момент часу t_n залежить від значення процесу в момент часу t_{n-1} і не залежить від значення процесу в попередні моменти часу ($k, \dots, i, j$ – номери станів).

Ймовірності

$$P_{ij}(s,t) = P\{\xi(t) = j / \xi(s) = i\}, \ t > s ,$$

називаються *ймовірностями переходу* із стану i в стан j за час від s до t.

Відмінність від ланцюгів Маркова полягає у тому, що час змінюється неперервно.

Для ймовірностей переходу виконуються умови: $0 \le P_{ij}(s,t) \le 1$,

$$\sum_j P_{ij}(s,t) = 1, \ t > s , \tag{8}$$

а

$$P_{ii}(s,s) = 1, \ P_{ij}(s,s) = 0, \ i \ne j . \tag{9}$$

Крім того перехідна ймовірність задовольняє таке рівняння Маркова: для довільних $s < u < t$

$$P_{ij}(s,t) = \sum_k P_{ik}(s,u) P_{kj}(u,t). \tag{10}$$

Дійсно, із означення перехідних ймовірностей випливає, що

$$P\{\xi(s) = i, \xi(u) = k, \xi(t) = j\} = P\{\xi(s) = i\} P_{ik}(s,u) P_{kj}(u,t),$$

$$P\{\xi(s) = i, \xi(t) = j\} = P\{\xi(s) = i\} P_{ij}(s,t).$$

Якщо підставимо ці величини в рівність

$$P\{\xi(s) = i, \xi(t) = j\} = \sum_k P\{\xi(s) = i, \xi(u) = k, \xi(t) = j\} ,$$

то одержимо

$$P\{\xi(s) = i\} P_{ij}(s,t) = \sum_k P\{\xi(s) = i\} P_{ik}(s,u) P_{kj}(u,t) ,$$

а це рівносильно (10).

Нехай час відраховується від деякого моменту часу t_0 і нехай $p_k^0 = P\left\{\xi\left(t_0\right)=k\right\}$, $k=1,2,\ldots$, – початковий розподіл процесу, а $P_k\left(t\right)=P\left\{\xi\left(t\right)=k\right\}$ – розподіл процесу в моменту часу t. Тоді $\sum_k p_k^0 = 1$, крім того

$$P_k\left(t\right)=\sum_i p_i^0 P_{ik}\left(t_0,t\right). \tag{11}$$

Якщо $P_{ij}\left(s,t\right)=P_{ij}\left(t-s\right)$ залежить лише від різниці $t-s$, то марковський процес називається *однорідним*. Покладемо $t_0=0$, тоді ймовірності $P_{ij}\left(t\right)$ є ймовірностями переходу із стану i в стан j за час t і

$$\sum_j P_{ij}\left(t\right)=1,$$

при цьому із (11), (10), (9) випливають відповідно такі формули:

$$P_k\left(t\right)=\sum_i p_i^0 P_{ik}\left(t\right), \tag{12}$$

$$P_{ij}\left(s+t\right)=\sum_k P_{ik}\left(s\right)P_{kj}\left(t\right) \tag{13}$$

$$P_{ii}\left(0\right)=1,\ P_{ij}\left(0\right)=0,\ i\neq j. \tag{14}$$

Будемо розглядати тепер лише однорідні марковські процеси.

Будемо вважати, що при всіх $i, j = 1, 2,\ldots$ перехідні ймовірності $P_{ij}\left(t\right)$ задовольняють наступним умовам:

$$1-P_{ii}\left(h\right)=\lambda_i h+o\left(h\right),\ P_{ij}\left(h\right)=\lambda_{ij}h+o\left(h\right),\ i\neq j, \tag{15}$$

при $h\to 0$. Величини λ_i називають щільністю виходу із стану *i*, а λ_{ij} називають щільністю переходу із стану *i* в стан *j*. Покладемо $\lambda_{ii}=-\lambda_i$.

Теорема. Якщо для марковського процесу із скінченною множиною станів виконуються умови (15), то ймовірності переходу $P_{ij}\left(t\right)$ задовольняють диференціальним рівнянням

$$P'_{ij}\left(t\right)=\sum_k \lambda_{ik}P_{kj}\left(t\right),\quad i,j=1,2,\ldots, \tag{16}$$

$$P'_{ij}\left(t\right)=\sum_k P_{ik}\left(t\right)\lambda_{kj},\quad i,j=1,2,\ldots \tag{17}$$

з початковими умовами (14), де $\lambda_{ii}\le 0$, а $\lambda_{ij}\ge 0$ при $i\neq j$ і

$$\sum_k \lambda_{ik}=0. \tag{18}$$

Рівняння (16) називаються першою або оберненою системою, а рівняння (17) – другою або прямою системою диференціальних рівнянь Колмогорова.

Доведення. Із рівності (13) і умов (15) маємо

$$P_{ij}\left(t+h\right)-P_{ij}\left(t\right)=\sum_k P_{ik}\left(h\right)P_{kj}\left(t\right)-P_{ij}\left(t\right)=P_{ij}\left(t\right)\left(P_{ii}\left(h\right)-1\right)+\sum_{k\neq i}P_{ik}\left(h\right)P_{kj}\left(t\right)=$$

$$=P_{ij}\left(t\right)\left(\lambda_{ii}h+o\left(h\right)\right)+\sum_{k\neq i}\left(\lambda_{ik}h+o\left(h\right)\right)P_{kj}\left(t\right)=\sum_k\left(\lambda_{ik}h+o\left(h\right)\right)P_{kj}\left(t\right).$$

Тому

$$\frac{P_{ij}\left(t+h\right)-P_{ij}\left(t\right)}{h}=\sum_k\left(\lambda_{ik}+\frac{o\left(h\right)}{h}\right)P_{kj}\left(t\right).$$

Оскільки число доданків у правій частині скінченне, тому існує скінченна границя правої частини при $h\to 0$, отже існує скінченна границя $P'_{ij}\left(t\right)$ і лівої частини і справедлива рівність (16). Аналогічно,

$$P_{ij}(t+h)-P_{ij}(t)=\sum_k P_{ik}(t)P_{kj}(h)-P_{ij}(t)=P_{ij}(t)\big(P_{jj}(h)-1\big)+\sum_{k\neq j}P_{ik}(t)P_{kj}(h)=$$
$$=P_{ij}(t)\big(\lambda_{jj}h+o(h)\big)+\sum_{k\neq j}P_{ik}(t)\big(\lambda_{kj}h+o(h)\big)=\sum_k P_{ik}(t)\big(\lambda_{kj}h+o(h)\big).$$

Після ділення на h і переходу до границі при $h\to 0$, одержуємо (17).

Рівність $\sum_j P_{ij}(h)=1$ перепишемо у вигляді $\sum_{j\neq i}P_{ij}(h)+P_{ii}(h)-1=0$ і використаємо умови (15), тоді $\sum_{j\neq i}\big(\lambda_{ij}h+o(h)\big)+\lambda_{ii}h+o(h)=0$ і $\sum_j\left(\lambda_{ij}+\frac{o(h)}{h}\right)=0$, звідки випливає (18). Теорема доведена.

Відзначимо, що при додаткових умовах, рівняння (16) і (17) виконуються і для випадку зліченної множини станів.

Продиференціюємо рівність $P_k(t)=\sum_i p_i^0P_{ik}(t)$ по параметру t, тоді із (17) одержуємо $P_j'(t)=\sum_i p_i^0P_{ij}'(t)=\sum_i p_i^0\sum_k P_{ik}(t)\lambda_{kj}=\sum_k\lambda_{kj}\sum_i p_i^0P_{ik}(t)=\sum_k P_k(t)\lambda_{kj}$, тобто ймовірності $P_j(t)$ задовольняють систему диференціальних рівнянь

$$P_j'(t)=\sum_k P_k(t)\lambda_{kj}\,. \tag{19}$$

Розв'язок цієї системи знаходиться при початкових умовах $P_j(0)=p_j^0$. Якщо P_k такі, що $P_j'(t)=0$, тобто вони не змінюються з часом, то вони називаються *стаціонарними*. Із (19) випливає така система для стаціонарних ймовірностей:

$$\sum_k P_k\lambda_{kj}=0\,.$$

Приклад. Нехай на деяку систему масового обслуговування поступають замовлення так, що $\xi(t)$ – число замовлень за час t – утворює однорідний марковський процес із зліченною множиною станів 0, 1, 2, … . Із стану i система безпосередньо може перейти тільки в $i+1$, i = 0, 1, 2, … . Нехай щільності виходу із кожного стану однакові і дорівнюють λ. Тому $\lambda_{ii+1}=\lambda$, інші $\lambda_{ij}=0$ при інших $j\neq i,i+1$. Із (18) випливає, що $\lambda_{ii}=-\lambda$. Система (19) буде мати вигляд

$$P_0'(t)=-\lambda P_0(t),$$
$$P_k'(t)=\lambda P_{k-1}(t)-\lambda P_k(t),\ k=1,2,\dots\,.$$

Розв'язок цієї системи задовольняє початковим умовам: $P_0(0)=1$, $P_k(0)=0$, $k=1,2,\dots$. Точно таку систему ми одержали при вивченні процесу Пуассона. Там ми знайшли і її розв'язок:

$$P_k(t)=e^{-\lambda t}\frac{(\lambda t)^k}{k!},\ k=0,1,2,\dots.$$

Отже, $\xi(t)$ є процесом Пуассона.

Розділ 9. Математична статистика

9.1. Основні поняття вибіркового методу

9.1.1. Вибірка. Емпірична функція розподілу. Математична статистика вивчає методи збору і обробки результатів спостережень для одержання наукових і практичних висновків. Методи розв'язування задач в теорії ймовірностей вимагають знання різних ймовірнісних характеристик. При розв'язуванні практичних задач ці характеристики не можуть бути відомими. Математична статистика розглядає методи знаходження необхідних ймовірнісних характеристик за статистичними даними. Тобто, основним предметом дослідження в математичній статистиці є випадкові величини, але висновки про властивості цих випадкових величин робляться на основі статистичних даних.

Одна із важливих задач математичної статистики пов'язана із знаходженням розподілу випадкової величини за статистичними даними. До цієї задачі зводяться багато інших задач. Основними задачами математичної статистики є розробка методів знаходження оцінок (знайдених на основі статистичних даних характеристик досліджуваної величини), дослідження їх точності, розробка методів перевірки гіпотез, аналіз зв'язку між величинами.

Нехай задано сукупність об'єктів, що об'єднані спільною ознакою ξ (кількісною або якісною, якою характеризуються однорідні об'єкти). Її значення для кожного об'єкту передбачити неможливо, тому досліджувана ознака ξ є випадковою величиною. Вивчення властивостей величини ξ пов'язане із проведенням статистичних спостережень – знаходження значень величини ξ для деяких об'єктів. Таким чином можна обстежити всю сукупність або деяку її частину. Обстеження всієї сукупності не завжди доцільне, а інколи просто неможливе. Таким методом, зокрема, здійснюють контроль за якістю продукції, проводять соціологічні дослідження. У всіх таких випадках проводиться обстеження тільки частини об'єктів і по його результатах роблять висновки про всю сукупність. Такий метод називається вибірковим.

Множина значень $x_1,...,x_n$ величини ξ, що одержані в результаті n експериментів, називається *вибіркою обсягу* n. Множину всіх значень величини ξ ми будемо називати генеральною сукупністю. Поняття вибірки вживається і більш широко – як n-вимірна випадкова величина. Вибірка є єдиним джерелом інформації про досліджувану величину. Ми будемо розглядати вибірку організовану методом випадкового відбору. При цьому кожен об'єкт сукупності має однакову ймовірність бути відібраним, така вибірка буде правильно відображати властивості всієї сукупності. На практиці частіше всього використовується вибір без повернення (безповторна вибірка), коли кожен відібраний об'єкт, перед вибором наступних, назад до сукупності не повертається. Вибір з поверненням розглядається частіше в теоретичних дослідженнях. Якщо обсяг вибірки значно менший обсягу досліджуваної сукупності, то обидві вибірки дають практично однакові результати.

При одночасному дослідженні двох (або більшої кількості) ознак вибірка буде складатись із упорядкованих пар (або упорядкованих наборів) чисел.

Для розв'язування багатьох задач важливо знати розподіл досліджуваної випадкової величини, бо на його основі приймають ті чи інші рішення.

Нехай задано вибірку (статистичні дані) $x_1,...,x_n$ для величини ξ. Безпосередньо із вибірки важко зробити якісь висновки про властивості досліджуваної величини, тому необхідно провести первинну обробку результатів спостережень.

Статистичні дані $x^{(1)},...,x^{(n)}$, що записані у порядку зростання ($x^{(i)} \le x^{(i+1)}$), називають *варіаційним рядом*, а його члени $x^{(i)}$ – варіантами.

У варіаційному ряді значення можуть повторюватись. Якщо ми будемо записувати тільки різні значення $x_{(1)},...,x_{(k)}$ $(k \le n)$ і, якщо значення $x_{(i)}$ зутрічається у вибірці n_i раз, то число n_i називають частотою, а $\frac{n_i}{n}$ – відносною частотою значення $x_{(i)}$.

Перелік різних варіант і відповідних їм частот (відносних частот) називають *емпіричним* або *статистичним* розподілом (статистичним рядом). Його можна подати у вигляді такої таблиці (надалі дужки в індексах будемо опускати):

Варіанти x_i	x_1	x_2	$\cdots$	x_k
Частоти n_i	n_1	n_2	$\cdots$	n_k

Відзначимо, що $n_1 + n_2 + ... + n_k = n$ – обсяг вибірки. Такий розподіл називають дискретним.

Для графічного зображення дискретного розподілу використовують полігон частот (або відносних частот) – це ламана із вершинами в точках (x_i, n_i) (або $(x_i, n_i / n)$).

Якщо досліджується дискретна випадкова величина, кількість можливих значень якої досить велике, або досліджується неперервна випадкова величина, то будують *інтервальний статистичний ряд* або *інтервальний розподіл*. Для цього множину всіх значень $x_1,...,x_n$, від найменшого ($x_{\min}$) до найбільшого ($x_{\max}$), розбивають на певну кількість інтервалів і для кожного із інтервалів вказують число статистичних даних, які в нього попадають. Інтервали можуть мати однакову або різну довжину. При розбитті на нерівні інтервали їх необхідно підбирати так, щоб розподіл ξ в кожному із них був приблизно рівномірним. Кількість інтервалів залежить від мети дослідження. Інтервальний розподіл можна подати у вигляді такої таблиці:

Інтервали	(a_0, a_1)	(a_1, a_2)	$\cdots$	(a_{s-1}, a_s)
Частоти n_i	n_1	n_2	$\cdots$	n_s

Тут n_i кількість статистичних даних, що належать *i*-му інтервалу, а $n_1 + n_2 + ... + n_s = n$. Якщо замість n_i поставити $\frac{n_i}{n}$, то будемо мати емпіричний розподіл відносних частот. Для подальшої обробки результатів спостережень від інтервального розподілу переходять до дискретного, замінюючи кожен інтервал його серединою.

Графічним зображенням інтервального розподілу є гістограма – сукупність прямокутників, основами яких є інтервали групування, а висоти дорівнюють відношенню частоти інтервалу до його довжини $\frac{n_i}{h_i}$. Якщо висоти рівні $\frac{n_i}{nh_i}$, то таку сукупність прямокутників називають гістограмою відносних частот. Площа *i*-го прямокутника дорівнює відносній частоті $\frac{n_i}{n}$, яка при $n \to \infty$ збігається за ймовірністю до ймовірності попадання значень випадкової величини у відповідний інтервал. Якщо довжина h інтервалу мала, то ця ймовірність приблизно рівна $p(x)h$, де $p(x)$ – щільність. Отже, верхню межу гістограми можна розглядати як статистичний аналог щільності розподілу досліджуваної випадкової величини.

Визначимо для кожного дійсного x випадкову величину $\mu_n(x)$, яка дорівнює числу елементів вибірки, що менші за x, і покладемо $F_n(x) = \frac{\mu_n(x)}{n}$. Функція $F_n(x)$ називається *емпіричною* функцією розподілу. Функцію розподілу $F(x)$, досліджуваної випадкової величини ξ, називають *теоретичною* функцією розподілу. Емпіричну функцію розподілу можна визначити рівністю $F_n(x) = \frac{1}{n} \sum\limits_{i : x_i < x} n_i$. Емпірична функція розподілу є однією із найважливіших характеристик вибірки. Оскільки $F_n(x)$ є відносною частотою події $\{\xi < x\}$ в n незалежних випробуваннях з ймовірністю настання цієї події $F(x)$, то за законом великих чисел (теорема Бернуллі)

$$F_n(x) \xrightarrow[n \to \infty]{p} F(x),$$

а із теореми Бореля (посилений закон великих чисел) випливає, що ця збіжність має місце із ймовірністю одиниця. Справедлива і більш загальна теорема, яку ми наведемо без доведення.

Теорема (Глівенко-Кантеллі). Нехай $F(x)$ – функція розподілу випадкової величини ξ, а $F_n(x)$ – емпірична функція розподілу, що знайдена за вибіркою обсягу *n*. Емпірична функція розподілу рівномірно по x з ймовірністю 1 збігається при $n \to \infty$ до теоретичної функції розподілу:

$$P\left\{ \lim_{n \to \infty} \sup_{x \in R} \left| F_n(x) - F(x) \right| = 0 \right\} = 1.$$

Таким чином, при великих n емпірична функція розподілу в кожній точці є наближеним значенням теоретичної функції розподілу.

9.1.2. Вибіркові характеристики. Будь-які характеристики, які знаходяться на основі статистичний даних, називаються *емпіричними* або статистичними характеристиками, а характеристики, які знаходяться на основі розподілу досліджуваної величини, називаються *теоретичними* характеристиками.

Використання всіх статистичних даних для аналізу ознаки не завжди доцільно. Для аналізу властивостей досліджуваної ознаки на основі статистичних даних використовують числові характеристики вибірки (статистичні характеристики). До основних характеристик вибірки відносять величини, які характеризують середнє значення та розсіювання можливих значень досліджуваної величини.

Однією із основних характеристик середнього значення є *вибіркова середня* $\bar{x}$ – *середнє арифметичне* результатів спостережень:

$$\bar{x} = \frac{1}{n} \sum_i x_i . \tag{1}$$

Попередню уяву про розсіювання статистичних даних дає *розмах варіювання* $R = x_{\max} - x_{\min}$, але ця величина є досить грубою характеристикою розсіювання.

До основних характеристик розсіювання статистичних даних відносять *вибіркову дисперсію* :(середнє арифметичне квадратів відхилень результатів спостережень від вибіркової середньої)

$$D_B = \frac{1}{n} \sum_i \left(x_i - \bar{x} \right)^2 \tag{2}$$

і вибіркове середнє квадратичне відхилення

$$\sigma_B = \sqrt{D_B} .$$

Для незгрупованих даних вибіркова середня і вибіркова дисперсія знаходяться за формулами (1) і (2). Якщо ж статистичні дані згруповані, то

$$\overline{x} = \frac{1}{n}\sum_i n_i x_i\,,\quad D_B = \frac{1}{n}\sum_i \left(x_i - \overline{x}\right)^2 n_i\,, \tag{3}$$

де x_i – різні результати спостережень випадкової величини ξ, а n_i – відповідні частоти.

Для вибіркової середньої і вибіркової дисперсії для довільних $c \in R$ і $h \neq 0$ справедливі формули

$$\overline{x} = h\overline{u} + c\,,\quad D_B = h^2\left(\overline{u^2} - \left(\overline{u}\right)^2\right), \tag{4}$$

де $\overline{u} = \frac{1}{n}\sum_i n_i u_i$, $\overline{u^2} = \frac{1}{n}\sum_i n_i u_i^2$, а $u_i = \frac{x_i - c}{h}$.

Дійсно, із (3) маємо $\overline{x} = \frac{1}{n}\sum_i n_i x_i = \frac{1}{n}\sum_i \left(\left(x_i - c\right) + c\right)n_i = \frac{1}{n}\sum_i \left(x_i - c\right)n_i + \frac{1}{n}\sum_i c n_i =$

$$= h\frac{1}{n}\sum_i \frac{\left(x_i - c\right)}{h} n_i + c = h\overline{u} + c\,,\quad D_B = \frac{1}{n}\sum_i \left(x_i - \overline{x}\right)^2 n_i = \frac{1}{n}\sum_i \left(\left(x_i - c\right) + \left(c - \overline{x}\right)\right)^2 n_i =$$

$$= \frac{1}{n}\left(\sum_i \left(x_i - c\right)^2 n_i + 2\left(c - \overline{x}\right)\sum_i \left(x_i - c\right)n_i + \sum_i \left(c - \overline{x}\right)^2 n_i\right) =$$

$$= h^2\frac{1}{n}\sum_i \left(\frac{x_i - c}{h}\right)^2 n_i + 2\left(c - \overline{x}\right)\left(\overline{x} - c\right) + \left(c - \overline{x}\right)^2 = h^2\overline{u^2} - \left(c - \overline{x}\right)^2 = h^2\left(\overline{u^2} - \left(\overline{u}\right)^2\right).$$

При $h = 1$ ми дістанемо:

$$\overline{x} = \frac{1}{n}\sum_i \left(x_i - c\right)n_i + c\,,\ D_B = \frac{1}{n}\sum_i \left(x_i - c\right)^2 n_i - \left(\overline{x} - c\right)^2,$$

а якщо $c = 0$, то

$$D_B = \overline{x^2} - \left(\overline{x}\right)^2,$$

де $\overline{x^2} = \frac{1}{n}\sum_i n_i x_i^2$.

Із одержаних формул випливає, що $\min\limits_c \frac{1}{n}\sum_i \left(x_i - c\right)^2 n_i = D_B$, тобто, середнє арифметичне квадратів відхилень від деякої сталої c $\frac{1}{n}\sum_i \left(x_i - c\right)^2 n_i \geq D_B$, рівність досягається тільки тоді, коли $c = \overline{x}$.

Нехай статистичні дані розбиті на s груп. Обсяг k-ї групи позначимо N_k, середнє арифметичне k-ї групи позначимо $\overline{x_{k\,гр}} = \frac{1}{N_k}\sum\limits_{i:x_i \in k-oї\,гр.} x_i$, а дисперсію – $D_{k\,гр} = \frac{1}{N_k}\sum\limits_{i:x_i \in k-oї\,гр} \left(x_i - \overline{x_{k\ гр}}\right)^2$. Обсяг всієї сукупності $n = N_1 + \ldots + N_s$. Тоді справедливі формули:

$$\overline{x} = \frac{1}{n}\sum_k N_k\overline{x_{k\ гр}}\,,\quad D_B = \frac{1}{n}\sum_k N_k D_{k\,гр} + \frac{1}{n}\sum_k N_k\left(\overline{x_{k\ гр}} - \overline{x}\right)^2.$$

У останній формулі перший доданок називають внутрігруповою дисперсією, а другий – міжгруповою дисперсією.

Для характеристики середнього значення розглядають також середнє геометричне, середнє гармонійне, моду, медіану, квантилі тощо. Розглянемо емпіричну функцію розподілу $F_n\left(x\right)$. Число $\hat{x}_p = \max\left\{x : F_n\left(x\right) \leq p\right\}$ називається *емпіричною* p*- квантиллю*.

Можна записати таку формулу: $\hat{x}_p = \begin{cases} x_{np}, & np - ціле, \\ x_{[np]+1}, & np - не\ ціле, \end{cases}$ де $x_{[np]}$ – член варіаційного ряду із номером $[np]$. *Медіана* – це квантиль $\hat{x}_{1/2}$. *Мода* – це член варіаційного ряду із найбільшою частотою.

Для більш детального вивчення властивостей розподілу досліджуваної величини використовують емпіричні моменти і центральні емпіричні моменти. Емпіричним моментом r - го порядку (r>0) називають величину

$$\hat{\alpha}_r = \frac{1}{n}\sum_i n_i x_i^r ,$$

а центральним емпіричним моментом r - го порядку –

$$\hat{\mu}_r = \frac{1}{n}\sum_i \left(x_i - \overline{x}\right)^r n_i .$$

Розглянемо приклад обчислення вибіркової середньої і вибіркової дисперсії.

Приклад. Знайти вибіркову середню і вибіркову дисперсію за заданим емпіричним розподілом

x_i	18,4	18,9	19,3	19,6
n_i	5	10	20	15

Розв’язування. Використаємо формули (4) $\overline{x} = h\overline{u} + c$, $D_B = h^2\left(\overline{u^2} - \left(\overline{u}\right)^2\right)$, де $\overline{u} = \frac{1}{n}\sum_i n_i u_i$, $\overline{u^2} = \frac{1}{n}\sum_i n_i u_i^2$, а $u_i = \frac{x_i - c}{h}$. Покладемо c = 19,3, h = 1. У відповідності до вибраних формул складемо розрахункову таблицю

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$
18,4	5	-0,9	-4,5	4,05
18,9	10	-0,4	-4	1,6
19,3	20	0	0	0
19,6	15	0,3	4,5	1,35
$\sum$	n=50		$\sum n_i u_i$ =4	$\sum n_i u_i^2$ =7

При заповненні третього стовпчика ($u_i = x_i - c$) знаходимо різниці чисел першого і другого стовпчиків, при заповненні четвертого стовпчика знаходимо добуток чисел другого і третього стовпчиків, при заповненні п’ятого стовпчика знаходимо добуток чисел третього і четвертого стовпчиків, що розміщені в одному рядку. В самому нижньому рядку записуємо суми чисел відповідних стовпчиків. Підставляємо одержані значення у формули (4). Тоді $\overline{u} = \frac{1}{50}\cdot(-4)$ = -0,08; $\overline{u^2} = \frac{1}{50}\cdot 7 = 0{,}14$. Отже,

$\overline{x}$ = -0,08+19,3 = 19,22; $D_B = 0{,}14 - (-0{,}08)^2 \approx 0{,}13$.

Вправи.

1. В результаті обстеження одержано дані про число пташенят в гніздах лісової ластівки (Iridoprocne bicolor): 4 5 4 5 5 4 5 4 3 5 4 5 6 1 6 4 4 4 5 5 3 5 5 4 6 4 6 2 3 4 5 5 5 5 5 5 4 5 5 6 4 6 2 5 5 3 5 5 4 5 5 6 4 6 2 5 5 3 5 5 5 7 5 5 5 5 4 3 7 6 4 4 5 5 6 6 4 4 6. Подати дані у вигляді варіаційного ряду, емпіричного розподілу, побудувати полігон частот. Знайти емпіричну функцію розподілу.
2. Довжина 100 листків садової суниці (в см) характеризується такими даними: 8.2 9.7 5.6 7.4 8.0 6.4 6.6 6.8 8.4 7.1 9.0 6.9 7.6 8.1 11.8 5.8 9.3 7.3 8.2 7.2

7.2 6.4 7.7 9.0 8.1 7.1 7.1 8.8 7.5 9.2 7.5 6.8 7.0 6.4 7.4 8.2 6.3 7.0 8.1 10.0
7.0 7.1 8.7 6.3 8.6 7.7 7.3 8.0 8.4 9.3 7.3 6.0 7.7 6.1 9.6 7.4 7.2 7.2 8.7 7.5
9.1 6.4 8.3 6.5 8.2 7.2 6.9 6.9 8.2 9.0 7.4 8.0 8.4 7.0 7.1 7.4 6.6 6.4 8.3 7.9
8.3 7.2 7.2 6.6 6.6 7.7 8.7 5.6 7.5 5.7 6.9 7.4 7.2 6.2 6.9 6.8 9.2 9.2 7.1 6.5.
Побудувати інтервальний розподіл, розбивши статистичні дані на 6 – 7 інтервалів однакової довжини, побудувати гістограму частот.

3. За статистичними даними вправ 1 і 2 (після їх виконання) знайти вибіркову середню і вибіркову дисперсію.
4. Знайти вибіркову середню і вибіркову дисперсію за заданим інтервальним розподілом розміру денного виробітку тканин 100 ткалями:

Денний виробіток, м	40-44	44-48	48-52	52-56	56-60
Кількість ткаль	12	28	36	16	8

9.2. Статистичні оцінки параметрів розподілу

9.2.1. Статистичні оцінки і їх властивості. Нехай функція розподілу досліджуваної величини належить до деякої множини розподілів $F(x,\theta)$, що залежить від параматра $\theta \in \Theta$, де Θ - множина на прямій або в скінченно-вимірному просторі R^s. Наприклад, нормальний розподіл на прямій залежить від двох параметрів – математичного сподівання і дисперсії. Вважаємо, що значення параметра θ невідоме і його необхідно знайти на основі результатів експерименту. Процедура знаходження невідомих параметрів на основі статистичних даних називається *статистичним оцінюванням*.

Якщо ми плануємо провести вибірку (повторну, методом випадкового відбору) обсягу n, то результат *i*-го спостереження наперед передбачити неможливо, тому його можна розглядати як випадкову величину X_i із тим самим розподілом, що і досліджувана величина ξ, бо при проведенні різних вибірок обсягу *n*, величина X_i може приймати будь-яке із значень величини ξ і так само часто, як його може приймати величина ξ. Якщо результати проводяться незалежно, то випадкові величини $X_1,...,X_n$ будуть незалежними. Таким чином, ми будемо розглядати вибірку обсягу n, як набір n незалежних однаково розподілених випадкових величин $X_1,...,X_n$, які мають такий самий розподіл *F*, як і досліджувана величина ξ. В цьому випадку говорять, що $X_1,...,X_n$ є вибірка обсягу n із генеральної сукупності із розподілом *F*. Простір R^n, в якому приймає значення вектор $X=(X_1,...,X_n)$, в математичній статистиці називають *вибірковим простором*. При цьому, якщо серія із *n* експериментів уже проведена, то набір чисел $x_1, x_2,..., x_n$ будемо розглядати як реалізацію вибірки ($X_1,...,X_n$).

Для описування властивостей досліджуваної величини на основі великої кількості результатів спостережень вводять порівняно невелику кількість характеристик, які залежать (є функціями) від результатів спостережень.

Будь-яка борелева функція $f(X_1,...,X_n)$ від результатів спостережень називається *статистикою*. Тому статистика є випадковою величиною, що визначена на вибірковому просторі R^n. Наприклад, $X^{(1)}=\min\{X_1,...,X_n\}$ і $X^{(n)}=\max\{X_1,...,X_n\}$ – крайні члени варіаційного ряду, є статистиками. Прикладами статистик є також вибіркова середня, вибіркова дисперсія, емпіричні моменти, емпірична функція розподілу.

Статистика, яка приймається за наближене значення невідомого параметру, називається *статистичною оцінкою*. Отже, статистична оцінка є функцією, що визначена на вибірковому просторі R^n і приймає значення в множині Θ. Статистичну оцінку невідомого параметра θ будемо позначати $\hat{\theta}$. Якщо розподіл $F(x,\theta)$ залежить від векторного параметра $\theta=(\theta_1,...,\theta_s)$, то статистична оцінка $\hat{\theta}=(\hat{\theta}_1,...,\hat{\theta}_s)$ є векторною випадковою величиною. При проведенні різних вибірок будемо одержувати, взагалі кажучи, різні наближені значення невідомого параметра. Ці значення повинні концентруватись біля невідомого параметра і мало від нього відрізнятись. Отже, якість оцінки визначається розподілом величини $\hat{\theta}-\theta$. Оцінка буде доброю, якщо цей розподіл зосереджений близько біля нуля.

Оцінка $\hat{\theta}$ називається *незміщеною* оцінкою невідомого параметру θ, якщо $M\hat{\theta}=\theta$. Якщо ж $M\hat{\theta}\neq\theta$, то оцінка називається зміщеною і величина $M\hat{\theta}-\theta$ називається зміщенням оцінки.

Нехай маємо послідовність статистичних оцінок $\hat{\theta}_1,...,\hat{\theta}_n,...$. Оцінка $\hat{\theta}_n$ називається *спроможною* (*консистентною*), якщо $\hat{\theta}_n \xrightarrow[n\to\infty]{p} \theta$.

Нехай $\hat{\theta}_n$ є незміщеною оцінкою, тобто $M\hat{\theta}_n=\theta$. Тоді до випадкової величини $\hat{\theta}_n$ можна застосувати нерівність Чебишова $P\left\{\left|\hat{\theta}_n-\theta\right|\geq\varepsilon\right\}\leq\frac{D\hat{\theta}_n}{\varepsilon^2}$. Якщо у цьому випадку $D\hat{\theta}_n \xrightarrow[n\to\infty]{} 0$, то і ліва частина буде прямувати до нуля, а за означенням збіжності за ймовірністю одержуємо, що $\hat{\theta}_n \xrightarrow[n\to\infty]{p} \theta$.

Тобто умова $D\hat{\theta}_n \xrightarrow[n\to\infty]{} 0$ є достатньою умовою спроможності.

Розглянемо тепер один із підходів до порівняння двох оцінок і визначення *оптимальної* (*ефективної*) оцінки. Для того, щоб оцінити величину різниці $\hat{\theta}-\theta$, будемо вважати, що задана деяка функція $r(\alpha,\theta), \alpha\in\Theta, \theta\in\Theta$, яку називають функцією втрат або функцією ризику, що характеризує величину помилки, якщо за значення параметра θ прийняти значення α. Якщо $\hat{\theta}$ – оцінка параматра θ, то величина

$$\bar{r}(\hat{\theta},\theta)=Mr(\hat{\theta},\theta)$$

характеризує середні втрати, коли за значення параметра θ прийняти $\hat{\theta}$.

Якщо для двох статистичних оцінок $\hat{\theta}^{(1)}$ і $\hat{\theta}^{(2)}$ параметра θ

$$\bar{r}(\hat{\theta}^{(1)},\theta)\leq\bar{r}(\hat{\theta}^{(2)},\theta),$$

то ми будемо вважати, що оцінка $\hat{\theta}^{(1)}$ є кращою, ніж $\hat{\theta}^{(2)}$.

Якщо існує $\min\limits_{\hat{\theta}}\bar{r}(\hat{\theta},\theta)=\bar{r}(\theta)$ (мінімум береться по деякому класу Ф оцінок), то те значення $\hat{\theta}$, для якого $\bar{r}(\hat{\theta},\theta)=\bar{r}(\theta)$, називається *оптимальною оцінкою* невідомого параметра θ. Отже, оптимальна оцінка параметра θ в класі Ф мінімізує функцію втрат.

Якщо покласти

$$r(\hat{\theta},\theta)=\left|\hat{\theta}-\theta\right|^2,$$

то

$$\delta^2=M\left|\hat{\theta}-\theta\right|^2$$

називають середньо квадратичною похибкою оцінки $\hat{\theta}$. Якщо ми розглядаємо незміщені оцінки ($M\hat{\theta}_n=\theta$), то тоді $\delta^2=M\left|\hat{\theta}-\theta\right|^2$ є сумою дисперсій компонент оцінки $\hat{\theta}$, а якщо θ – одновимірний параметр, то $\delta^2=M(\hat{\theta}-\theta)^2=D\hat{\theta}$ (є дисперсією оцінки $\hat{\theta}$).

Ефективною оцінкою (або оцінкою мінімальної дисперсії) параметра θ називається така статистична оцінка $\hat{\theta}_e$ із класу незміщених оцінок, для якої $\inf\limits_{\hat{\theta}:M\hat{\theta}=\theta} D\hat{\theta} = D\hat{\theta}_e$.

Якщо $M\hat{\theta}_n \neq \theta$ але $M\hat{\theta}_n \to \theta$ при $n \to \infty$, то оцінку $\hat{\theta}_n$ називають *асимптотично незміщеною* оцінкою параметру θ.

Якщо $\hat{\theta}_n \to \theta$ з ймовірністю 1, то $\hat{\theta}_n$ називають *сильно спроможною* оцінкою.

Приклад. Нехай досліджувана випадкова величина ξ має рівномірний розподіл на $[a,b]$. Розглянемо оцінку $\hat{\theta} = \min\{X_1,...,X_n\}$. Чи буде $\hat{\theta}$ незміщеною, спроможною оцінкою для деякого із параметрів a або b досліджуваної випадкової величини ξ. Знайти розподіл цієї оцінки.

Розв'язування. Щільність рівномірного розподілу на $[a,b]$ має вигляд

$p(x) = \begin{cases} \dfrac{1}{b-a}, x \in (a,b), \\ 0, \quad x \notin (a,b), \end{cases}$ а функція розподілу $F(x) = \begin{cases} 0, \quad x \leq a, \\ \dfrac{x-a}{b-a}, a < x \leq b, \\ 1, \quad x > b. \end{cases}$

Знайдемо функцію розподілу $\hat{\theta}$

$$F_{\hat{\theta}}(x) = P\{\min(X_1,...,X_n) < x\} = 1 - P\{\min(X_1,...,X_n) \geq x\} = 1 - P\left\{\bigcap_{k=1}^{n}(X_k \geq x)\right\} =$$

$$= 1 - \prod_{k=1}^{n} P\{X_k \geq x\} = 1 - \prod_{k=1}^{n}\left(1 - P\{X_k < x\}\right) = 1 - \left(1 - F(x)\right)^n.$$

Тобто,

$$F_{\hat{\theta}}(x) = 1 - \left(1 - F(x)\right)^n.$$

Тому $p_{\hat{\theta}}(x) = F'_{\hat{\theta}}(x) = n\left(1 - F(x)\right)^{n-1} F'(x) = \begin{cases} 0, \quad x \notin (a,b), \\ n\dfrac{(b-x)^{n-1}}{(b-a)^n}, x \in (a,b). \end{cases}$ Знайдемо тепер

$$M\hat{\theta} = \int_{-\infty}^{+\infty} x p_{\hat{\theta}}(x)\,dx = \int_a^b nx\frac{(b-x)^{n-1}}{(b-a)^n}\,dx = \frac{n}{(b-a)^n}\int_a^b x(b-x)^{n-1}\,dx =$$

$$= \frac{1}{(b-a)^n}\left(-x(b-x)^n\Big|_a^b + \int_a^b (b-x)^n\,dx\right) = a + \frac{b-a}{n+1}.$$

Отже $M\hat{\theta} \neq a$, тому $\hat{\theta}$ є зміщеною оцінкою параметра a. Але $M\hat{\theta} \to a$ при $n \to \infty$, тому $\hat{\theta}$ є асимптотично незміщеною оцінкою параметра a.

Оцінка $\hat{\theta}$ буде спроможною оцінкою параметра $\theta = a$, якщо при $n \to \infty$ $\hat{\theta} \xrightarrow{p} \theta$ або $\forall \varepsilon > 0$ $\lim\limits_{n\to\infty} P\left\{\left|\hat{\theta} - \theta\right| < \varepsilon\right\} = 1$. Нехай $\varepsilon < b - a$, тоді

$$P\left\{\left|\hat{\theta} - a\right| < \varepsilon\right\} = P\left\{a - \varepsilon < \hat{\theta} < a + \varepsilon\right\} = F_{\hat{\theta}}(a+\varepsilon) - F_{\hat{\theta}}(a-\varepsilon).$$

Оскільки

$$F_{\hat{\theta}}(a+\varepsilon) = 1 - \left(1 - F(a+\varepsilon)\right)^n = 1 - \left(1 - \frac{b-a-\varepsilon}{b-a}\right)^n = 1 - \left(\frac{\varepsilon}{b-a}\right)^n,$$

а

$$F_{\hat{\theta}}(a-\varepsilon) = \left(1 - \left(1 - F(a-\varepsilon)\right)^n\right) = 0,$$

бо $F(a-\varepsilon)=0$, то

$$P\left\{\left|\hat{\theta}-a\right|<\varepsilon\right\}=1-\left(\frac{\varepsilon}{b-a}\right)^n \xrightarrow[n\to\infty]{} 1.$$

Отже, $\hat{\theta}$ є спроможною оцінкою параметра a рівномірного розподілу.

Вправи.

5. Нехай $X_1,...,X_n$ – вибірка із розподілу зі щільністю $p(x,\alpha)=\begin{cases} 0, & x\le 0, \\ \alpha^{-1}e^{-x/\alpha}, & x>0, \end{cases}$ $(\alpha>0)$. Довести, що $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ є незміщеною і спроможною оцінкою параметра α.
6. Нехай досліджувана випадкова величина ξ має рівномірний розподіл на $[a,b]$. Показати, що оцінка $\hat{\theta}=\max\{X_1,...,X_n\}$ буде асимптотично незміщеною та спроможною оцінкою для параметра b. Знайти розподіл цієї оцінки.

9.2.2. Статистичні оцінки для математичного сподівання, дисперсії, моментів. Асимптотична нормальність емпіричних моментів. Нехай досліджувана величина ξ має скінченне математичне сподівання $M\xi$ і скінченну дисперсію $D\xi$. Тоді *вибіркова середня є незміщеною і спроможною оцінкою математичного сподівання*.

Нехай невідомий параметр $\theta=M\xi$ і нехай вибіркова середня є статистичною оцінкою цього параметра: $\hat{\theta}=\overline{X}=\frac{X_1+\cdots+X_n}{n}$. Знайдемо її математичне сподівання

$$M\hat{\theta}=M\overline{X}=M\left(\frac{X_1+\cdots+X_n}{n}\right)=\frac{1}{n}\sum_{i=1}^{n}MX_i=M\xi=\theta.$$

Отже, $M\hat{\theta}=\theta$, тому $\hat{\theta}$ є незміщеною оцінкою параметра $\theta=M\xi$. Із закону великих чисел випливає, що при $n\to\infty$ $\hat{\theta}\to\theta$ з ймовірністю 1. Тому оцінка $\hat{\theta}$ є сильно спроможною.

Покажемо, що *вибіркова дисперсія є зміщеною оцінкою дисперсії досліджуваної випадкової величини*.

Покладемо $\theta=D\xi$, $\hat{\theta}=D_B=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$. Необхідно довести, що $M\hat{\theta}=MD_B\ne D\xi=\theta$. Врахуємо, що $MX_i=M\xi$, $M\left(\overline{X}\right)=M\xi$, тоді $M\left(X_i-\overline{X}\right)=0$ і $M\left(X_i-\overline{X}\right)^2=D\left(X_i-\overline{X}\right)$. Тому

$$MD_B=M\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)=\frac{1}{n}\sum_{i=1}^{n}M\left(X_i-\overline{X}\right)^2=\frac{1}{n}\sum_{i=1}^{n}D\left(X_i-\overline{X}\right).$$

Але

$$D\left(X_i-\overline{X}\right)=D\left(X_i-\frac{X_1+\cdots+X_n}{n}\right)=D\left(\frac{n-1}{n}X_i-\frac{1}{n}\sum_{k\ne i}X_k\right)=$$

$$=\left(\frac{n-1}{n}\right)^2 D(X_i)+\frac{1}{n^2}\sum_{k\ne i}D(X_k)=D(\xi)\frac{(n-1)^2}{n^2}+\frac{1}{n^2}(n-1)D(\xi)=\frac{n-1}{n}D(\xi)$$

є однаковими для всіх *i*, тому

$$MD_B=\frac{1}{n}\sum_{i=1}^{n}D\left(X_i-\overline{X}\right)=\frac{n-1}{n}D(\xi).$$

Отже, $M\hat{\theta} = MD_B \neq D\xi = \theta$. Але $MD_B \to D\xi$ при $n \to \infty$.

Розглянемо нову характеристику

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2,$$

яку називають *виправленою дисперсією*. Тоді $s^2 = \frac{n}{n-1}D_B$ і $Ms^2 = \frac{n}{n-1}MD_B = \frac{n}{n-1}\frac{n-1}{n}D(\xi) = D(\xi)$. Отже s^2 є незміщеною оцінкою дисперсії досліджуваної величини.

Розглянемо один приклад знаходження оцінок за допомогою емпіричної функції розподілу. Нехай досліджувана величина ξ має функцію розподілу $F(x)$ і існує $M\left(g(\xi)\right)^2$. Покладемо

$$G = M\left(g(\xi)\right) = \int_{-\infty}^{+\infty} g(x)\,dF(x).$$

Для знаходження оцінки параметра G замінимо теоретичну функцію розподілу $F(x)$ на емпіричну функцію розподілу $F_n(x)$. Тоді $\int_{-\infty}^{+\infty} g(x)\,dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$. Розглянемо статистичну оцінку

$$\hat{G}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i). \tag{5}$$

Оцінка $\hat{G}_n$ є незміщеною оцінкою параметра G, бо

$$M\hat{G}_n = \frac{1}{n}\sum_{i=1}^{n} Mg(X_i) = Mg(\xi) = G.$$

За посиленим законом великих чисел при $n \to \infty$

$$\hat{G}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \to G \tag{6}$$

з ймовірністю 1, тому оцінка $\hat{G}_n$ є сильно спроможною оцінкою величини G.

Випадкові величини $g(X_i)$ незалежні однаково розподілені і мають скінченну дисперсію $Dg(X_i) = Dg(\xi) = \sigma_g^2$, тому із центральної граничної теореми для однаково розподілених випадкових величин випливає, що величина

$$\eta_n = \frac{\sqrt{n}}{\sigma_g}\left(\hat{G}_n - G\right) = \frac{\sqrt{n}}{\sigma_g}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i) - G\right) = \frac{1}{\sigma_g\sqrt{n}}\sum_{i=1}^{n}\left(g(X_i) - G\right) \tag{7}$$

асимптотично нормальна (0,1), а величина $\hat{G}_n$ асимптотично нормальна з параметрами G і $\frac{\sigma_g^2}{n}$.

Випадкова величина η називається асимптотично нормальною з параметрами a і b, якщо $\frac{\eta - a}{\sqrt{b}}$ має функцію розподілу, яка збігається до функції розподілу стандартного нормального закону.

Нехай для досліджуваної випадкової величини ξ існує $M\left(|\xi|^{2k}\right) < \infty$. Якщо в (5) покладемо $g(x) = x^r$, то

$$\hat{G}_n = \hat{\alpha}_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$$

є емпіричним моментом r – го порядку (r>0). Тому *емпіричні моменти є незміщеними і спроможними оцінками відповідних теоретичних моментів*. Оскільки $D(\hat{\alpha}_r) = \frac{1}{n^2}\sum_i D(X_i^r) = \frac{1}{n}\left(M(\xi^{2r}) - (M(\xi^r))^2\right) = \frac{1}{n}(\alpha_{2r} - \alpha_r^2)$, де $\alpha_r = M(\xi^r)$ – теоретичний момент r–го порядку, то *емпіричні моменти є асимптотично нормальними* $\left(\alpha_r, \frac{1}{n}(\alpha_{2r} - \alpha_r^2)\right)$.

Вправа.

7. Показати, що центральні емпіричні моменти $\hat{\mu}_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^r$ є спроможними оцінками відповідних теоретичних моментів.

9.2.3. Ефективні оцінки. Нерівність Крамера – Рао. Ми уже раніше визначили поняття ефективної оцінки. Нагадаємо його. *Ефективною оцінкою* (або оцінкою мінімальної дисперсії) параметра θ називається така статистична оцінка $\hat{\theta}_e$ із класу незміщених оцінок, для якої $\inf\limits_{\hat{\theta}: M\hat{\theta}=\theta} D\hat{\theta} = D\hat{\theta}_e$. Але ефективна оцінка не завжди існує. В даному пункті ми розглянемо умови, при виконанні яких для сім'ї функцій розподілу $\{F(x,\theta), \theta \in \Theta\}$ можна вказати нижню межу дисперсій всіх незміщених оцінок параметра θ. В деяких випадках існують оцінки, на яких нижня межа досягається, тоді вони є ефективними. Порівняння дисперсії даної незміщеної оцінки $\hat{\theta}$ з нижньою межою дисперсії незміщених оцінок $\inf\limits_{\hat{\theta}: M\hat{\theta}=\theta} D\hat{\theta} = D\hat{\theta}_e$ дозволяє робити висновок про те, наскільки дана оцінка близька до ефективної.

Розглянемо вибірку (вибірковий вектор) $X = (X_1, ..., X_n)$ фіксованого обсягу n. Нехай θ - одновимірний параметр, а Θ – інтервал в множині дійсних чисел. Позначимо через $f(x,\theta) = f(x_1, ..., x_n, \theta)$ – щільність розподілу X, якщо досліджувана випадкова величина ξ – неперервна. Якщо розподіл ξ – дискретний, то $f(x,\theta) = P_\theta\{X = x\}$, при цьому, $f(x,\theta) > 0$ тільки для скінченної або зліченної множини точок $x^{(k)} = (x_1^{(k)}, ..., x_n^{(k)})$. Нехай $p(x,\theta)$ – щільність розподілу досліджуваної випадкової величини ξ, якщо ξ – неперервна, і $p(x,\theta) = P_\theta\{\xi = x\}$, якщо розподіл ξ – дискретний. Оскільки результати спостережень незалежні, то $f(x,\theta) = \prod_{k=1}^{n} p(x_k, \theta)$. Доведення результатів будемо проводити тільки для неперервного випадку, вони легко переносяться і на дискретні розподіли.

Надалі ми будемо розглядати функцію $f(X,\theta)$, що залежить від випадкових вибіркових значень та оцінюваного параметра θ. Цю функцію позначають також через $L(X,\theta)$ і називають функцією вірогідності (правдоподібності). Для одержання нерівності Крамера – Рао на функцію $f(X,\theta)$ будемо накладати деякі обмеження.

Лема. Нехай для будь-якого $\theta \in \Theta$ існують похідні $\frac{\partial}{\partial\theta} f(x,\theta)$, $\frac{\partial^2}{\partial\theta^2} f(x,\theta)$ і виконуються умови

$$M\left|\frac{\partial \ln f(X,\theta)}{\partial\theta}\right|<\infty,\quad M\left|\frac{\partial^2 \ln f(X,\theta)}{\partial\theta^2}\right|<\infty,\quad M\left|\frac{\partial \ln f(X,\theta)}{\partial\theta}\right|^2<\infty.$$

Тоді для всіх $\theta\in\Theta$

$$M\frac{\partial \ln f(X,\theta)}{\partial\theta}=0;\ M\left|\frac{\partial \ln f(X,\theta)}{\partial\theta}\right|^2=-M\frac{\partial^2 \ln f(X,\theta)}{\partial\theta^2}. \tag{8}$$

Доведення. Оскільки $f(x,\theta)$ є щільністю розподілу в R^n, то

$$\int\limits_{R^n} f(x,\theta)\,dx=1.$$

Із умов леми випливає, що можна диференціювати під знаком інтеграла по параметру θ. Продиференціювавши попередню рівність по параметру θ, одержимо $\int\limits_{R^n}\frac{\partial f(x,\theta)}{\partial\theta}dx=0$.

Після перетворення підінтегрального виразу, одержуємо першу із рівностей (8):

$$0=\int\limits_{R^n} f'_\theta(x,\theta)\,dx=\int\limits_{R^n}\frac{f'_\theta(x,\theta)}{f(x,\theta)}f(x,\theta)\,dx=\int\limits_{R^n}\frac{\partial}{\partial\theta}\ln f(x,\theta)f(x,\theta)\,dx=M\frac{\partial}{\partial\theta}\ln f(X,\theta).$$

Умови леми дають можливість знову диференціювати рівність $\int\limits_{R^n}\frac{\partial \ln f(x,\theta)}{\partial\theta}f(x,\theta)\,dx=0$ по параметру θ. Тоді одержимо

$$\int\limits_{R^n}\left(\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}f(x,\theta)+\frac{\partial \ln f(x,\theta)}{\partial\theta}\frac{\partial f(x,\theta)}{\partial\theta}\right)dx=0.$$

Другий доданок у останній рівності помножимо і поділимо на $f(x,\theta)$, тоді її можна подати у вигляді

$$\int\limits_{R^n}\frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}f(x,\theta)\,dx+\int\limits_{R^n}\left(\frac{\partial \ln f(x,\theta)}{\partial\theta}\right)^2 f(x,\theta)\,dx=0,$$

або

$$M\left(\frac{\partial \ln f(X,\theta)}{\partial\theta}\right)^2+M\frac{\partial^2 \ln f(X,\theta)}{\partial\theta^2}=0.$$

Звідки випливає друга рівність із (8). Лема доведена.

Функцію

$$I_n(\theta)=M\left(\frac{\partial \ln f(X,\theta)}{\partial\theta}\right)^2$$

називають інформацією Фішера, що міститься у вибірці обсягу n, про значення параметра θ. Із леми випливає, що інформацію Фішера можна подати у вигляді

$$I_n(\theta)=-M\left(\frac{\partial^2 \ln f(X,\theta)}{\partial\theta^2}\right).$$

Якщо $f(X,\theta)=\prod_{i=1}^{n}p(X_i,\theta)$, то $\frac{\partial}{\partial\theta}\ln f(X,\theta)=\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln p(X_i,\theta)$, $\frac{\partial^2}{\partial\theta^2}\ln f(X,\theta)=\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln p(X_i,\theta)$. Тому

$$I_n(\theta) = -M\left(\frac{\partial^2 \ln f(X,\theta)}{\partial\theta^2}\right) = -\sum_{i=1}^{n} M\frac{\partial^2}{\partial\theta^2}\ln p(X_i,\theta) = nI(\theta),$$

тобто,

$$I_n(\theta) = nI(\theta),$$

де

$$I(\theta) = M\left(\frac{\partial \ln p(X_i,\theta)}{\partial\theta}\right)^2 = -M\left(\frac{\partial^2 \ln p(X_i,\theta)}{\partial\theta^2}\right).$$

Теорема (Крамера – Рао). Нехай виконуються умови леми і $h(\theta)$ – диференційовна функція на Θ, для якої існує незміщена оцінка $\hat{h}(X)$ із скінченною дисперсією, що задовольняє умові

$$\int_{R^n}\left|\hat{h}(x)\frac{\partial}{\partial\theta}f(x,\theta)\right|dx < \infty,\ \forall\theta\in\Theta.$$

Тоді

$$M\left(\hat{h}(X) - h(\theta)\right)^2 \geq \frac{\left(h'(\theta)\right)^2}{I_n(\theta)}, \tag{9}$$

при цьому знак рівності має місце тоді і тільки тоді, коли

$$\frac{\partial \ln f(x,\theta)}{\partial\theta} = c(\theta)\left(\hat{h}(x) - h(\theta)\right). \tag{10}$$

Доведення. За умовою $\hat{h}(X)$ є незміщеною оцінкою $h(\theta)$: $M\hat{h}(X) = h(\theta)$. А це в свою чергу означає, що

$$\int_{R^n}\hat{h}(x)f(x,\theta)dx = h(\theta).$$

Із умов теореми випливає, що даний інтеграл можна продиференціювати по параметру θ. Тому

$$h'(\theta) = \int_{R^n}\hat{h}(x)\frac{\partial f(x,\theta)}{\partial\theta}dx = \int_{R^n}\hat{h}(x)\frac{\partial \ln f(x,\theta)}{\partial\theta}f(x,\theta)dx = M\left(\hat{h}(X)\frac{\partial}{\partial\theta}\ln f(X,\theta)\right).$$

Помножимо першу рівність із (8) на $h(\theta)$ і віднімемо від попередньої. Тоді

$$M\left(\hat{h}(X) - h(\theta)\right)\frac{\partial \ln f(X,\theta)}{\partial\theta} = h'(\theta).$$

Використовуючи нерівність Коші – Буняковського, одержимо

$$\left(h'(\theta)\right)^2 \leq M\left(\hat{h}(X) - h(\theta)\right)^2 M\left(\frac{\partial \ln f(X,\theta)}{\partial\theta}\right)^2 = M\left(\hat{h}(X) - h(\theta)\right)^2 I_n(\theta),$$

де знак рівності можливий тоді і тільки тоді, коли з ймовірністю 1 величини $\hat{h}(X) - h(\theta)$ і $\frac{\partial}{\partial\theta}\ln f(X,\theta)$ зв'язані лінійною функціональною залежністю. Враховуючи першу рівність із (8) і рівність $M\hat{h}(X) = h(\theta)$ бачимо, що останнє твердження еквівалентне (10).

Теорема доведена.

Нерівність (9) називають *нерівністю Крамера – Рао*. Якщо $h(\theta) = \theta$, то нерівність Крамера – Рао набуває вигляду

$$M\left(\hat{h}(X)-\theta\right)^2 \ge \frac{1}{I_n(\theta)}.$$

Нерівність Крамера – Рао встановлює нижню межу для дисперсій всіх незміщених оцінок. Якщо для деякої оцінки $\hat{h}(X)$ нерівність Крамера – Рао перетворюється в рівність

$$D\left(\hat{h}(X)\right)=\frac{\left(h'(\theta)\right)^2}{I_n(\theta)},$$

то оцінка $\hat{h}(X)$ буде ефективною оцінкою $h(\theta)$.

Із теореми випливає, що рівність (10) є необхідною і достатньою умовою того, що оцінка $\hat{h}(X)$ є ефективною оцінкою $h(\theta)$. Якщо знайдемо математичне сподівання обох частин рівності (10) і використаємо лему, то одержимо $M\left(\hat{h}(X)-h(\theta)\right)=0$. Це означає, що оцінка $\hat{h}(X)$ є незміщеною оцінкою $h(\theta)$. Якщо ж рівність (10) піднесемо до квадрату і знайдемо математичне сподівання обох частин, то одержимо

$$I_n(\theta)=c^2(\theta)D\left(\hat{h}(X)\right),$$

Тому для ефективної оцінки $\dfrac{I_n(\theta)}{c^2(\theta)}=\dfrac{\left(h'(\theta)\right)^2}{I_n(\theta)}$, звідки

$$c(\theta)=\frac{I_n(\theta)}{h'(\theta)}.$$

Якщо $\hat{h}(X)$ є ефективною оцінкою параметра θ, то $D\left(\hat{h}(X)\right)=\dfrac{1}{nI(\theta)}=O\left(\dfrac{1}{n}\right)$. Тобто ефективна оцінка завжди спроможна і незміщена.

На основі нерівності Крамера – Рао можна занаходити нижню межу і для зміщених оцінок. Нехай $\hat{h}(X)$ є зміщеною оцінкою невідомого параметра θ, тоді $M\hat{h}(X)-\theta=b(\theta)$, де $b(\theta)$ – зміщення оцінки. Оскільки $M\left(\hat{h}(X)\right)=\theta+b(\theta)$, то $\hat{h}(X)$ є незміщеною оцінкою параметра $\theta+b(\theta)=h(\theta)$. Тоді (при врахуванні рівності $M\left(\hat{h}(X)-\theta-b(\theta)\right)=0$)

$$M\left(\hat{h}(X)-\theta\right)^2=M\left(\hat{h}(X)-\theta-b(\theta)+b(\theta)\right)^2=M\left(\hat{h}(X)-\theta-b(\theta)\right)^2+$$

$$+2b(\theta)M\left(\hat{h}(X)-\theta-b(\theta)\right)+b^2(\theta)=b^2(\theta)+M\left(\hat{h}(X)-\theta-b(\theta)\right)^2\ge b^2(\theta)+\frac{\left(1+b'(\theta)\right)^2}{I_n(\theta)}.$$

Тобто,

$$M\left(\hat{h}(X)-\theta\right)^2\ge b^2(\theta)+\frac{\left(1+b'(\theta)\right)^2}{I_n(\theta)}.$$

Для порівняння якості двох оцінок можна ввести поняття відносної ефективності двох оцінок. Нехай $\hat{\theta}_1$ і $\hat{\theta}_2$ – незміщенені оцінки параметра θ, тоді $\dfrac{M\left(\hat{\theta}_1-\theta\right)^2}{M\left(\hat{\theta}_2-\theta\right)^2}=eff\left(\hat{\theta}_1,\hat{\theta}_2\right)$

називають *відносною ефективністю* оцінок. Якщо ця величина менша за 1, то вважають, що оцінка $\hat{\theta}_1$ краща, ніж оцінка $\hat{\theta}_2$.

Послідовність незміщених оцінок $\hat{\theta}_n$ параметра θ, що побудовані за вибіркою обсягу n, називається *асимптотично ефективною*, якщо відношення дисперсії оцінки до нижньої межі Крамера – Рао дисперсій прямує до 1:

$$I_n(\theta)D\left(\hat{\theta}_n\right)\xrightarrow[n\to\infty]{}1.$$

Приклад 1. Задано вибірку $x_1,...,x_n$ із нормального розподілу. Довести, що вибіркова середня є ефективною оцінкою математичного сподівання нормально розподіленої випадкової величини. Знайти ефективну оцінку дисперсії нормального закону при відомому математичному сподіванні.

Розв'язування. Невідомим параметром є $M\xi = a = \theta$ ($h(\theta)=\theta$). Розглянемо статистичну оцінку $\hat{\theta} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Ми розглянемо два підходи. Перший оснований на нерівності Крамера – Рао $D\left(\hat{h}(X)\right) \geq \frac{1}{I_n(\theta)}$. Для цього знайдемо спочатку дисперсію оцінки $D\left(\hat{\theta}\right) = D\left(\overline{X}\right) = D\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{D\xi}{n} = \frac{\sigma^2}{n}$. А тепер знайдемо інформацію Фішера, для цього вигідніше використовувати таку формулу $I_n(\theta) = -M\frac{\partial^2 \ln f(X,\theta)}{\partial\theta^2}$, бо друга похідна частіше має простіший вигляд і легше знаходити від неї математичне сподівання. Щільність нормального розподілу має вигляд

$$p(x,\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right).$$

Тоді

$$f(x,\theta) = \prod_{i=1}^{n} p(x_i,\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right),$$

а

$$\ln f(X,\theta) = -n\ln\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\theta)^2.$$

Продиференціюємо двічі цю рівність по θ: $\frac{\partial \ln f(X,\theta)}{\partial\theta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\theta)$, $\frac{\partial \ln^2 f(X,\theta)}{\partial\theta^2} = -\frac{n}{\sigma^2}$. Звідки, $I_n(\theta) = -M\frac{\partial^2 \ln f(X,\theta)}{\partial\theta^2} = \frac{n}{\sigma^2}$. Отже,

$$\frac{\sigma^2}{n} = D\left(\overline{X}\right) = \frac{1}{I_n(\theta)} = \frac{\sigma^2}{n}.$$

Тобто, нерівність Крамера – Рао перетворилась у рівність, а це означає, що вибіркова середня є ефективною оцінкою математичного сподівання нормального розподілу.

Другий підхід базується на рівності (10). Першій похідній надамо вигляду

$$\frac{\partial \ln f(X,\theta)}{\partial\theta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\theta) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}X_i - n\theta\right) = \frac{n}{\sigma^2}\left(\frac{1}{n}\sum_{i=1}^{n}X_i - \theta\right).$$

Із рівності (10) випливає, що $\hat{\theta} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n}X_i$ є ефективною оцінкою параметра $\theta = a = M\xi$ нормального розподілу.

Нехай невідомим є $\theta = \sigma^2$, а a – відоме. Тоді

$$f(X,a,\theta)=\prod_{k=1}^{n}\frac{1}{\sqrt{2\pi\theta}}\exp\left(-\frac{(X_k-a)^2}{2\theta}\right)=\left(\frac{1}{\sqrt{2\pi\theta}}\right)^n\exp\left(-\sum_{k=1}^{n}\frac{(X_k-a)^2}{2\theta}\right),$$

$$\ln f(X,a,\theta)=-n\ln\sqrt{2\pi}-\frac{n}{2}\ln\theta-\frac{1}{2\theta}\sum_{k=1}^{n}(X_k-a)^2,$$

$$\frac{\partial\ln f(X,a,\theta)}{\partial\theta}=-\frac{n}{2\theta}+\frac{1}{2\theta^2}\sum_{i=1}^{n}(X_k-a)^2=\frac{n}{2\theta^2}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k-a)^2-\theta\right).$$

Звідки, $c(\theta)=\frac{n}{2\theta^2}$, $\hat{h}(X)=\frac{1}{n}\sum_{i=1}^{n}(X_i-a)^2$. Умова (10) існування ефективної оцінки виконується і оцінка $\hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{n}(X_k-a)^2$ є ефективною оцінкою дисперсії σ^2 нормального закону.

Приклад 2. Розглянемо біномний розподіл. За вибіркою $x_1,...,x_n$, де $x_i=0$ або 1, оцінити ймовірність $p=\theta$ настання події в кожному експерименті.

Розв'язування. Оскільки,

$$p(x,\theta)=P\{\xi=x\}=\theta^x(1-\theta)^{1-x},$$

то

$$f(X,\theta)=\prod_{k=1}^{n}\theta^{X_k}(1-\theta)^{1-X_k}=\theta^{\sum_{k=1}^{n}X_k}(1-\theta)^{n-\sum_{k=1}^{n}X_k},$$

$$\ln f(X,\theta)=\sum_{k=1}^{n}X_k\ln\theta+\left(n-\sum_{k=1}^{n}X_k\right)\ln(1-\theta),$$

$$\frac{\partial\ln f(X,\theta)}{\partial\theta}=\sum_{k=1}^{n}X_k\frac{1}{\theta}-\left(n-\sum_{k=1}^{n}X_k\right)\frac{1}{1-\theta}=\frac{n}{\theta(1-\theta)}\left(\frac{1}{n}\sum_{k=1}^{n}X_k-\theta\right).$$

Звідки $\hat{\theta}=\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$. Оскільки $\sum_{k=1}^{n}X_k=\mu_n$ є число появ події в n експериментах, то $\hat{\theta}=\frac{\mu_n}{n}$ є відносною частотою події в n експериментах. Отже, з теореми Крамера-Рао випливає, що відносна частота є ефективною оцінкою ймовірності появи події в схемі Бернуллі.

Вправи.

8. Нехай $x_1,...,x_n$ – вибірка із розподілу Пуассона з параметром λ. Показати, що вибіркова середня є ефективною оцінкою параметра λ.
9. Нехай $x_1,...,x_n$ – вибірка із показникового розподілу із щільністю $p(x,\theta)=\begin{cases}\frac{1}{\theta}e^{-\frac{x-m}{\theta}}, & x\ge m,\\ 0, & x<m,\end{cases}$ $\theta>0$, m - відоме. Чи є $\hat{\theta}=\bar{X}-m$ ефективною оцінкою параметра θ.

9.2.4. Достатні статистики. Одним із методів побудови ефективних оцінок є використання достатніх статистик. Нехай $X=(X_1,...,X_n)$ – випадковий вектор із значеннями в R^n, щільність розподілу якого $f(x,\theta)$ ($f(x,\theta)=P(X=x)$ є розподілом вектора X, якщо розподіл X – дискретний) залежить від невідомого параметра θ, $\theta\in\Theta\subset R^s$.

Статистика $T = T(X)$ (може бути і векторна) називається *достатньою* для параметра θ, якщо умовна щільність (або ймовірність в дискретному випадку) $L(x/t;\theta)$ випадкового вектора X, при умові $T(X) = t$, не залежить від параметра θ.

Звідси можна зробити висновок, що статистика T містить в собі всю інформацію про параметр θ, що міститься у векторі X, всі висновки формулюються тільки в термінах цієї статистики. Прикладом достатньої статистики є вся вибірка. Але важливо знайти статистику мінімальної розмірності.

Розглянемо приклад достатньої статистики. Нехай X_i - число успіхів у i – му випробуванні Бернуллі, $\mu_n = X_1 + ... + X_n$ – число успіхів у n випробуваннях. Статистика μ_n є достатньою для параметра $\theta = p$ – ймовірності успіху в кожному випробуванні. Дійсно,

$$P(X_1 = x_1, ..., X_n = x_n / \mu_n = t) = \frac{P(X_1 = x_1, ..., X_n = x_n, \mu_n = t)}{P(\mu_n = t)}.$$

Якщо $x_1 + ... + x_n = t$, то

$$P(X_1 = x_1, ..., X_n = x_n, \mu_n = t) = P(X_1 = x_1, ..., X_n = x_n) = \theta^t (1-\theta)^{n-t},$$

а

$$P(\mu_n = t) = C_n^t \theta^t (1-\theta)^{n-t}.$$

Тому

$$P(X_1 = x_1, ..., X_n = x_n / \mu_n = t) = \frac{\theta^t (1-\theta)^{n-t}}{C_n^t \theta^t (1-\theta)^{n-t}} = \frac{1}{C_n^t}.$$

Якщо $x_1 + ... + x_n \neq t$, то $P(X_1 = x_1, ..., X_n = x_n, \mu_n = t) = 0$, тому і

$$P(X_1 = x_1, ..., X_n = x_n / \mu_n = t) = 0.$$

Отже, умовний розподіл $P(X_1 = x_1, ..., X_n = x_n / \mu_n = t)$ не залежить від параметра θ, тому статистика μ_n є достатньою для цього параметра.

Доведемо тепер теорему (критерій факторизації), яка дає можливість встановити існування достатньої статистики і одночасно знайти її вигляд.

Теорема (Неймана – Фішера). Статистика $T(X)$ буде достатньою для параметра θ тоді і тільки тоді, коли функція вірогідності $L(x,\theta)$ має вигляд

$$L(x,\theta) = g(T(x);\theta)h(x), \tag{11}$$

де функція g може залежати від θ, а від x залежить тільки через $T(x)$, функія h від параметра θ не залежить.

Доведення проведемо для дискретного випадку. Нехай статистика T достатня, тоді для довільного t із множини значень $T(x)$ функція $L(x/t;\theta)$ не залежить від θ і її можна записати у вигляді $h(x,t)$. Якщо $T(x) = t$, тоді $\{X = x\} \subset \{T(X) = t\}$ і

$$L(x,\theta) = P(X = x) = P(X = x, T(X) = t) = P(T(X) = t)P(X = x / T(X) = t) =$$
$$= g(t,\theta)L(x/t;\theta) = g(T(x),\theta)h(x,t),$$

тобто має місце рівність (11).

Навпаки, нехай має місце рівність (11). Тоді

$$P(T(X) = t) = \sum_{y:T(y)=t} L(y,\theta) = \sum_{y:T(y)=t} g(T(y),\theta)h(y) = g(t,\theta) \sum_{y:T(y)=t} h(y)$$

і для будь-якого x, для якого $T(x) = t$,

$$L(x/t;\theta)=P_\theta\left(X=x/T(X)=t\right)=\frac{P_\theta\left(X=x,T(X)=t\right)}{P_\theta\left(T(X)=t\right)}=\frac{L(x,\theta)}{P_\theta\left(T(X)=t\right)}=$$

$$=\frac{g(t,\theta)h(x)}{g(t,\theta)\sum\limits_{y;T(y)=t}h(y)}=\frac{h(x)}{\sum\limits_{y:T(y)=t}h(y)},$$

тобто не залежить від θ. Якщо ж *x* таке, що $T(x)\neq t$, то $L(x/t;\theta)=0$ і знову не залежить від θ.

Якщо параметр θ одновимірний, то із теореми Крамера – Рао випливає, що будь-яка ефективна оцінка є достатньою статистикою. Оскільки із рівності (10)

$$\frac{\partial \ln f(x,\theta)}{\partial\theta}=c(\theta)\left(\hat{h}(x)-\theta\right)$$

одержуємо $L(x,\theta)=f(x,\theta)=\exp\left\{\varphi(\theta)\hat{h}(x)+\psi(\theta)\right\}K(x)$, а за теоремою це еквівалентно тому, що $\hat{h}(x)$ є достатньою статистикою для параметра θ.

Наведемо без доведення і більш загальний результат.

Якщо для параметра θ існує ефективна оцінка, то вона є функцією від достатньої статистики.

Вправа.

10. Нехай $(X_1,...,X_n)$ – вибірка із нормального закону $N(a,\sigma^2)$. Побудувати достатні статистики для параметра $\theta=(\theta_1,\theta_2)$, де $\theta_1=a$, $\theta_2=\sigma^2$.

9.2.5. Методи одержання статистичних оцінок. Ми розглянемо два методи одержання статистичних оцінок: метод моментів і метод максимальної вірогідності.

Метод моментів одержання статистичних оцінок базується на тому факті, що емпіричні моменти є незміщеними і спроможними оцінками відповідних теоретичних моментів.

Нехай розподіл $F(x,\theta)$ залежить від векторного параметра $\theta=(\theta_1,...,\theta_s)$. Тоді на основі розподілу $F(x,\theta)$ можна знайти, за умови їх існування, теоретичні моменти $M\xi,...,M\xi^s$. Ці теоретичні моменти будуть залежати від невідомих параметрів, тобто $M\xi^k=\alpha_k(\theta_1,...,\theta_s)$. За результатами спостережень $X_1,...,X_n$ можна знайти відповідні емпіричні моменти $\hat{\alpha}_k$: $\hat{\alpha}_k=\overline{X^k}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i^k$. Крім того $\hat{\alpha}_k$ є незміщеними оцінками відповідних теоретичних моментів α_k і $\hat{\alpha}_k\xrightarrow[n\to\infty]{p}\alpha_k$, тобто при великих n має місце наближена рівність $\alpha_k\approx\hat{\alpha}_k$. Тому *оцінками за методом моментів* для невідомих параметрів $\theta_1,...,\theta_s$ називають розв'язок системи рівнянь

$$\begin{cases}\alpha_1(\theta_1,...,\theta_s)=\overline{X},\\ \alpha_2(\theta_1,...,\theta_s)=\overline{X^2},\\ ....\\ \alpha_s(\theta_1,...,\theta_s)=\overline{X^s}.\end{cases}$$

Якщо теоретичні моменти є неперервними функціями параметрів, то оцінки за методом моментів будуть спроможними.

Приклад. Нехай $x_1,...,x_n$ – вибірка із рівномірного розподілу із щільністю $p(x)=\begin{cases}\dfrac{1}{b-a}, x\in(a,b),\\ 0,\quad x\notin(a,b).\end{cases}$ За методом моментів знайти оцінки параметрів a і b.

Розв'язування. На основі щільності рівномірного розподілу ми раніше уже знайшли $M\xi=\dfrac{a+b}{2}$, $M\left(\xi^2\right)=\dfrac{b^2+ab+a^2}{3}$. Тому для знаходження оцінок одержуємо систему рівнянь: $\dfrac{a+b}{2}=\overline{X}$, $\dfrac{b^2+ab+a^2}{3}=\overline{X^2}$. Піднесемо перше рівняння до квадрату і віднімемо від другого. Тоді система набуде вигляду

$$\begin{cases}\dfrac{a+b}{2}=\overline{X},\\ \dfrac{(b-a)^2}{12}=D_B\end{cases} \text{ або } \begin{cases}a+b=2\overline{X},\\ b-a=2\sqrt{3}\sigma_B.\end{cases} \text{ Звідки, } \hat{a}=\overline{X}-\sqrt{3}\sigma_B,\ \hat{b}=\overline{X}+\sqrt{3}\sigma_B.$$

Переходимо до розгляду методу *максимальної вірогідності*. Нехай $p(x,\theta)$ – щільність розподілу випадкової величини ξ, якщо величина неперервна, і $p(x,\theta)=P_\theta\{\xi=x\}$, якщо випадкова величина ξ – дискретна. Параметр θ приймає значення із деякої множини $\Theta\subset R^s$, $\theta=(\theta_1,...,\theta_s)$. Нехай маємо вибірку $X=(X_1,...,X_n)$ обсягу n. Позначимо через $f(x,\theta)$ – щільність розподілу вибірки X, тоді $L(x,\theta)=f(x,\theta)=p(x_1,\theta)\cdot...\cdot p(x_n,\theta)$. Таку функцію називають *функцією вірогідності*.

Метод оцінки параметра θ, що одержав назву методу максимальної вірогідності, грунтується на тому, що в більшій частині випадків у експерименті спостерігається таке значення вектора X, при якому щільність $f(x,\theta)$ близька до максимального значення. Отже, за оцінку $\hat{\theta}$ невідомого параметру θ природно прийняти розв'язок рівняння $f\left(X,\hat{\theta}\right)=\max\limits_{\theta} f(X,\theta)$, якщо це рівняння має зміст і воно має розв'язки.

Розв'язки рівняння

$$f\left(X,\hat{\theta}\right)=\max_{\theta} f(X,\theta)$$

називають *оцінками за методом максимальної вірогідності*.

Якщо $\theta=(\theta_1,...,\theta_s)$, то $\hat{\theta}=\left(\hat{\theta}_1,...,\hat{\theta}_s\right)$ і $L(x,\theta)$ є функцією s змінних. Якщо функція $L(x,\theta)$ неперервно диференційовна і в деякій точці має максимум, то частинні похідні в цій точці дорівнюють 0. Але похідна добутку має складний вигляд, тому краще користуватися функцією $U(x,\theta)=\ln L(x,\theta)$. Функцію $U(x,\theta)$ називають логарифмічною функцією вірогідності. При цьому, якщо в деякій точці функція $L(x,\theta)$ має максимум, то в цій точці буде мати максимум і функція $U(x,\theta)$, і навпаки. Для знаходження оцінок за методом максимальної вірогідності, знаходимо спочатку стаціонарні точки функції $U(x,\theta)$ і вибираємо ті із них, в яких ця функція має максимум. Якщо стаціонарна точка одна, то у більшості випадків вона і буде точкою максимуму. Таким чином, одержуємо систему:

$$\begin{cases} \dfrac{\partial \ln L\left(x,\theta_1,...,\theta_s\right)}{\partial \theta_1} = 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \dfrac{\partial \ln L\left(x,\theta_1,...,\theta_s\right)}{\partial \theta_s} = 0. \end{cases}$$

Ці рівняння називають рівняннями максимальної вірогідності. У більшості випадків, розв'язки цієї системи і є оцінками максимальної вірогідності невідомих параметрів $\theta_1,...,\theta_s$. При розв'язуванні рівнянь вірогідності необхідно відкидати розв'язки вигляду $\theta = const$ і розглядати лише ті розв'язки, які залежать від $x_1,...,x_n$ і попадають в область допустимих значень параметра Θ.

Звернемо увагу на одну властивість методу максимальної вірогідності.

Якщо виконуються умови теореми Крамера – Рао і існує ефективна оцінка $\hat{h}\left(X\right)$ одновимірного параметра θ, то вона буде також оцінкою максимальної вірогідності.

Дійсно, нехай виконуються умови теореми Крамера – Рао і існує ефективна оцінка $\hat{h}\left(X\right)$ параметра θ, тоді $\dfrac{\partial \ln f\left(X,\theta\right)}{\partial \theta} = c\left(\theta\right)\left(\hat{h}\left(X\right) - \theta\right)$ і рівняння максимальної вірогідності буде мати вигляд:

$$\frac{\partial \ln f\left(X,\theta\right)}{\partial \theta} = c\left(\theta\right)\left(\hat{h}\left(X\right) - \theta\right) = 0 .$$

Отже, $\hat{h}\left(X\right)$ є оцінкою за методом максимальної вірогідності параметра θ. За теоремою Крамера – Рао $\hat{h}\left(X\right)$ є незміщеною і ефективною оцінкою параметра θ.

Приклад. За вибіркою $x_1,...,x_n$ із нормального закону знайти оцінки максимальної вірогідності параметрів $a = \theta_1$ і $\sigma^2 = \theta_2$.

Розв'язування. На основі щільності нормального закону

$$p\left(x,\theta_1,\theta_2\right) = \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{\left(x-\theta_1\right)^2}{2\theta_2} \right)$$

запишемо функцію вірогідності

$$L\left(X,\theta_1,\theta_2\right) = f\left(X,\theta_1,\theta_2\right) = \prod_{i=1}^{n} p\left(X_i,\theta_1,\theta_2\right) = \left(\frac{1}{\sqrt{2\pi\theta_2}} \right)^n \exp\left(-\frac{1}{2\theta_2} \sum_{i=1}^{n} \left(X_i - \theta_1\right)^2 \right).$$

Тоді

$$\ln L\left(X,\theta_1,\theta_2\right) = -n\ln\left(\sqrt{2\pi}\right) - \frac{n}{2}\ln\theta_2 - \frac{1}{2\theta_2}\sum_{i=1}^{n}\left(X_i - \theta_1\right)^2 .$$

Знаходимо частинні похідні по θ_1 і θ_2:

$$\frac{\partial \ln L\left(X,\theta_1,\theta_2\right)}{\partial \theta_1} = \frac{1}{\theta_2}\sum_{i=1}^{n}\left(X_i - \theta_1\right), \quad \frac{\partial \ln L\left(X,\theta_1,\theta_2\right)}{\partial \theta_2} = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2}\sum_{i=1}^{n}\left(X_i - \theta_1\right)^2 .$$

Прирівнявши похідні до нуля, записуємо систему рівнянь вірогідності:

$$\begin{cases} \dfrac{1}{\theta_2}\displaystyle\sum_{i=1}^{n}\left(X_i - \theta_1\right) = 0, \\ -\dfrac{n}{2\theta_2} + \dfrac{1}{2\theta_2^2}\displaystyle\sum_{i=1}^{n}\left(X_i - \theta_1\right)^2 = 0. \end{cases}$$

Із першого рівняння знаходимо

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}.$$

Підставивши це значення в друге рівняння, одержимо

$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = D_B.$$

Тобто оцінками максимальної вірогідності математичного сподівання і дисперсії нормального закону є відповідно вибіркова середня і вибіркова дисперсія.

Вправи.

11. Нехай $x_1,...,x_n$ – вибірка із геометричного розподілу $p(k,p)=(1-p)^k p$, $k=0,1,...$ Оцінити параметр p за методом моментів.
12. Нехай $x_1,...,x_n$ – вибірка із гамма-розподілу $p(x,\alpha,\beta)=\frac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)}x^{\alpha}e^{-x/\beta}$, ($x>0$, $\alpha>-1$, $\beta>0$). Оцінити параметри α і β за методом моментів.
13. Нехай $x_1,...,x_n$ – вибірка із гамма-розподілу $p(x,\alpha,\beta)=\frac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)}x^{\alpha}e^{-x/\beta}$, ($x>0$, $\alpha>-1$, $\beta>0$). Знайти оцінку параметра β за методом максимальної вірогідності.
14. Методом максимальної вірогідності знайти оцінку параметра λ показникового розподілу $p(x,\lambda)=\lambda e^{-\lambda x}$ ($\lambda>0$, $x>0$), якщо в результаті n експериментів випадкова величина прийняла значення $x_1,...,x_n$.
15. За вибіркою $x_1,...,x_n$ знайти методом максимальної вірогідності оцінку ймовірності успіху в схемі Бернуллі ($\theta = p$).

9.3. Інтервальні оцінки невідомих параметрів розподілу

9.3.1. Поняття довірчого інтервалу. Довірчий інтервал для невідомої ймовірності у схемі Бернуллі. Оскільки невідомий параметр θ і його статистична оцінка $\hat{\theta}$ містяться в множині $\Theta \subset R^s$, то оцінці відповідає деяка точка в R^s, тому такі оцінки називають точковими. При реалізації вибірки в кожному конкретному випадку значення оцінки може відрізнятись від оцінюваного параметра, тому необхідно знати похибку, яку ми допускаємо при заміні невідомого параметра його оцінкою. Тобто, в багатьох випадках важливо знати не тільки оцінку невідомого параметра θ, але знати і область $S \subset \Theta$, що містить в собі значення параметра θ. Така область будується за статистичними даними і змінюється від вибірки до вибірки, тому є випадковою. Отже можна говорити про ймовірність того, що дана область містить невідомий параметр. Якщо існує область S, яка з ймовірністю, що не менше γ, містить невідомий параметр θ, то область S називають *довірчою областю*, а число γ називають *надійністю* (і воно має бути близьким до одиниці). Якщо параметр θ одновимірний, то для параметра θ знаходять довірчий інтервал, який задається двома статистичними оцінками, що є кінцями інтервалу. В цьому випадку говорять про інтервальне оцінювання.

Розглянемо дві оцінки $\hat{\theta}_H$ і $\hat{\theta}_B$ невідомого параметра θ і розглянемо інтервал $\left(\hat{\theta}_H, \hat{\theta}_B\right)$, де $\hat{\theta}_H < \hat{\theta}_B$. Нехай $P\left\{\theta \in \left(\hat{\theta}_H, \hat{\theta}_B\right)\right\} = \gamma$, тобто $\gamma = P\left\{\hat{\theta}_H < \theta < \hat{\theta}_B\right\}$, тоді γ називається *надійністю* оцінки, а інтервал $\left(\hat{\theta}_H, \hat{\theta}_B\right)$ – *надійним* або *довірчим інтервалом*, який з надійністю γ оцінює невідомий параметр θ (γ-довірчим інтервалом).

Для забезпечення однозначності визначення довірчого інтервалу $\left(\hat{\theta}_H, \hat{\theta}_B\right)$, межі інтервалу $\hat{\theta}_H$ і $\hat{\theta}_B$ вибирають так, щоб виконувались умови $P\left\{\theta \le \hat{\theta}_H\right\} = P\left\{\theta \ge \hat{\theta}_B\right\} = \frac{1-\gamma}{2}$.

Якщо $\left|\hat{\theta} - \theta\right| < \delta$, то в цьому випадку δ називається точністю оцінки. Виберемо δ так, щоб $P\left\{\left|\hat{\theta} - \theta\right| < \delta\right\} = \gamma$. Тоді інтервал

$$\hat{\theta} - \delta < \theta < \hat{\theta} + \delta$$

буде довірчим інтервалом, який із надійністю γ оцінює невідомий параметр θ. Точність оцінки δ залежить від надійності γ і обсягу вибірки n.

У випадку, коли вибірка побудована за великою кількістю незалежних однаково розподілених випадкових величин, для побудови довірчих інтервалів можна запропонувати такий метод. Якщо для функції правдоподібності $f\left(X, \theta\right)$ виконуються умови леми із пункту 9.2.3, то величина

$$\eta = \frac{\frac{\partial}{\partial\theta}\ln f\left(X,\theta\right)}{\sqrt{D\left(\frac{\partial}{\partial\theta}\ln f\left(X,\theta\right)\right)}}$$

асимптотично нормальна $(0,1)$. Нехай c_γ визначається із умови $P\left\{\left|\eta\right| < c_\gamma\right\} = \gamma$, де η має стандартний нормальний розподіл. Якщо $c_\gamma = c_\gamma\left(\theta\right)$ є монотонною функцією θ, то розв'язуючи нерівність $\left|\eta\right| < c_\gamma$ відносно θ, одержимо $\gamma-$довірчий інтервал для θ.

Розглянемо один із методів знаходження інтервальних оцінок за допомогою емпіричної функції розподілу. Нехай досліджувана величина ξ має функцію розподілу $F\left(x\right)$ і існує $M\left(g\left(\xi\right)\right)^2$. Покладемо

$$G = M\left(g\left(\xi\right)\right) = \int_{-\infty}^{+\infty} g\left(x\right) dF\left(x\right).$$

Тоді оцінка (5)

$$\hat{G}_n = \frac{1}{n}\sum_{i=1}^{n} g\left(X_i\right)$$

є незміщеною і сильно спроможною оцінкою параметра G. Із (7) випливає, що величина

$$\eta_n = \frac{\sqrt{n}}{\sigma_g}\left(\hat{G}_n - G\right),$$

де $\sigma_g^2 = Dg\left(\xi\right)$, асимптотично нормальна (0,1). Тому

$$P\left\{\left|\eta_n\right| < t\right\} \approx 2\Phi_0\left(t\right),$$

де $\Phi_0\left(t\right)$ – функція Лапласа.

По надійності γ можна знайти таке число t, щоб $2\Phi_0(t)=\gamma$. Тоді при великих n приблизно з ймовірністю γ буде виконуватись нерівність $|\eta_n|<t$ або

$$\hat{G}_n - t\frac{\sigma_g}{\sqrt{n}} < G < \hat{G}_n + t\frac{\sigma_g}{\sqrt{n}}. \qquad (12)$$

Побудована інтервальна оцінка параметра G є незручною, бо містить величину σ_g, яка може бути невідомою. Але при великих n

$$\sigma_g^2 \approx \frac{1}{n}\sum_{k=1}^{n}\left(g(X_k)-\hat{G}_n\right)^2 = \hat{\sigma}_g^2.$$

Замінимо в (12) σ_g на $\hat{\sigma}_g$. Тоді знову при великих n приблизно з ймовірністю γ буде виконуватись нерівність

$$\hat{G}_n - t\frac{\hat{\sigma}_g}{\sqrt{n}} < G < \hat{G}_n + t\frac{\hat{\sigma}_g}{\sqrt{n}}. \qquad (13)$$

Подібний метод ми уже використовували при оцінці ймовірності в схемі Бернуллі (розділ 2). Знайдемо тепер довірчий інтервал для невідомої ймовірності $P(A)=p$ у схемі Бернуллі. Із теореми Бореля випливає, що оцінка $\frac{\mu_n}{n}=\hat{p}$ (відносна частота події A, а $\mu_n=\xi_1+\dots+\xi_n$, де ξ_i – число появ події в $i-$му експерименті) є сильно спроможною і незміщеною оцінкою параметра $P(A)=p$. Оскільки $M\eta_n=np$, а $D\mu_n=npq$, то величина $Z=\frac{\mu_n-np}{\sqrt{npq}}=\frac{\hat{p}-p}{\sqrt{pq}}\sqrt{n}$ за ценральною граничною теоремою при великих n має розподіл близький до стандартного нормального закону, тобто $P\{|Z|<t\}\approx 2\Phi_0(t)$. Задамо надійність γ близьку до 1 і знайдемо t із рівняння $2\Phi_0(t)=\gamma$. Тоді при великих n із надійністю, що близька до γ, виконується нерівність $|Z|<t$ або

$$\hat{p}-t\sqrt{\frac{p(1-p)}{n}} < p < \hat{p}+t\sqrt{\frac{p(1-p)}{n}}. \qquad (14)$$

Якщо розв'яжемо нерівність (14) відносно p, то одержимо

$$\frac{\hat{p}n+\frac{1}{2}t^2-t\sqrt{\hat{p}(1-\hat{p})n+\frac{1}{4}t^2}}{n+t^2} < p < \frac{\hat{p}n+\frac{1}{2}t^2+t\sqrt{\hat{p}(1-\hat{p})n+\frac{1}{4}t^2}}{n+t^2}.$$

Якщо, аналогічно переходу від (12) до (13), у лівій і правій частинах нерівності (14) p замінити на $\hat{p}$, то одержимо більш простий вигляд довірчого інтервалу для невідомої ймовірності:

$$\hat{p}-t\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p}+t\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}. \qquad (15)$$

Якщо врахувати, що $p(1-p)\le\frac{1}{4}$, то із ймовірністю, що близька до γ, при великих n виконується нерівність

$$\hat{p}-\frac{t}{2\sqrt{n}} < p < \hat{p}+\frac{t}{2\sqrt{n}}.$$

Але остання оцінка є досить грубою. Частіше використовують інтервал у вигляді (15). У цьому інтервалі величина $t\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \delta$ є точність оцінки невідомої ймовірності із надійністю γ.

9.3.2. Довірчі інтервали для параметрів нормального розподілу. Щільність нормального розподілу з параметрами a і σ^2 має вигляд: $f(x,a,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$, де $a = M(\xi)$, $\sigma^2 = D(\xi)$.

Розглянемо окремі випадки знаходження довірчих інтервалів для параметрів a і σ^2. Для цього використаємо наступну теорему, яку ми приймемо без доведення.

Теорема. Нехай випадкові величини $X_1,...,X_n$ – незалежні і нормально розподілені з параметрами $(0,1)$. Тоді вибіркова середня $\overline{X}$ і вибіркова дисперсія D_B незалежні і величини $nD_B = (n-1)s^2$ мають розподіл χ^2 з ($n-1$) ступенями вільності.

а) Знайдемо довірчі інтервали для невідомого метематичного сподівання $a = M(\xi)$, коли σ^2 – відоме. За оцінку невідомого математичного сподівання виберемо $\overline{X}$. Оскільки $\overline{X} = \frac{X_1 + \cdots + X_n}{n}$, де X_i – нормально розподілені випадкові величини, то $\overline{X}$ є також номально розподіленою величиною із параметрами $M\overline{X} = a$ і $D\overline{X} = \frac{\sigma^2}{n}$. Тому величина $Z = (\overline{X} - a)\frac{\sqrt{n}}{\sigma}$ буде мати стандартний нормальний розподіл і $P\{|Z| < t\} = 2\Phi_0(t)$. За надійністю γ можна знайти t_γ як розв'язок рівняння $2\Phi_0(t) = \gamma$. Тоді із надійністю γ буде виконуватись нерівність $\left|\overline{X} - a\right|\frac{\sqrt{n}}{\sigma} < t_\gamma$ або

$$\overline{X} - t_\gamma\frac{\sigma}{\sqrt{n}} < a < \overline{X} + t_\gamma\frac{\sigma}{\sqrt{n}},$$

де величина

$$t_\gamma\frac{\sigma}{\sqrt{n}} = \delta$$

є точністю оцінки.

Тобто, довірчий інтервал, який із надійністю γ оцінює невідоме метематичне сподівання $a = M(\xi)$, коли σ^2 – відоме, має вигляд

$$\overline{X} - t_\gamma\frac{\sigma}{\sqrt{n}} < a < \overline{X} + t_\gamma\frac{\sigma}{\sqrt{n}},$$

де число t_γ знаходиться із рівняння $2\Phi_0(t_\gamma) = \gamma$.

б) Нехай σ^2 – невідоме. Величина $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ є незміщеною і спроможною оцінкою для σ^2. Тоді величина $\left(\overline{X} - a\right)\frac{\sqrt{n}}{\sigma}$ має стандартний нормальний розподіл, а за наведеною теоремою, величина $\frac{(n-1)s^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2$ буде мати розподіл χ^2 із $(n-1)$ ступенями вільності, при цьому, ці величини є незалежними. Тому величина

$T=\frac{\left(\overline{X}-a\right)\frac{\sqrt{n}}{\sigma}}{\sqrt{\frac{1}{n-1}\frac{(n-1)s^2}{\sigma^2}}}=\frac{\left(\overline{X}-a\right)\sqrt{n}}{s}$ буде мати розподіл Стьюдента із $(n-1)$ ступенями вільності (s – виправлене середнє квадратичне відхилення).

Якщо надійність γ задана, то за розподілом Стьюдента можна знайти таке число $t_{\gamma,n}$, яке задовольняє умову $P\left\{|T|<t_{\gamma,n}\right\}=\gamma$. Тоді із надійністю γ буде виконуватись нерівність $\frac{\left|\overline{X}-a\right|}{s}\sqrt{n}<t_{\gamma,n}$ або

$$\overline{X}-t_{\gamma,n}\frac{s}{\sqrt{n}}<a<\overline{X}+t_{\gamma,n}\frac{s}{\sqrt{n}},$$

де

$$t_{\gamma,n}\frac{s}{\sqrt{n}}=\delta$$

– точність оцінки.

Тобто, довірчий інтервал, який із надійністю γ оцінює невідоме метематичне сподівання $a=M(\xi)$, коли σ^2 – невідоме, має вигляд

$$\overline{X}-t_{\gamma,n}\frac{s}{\sqrt{n}}<a<\overline{X}+t_{\gamma,n}\frac{s}{\sqrt{n}},$$

де число $t_{\gamma,n}$ знаходиться за надійністю γ і числом ступенів вільності ($n-1$) за розподілом Стьюдента із умови $P\left\{|T|<t_{\gamma,n}\right\}=\gamma$.

в) Довірчі інтервали для невідомої дисперсії σ^2. Якщо a відоме, то оцінкою дисперсії є величина $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-a\right)^2$. Оскільки величини $\frac{X_i-a}{\sigma}$ мають стандартний нормальний розподіл, то величина $\frac{n\hat{\sigma}^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-a}{\sigma}\right)^2$ має розподіл χ^2 з n ступенями вільності ($\chi^2(n)$). При заданій надійності γ за розподілом $\chi^2(n)$ можна знайти такі числа h' і h'', щоб виконувались умови $P\left\{\chi^2(n)<h'\right\}=\frac{1-\gamma}{2}$ і $P\left\{\chi^2(n)<h''\right\}=\frac{1+\gamma}{2}$. Тоді із надійністю γ буде виконуватись нерівність $h'<\frac{n\hat{\sigma}^2}{\sigma^2}<h''$ або $\frac{n\hat{\sigma}^2}{h''}<\sigma^2<\frac{n\hat{\sigma}^2}{h'}$.

Отже, довірчий інтервал, який із надійністю γ оцінює невідому дисперсію σ^2, коли a – відоме, має вигляд

$$\frac{n\hat{\sigma}^2}{h''}<\sigma^2<\frac{n\hat{\sigma}^2}{h'},$$

де числа h' і h'' вибираються так, щоб виконувались умови $P\left\{\chi^2(n)<h'\right\}=\frac{1-\gamma}{2}$ і $P\left\{\chi^2(n)<h''\right\}=\frac{1+\gamma}{2}$.

Нехай a – невідоме. Тоді величина $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ є оцінкою для σ^2. За теоремою (стор. 150), величина $\frac{s^2(n-1)}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2$ має розподіл $\chi^2(n-1)$. При заданій надійності γ за розподілом $\chi^2(n-1)$ можна знайти такі числа h' і h'', щоб виконувались умови $P\left\{\chi^2(n-1) < h'\right\} = \frac{1-\gamma}{2}$ і $P\left\{\chi^2(n-1) < h''\right\} = \frac{1+\gamma}{2}$. Тоді із надійністю γ буде виконуватись нерівність $h' < \frac{(n-1)s^2}{\sigma^2} < h''$ або $\frac{(n-1)s^2}{h''} < \sigma^2 < \frac{(n-1)s^2}{h'}$.

Отже, довірчий інтервал, який із надійністю γ оцінює невідому дисперсію σ^2, коли a – невідоме, має вигляд

$$\frac{(n-1)s^2}{h''} < \sigma^2 < \frac{(n-1)s^2}{h'},$$

де числа h' і h'' вибираються так, щоб виконувались умови $P\left\{\chi^2(n-1) < h'\right\} = \frac{1-\gamma}{2}$ і $P\left\{\chi^2(n-1) < h''\right\} = \frac{1+\gamma}{2}$.

Вправи.

16. При проведенні обстеження стану очей у 265 учнів, 131 були хворі (49,4%). Встановити межі довірчого інтервалу для долі хворих дітей при надійності $\gamma = 0{,}95$.
17. За вибіркою об'єму $n = 25$ із нормального розподілу знайдені $\overline{x} = 18{,}2$ і $s^2 = 96{,}04$. Знайти довірчі інтервали для параметрів a і σ^2 при надійності $\gamma = 0{,}95$.

9.4. Перевірка статистичних гіпотез

9.4.1. Поняття статистичної гіпотези і статистичного критерію. Критерій Неймана-Пірсона. Однією із важливіших задач математичної статистики є задача перевірки статистичних гіпотез на основі статистичних даних, перевірка відповідності результатів експериментів сформульованій гіпотезі. *Статистичною гіпотезою* (або просто *гіпотезою*) називають будь-яке твердження про вигляд або властивості розподілу досліджуваних в експерименті випадкових величин. Тобто гіпотези, у яких іде мова про невідомий розподіл або невідомі параметри розподілу, є статистичними. Наприклад, твердження, досліджувана величина має нормальний розподіл, є статистичною гіпотезою. Часто розподіл досліджуваної величини відомий, а необхідно перевірити гіпотези про значення параметрів цього розподілу. Такі гіпотези називають *параметричними*.

Висунуту гіпотезу, яка має піддаватися перевірці, позначають H_0 і називають *основною* або *нульовою* гіпотезою. Поряд із основною розглядають гіпотезу, що протирічить основній, її називають *альтернативною* (альтернативою до гіпотези H_0) або *конкурентною*. Наприклад, нехай $H_0 : \theta \in T \subset \Theta$, $H_1 : \theta \in T_1 \subset \Theta$. Гіпотеза H_1 буде альтернативною, якщо $T \bigcap T_1 = \varnothing$. Якщо при цьому множина точок T складається лише із одного елемента, то гіпотеза називається *простою*. Тобто гіпотеза є простою, якщо їй відповідає один розподіл або одна точка в просторі параметрів, в іншому разі – *складною*.

Статистичні гіпотези перевіряються на основі статистичних даних. Для перевірки статистичних гіпотез вибирається деяка випадкова величина, яка залежить від результатів спостережень, тобто є статистикою. Статистики, які вибирають для перевірки гіпотез, називають статистиками критерію або *критеріями*. Нехай вибраний деякий критерій K.

Множину значень критерію розбивають на дві області: *область прийняття гіпотези* – множина значень критерію, при яких гіпотеза приймається; *критичну область* S – множина значень критерію, при яких гіпотеза відкидається. Це рівносильно тому, що у вибірковому просторі R^n виділяється дві множини: множина тих точок із R^n, при яких критерій приймає значення із області прийняття гіпотези, і множина V тих точок із R^n, при яких критерій приймає значення із критичної області.

Ймовірність того, що критерій прийме значення із критичної області, називається *рівнем значущості* критерію і позначається α. Частіше всього α задають близьким до 0 і покладають рівним 0,1; 0,05; 0,01; 0,005; 0,001.

Застосування процедури перевірки гіпотези пов'язано із такими помилками: відкинути гіпотезу H_0, якщо вона правильна (*помилка першого роду*); прийняти гіпотезу H_0, якщо вона неправильна (*помилка другого роду*). Тобто, у двох випадках ми приймаємо правильне рішення – правильна гіпотеза приймається або неправильна відкидається; і у двох випадках неправильне рішення – правильна гіпотеза відкидається або неправильна приймається. Ймовірність того, що правильна гіпотеза відкидається – це рівень значущості критерію. При побудові процедури перевірки гіпотези необхідно домагатися мінімальних значень ймовірностей помилок обох родів.

Якщо експериментальні дані не узгоджуються із гіпотезою за вибраним критерієм, то це означає, що критерій приймає значення із критичної області. Тоді з ймовірністю помилки першого роду спостерігається подія, яка протирічить гіпотезі. Якщо ймовірність такої події мала, то це означає, що спостерігається практично неможлива подія. В цьому випадку гіпотеза має бути відкинута з практичною достовірністю. Якщо експериментальні дані узгоджуються із висунутою гіпотезою, то це ще не означає, що вони не будуть узгоджуватися із іншою гіпотезою. При перевірці статистичних гіпотез за результатами спостережень неможливо довести справедливість гіпотези, можна лише стверджувати, що статистичні дані не протирічать висунутій гіпотезі. Отже, висновки, які приймаються на основі статистичних даних, формулюються у такому вигляді: експериментальні дані узгоджуються із даною гіпотезою, протирічать даній гіпотезі.

Із сказаного випливає, що перевірку статистичної гіпотези можна проводити за такою схемою: вияснення змісту гіпотези і вибір критерію для перевірки статистичної гіпотези; знаходження розподілу критерію, якщо розподіл спостережень співпадає з гіпотетичним; за заданим рівнем значущості α виділяють область прийняття гіпотези і критичну область або знаходять критичні точки критерію, тобто точки, які відділяють критичну область і область прийняття гіпотези; за результатами спостережень $x_1,...,x_n$ знаходять спостережуване значення критерію K_{cn}, якщо спостережуване значення критерію попадає в область прийняття гіпотези, то гіпотезу приймають, у протилежному випадку – відкидають.

Критерієм називають також і правило, яке дозволяє за вибіркою $(x_1,\ldots,x_n)$ прийняти або основну гіпотезу H_0, або альтернативну H_1. Кожен критерій однозначно визначається заданням критичної множини $V\subset R^n$, де V – прообраз області S при відображенні $K=K(x_1,\ldots,x_n)$. Попадання вибіркової точки $(x_1,\ldots,x_n)$ у V свідчить про те, що статистичні дані не узгоджуються із гіпотезою H_0. Можна різними способами вибирати множину V, щоб виконувалась умова (ймовірність помилки першого роду дорівнювала α)

$$P\left((x_1,...,x_n)\in V/H_0\right)=P\left(V/H_0\right)=\alpha .$$

Розглянемо тепер ймовірність помилки другого роду (прийняти гіпотезу H_0, якщо вона неправильна)

$$P\left((x_1,...,x_n)\in R^n\setminus V/H_1\right)=P\left(R^n\setminus V/H_1\right)=\beta .$$

Оскільки

$$P\left(V/H_1\right)+P\left(R^n \setminus V/H_1\right)=1,$$

то

$$P\left(V/H_1\right)=1-\beta$$

називають *потужністю критерію*. Отже, задача знаходження критерію (знаходження критичної множини V, множину V також називають критерієм), що має найменшу ймовірність помилки другого роду, є задачею знаходження критерію, що має найбільшу потужність (найбільш потужного критерію).

Критерій V називається *найбільш потужним*, якщо для довільної множини V' такої, що $P\left(V/H_0\right)=P\left(V'/H_0\right)=\alpha$,

$$P\left(V/H_1\right)\ge P\left(V'/H_1\right).$$

Розглянемо задачу вибору із двох простих гіпотез. Нехай $H_0:\theta=\theta_0$ і $H_1:\theta=\theta_1\ne\theta_0$. Множину V необхідно вибрати так, щоб ймовірність попадання вибіркової точки у V була малою, якщо H_0 правильна, і великою, в протилежному випадку. Задамо $\alpha=P\left(V/H_0\right)$. Множину V необхідно вибрати так, щоб потужність критерію $P\left(V/H_1\right)$ була максимальною для довільного допустимого значення $\theta_1\ne\theta_0$.

Нехай $p(x,\theta_i)$ – щільність розподілу (або розподіл у дискретному випадку) досліджуваної випадкової величини у випадку справедливості гіпотези H_i (i=0;1). Розглянемо множину X тих точок $(x_1,\ldots,x_n)$, для яких

$$\prod_{k=1}^{n} p(x_k,\theta_1)\ge c\prod_{k=1}^{n} p(x_k,\theta_0), \tag{16}$$

де $c\ge 0$ – довільна стала.

Щільність $\prod_{k=1}^{n} p(x_k,\theta_i)$ називається *функцією вірогідності гіпотези* H_i (i=0;1), а відношення $\dfrac{\prod_{k=1}^{n} p(x_k,\theta_1)}{\prod_{k=1}^{n} p(x_k,\theta_0)}=l(x)$ називають відношенням вірогідності. Критерій із критичною множиною X називають *критерієм відношення вірогідності*. Множина X залежить від c, тому $P\left(X/H_0\right)=g(c)$ є функцією від c. Функція $g(c)$ є незростаючою (при $c_1<c_2$ $X_{c_1}\supset X_{c_2}$), $g(c)\ge 0$, $g(0)=1$. Оскільки

$$cg(c)=cP\left(X/H_0\right)=c\int_X\prod_{k=1}^{n} p(x_k,\theta_0)dx_1\cdot\ldots\cdot dx_n\le\int_X\prod_{k=1}^{n} p(x_k,\theta_1)dx_1\cdot\ldots\cdot dx_n\le 1,$$

то $g(c)\le\dfrac{1}{c}$. Отже при $c\to\infty$ $g(c)\to 0$. Будемо вважати, що існує таке c, для якого $g(c)=\alpha$ (ця умова не є необхідною). Ця умова виконується, якщо $g(c)$ неперервна. Якщо ж $g(c)$ має в точці c розрив, то визначення X змінюємо, виключаючи з неї ті точки $(x_1,\ldots,x_n)$, для яких в (16) виконується рівність. Для множини X, що відповідає вибраному c $P\left(X/H_0\right)=g(c)=\alpha$.

Теорема (Неймана – Пірсона). Серед всіх критеріїв із заданим рівнем значущості α, що перевіряють дві прості гіпотези H_0 і H_1, критерій відношення вірогідності є найбільш потужним.

Доведення. Позначимо через Y критичну множину якого-небудь іншого критерію, для якого $P\left(Y/H_0\right)=\alpha$, і покажемо, що $P\left(X/H_1\right)\ge P\left(Y/H_1\right)$, тобто, потужність критерію відношення вірогідності більша від потужності будь-якого іншого критерію. Оскільки

$$P\big(X\setminus(X\cap Y)/H_0\big)=P\big(X/H_0\big)-P\big(X\cap Y/H_0\big)=\alpha-P\big(X\cap Y/H_0\big),$$
$$P\big(Y\setminus(X\cap Y)/H_0\big)=P\big(Y/H_0\big)-P\big(X\cap Y/H_0\big)=\alpha-P\big(X\cap Y/H_0\big),$$

то

$$P\big(X\setminus(X\cap Y)/H_0\big)=P\big(Y\setminus(X\cap Y)/H_0\big).$$

Із визначення X випливає, що для довільної точки $(x_1,\ldots,x_n)\notin X$ виконується нерівність

$$c\prod_{k=1}^{n}p(x_k,\theta_0)>\prod_{k=1}^{n}p(x_k,\theta_1),$$

тому

$$P\big(X\setminus(X\cap Y)/H_1\big)=\int\limits_{X\setminus(X\cap Y)}\prod_{k=1}^{n}p(x_k,\theta_1)dx_1\cdot\ldots\cdot dx_n\geq c\int\limits_{X\setminus(X\cap Y)}\prod_{k=1}^{n}p(x_k,\theta_0)dx_1\cdot\ldots\cdot dx_n=$$

$$=cP\big(X\setminus(X\cap Y)/H_0\big)=cP\big(Y\setminus(X\cap Y)/H_0\big)=c\int\limits_{Y\setminus(X\cap Y)}\prod_{k=1}^{n}p(x_k,\theta_0)dx_1\cdot\ldots\cdot dx_n>$$

$$>\int\limits_{Y\setminus(X\cap Y)}\prod_{k=1}^{n}p(x_k,\theta_1)dx_1\cdot\ldots\cdot dx_n=P\big(Y\setminus(X\cap Y)/H_1\big).$$

Додаючи до обох частин попередньої нерівності величину $P\big(X\cap Y/H_1\big)$, одержуємо $P\big(X/H_1\big)\geq P\big(Y/H_1\big)$. Теорема доведена.

Розглянемо приклад застосування доведеної теореми. Нехай досліджується випадкова величина, що має нормальний розподіл із параметрами a і σ^2 (σ^2 – відома). Розглянемо гіпотези $H_i: a=\theta_i\ (i=0;1)$, $\theta_1>\theta_0$. Тоді

$$l(x)=\frac{\prod_{k=1}^{n}p(x_k,\theta_1)}{\prod_{k=1}^{n}p(x_k,\theta_0)}=\frac{\frac{1}{(2\pi\sigma)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}\big(x_k-\theta_1\big)^2\right\}}{\frac{1}{(2\pi\sigma)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}\big(x_k-\theta_0\big)^2\right\}}=\exp\left\{\frac{n}{2\sigma^2}\big(\theta_1-\theta_0\big)\big(2\bar{x}-\theta_1-\theta_0\big)\right\}$$

і нерівність $l(x)>c$ еквівалентна нерівності

$$\bar{x}\geq\frac{\sigma^2\ln c}{n(\theta_1-\theta_0)}+\frac{\theta_1+\theta_0}{2},$$

яку можна подати у вигляді

$$\frac{\sqrt{n}}{\sigma}\big(\bar{x}-\theta_0\big)\geq\frac{\sigma\ln c}{\sqrt{n}\big(\theta_1-\theta_0\big)}+\frac{\sqrt{n}}{2\sigma}\big(\theta_1-\theta_0\big)=t(c).$$

При справедливості основної гіпотези, величина $\frac{\sqrt{n}}{\sigma}\big(\bar{x}-\theta_0\big)$ має стандартний нормальний розподіл ($\Phi(x)$ – відповідна функція розподілу), тому

$$\psi(c)=P\big(l(x)\geq c\big)=P\left(\frac{\sqrt{n}}{\sigma}\big(\bar{x}-\theta_0\big)\geq t(c)\right)=1-\Phi\big(t(c)\big)=\Phi\big(-t(c)\big).$$

При $c>0$ функція $t(c)$ – неперервна, тому $\psi(c)$ також неперервна і для заданого $\alpha\in(0,1)$ однозначно визначається величина c_α така, що $t(c_\alpha)=t_\alpha$, $\Phi\big(-t_\alpha\big)=\alpha$.

Отже, найбільш потужний критерій для перевірки гіпотези $H_0: a=\theta_0$ при альтернативній $H_1: a=\theta_1>\theta_0$ задається критичною областю

$$V=\left\{x:\frac{\sqrt{n}}{\sigma}\big(\bar{x}-\theta_0\big)\geq t_\alpha\right\},\ \ \Phi\big(-t_\alpha\big)=\alpha.$$

Знайдемо потужність одержаного критерію

$$P(V/H_1)=P\left(\frac{\sqrt{n}}{\sigma}\left(\bar{x}-\theta_0\right)\geq t_\alpha\right)=P\left(\frac{\sqrt{n}}{\sigma}\left(\bar{x}-\theta_1\right)\geq-\frac{\sqrt{n}}{\sigma}\left(\theta_1-\theta_0\right)+t_\alpha\right)=\Phi\left(\frac{\sqrt{n}}{\sigma}\left(\theta_1-\theta_0\right)-t_\alpha\right).$$

Крім критеріїв перевірки гіпотез, що основані на вибірках фіксованого обсягу, існують послідовні критерії. При використанні цих критеріїв, випробування проводяться послідовно і після кожного випробування робиться висновок про прийняття однієї із гіпотез або продовження випробувань.

9.4.2. Перевірка гіпотез про рівність ймовірностей. Нехай задано деяку сукупність однорідних об'єктів. Необхідно перевірити гіпотезу – доля об'єктів із заданою властивістю дорівнює заданому числу p_0. У зв'язку з цим розглянемо просту гіпотезу $H_0: p=p_0$ при альтернативній гіпотезі $H_1: p\neq p_0$. Нехай проведено n експериментів, у яких подія настала μ_n разів. Ймовірність p є невідомою, але відносна частота $\frac{\mu_n}{n}$ є незміщеною і спроможною оцінкою невідомої ймовірності p, тому ми можемо порівняти відносну частоту $\frac{\mu_n}{n}$ з p_0, де $\frac{\mu_n}{n}=\frac{X_1+\cdots+X_n}{n}$, X_i – число появ події в i-му експерименті. При справедливості висунутої гіпотези $MX_i=p_0$, $DX_i=p_0\left(1-p_0\right)$, $D\frac{\mu_n}{n}=\frac{p_0\left(1-p_0\right)}{n}$. Для перевірки гіпотези використаємо критерій – випадкову величину $Z=\frac{\frac{\mu_n}{n}-p_0}{\sqrt{\frac{p_0\left(1-p_0\right)}{n}}}=\frac{\mu_n-np_0}{\sqrt{np_0\left(1-p_0\right)}}$. При справедливості гіпотези H_0 $MZ=0$, $DZ=1$, тому при великих n величина Z є асимптотично нормальною (0,1).

Нехай задано рівень значущості α. Виберемо критичну точку z_α так, щоб виконувалась умова $P\left\{|Z|<z_\alpha\right\}\approx 2\Phi_0\left(z_\alpha\right)=1-\alpha$. Знайшовши z_α із рівняння $2\Phi_0\left(z_\alpha\right)=1-\alpha$, ми поділяємо множину всіх значень Z на область прийняття гіпотези і критичну область. Критична область – це множина тих значень Z, для яких $|Z|\geq z_\alpha$. За результатами спостережень знаходимо спостережуване значення критерію $z_{сп}$. Якщо $|z_{сп}|<z_\alpha$, то гіпотезу приймають, у протилежному випадку – відкидають.

Розглянемо просту гіпотезу $H_0: p=p_0$ при альтернативній гіпотезі $H_1: p>p_0$. В цьому випадку критична область буде односторонньою: це множина тих значень Z, для яких $Z\geq z_\alpha$, а критична точка z_α знаходиться із умови $P\left\{Z\geq z_\alpha\right\}=\alpha$. Оскільки $\alpha=P\left\{Z\geq z_\alpha\right\}=\frac{1}{2}P\left\{|Z|\geq z_\alpha\right\}=\frac{1}{2}\left(1-P\left\{|Z|<z_\alpha\right\}\right)\approx\frac{1}{2}\left(1-2\Phi_0\left(z_\alpha\right)\right)=\frac{1}{2}-\Phi_0\left(z_\alpha\right)$, то критичну точку z_α знаходимо із умови $\Phi_0\left(z_\alpha\right)=0,5-\alpha$. За результатами спостережень знаходимо спостережуване значення критерію $z_{сп}$. Якщо $z_{сп}<z_\alpha$, то гіпотезу приймають, у протилежному випадку – відкидають.

Нехай розглядаються дві сукупності. Ймовірність настання події A в першій сукупності дорівнює p_1, а в другій – p_2. Часто виникає необхідність перевіряти гіпотезу $H_0: p_1=p_2$ при альтернативі $H_1: p_1\neq p_2$. Нехай в першій сукупності проведено n_1 експериментів, в яких подія настала m_1 разів, в другій сукупності проведено n_2

експерименти, в яких подія настала m_2 разів. Тоді оцінками невідомих ймовірностей p_1 і p_2 будуть відповідні відносні частоти $\frac{m_1}{n_1}$ і $\frac{m_2}{n_2}$. Розглянемо різницю $\frac{m_1}{n_1}-\frac{m_2}{n_2}$. $M\left(\frac{m_1}{n_1}\right)=p_1$, $M\left(\frac{m_2}{n_2}\right)=p_2$, $D\left(\frac{m_1}{n_1}-\frac{m_2}{n_2}\right)=\frac{p_1\left(1-p_1\right)}{n_1}+\frac{p_2\left(1-p_2\right)}{n_2}$. При правильності основної гіпотези $p_1=p_2=p$, $D\left(\frac{m_1}{n_1}-\frac{m_2}{n_2}\right)=p\left(1-p\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$. Тому величина $\frac{\frac{m_1}{n_1}-\frac{m_2}{n_2}}{\sqrt{p\left(1-p\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ буде асимптотично нормальною з параметрами 0 і 1. Якщо гіпотеза правильна, то ми можемо вважати, що проведено n_1+n_2 спостережень, в яких подія настала m_1+m_2 раз. Замінимо p його оцінкою $\frac{m_1+m_2}{n_1+n_2}$ і для перевірки гіпотези використаємо критерій $Z=\frac{\frac{m_1}{n_1}-\frac{m_2}{n_2}}{\sqrt{\frac{m_1+m_2}{n_1+n_2}\left(1-\frac{m_1+m_2}{n_1+n_2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$. Величина Z є асимптотично нормальною з параметрами 0 і 1. Тому за рівнем значущості α можна знайти z_α із умови $P\left\{\left|Z\right|<z_\alpha\right\}\approx 2\Phi_0\left(z_\alpha\right)=1-\alpha$. За результатами спостережень знаходимо спостережуване значення критерію $z_{сп}$. Якщо $\left|z_{сп}\right|<z_\alpha$, то гіпотезу приймають, у протилежному випадку – відкидають.

9.4.3. Перевірка статистичних гіпотез про рівність середніх двох нормально розподілених випадкових величин. Нехай $X_1,...,X_n$ і $Y_1,...,Y_m$ дві незалежні нормально розподілені вибірки із параметрами $\left(a_1,\sigma_1^2\right)$ і $\left(a_2,\sigma_2^2\right)$ відповідно. Нехай параметри $MX=a_1$ і $MY=a_2$ невідомі. Часто на практиці виникає потреба встановити, чи суттєво відрізняються середні в цих вибірках. Тобто необхідно перевірити гіпотезу $H_0:M\left(X\right)=M\left(Y\right)$. Нехай альтернативна гіпотеза має вигляд $H_1:M\left(X\right)\neq M\left(Y\right)$. У цьому випадку критична область буде симетричною. Нехай дисперсії σ_1^2 і σ_2^2 відомі. Оскільки величини $\overline{X}=\frac{X_1+\cdots+X_n}{n}$ і $\overline{Y}=\frac{Y_1+\cdots+Y_m}{m}$ незалежні і нормальні $\left(a_1,\sigma_1^2/n\right)$, $\left(a_2,\sigma_2^2/m\right)$, то при правильності основної гіпотези $M\left(\overline{X}-\overline{Y}\right)=MX-MY=0$, а $D\left(\overline{X}-\overline{Y}\right)=\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}$, тому величина

$$Z=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}$$

буде мати стандартний нормальний розподіл. За рівнем значущості α із рівняння $1-\alpha=P\left\{\left|Z\right|<z_\alpha\right\}=2\Phi_0\left(z_\alpha\right)$ можна знайти z_α (із рівняння $1-\alpha=2\Phi_0\left(z_\alpha\right)$), а за

вибірками $x_1,...,x_n$ і $y_1,...,y_m$ знаходимо спостережуване значення критерію z_{cn}. Якщо $|z_{cn}| < z_\alpha$, то гіпотеза приймається, інакше - відкидається.

Розглянемо тепер випадок, коли дисперсії σ_1^2 і σ_2^2 невідомі, але рівні між собою, тобто $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Тоді величина $\dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{\sigma^2}{n}+\dfrac{\sigma^2}{m}}} = \dfrac{\overline{X}-\overline{Y}}{\sigma}\sqrt{\dfrac{mn}{m+n}} = Z$ має стандартний нормальний розподіл. Величини $s_X^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ і $s_Y^2 = \dfrac{1}{m-1}\sum_{i=1}^{m}\left(Y_i-\overline{Y}\right)^2$ є незміщеними та спроможними оцінками дисперсій $D(X)$ і $D(Y)$. Тоді величина $\dfrac{(n-1)s_X^2}{\sigma^2} = \sum_{i=1}^{n}\dfrac{\left(X_i-\overline{X}\right)^2}{\sigma^2}$ буде мати розподіл $\chi^2(n-1)$, а величина $\dfrac{(m-1)s_Y^2}{\sigma^2} = \sum_{i=1}^{m}\dfrac{\left(Y_i-\overline{Y}\right)^2}{\sigma^2}$ буде мати розподіл $\chi^2(m-1)$. Тому величина $\dfrac{(n-1)s_X^2}{\sigma^2}+\dfrac{(m-1)s_Y^2}{\sigma^2} = \chi^2(m+n-2)$ буде мати розподіл $\chi^2(m+n-2)$. Звідси випливає, що величина $T = \dfrac{Z}{\sqrt{\dfrac{1}{m+n-2}\chi^2(m+n-2)}} =$

$$= \frac{\dfrac{\overline{X}-\overline{Y}}{\sigma}\sqrt{\dfrac{mn}{m+n}}}{\sqrt{\dfrac{1}{m+n-2}\left(\dfrac{(n-1)s_X^2}{\sigma^2}+\dfrac{(m-1)s_Y^2}{\sigma^2}\right)}} = \frac{\overline{X}-\overline{Y}}{\sqrt{(n-1)s_X^2+(m-1)s_Y^2}}\sqrt{\frac{mn(m+n-2)}{m+n}}$$

буде мати розподіл Стьюдента із $(m+n-2)$ ступенями вільності. Отже, для перевірки гіпотези одержуємо критерій

$$T = \frac{\overline{X}-\overline{Y}}{\sqrt{(n-1)s_X^2+(m-1)s_Y^2}}\sqrt{\frac{mn(m+n-2)}{m+n}},$$

що має розподіл Стьюдента з ($n+m-2$) ступенями вільності.

За рівнем значущості α і числом ступенів вільності $(m+n-2)$ можна знайти таку точку T_α, щоб $P\{|T| > T_\alpha\} = \alpha$. А за вибірками $x_1,...,x_n$ і $y_1,...,y_m$ знаходимо $\overline{X}$, $\overline{Y}$, s_X^2, s_Y^2 і спостережуване значення критерію T_{cn}. Якщо $|T_{cn}| < T_\alpha$, то гіпотеза приймається, в протилежному випадку – відкидається.

9.4.4. Перевірка гіпотез про рівність дисперсій двох нормально розподілених випадкових величин. У попередньому пункті ми робили припущення, що дисперсії σ_1^2 і σ_2^2 невідомі, але рівні між собою: $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Тому для застосування критерію Стьюдента необхідно спочатку перевірити гіпотезу про рівність дисперсій. Нехай знову $X_1,...,X_n$ і $Y_1,...,Y_m$ – дві незалежні нормально розподілені вибірки із параметрами $\left(a_1,\sigma_1^2\right)$ і $\left(a_2,\sigma_2^2\right)$ відповідно. Розглянемо гіпотезу $H_0 : D(X) = D(Y)$ при альтернативній гіпотезі $H_1 : D(X) > D(Y)$. Незміщеними і спроможними оцінками для дисперсій є відповідні виправлені дисперсії $s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ і $s_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)^2$. Оскільки досліджувані величини X і Y мають нормальний розподіл, то величини $\frac{(n-1)s_X^2}{\sigma^2} = \chi^2(n-1)$ і $\frac{(m-1)s_Y^2}{\sigma^2} = \chi^2(m-1)$ мають розподіли χ^2 відповідно з $(n-1)$ та $(m-1)$ ступенями вільності. Тоді величина $F = \frac{\frac{1}{n-1}\chi^2(n-1)}{\frac{1}{m-1}\chi^2(m-1)} = \frac{s_X^2}{s_Y^2}$ буде мати розподіл Фішера з $(n-1, m-1)$ ступенями вільності. Тобто, для перевірки гіпотез про рівність дисперсій, використовуємо критерій Фішера

$$F = \frac{s_X^2}{s_Y^2}$$

з $(n-1, m-1)$ ступенями вільності.

За результатами спостережень знаходимо s_X^2 і s_Y^2, якщо поділимо більшу із виправлених дисперсій на меншу, то одержимо спостережуване значення критерію $F_{сп}$. За рівнем значущості α і ступенями вільності $(n-1, m-1)$ можна знайти критичну точку $F_{кр}$ із умови $P\left\{F > F_{кр}\right\} = \alpha$, де n - обсяг вибірки, за якою знайдена більша виправлена дисперсія, а m – менша. У випадку $F_{сп} < F_{кр}$ гіпотеза приймається, інакше – відкидається.

9.4.5. Перевірка гіпотез про вигляд розподілу. Критерії перевірки гіпотез про вигляд розподілу називають критеріями згоди. Ми розглянемо критерій Колмогорова і критерій χ^2. Критерій Колмогорова оснований на теоремі Колмогорова.

Теорема. Нехай $F(x)$ – неперервна функція розподілу, а $F_n(x)$ – емпірична функція розподілу, що знайдена за вибіркою із розподілу $F(x)$. Позначимо $D_n = \sup\limits_x \left|F_n(x) - F(x)\right|$. Тоді

$$\lim_{n\to\infty} P\left\{\sqrt{n}D_n < t\right\} = K(t) = \begin{cases} 0, & t \le 0, \\ 1 - 2\sum\limits_{n=1}^{\infty}(-1)^{n-1} e^{-2n^2t^2}, & \end{cases}$$

де $K(t)$ – функція Колмогорова.

Нехай перевіряється гіпотеза про те, що $x_1, ..., x_n$ є вибіркою із розподілу $F(x)$. За вибіркою $x_1, ..., x_n$ знаходимо емпіричну функцію розподілу $F_n(x)$. При великих n граничну функцію $K(t)$ можна використовувати для практичних розрахунків, тобто $K(t)$ досить добре наближує ймовірність $P\left\{\sqrt{n}D_n < t\right\}$, тому для перевірки гіпотези можна використати статистику $K_n = \sqrt{n}D_n$. Граничний розподіл $K(t)$ не залежить від вигляду розподілу $F(x)$. Важливим є і те, що розподіл величини K_n при великих n (уже при $n \ge 20$) практично не залежить від n. Оскільки емпірична функція розподілу збігається до відповідної теоретичної функції розподілу, то критична точка t_α за рівнем значущості α

визначається із умови $P\left\{\sqrt{n}D_n \ge t_\alpha\right\} = \alpha$. Враховуючи наближену рівність $P\left\{\sqrt{n}D_n < t\right\} \approx K(t)$, критичну точку t_α можна знайти із рівняння $K(t_\alpha) = 1 - \alpha$ (наприклад, $\alpha = 0{,}1$ $t_\alpha = 1{,}22$, $\alpha = 0{,}05$ $t_\alpha = 1{,}36$, $\alpha = 0{,}01$ $t_\alpha = 1{,}63$). Для знаходження спостережуваного значення критерію K_n за вибіркою $x_1,...,x_n$, для кожної із точок знайдемо спочатку різниці $\left|F_n(x_i) - F(x_i)\right|$, $\left|F_n(x_i + 0) - F(x_i)\right| = \left|F_n(x_{i+1}) - F(x_i)\right|$. Значення D_n дорівнює найбільшій із цих $2n$ різниць. Тоді спостережуване значення критерію $K_{cn} = \sqrt{n}D_n$. Якщо $K_{cn} < t_\alpha$, то гіпотезу приймають, інакше – відкидають.

Досить широке використання в статистиці знайшов *критерій* χ^2 (*критерій Пірсона*), що оснований на порівнянні емпіричних та теоретичних частот.

Нехай перевіряється гіпотеза про те, що $x_1,...,x_n$ є вибіркою із розподілу $F(x)$. Нехай множина значень досліджуваної величини X розбита на s підмножин: $\Delta_1,...,\Delta_s$. Позначимо через n_i число результатів спостережень x_j, які попадають до Δ_i, $n_1 + ... + n_s = n$, тоді n_i називають *емпіричними частотами*. Оскільки гіпотеза стверджує, що досліджувана величина X має розподіл $F(x)$, то за розподілом $F(x)$ можна знайти ймовірності $p_i = P\left\{X \in \Delta_i\right\}$. Тоді $n_i' = np_i$ називають *теоретичними частотами*. Із закону великих чисел, теорема Бореля, випливає, що при справедливості основної гіпотези $\frac{n_i}{n} \to p_i$ з ймовірністю одиниця. Тобто, при великих n, різниці $n_i - np_i = n_i - n_i'$ будуть невеликими.

Розглянемо випадкову величину $\chi^2 = \sum_{i=1}^{s} \frac{(n_i - np_i)^2}{np_i}$. Можна довести, що при $n \to \infty$ розподіл величини χ^2 збігається до розподілу χ^2 з $(s-1)$ ступенями вільності. На практиці граничний розподіл можна використовувати із гарним наближенням уже при $n \ge 50$ і $n_j \ge 5$.

Величина χ^2 є невід’ємною. За рівнем значущості α і числом ступенів вільності $(s-1)$ із рівняння $P\left\{\chi^2 > \chi^2_\alpha\right\} = \alpha$ знаходимо χ^2_α. За результатами спостережень знаходимо спостережуване значення критерію $\chi^2_{cn} = \sum_{i=1}^{s} \frac{\left(n_i - n_i'\right)^2}{n_i'}$. Якщо $\chi^2_{cn} < \chi^2_\alpha$, то гіпотеза приймається, інакше – відкидається.

Приклад. В експериментах з селекцією гороху Мендель спостерігав частоти різного вигляду насіння, що одержані при схрещуванні рослин з круглим жовтим і зморшкуватим зеленим насінням. Ці дані та значення теоретичних ймовірностей, що визначаються за теорією спадковості Менделя, наведені в таблиці:

Насіння	Частота n_i	Ймовірність p_i
Кругле жовте	315	9/16
Зморшкувате жовте	101	3/16
Кругле зелене	108	3/16
Зморшкувате зелене	32	1/16
Всього	$n = 556$	1

При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про узгодження експериментальних даних із теоретичними ймовірностями.

За результатами спостережень знаходимо спостережуване значення критерію

$$\chi_{cn}^2=\sum_{i=1}^{s}\frac{\left(n_i-n_i'\right)^2}{n_i'}=$$

$$=\frac{\left(315-556\cdot 9/16\right)^2}{556\cdot 9/16}+\frac{\left(101-556\cdot 3/16\right)^2}{556\cdot 3/16}+\frac{\left(108-556\cdot 3/16\right)^2}{556\cdot 3/16}+\frac{\left(32-556\cdot 1/16\right)^2}{556\cdot 1/16}=0,47\,.$$

За рівнем значущості $\alpha=0,05$ і числом ступенів вільності 4-1=3 за таблицею розподілу χ^2 знаходимо $\chi_\alpha^2=7,8$. Оскільки $\chi_{cn}^2=0,47<7,8=\chi_\alpha^2$, то можна зробити висновок, що експериментальні дані добре узгоджуються із теоретичними ймовірностями.

При великих значеннях s, розподіл χ^2 можна наближено замінити нормальним розподілом із середнім ($s-1$) і дисперсією $2\left(s-1\right)$.

Нехай розподіл $F\left(x,\theta_1,...,\theta_r\right)$ залежить від r параметрів. Тоді на основі результатів спостережень ми можемо замінити невідомі параметри $\theta_1,...,\theta_r$ їх відповідними оцінками $\hat{\theta}_1,...,\hat{\theta}_r$ і на основі функції $F\left(x,\hat{\theta}_1,...,\hat{\theta}_r\right)$ знаходимо теоретичні частоти n_i'. Тоді при $n\to\infty$ розподіл величини χ^2 збігається до розподілу χ^2 з ($s-1-r$) ступенями вільності, де r – число оцінюваних параметрів розподілу. А далі процедура перевірки гіпотези така ж сама.

Вправи.

18. За вибірками обсягів $n=30$ і $m=40$, вибраних із нормально розподілених генеральних сукупностей із відомими дисперсіями $D\left(X\right)=60$ і $D\left(Y\right)=80$, знайдені вибіркові середні $\overline{x}=130$ і $\overline{y}=125$. Перевірити гіпотезу про рівність середніх при рівні значущості $\alpha=0,05$.
19. За вибірками обсягів $n=12$ і $m=15$, вибраних із нормально розподілених генеральних сукупностей X і Y, знайдені виправлені дисперсії $s_x^2=11,41$ і $s_y^2=6,52$. Перевірити гіпотезу про рівність дисперсій при рівні значущості $\alpha=0,05$.
20. При 4040 киданнях монети Бюфон одержав 2048 випадань герба. При рівні значущості $\alpha=0,01$ перевірити гіпотезу H_0: монета симетрична.
21. За вибіркою обсягом $n=1000$ підраховано число обривів ниток. При рівні значущості $\alpha=0,05$ перевірити гіпотезу: число обривів ниток X має розподіл Пуассона.

Число обривів x_i	0	1	2	3
Частота n_i	600	320	70	10

9.5. Вибіркова кореляція і регресія

9.5.1. Основні поняття. Для аналізу явищ і процесів, що виникають в економіці, фінансах, соціології, біології і інших галузях виникає потреба у вивченні зв'язків (залежності) між різними показниками (випадковими величинами). Наприклад, залежність обсягу виробництва від розміру основних фондів, рівня освіченості від статі. Для таких зв'язків характерно, що кожному значенню однієї величини відповідає деяка множина значень іншої величини.

Залежність, при якій середнє значення однієї величини залежить від значень, які приймає інша величина, називається *кореляційною*. Закон зміни середнього значення однієї величини в залежності від значень, які приймає інша величина, називається *регресією*.

Сукупність методів дослідження таких залежностей називають кореляційним аналізом. Завдання кореляційного аналізу полягає у встановленні на основі статистичних даних форми кореляційної залежності і сили кореляційної залежності. Коли залежність досліджується між двома величинами (X, Y), то говорять про парну кореляцію і регресію. Якщо величин більше як дві, то говоримо про множинну кореляцію і регресію. Нехай досліджується вектор $(X_1, ..., X_m, Y)$, де $X_1, ..., X_m$ – фактори, а Y – результуюча ознака. Якщо $\overline{Y_x} = f(X_1, ..., X_m)$, то ця функція називається регресією, а така регресія називається множинною. $\overline{Y_x}$ – середнє арифметичне тих значень Y, які відповідають значенням $X = (X_1, ..., X_m)$.

Обмежимось випадком, коли залежність досліджується між двома величинами (X, Y). Вибіркою обсягом n буде набір n пар чисел $(x_1, y_1), ..., (x_n, y_n)$.

Допоміжним засобом аналізу статистичних даних за двома ознаками є побудова кореляційної таблиці і кореляційного поля. При нанесенні на координатну площину точок (x_i, y_i) одержуємо *кореляційне поле*. По характеру розміщення точок поля можна зробити попередній висновок про форму залежності величин (наприклад, що одна із величин в середньому зростає або спадає при зростанні іншої).

Для числової обробки, результати спостережень групують і зображають у вигляді кореляційної таблиці. При дослідженні за двома ознаками, статистичні дані подають у вигляді кореляційної таблиці (інтервальної або дискретної).

$X \setminus Y$	y_1	...	y_j	...	y_l	n_x
x_1	n_{11}		n_{1j}		n_{1l}	n_{x_1}
...						
x_i	n_{i1}		n_{ij}		n_{il}	n_{x_i}
...						
x_m	n_{m1}		n_{mj}		n_{ml}	n_{x_m}
n_y	n_{y_1}	...	n_{y_j}	...	n_{y_l}	n

Для побудови кореляційної таблиці різні значення $x_1, ..., x_m$ величини X (або інтервали групування) запишемо у перший стовпчик, а значення $y_1, ..., y_l$ величини Y (або відповідні інтервали групування) запишемо у перший рядок. На перетині i-го рядка і j-го стовпчика записуємо частоту n_{ij} пари (x_i, y_j) у даній вибірці (або число тих пар (x, y), компоненти яких попадають у відповідні інтервали групування за кожною змінною). В останньому стовпчику записуємо частоти n_{x_i} подій $\{X = x_i\}$, а у останньому рядку – частоти n_{y_j} подій $\{Y = y_j\}$. Очевидно, сума всіх частот $\sum_{i,j} n_{ij} = n$ дає обсяг вибірки, сума по рядках $\sum_{j=1}^{l} n_{ij} = n_{x_i}$, а – стовпцях $\sum_{i=1}^{m} n_{ij} = n_{y_j}$. Тому перший і останній стовпчики дають емпіричний розподіл величини X, а перший і останній рядки дають емпіричний розподіл величини Y.

При вивченні вигляду залежності Y від X, результати експериментів по відношенню до змінної X інтерпретуються по різному: значення величини Y знаходять при фіксованих значеннях величини X; результати спостережень (x_i, y_i) ($i = 1, ..., n$) є вибірка із двовимірної сукупності. Методи дослідження однакові в обох випадках.

Умовне математичне сподівання $M(Y / X = x) = f(x)$ називають теоретичною регресією. Середнє арифметичне значень величини Y, які зустрічаються у парі із заданим значенням $X = x$ називається *умовною середньою* Y на X і позначають $\overline{Y_x}$. Тоді $\overline{Y_x}$

змінюється із зміною x, тобто є функцією від x. Позначимо її $f(x)$. Функція $f(x)$ (закон зміни $\overline{Y_x} = f(x)$) називається *емпіричною регресією* Y на X, а рівняння $\overline{Y_x} = f(x)$ називають емпіричним рівнянням регресії Y на X. Графік функції $f(x)$ називають лінією регресії Y на X. Наближене уявлення про вигляд залежності Y від X дає емпірична ламана регресії Y на X – це ламана із вершинами в точках $\left(x_i, \overline{Y_{x_i}}\right)$. При зростанні обсягу вибірки, емпірична ламана регресії наближається до деякої згладженої лінії – лінії регресії. На практиці вигляд регресії $f(x)$ нам невідомий, його можна оцінити за статистичними даними, можна підібрати за виглядом емпіричної ламаної регресії. Наприклад, якщо емпірична ламана регресії близька до деякої прямої, то $f(x)$ знаходять у вигляді лінійної функції. При цьому параметри, якими визначається $f(x)$, знаходяться (оцінюються) за статистичними даними.

Аналогічно можна розглядати регресію X на Y.

Встановлення сили кореляційного зв'язку полягає у оцінці розсіювання значень величини Y біля лінії регресії. Найуживанішими характеристиками сили кореляційної залежності є вибірковий коефіцієнт кореляції і вибіркове кореляційне відношення.

9.5.2. Вибірковий коефіцієнт кореляції, вибіркове кореляційне відношення. *Вибірковим коефіцієнтом кореляції* називається величина

$$r_B = \frac{\frac{1}{n}\sum_i \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sigma_X \sigma_Y},$$

де $\overline{x}$ і $\overline{y}$ – вибіркові середні, а σ_X і σ_Y – вибіркові середні квадратичні відхилення величин X і Y. Нескладними перетвореннями можна одержати формулу

$$r_B = \frac{\frac{1}{n}\sum_i x_i y_i - \overline{x}\,\overline{y}}{\sigma_X \sigma_Y}.$$

Вибірковий коефіцієнт кореляції є оцінкою теоретичного коефіцієнта кореляції r і має аналогічні властивості. $|r_B| \le 1$. Якщо $r_B = \pm 1$, то між величинами X і Y існує лінійна функціональна залежність, лінія регресії буде прямою. Тому коефіцієнт кореляції є мірою сили лінійної залежності між величинами X і Y. Якщо $|r_B|$ близький до одиниці, то це говорить про те, що залежність є близькою до лінійної. Випадок $r_B = 0$ вказує на відсутність лінійної залежності, але не заперечує наявності іншого вигляду залежності.

Для обчислення коефіцієнта кореляції можна перейти до нових величин $U = \frac{X - c_1}{h_1}$ і $V = \frac{Y - c_2}{h_2}$, при цьому для довільних c_1, c_2, $h_1 \ne 0$, $h_2 \ne 0$ справедлива формула

$$r_B = \frac{\sum n_{uv} uv - n\overline{u}\,\overline{v}}{n\sigma_u \sigma_v},$$

де $\overline{u} = \frac{1}{n}\sum n_u u$, $\overline{v} = \frac{1}{n}\sum n_v v$, $\sigma_u^2 = \left(\overline{u^2} - \left(\overline{u}\right)^2\right)$, $\sigma_v^2 = \left(\overline{v^2} - \left(\overline{v}\right)^2\right)$, $\overline{u^2} = \frac{1}{n}\sum n_u u^2$, $\overline{v^2} = \frac{1}{n}\sum n_v v^2$, n_{uv} є частота відповідної пари (x, y).

Розглянемо гіпотезу $H_0 : r = 0$. Це гіпотеза значущості зв'язку між величинами, тобто, чи суттєво r_B відрізняється від нуля. Для перевірки цієї гіпотези використовується статистика

$$T = \frac{r_B\sqrt{n-2}}{\sqrt{1-r_B^2}},$$

яка у випадку нормально розподіленої вибірки, має розподіл Стьюдента із $(n-2)$ ступенями вільності. Із даної рівності виразимо вибірковий коефіцієнт кореляції

$$r_B = \left(1+\frac{n-2}{T^2}\right)^{-\frac{1}{2}}.$$

За рівнем значущості α і числом ступенів вільності $(n-2)$ можна знайти $T_{n,\alpha}$. Якщо $r_B > \left(1+\frac{n-2}{T_{n,\alpha}^2}\right)^{-\frac{1}{2}}$, то залежність між величинами буде сильною, тобто, коефіцієнт кореляції суттєво відрізняється від нуля.

У випадку $r_B \neq 0$ часто використовують так зване z-перетворення Фішера, замінюючи r_B на z за формулою

$$z = \frac{1}{2}\ln\frac{1+r_B}{1-r_B}.$$

Уже при порівняно невеликих n розподіл величини z добре наближується нормальним розподілом з математичним сподіванням, що дорівнює $\frac{1}{2}\ln\frac{1+r}{1-r}+\frac{r}{2(n-1)}$ і дисперсією, що дорівнює $\frac{1}{n-3}$. Виходячи із цього, можна визначати наближені інтервали для коефіцієнта кореляції.

Вибіркове кореляційне відношення Y на X є статистичною оцінкою теоретичного кореляційного відношення (позначимо його $\eta_{y/x}$) і визначається за формулою

$$\eta_{y/x} = \frac{\sigma_{\overline{Y_x}}}{\sigma_Y},$$

де $\sigma_Y^2 = \frac{1}{n}\sum_i n_{y_i}\left(y_i - \overline{y}\right)^2$, $\sigma_{\overline{Y_x}}^2 = \frac{1}{n}\sum_i n_{x_i}\left(\overline{Y_{x_i}} - \overline{y}\right)^2$.

Вибіркове кореляційне відношення використовується для характеристики довільного зв'язку між величинами і має властивості, що аналогічні до властивостей теоретичного кореляційного відношення. Зокрема, $0 \le \eta_{y/x} \le 1$, $|r_B| \le \eta_{y/x}$. Якщо $\eta_{y/x} = 1$, то величина Y пов'язана із X точною функціональною залежністю, якщо ж при цьому $r_B^2 = 1$, то ця залежність буде лінійною. У випадку $\eta_{y/x} < 1$ функціональна залежність відсутня. Якщо $\eta_{y/x} = 0$, то $\overline{Y_x} = \overline{Y}$.

9.5.3. Знаходження рівняння прямої лінії регресії методом найменших квадратів. Ми уже відзначали, що вигляд функції регресії є невідомим. Будемо знаходити регресію у вигляді $g(x,\beta)$, де β, взагалі кажучи, є векторним параметром, що повністю визначає функцію $g(x,\beta)$. Тобто вигляд функції $g(x,\beta)$ відомий і повністю визначається параметром β. Вигляд функції $g(x,\beta)$ можна вибирати, наприклад, за виглядом емпіричної ламаної регресії. Тоді знаходження рівняння регресії, вигляд функції регресії якої вважається відомим, означає перш за все визначення на основі статистичних даних $(x_1, y_1),...,(x_n, y_n)$ параметрів, від яких залежить регресія. Ці параметри часто оцінюють методом найменших квадратів, який полягає у тому, що за оцінки параметрів вибирають такі значення β, при яких вираз

$$\sum_{i=1}^{n}\left(y_i - g\left(x_i, \beta\right)\right)^2 = F\left(\beta\right)$$

є мінімальним. Якщо знайдемо частинні похідні від $F\left(\beta\right)$ по параметрах і прирівняємо їх до нуля, то одержимо систему нормальних рівнянь, розв'язок якої дає оцінки невідомих параметрів за методом найменших квадратів. В цьому випадку рівняння $\overline{Y_x} = g\left(x, \beta\right)$ називають емпіричним рівнянням регресії Y на X. Таку процедуру називають ще згладжуванням статистичних даних.

У найпростішому випадку лінійної регресії емпіричне рівняння регресії Y на X знаходять у вигляді $\overline{Y_x} = \rho x + \beta$. Тобто необхідно знайти пряму, яка б була найближче розміщена до точок $\left(x_i, y_i\right)$ у розумінні методу найменших квадратів. У цьому випадку параметри ρ і β вибираються так, щоб сума

$$\sum_{i=1}^{n}\left(y_i - \rho x_i - \beta\right)^2 = F\left(\rho, \beta\right)$$

була мінімальною. Для цього утворимо систему нормальних рівнянь

$$\begin{cases} \dfrac{\partial F}{\partial \rho} = 0, \\ \dfrac{\partial F}{\partial \beta} = 0. \end{cases}$$

Після знаходження частинних похідних одержуємо систему

$$\begin{cases} \dfrac{\partial F}{\partial \rho} = \sum\limits_{i=1}^{n} 2\left(y_i - \rho x_i - \beta\right)\left(-x_i\right) = 0, \\ \dfrac{\partial F}{\partial \beta} = \sum\limits_{i=1}^{n} 2\left(y_i - \rho x_i - \beta\right)\left(-1\right) = 0, \end{cases}$$

яку можна записати у вигляді $\begin{cases} \sum\limits_{i=1}^{n}\left(x_i y_i - \rho x_i^2 - \beta x_i\right) = 0, \\ \sum\limits_{i=1}^{n}\left(y_i - \rho x_i - \beta\right) = 0 \end{cases}$ або $\begin{cases} \rho\sum\limits_{i=1}^{n} x_i^2 + \beta\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i y_i, \\ \rho\sum\limits_{i=1}^{n} x_i + n\beta = \sum\limits_{i=1}^{n} y_i. \end{cases}$

Введемо позначення $\overline{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$, $\overline{x^2} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i^2$, $\overline{y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$ тоді система набуває вигляду

$$\begin{cases} \rho\overline{x^2} + \beta\overline{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i y_i, \\ \rho\overline{x} + \beta = \overline{y}, \end{cases}$$

звідки

$$\beta = \overline{y} - \rho\overline{x}, \quad \rho = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} x_i y_i - \overline{x}\,\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} x_i y_i - \overline{x}\,\overline{y}}{\sigma_X^2}.$$

Коефіцієнт ρ називається *коефіцієнтом регресії* Y на X. Враховуючи вигляд вибіркового коефіцієнта кореляції, одержимо $\rho = r_B \frac{\sigma_Y}{\sigma_X}$. Підставимо знайдені значення параметрів у рівняння прямої лінії регресії, тоді емпіричне рівняння прямої регресії Y на X буде мати вигляд

$$\overline{Y_x} - \overline{y} = r_B \frac{\sigma_Y}{\sigma_X}\left(x - \overline{x}\right).$$

У деяких випадках знаходження рівняння лінії регресії можна звести до знаходження рівняння прямої лінії регресії. Це легко можна зробити, коли рівняння лінії регресії має вигляд $y = a\exp\left(x\right) + b$ або $y = a\frac{1}{x} + b$.

У першому випадку ми вводимо нову заміну $t = \exp\left(x\right)$, тобто від пари статистичних даних $\left(x_i, y_i\right)$ ми переходимо до нової пари $\left(t_i, y_i\right)$; $t_i = \exp\left(x_i\right)$. Аналогічно, у другому випадку статистичні дані перетворюємо за формулою $\frac{1}{x} = t$. Тоді знаходимо рівняння прямої лінії регресії у вигляді $y = at + b$ і повертаємось до старої змінної.

Вправа.

22. Знайти вибіркове рівняння прямої лінії регресії за такими статистичними даними для випадкового вектора $\left(X, Y\right)$:

x_i	18	19	25	20	25	21	23	22	23	24
y_i	20	20	35	20	30	25	25	25	30	30

Таблиці

Таблиця 1. Значення функції $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ **.**

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3969	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1.0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2.7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

x	4,0	4,2	4,4	4,6	4,8	5,0
$\varphi(x)$	0,00013	0,0000589	0,0000249	0,0000101	0,0000040	0,0000015

Таблиця 2. Значення функції Лапласа $\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$ **.**

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41308	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49638	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997

x	4,0	4,2	4,4	4,6	4,8	5,0
$\Phi_0(x)$	0,4999683	0,4999867	0,4999946	0,4999979	0,4999992	0,4999997

Таблиця 3. Квантилі хі-квадрат розподілу $\chi^2_{p,k}$

k	*P*										
	0,005	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995
1	$0{,}0^4 4$	$0{,}0^3 2$	0,001	0,004	0,016	0,46	2,71	3,84	5,02	6,63	7,88
2	0,010	0,020	0,051	0,103	0,211	1,39	4,61	5,99	7,38	9,21	10,6
3	0,072	0,115	0,216	0,352	0,584	2,37	6,25	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	3,36	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	4,35	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	5,35	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	6,35	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	7,34	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	8,34	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	9,34	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	10,3	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	11,3	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	12,3	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	13,3	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	14,3	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	15,3	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	16,3	24,8	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	10,9	17,3	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,4	11,7	18,3	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	19,3	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,3	11,6	13,2	20,3	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	21,3	30,8	33,9	36,8	40,3	42,8
23	9,26	10,2	11,7	13,1	14,8	22,3	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	23,3	33,2	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	24,3	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	25,3	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	26,3	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	27,3	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	28,3	39,1	42,6	45,7	49,6	52,3
30	13,8	15,0	16,8	18,5	20,6	29,3	40,3	43,8	47,0	50,9	53,7
35	17,2	18,5	20,6	22,5	24,8	34,3	46,1	49,8	53,2	57,3	60,3
40	20,7	22,2	24,4	26,5	29,1	39,3	51,8	55,8	59,3	63,7	66,8
45	24,3	25,9	28,4	30,6	33,4	44,3	57,5	61,7	65,4	70,0	73,2
50	28,0	29,7	32,4	34,8	37,7	49,3	63,2	67,5	71,4	76,2	79,5
75	47,2	49,5	52,9	56,1	59,8	74,3	91,1	96,2	100,8	106,4	110,3
100	67,3	70,1	74,2	77,9	82,4	99,3	118,5	124,3	129,6	135,6	140,2

Таблиця 4. Квантилі розподілу Стьюдента $t_{p,k}$

k	p						
	0,750	0,900	0,950	0,975	0,990	0,995	0,999
1	1,000	3,078	6,314	12,71	31,82	63,66	318
2	0,816	1,886	2,920	4,303	6,965	9,925	22,3
3	0,765	1,638	2,353	3,182	4,541	5,841	10,2
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297
10	0,700	1,372	1.812	2,228	2,764	3,169	4,144
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307
60	0,679	1,296	1,671	2,000	2,390	2,660	3,232
120	0,677	1,289	1,658	1,980	2,358	2,617	3,160
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,090

Таблиця 5. Квантилі розподілу Фішера-Снедекора $F_p\left(k_1, k_2\right)$ (k_1 – число ступенів вільності більшої дисперсії, k_2 – число ступенів вільності меншої дисперсії)

p=0,95

k_2	k_1								
	1	2	3	4	5	6	7	8	9
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38
3	16,13	9,55	9,78	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,75	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3.86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,85	3,98	3,59	3,36	3,20	3,09	3,01	2,93	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,30	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,37	2,30
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88

Продовження таблиці (p=0,95)

k_2	k_1								
	10	12	15	20	24	30	40	60	120
1	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3
2	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49
3	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55
4	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66
5	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40
6	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70
7	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27
8	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97
9	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75
10	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58
11	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45
12	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34
13	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25
14	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18
15	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11
16	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06
17	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01
18	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97
19	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93
20	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90
21	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87
22	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84
23	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81
24	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79
25	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77
26	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75
27	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73
28	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71
29	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70
30	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68
40	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58
60	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47
120	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35
∞	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22

p=0,99

k_2	k_1								
	1	2	3	4	5	6	7	8	9
1	4052	4999,5	5403	5625	5764	5859	5928	5982	6022
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07
40	7,31	5,18	4,31	3,53	3,51	3,29	3,12	2,99	2,89
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41

Продовження таблиці (p=0,99)

k_2	k_1								
	10	12	15	20	24	30	40	60	120
1	6056	6106	6157	6209	6235	6261	6287	6313	6339
2	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49
3	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22
4	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56
5	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11
6	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97
7	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74
8	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95
9	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40
10	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00
11	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69
12	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45
13	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25
14	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09
15	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96
16	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84
17	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75
18	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66
19	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58
20	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52
21	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46
22	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40
23	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35
24	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31
25	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27
26	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23
27	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20
28	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17
29	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14
30	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11
40	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92
60	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73
120	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53
∞	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32

Література

1. Анісімов В.В., Черняк О.І. Математична статистика.- К.: МП “Леся”, 1995.
2. Барра Ж.-Р. Основные понятия математической статистики.- М.: Мир,1974.-280 с.
3. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики.-М.: Наука, 1983.- 416 с.
4. Боровков А.А. Теория вероятностей.- М.: Наука, 1986.- 432 с.
5. Боровков А.А. Математическая статистика.- М.: Наука, 1984.- 472 с.
6. Бочаров П.П., Печинкин Ф.В. Теория вероятностей. Математическая статистика.- М.: Гардарика, 1998.- 328 с.
7. Ван дер Варден Б.Л. Математическая статистика.-М.: Иностр. лит., 1960.- 436 с.
8. Вентцель Е.С., Овчаров Л.Н. Теория вероятностей и ее инженерные приложения.- М.: Наука, 1988.- 480 с.
9. Гихман И.И., Скороход А.В., Ядренко М.И. Теория вероятностей и математическая статистика.-К.: Вища школа, 1979.- 320 с.
10. Гмурман В.Е. Теория вероятностей и математическая статистика.- М.: Высш. шк., 1977.-479 с.
11. Гнєденко Б.В. Курс теорії імовірностей.- Київ-Львів: Радянська школа, 1949.- 360 с.
12. Гнеденко Б.В. Курс теории вероятностей.- М.: Наука, 1988.- 400 с.
13. Гнеденко Б.В., Колмогоров А.Н. Предельные распределения для сумм независимых случайных величин.- М.: ГИТТЛ, 1949.-264 с.
14. Гурский Е.И. Теория вероятностей с элементами математической статистики.- М.: Высш. шк., 1971.- 328 с.
15. Зінченко Н.М., Оленко А.Я. Аналітичні моделі та методи соціології.- К.: Видав. центр “Київський ун-т”, 2000.- 106 с.
16. Ивченко Г.И., Медведев Ю.И. Математическая статистика. - М.: Высш. шк., 1984. - 248 с.
17. Карташов М.В. Конспект лекцій з курсу теорії ймовірностей.- К.: Видавничо-поліграфічний центр “Київський університет”, 2001.- 107 с.
18. Кендал М., Стьюарт А. Статистические выводы и связи.- М.: Наука, 1973.- 817 с.
19. Коваленко И.Н., Филипова А.А. Теория вероятностей и математическая статистика.- М.: Высш. шк., 1973.- 368 с.
20. Коваленко И.Н., Гнеденко Б.В. Теория вероятностей.- К.: Вища школа, 1990.- 328 с.
21. Колемаев В.А., Староверова О.В., Турундаевский В.Б. Теория вероятностей и математическая статистика.- М.: Высш. Шк., 1991.- 400 с.
22. Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф. Справочник по теории вероятностей и математической статистике. К.: Наукова думка, 1978.- 584 с.
23. Крамер Г. Математические методы статистики.- М.: Мир, 1975.- 648 с.
24. Ландкоф Н.С. Введение в теорию вероятностей.- Харьков: Изд-во Харьк. ун-та, 1968.- 236 с.
25. Леоненко М.М., Мішура Ю.С., Пархоменко В.М., Ядренко М.Й. Теоретико-ймовірнісні та статистичні методи в економетриці та фінансовій математиці.- К.: Інформтехніка, 1995.
26. Лоэв М. Теория вероятностей.- М.: Изд-во иностр. лит., 1962.- 720 с.
27. Пархоменко В.М. Методи вибіркових обстежень.- К.: Видав. центр “Київський ун-т”, 2001.- 148 с.
28. Пугачев В.С. Теория вероятностей и математическая статистика.- М.: Наука, 1979.- 496 с.
29. Пытьев Ю.П., Шишмарев И.А. Курс теории вероятностей и математической статистики для физиков.- М.: Изд-во Москов. ун-та, 1983.-256 с.
30. Розанов Ю.А. Теория вероятностей, случайные процессы и математическая статистика.- М.: Наука, 1985.- 320 с.

31. Розанов Ю.А. Случайные процессы (краткий курс).- М.: Наука, 1971.- 288 с.
32. Севастьянов Б.А. Курс теории вероятностей и математической статистики.- М.: Наука, 1982.- 256 с.
33. Скороход А.В. Елементи теорії ймовірностей та випадкових процесів.- К.: Вища школа, 1975.- 296 с.
34. Скороход А.В. Лекції з теорії випадкових процесів.- К.: Либідь, 1990.- 168 с.
35. Слюсарчук П.В. Теорія ймовірностей і математична статистика. Текст лекцій.- Ужгород: Вид-во УжДУ, 1984.- 66 с.
36. Смирнов Н.В., Дудин-Барковский И.В. Курс теории вероятностей и математической статистики для инженерных приложений.- М.: Наука, 1969.- 512 с.
37. Солодовников А.С. Теория вероятностей.- М.: Просвещение, 1983.- 207 с.
38. Тутубалин В.Н. Теория вероятностей.- М.: Изд-во Москов. ун-та, 1972.- 230 с.
39. Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х т.- М.: Мир, 1984.- Т.1.- 527 с. Т.2.- 751 с.
40. Чистяков В.П. Курс теории вероятностей.- М.: Наука, 1978.- 224 с.
41. Шефтель З.Г. Теорія ймовірностей.- К.: Вища школа, 1994.- 192 с.
42. Ширяев А.Н. Вероятность.- М.: Наука, 1980.- 576 с.
43. Шметтерер Л. Введение в математическую статистику.- М. Наука, 1976.- 520 с.

Збірники задач

44. Вентцель Е.С., Овчаров Л.А. Теория вероятностей. Задачи и упражнения.- М.: Наука, 1969.- 368 с.
45. Вишневецкий Л.Д., Гусак Д.В., Погребецкая Т.А., Тер-Саакянц Г.Л. Математическая статистика и случайные процессы: Практикум.- К.: Вища школа, 1992.- 143 с.
46. Вуколов Э.А., Ефимов А.В., Земсков В.Н., Золотарев Ю.Г., Каракулин А.Ф., Поспелов А.С., Терещенко А.М. Сборник задач по математике для втузов. Специальные курсы.- М.: Наука, 1984.-608 с.
47. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике.- М.: Высш. шк., 1975.- 333 с.
48. Дороговцев А.Я., Сільвестров Д.С., Скороход А.В., Ядренко М.Й. Теорія ймовірностей. Збірник задач.- К.: Вища школа, 1976.- 384 с.
49. Емельянов Г.В., Скитович В.П. Задачник по теории вероятностей и математической статистике.- Ленинград: Изд-во Ленинград. ун-та, 1967.- 332 с.
50. Ивченко Г.И., Медведев Ю.И., Чистяков А.В. Сборник задач по математической статистике.- М.: Высш. шк., 1989.- 255 с.
51. Климов Г.П., Кузьмин А.Д. Вероятность, процессы, статистика: задачи с решениями.- М.: Изд-во Москов. ун-та, 1985.- 232 с.
52. Козлов М.В. Элементы теории вероятностей в примерах и задачах.-М.: Изд-во Москов. ун-та, 1990.- 344 с.
53. Мешалкин Л.Д. Сборник задач по теории вероятностей.- М.: Изд-во Москов. ун-та, 1963.- 155 с.
54. Оленко А.Я. Ймовірність і статистика. Задачі.- К.: НаУКМА, 2002.- 53 с.
55. Прохоров А.В., Ушаков В.Г., Ушаков Н.Г. Задачи по теории вероятностей: Основные понятия. Предельные теоремы. Случайные процессы.- М.: Наука, 1986.-328 с.
56. Сборник задач по теории вероятностей, математической статистике и теории случайных функций. Под общей редакцией А.А. Свешникова.- М.: Наука, 1970.- 656 с.
57. Севастьянов Б.А., Чистяков В.П., Зубков А.М. Сборник задач по теории вероятностей.- М.: Наука, 1980.- 224 с.
58. Турчин В.М. Теорія ймовірностей. Основні поняття, приклади, задачі.- К.: Видав-во А.С.К., 2004.- 208 с.

59. Турчин В.М. Математична статистика в прикладах і задачах. У 2-х ч.- Дніпропетровськ: ДДУ, 1998.- Ч.1 – 68 с. Ч.2 - 228 с.
60. Черняк О.І., Обушна О.М., Ставицький А.В. Теорія ймовірностей та математична статистика: збірник задач.- К.: Т-во "Знання", КОО, 2002.- 199 с.
61. Чибисов Д.М., Пагурова В.И. Задачи по математической статистике.- М.: Изд-во МГУ, 1990.- 171 с.

Предметний покажчик