МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

Теорія ймовірностей, імовірнісні процеси та математична статистика

Методичні вказівки до вивчення навчальної дисципліни
Для студентів напряму підготовки
6.050101 «Комп'ютерні науки»

Кафедра автоматизованих систем обробки інформації та управління

Всіх форм навчання

Рекомендовано
Вченою радою факультету
Інформатики та обчислювальної
Техніки НТУУ «КПІ»
Протокол №12 від 28.05.2012р.

Київ

ФІОТ НТУУ «КПІ»

2012

Теорія ймовірностей, імовірнісні процеси та математична статистика. Методичні вказівки до вивчення навчальної дисципліни. /Уклад.: О.А.Павлов, В.М.Кузнецов –К.: НТУУ «КПІ», 2012.-329с.

Методичні вказівки призначені для студентів напряму підготовки 6.050101 «Комп'ютерні науки» кафедра автоматизованих систем обробки інформації і управління всіх форм навчання. У вказівках наведено систематизоване викладення багатьох основних відомостей з теорії ймовірностей, імовірнісних процесів та математичної статистики з орієнтацією на поглиблене вивчення вказаної дисципліни студентами технічних спеціальностей. Метою авторів було з одного боку закласти фундамент для можливості в майбутньому знайомлення з сучасною ймовірнісною літературою (аксіоматичний підхід до викладення теорії ймовірностей, строге доведення результатів тощо), а з іншого боку дати більш менш широкий набір результатів для їх застосування в інженерній практиці.

Укладачі	О.А. Павлов, д.т.н., професор
	В.М.Кузнецов, к.ф.-м.н., доцент
Відповідальний редактор	О.Г.Жданова, к.т.н., доцент
Рецензент	Ю.О.Кулаков, д.т.н., професор

За редакцією укладачів

Зміст

1. Вихідні поняття теорії ймовірностей

1.1 Випробування, події

Теорія ймовірностей виникла як наука з переконання у тому, що в основі масових випадкових явищ лежать детерміновані закономірності. Теорія ймовірностей вивчає вказані закономірності.

Наприклад: передбачити наперед, що випаде внаслідок підкидання симетричної монети, “орел” чи “решка”, неможливо, але при великій кількості підкидань монети “орли” та “решки” випадають приблизно однакову кількість разів.

В основі теорії ймовірностей, як і в основі будь-якої математичної дисципліни, лежать деякі поняття, що не мають чіткого означення, а лише тлумачаться. При їх введені в теперішній час прийнято теоретико-множинний підхід до викладання теорії ймовірностей, якого ми будемо дотримуватись і в цій книзі.

Вихідним поняттям теорії ймовірностей є поняття випробування або ймовірнісного (стохастичного) експерименту. Випробування – це реалізація певного комплексу вимог, що може повторюватись необмежену кількість разів. При цьому комплекс вимог містить у собі випадкові фактори, реалізація яких у кожному випробуванні призводить до неоднозначності результату випробування, його не можна передбачити заздалегідь.

Теорія ймовірностей вивчає випробування, які володіють вказаною властивістю невизначеності результату, а також властивістю статистичної стійкості, про яку буде сказано пізніше після введення поняття частості.

Наприклад: випробування – одноразове підкидання симетричної монети. Конкретним результатом випробування є елементарна подія.

Наприклад: випробування – одноразове підкидання шестигранного грального кубика, елементарна подія – випадає грань з "2".

Сукупність всіх можливих, різних, конкретних результатів випробування складає простір елементарних подій (множину усіх елементарних подій). Його ми будемо позначати Ω, а його елементи (елементарні події) – маленькими літерами, наприклад, ω з індексами. Таким чином, $\Omega = \{\omega\}$.

Приклад: випробування – одноразове підкидання грального кубика, простір елементарних подій - $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$, де ω_i – елементарна подія: випадання грані з і очками, $i=1...6$.

У попередньому прикладі простір елементарних подій є скінченна множина, але в багатьох задачах теорії ймовірностей доводиться мати справу з випробуваннями, що мають нескінченну (більш того, незліченну) множину можливих конкретних результатів.

Приклад: випробування – стрільба у круглу мішень. Якщо нас цікавить точка, у яку влучає куля, то як простір елементарних подій можна взяти множину всіх точок кругу К та ще одну додаткову точку θ, що означає невлучення стрільця у мішень.

Приклад: випробування – броунівський рух частинки, простір елементарних подій – множина всіх можливих траєкторій руху частинки.

Як складну подію розуміють деяку підмножину простору елементарних подій $A \subseteq \Omega$. Складна подія внаслідок випробування настає тоді і тільки тоді, коли внаслідок випробування відбувається деяка елементарна подія, що належить складній. Таким чином, якщо внаслідок випробування відбувається деяка елементарна подія, то відбуваються й усі складні події, до складу яких належить ця елементарна подія.

<u>Приклад:</u> випробування – одноразове підкидання грального кубика, елементарна подія ω_i – випадання грані з номером i, складна подія $A_1=\{\omega_1, \omega_3, \omega_5\}$ – випадання грані з непарним номером, складна подія $A_2 = \{\omega_1, \omega_2, \omega_3\}$ – випадання грані, номер якої не перевищує 3.

Серед складних подій відрізняють <u>вірогідну подію</u> $U = \Omega$ - подію, яка відбудеться завжди внаслідок випробування, тому що вона складається з усіх можливих елементарних подій.

Відрізняють також <u>неможливу подію</u> $V = \varnothing$, яка ніколи не відбудеться внаслідок випробування (не містить жодної елементарної події). За змістом, з нею ототожнюється все, що не може відбутися внаслідок випробування.

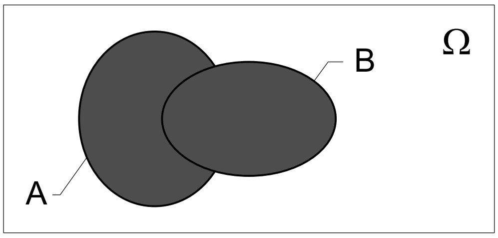
<u>Наприклад:</u> випробування – дворазове підкидання грального кубика, неможлива подія – сумарна кількість очок менше двох.

Для зручності будемо вважати, що $\varnothing\subset\Omega$.

1.2<u>Операції над подіями</u>

Різноманітні операції над подіями зручно ототожнювати із звичайними теоретико-множинними операціями.

Подія C називається <u>сумою</u> (об’єднанням) подій A і B та позначається $C=A\cup B$, якщо вона складається з усіх елементарних подій, які входять до складу A або B (або до A та B одночасно). При цьому, якщо елементарна подія ω входить до A та B, то до C вона входить лише один раз. Можна сказати, що внаслідок випробування відбувається подія $C=A\cup B$ тоді та й тоді, коли відбувається або A, або B (або обидві одночасно).



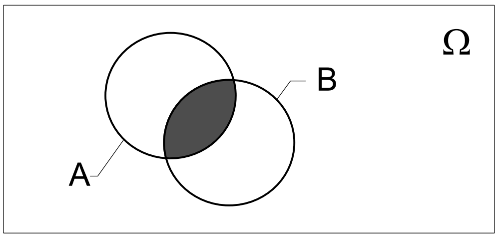
A∪B

Аналогічний зміст має сума будь-якої (в тому числі незліченної) кількості подій. Коли I – довільна множина значень деякого індексу i, A_i – деяка множина елементарних подій (i∈I), сума (об’єднання) $\underset{i\in I}{\text{¿}} A_i$ є подія, яка відбувається тоді, коли відбувається принаймні одна з подій A_i.

Для будь-якого A

$$A\cup V = A\cup\varnothing = A,\ A\cup U = A\cup\Omega = U = \Omega,\ A\cup A = A.$$

Подія C називається <u>добутком</u> (перетином) подій A і B та позначається C=A∩B (або C=A*B), якщо вона містить у собі елементарні події, що входять до A і B одночасно. Говорять ще, що внаслідок випробування подія C=A∩B відбувається тоді та тільки тоді, коли внаслідок випробування відбувається A та B одночасно.



A∩B

Аналогічне означення має добуток довільної кількості подій $\underset{i\in I}{\text{intersect}} A_i$ як подія, що відбувається тоді та тільки тоді, коли внаслідок випробування відбуваються одночасно всі події A_i, i∈I.

Для кожного A маємо

$$A\cap V = A\cap\varnothing = V=\varnothing,\ A\cap U = A\cap\Omega = A,\ A\cap A = A.$$

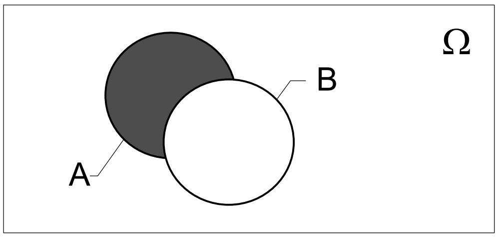
Зауважимо, що операції додавання і множення подій задовольняють закони:

1 комутативності: $AB = BA$; $A\cup B = B\cup A$;

2 асоціативності: $A(BC) = (AB)C$; $A\cup(B\cup C) = (A\cup B)\cup C$;

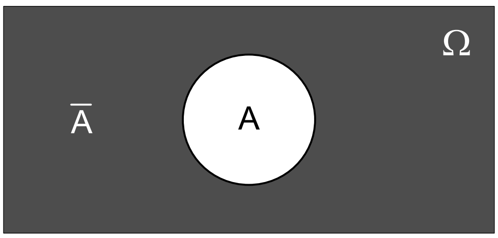
3 дистрибутивності: $A(B\cup C) = (AB)\cup(AC)$.

Подія C називається <u>різницею</u> подій A і B та позначається $C = A\backslash B$, якщо C містить в собі ті й тільки ті елементарні події, які входять до складу A та не входять до складу B. Говорять ще, що подія $C = A\backslash B$ відбувається тоді й тільки тоді, коли внаслідок випробування відбувається подія A, та не відбувається подія B.



$A\backslash B$

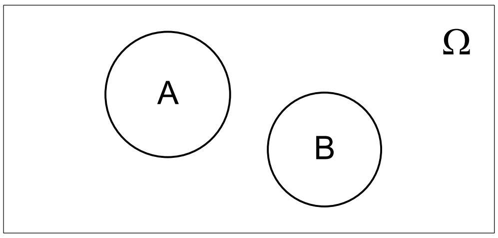
<u>Протилежною</u> до події A називається подія $\bar{A} = \Omega\backslash A$. Говорять, що внаслідок випробування відбувається протилежна подія $\bar{A}$ тоді й тільки тоді, коли у цьому випробуванні не відбувається подія A.



$\bar{A}$

Для кожного A має місце $A\cup\bar{A} = U = \Omega$, $A\cap\bar{A} = V = \varnothing$.

Події A і B називаються <u>несумісними</u>, якщо вони не мають спільних елементарних подій. Говорять, що події A і B несумісні, якщо вони ніколи не відбуваються разом внаслідок одного випробування.



Для несумісних подій $A \cap B = V = \varnothing$.

Система подій A_i, $i=\overline{1,n}$ називається парно-несумісною, якщо для кожних $i \neq j$ має місце $A_i \cap A_j = \varnothing$.

Суму парно-несумісних подій іноді позначають як $\sum_{i=1}^{n} A_i$ (n тут може бути нескінченним).

<u>Приклад:</u> для розглянутого раніше одноразового підкидання грального кубика, коли $\Omega=\{\omega_1, \omega_2, \omega_3, ..., \omega_6\}$, $A_1=\{\omega_1, \omega_3, \omega_5\}$, $A_2=\{\omega_1, \omega_2, \omega_3\}$, маємо:

$$A_1 \cup A_2=\{\omega_1, \omega_2, \omega_3, \omega_5\},\ A_1 \cap A_2=\{\omega_1, \omega_3\},\ \overline{A}_1=\{\omega_2, \omega_4, \omega_6\},$$

$$\overline{A}_2=\{\omega_4, \omega_5, \omega_6\},\ A_1 \backslash A_2=\{\omega_5\}.$$

Мають місце формули де Моргана[1] :

$$\overline{A \cup B}=\overline{A} \cdot \overline{B},\ \overline{A \cdot B}=\overline{A} \cup \overline{B}.$$

Доведемо першу з них. Візьмемо довільне $\omega \in \overline{A \cup B}$. Тоді очевидна послідовність висновків: $\omega \in \overline{A \cup B} \Rightarrow \omega \notin A \cup B \Rightarrow \omega \notin A, \omega \notin B \Rightarrow \omega \in \overline{A}$, $\omega \in \overline{B} \Rightarrow \omega \in \overline{A} \cdot \overline{B}$.

1 Морган де Огастес (Августус) (1806-1871) – шотландський математик і логік, секретар Королівського астрономічного товариства, чл. Лондонського королівського товариства, перший президент Лондонського математичного товариства.

Аналогічно, для кожного $\omega\in\overline{A}\cdot\overline{B}$ маємо: $\omega\in\overline{A}\cdot\overline{B} \Rightarrow \omega\in\overline{A}$, $\omega\in\overline{B} \Rightarrow \omega\notin A, \omega\notin B \Rightarrow \omega\notin A\cup B \Rightarrow \omega\in\overline{A\cup B}$.

За означенням рівності двох множин маємо першу рівність де Моргана.

Самостійно довести другу рівність, а також узагальнити рівність на довільну кількість доданків та співмножників

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}, \quad \overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A_i}.$$

Розглянемо нескінченну послідовність подій $A_1, A_2, \ldots, A_n, \ldots$, або коротко $\{A_n\}$, $A_n\subset\Omega$. Важливим є поняття границі послідовностей подій.

Множина всіх тих елементарних подій ω, що належать всім подіям A_n, за виключенням тільки скінченного їх числа, називається <u>нижньою границею послідовності</u> $\{A_n\}$ та позначається $\underline{\lim}_{n\to\infty} A_n$, або $\lim\inf_{n\to\infty} A_n$.

Переконаємось у тому, що $\underline{\lim}_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k$.

Дійсно, якщо $\omega\in\underline{\lim}_{n\to\infty} A_n$, то існує таке n, що $\omega\in\bigcap_{k=n}^{\infty} A_k$, а тому $\omega\in\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k$, тобто $\underline{\lim}_{n\to\infty} A_n\subset\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k$. Навпаки, якщо $\omega\in\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k$, то знайдеться таке n, що

$\omega\in\bigcap_{k=n}^{\infty} A_k$, тобто ω належить всім членам послідовності, починаючи з A_n.

І, таким чином, $\omega\in\lim_{n\to\infty} A_n$, а тому $\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k\subset\underline{\lim}_{n\to\infty} A_n$. Рівність множин доведено.

Множина всіх тих елементарних подій ω, що належить нескінченній кількості A_n, називається <u>верхньою границею послідовності</u> $\{A_n\}$ та позначається $\overline{\lim\limits_{n\to\infty}} A_n$, або $\limsup\limits_{n\to\infty} A_n$.

Тому, що кожне ω, що належить всім $\overline{A_n}$, за виключенням скінченного числа подій послідовності $\{\overline{A_n}\}$, повинна належати тільки скінченній кількості A_n, та навпаки, за формулами де Моргана одержуємо

$$\overline{\lim_{n\to\infty}} A_n = \overline{\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \overline{A_k}\right)} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k .$$

Кожна елементарна подія ω, що належить всім A_n, за виключенням їх скінченного числа, повинна міститься в нескінченній кількості A_n, тому

$$\underline{\lim_{n\to\infty}} A_n \subset \overline{\lim_{n\to\infty}} A_n .$$

Коли має місце і обернене включення, тобто $\underline{\lim\limits_{n\to\infty}} A_n = \overline{\lim\limits_{n\to\infty}} A_n = A$, то спільна границя A називається <u>границею послідовності</u> $\{A_n\}$ та позначається $\lim\limits_{n\to\infty} A_n$, або пишуть $A_n \to A$. Ще говорять, що A_n <u>збігається</u> до A.

Прикладом збіжних послідовностей є так звані монотонні послідовності.

Нехай $\{A_n\}$ – послідовність, що не зростає, тобто така послідовність, що

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

Тоді з означення виходить, що

$$\overline{\lim_{n\to\infty}} A_n = \bigcap_{n=1}^{\infty} A_n ,$$

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n ,$$

і ,таким чином, границя послідовності $\{A_n\}$ існує і дорівнює

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

У цьому випадку ще пишуть $A_n \downarrow A$.

Якщо $\{A_n\}$ – послідовність, що не спадає, тобто має місце

$$A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq A_{n+1} \subseteq \ldots\ldots ,$$

то

$$\overline{\lim_{n\to\infty}} A_n=\bigcup_{n=1}^{\infty} A_n ,$$

$$\underline{\lim_{n\to\infty}} A_n=\bigcup_{n=1}^{\infty} A_n ,$$

і існує границя послідовності A, яка дорівнює

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

У цьому випадку ще пишуть $A_n \uparrow A$.

1.3 <u>Частість наставання подій</u>

Нехай простір елементарних подій Ω складається з обмеженої кількості (m) елементарних подій $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$.

Розглянемо F – клас всіх підмножин множини Ω (включаючи порожню множину $\varnothing$). Цей клас містить 2^m елементів.

<u>Приклад:</u> $\Omega = \{\omega_1, \omega_2, \omega_3\}$.

Тоді $F=\{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_3\}, \varnothing\}$.

Будемо тлумачити клас F як всі події, які можуть настати внаслідок деякого випробування. Візьмемо деяку подію $A \in F$ і проведемо n випробувань.

Позначимо n(A) – кількість випробувань, у яких подія A настає (назвемо її <u>частотою</u> наставання події A у n випробуваннях). Назвемо відношення

$$W_n(A)=\frac{n(A)}{n}$$

<u>частістю</u> наставання події A у n випробуваннях (його також іноді називають відносною частотою).

Мають місце такі властивості частості:

1. $0 \le W_n \le 1$.
2. $W_n(\Omega) = 1$, $W_n(\varnothing) = 0$.
3. Частість суми парно-несумісних подій дорівнює сумі частостей цих подій. Тобто, якщо

$$A_i \cap A_j = \varnothing,\ i \ne j$$

і $A=\sum_{i=1}^{k} A_i$, то $W_n(A) = \sum_{i=1}^{k} W_n(A_i)$.

Дійсно, проведемо n випробувань. Якщо в деякому випробуванні настає подія A_i, то в ньому не може відбутися ніяка інша несумісна подія A_j, $j \ne i$. Тому $n(A)=\sum_{i=1}^{k} n(A_i)$. Тоді $W_n(A)=\frac{n(A)}{n}=\sum_{i=1}^{k}\frac{n(A_i)}{n}=\sum_{i=1}^{k} W(A_i)$.

Попрямуємо n до нескінченності, і будемо цікавитися існування границі

$$P(A)=\lim_{n\to\infty} W_n(A)=\lim_{n\to\infty}\frac{n(A)}{n}$$

для деякої події $A \in F$.

Теорія ймовірностей використовується для моделювання випадкових явищ, для яких виконується така властивість стійкості частості: для довільної події A частість відбування цієї події практично в довільній достатньо тривалій серії випробувань мало відрізняється від деякого числа, яке слушно вважати мірою можливості події (ймовірністю).

Спочатку аксіоматика теорії ймовірностей і будувалася на "емпіричному" означенні ймовірності як границі частості при кількості випробувань, що прямує до нескінченності (Р.Е.Мізес[2]). Однак математики відразу зіткнулися з серйозними логічними труднощами при побудові загальної теорії.

У подальшому (1933 р.) аксіоматика теорії ймовірностей була сформульована А.М. Колмогоровим[3] і базувалась на іншому принципі означення ймовірностей. Це виявило плодотворний вплив на весь наступний розвиток теорії ймовірностей, і тому аксіоматика А.М. Колмогорова прийнята нині повсюди. При цьому властивість стійкості частості (як ми побачимо далі) у певному розумінні зберігається.

1.4 Алгебри та σ-алгебри множин

До цих пір ми як події розглядали довільні підмножини множини Ω і, таким чином, неявно припускалося, що клас подій є клас усіх підмножин множини Ω. Це зручно тому, що запроваджені нами раніше операції не виводять нас з цього класу; можна не хвилюватися, що застосувавши над подіями яке завгодно число разів згадані операції, ми одержимо деякий новий об'єкт (вже не подію). Останнє дуже важливо з точки зору необхідності здійснювати різні граничні переходи, зокрема, при доведенні граничних теорем.

2 Мізес Ріхард Едлер (1883-1958) – німецький математик та механік (народився у Львові). Професор Страсбурзького, Берлінського, Стамбульського та Гарвардського університетів.

3 Колмогоров Андрій Миколайович (1903-1987) – видатний математик сучасності, академік АН СРСР, професор Московського університету. Був обраний членом більш ніж двадцяти іноземних академій та наукових товариств.

Але так клас всіх подій, як правило, визначають тільки у випадку, коли Ω є скінченна множина, тобто містить скінченну кількість елементарних подій. Коли ж Ω є нескінченна множина (більш того, незліченна), то через ряд причин, пов'язаних з наступним введенням імовірнісної міри, доводиться оперувати не з усіма підмножинами множини Ω, а лише з деякими більш вузькими класами цих підмножин, що мають назву алгебр та σ-алгебр. Останні замкнені з точки зору введених теоретико-множинних операцій.

<u>Означення 1.</u> Клас F підмножин простору Ω зветься <u>алгеброю множин,</u> якщо:

1. $\Omega \in F$,
2. з $A \in F$ виходить $\overline{A} \in F$,
3. з $A \in F$ та $B \in F$ виходить $A \cup B \in F$.

   Наслідком означення алгебри є її наступні властивості:

4. з властивостей 1) та 2) виходить $\varnothing \in F$;
5. якщо $A \in F$ та $B \in F$, то $A \cap B \in F$. дійсно, завдяки 2) маємо $\overline{A} \in F$, $\overline{B} \in F$, а завдяки 3) $\overline{A} \cup \overline{B} \in F$; далі, за властивістю 2) та за формулою де Моргана $\overline{\overline{A} \cup \overline{B}} = A \cap B \in F$;
6. якщо $A \in F$ та $B \in F$, то $A \backslash B \in F$ дійсно, завдяки 2) та 5) маємо $A \cap \overline{B} \in F$; але $A \cap \overline{B} = A \backslash B$;

властивості 3) та 5) розповсюджуються на довільну скінченну кількість доданків та співмножників (вказівка: для доведення використайте принцип математичної індукції).

<u>Означення 2.</u> Клас F підмножин простору Ω називається <u>σ-алгеброю множин</u>, якщо

1. $\Omega \in F$;

2. з $A \in F$ виходить $\overline{A} \in F$;

3. з $A_i \in F$, $i=\overline{1,\infty}$, виходить $\bigcup_{i=1}^{\infty} A_i \in F$ (і, таким чином, властивість алгебри

поширюється на нескінченне число доданків).

Очевидно, σ-алгебра є алгеброю, але зворотне твердження, взагалі кажучи, не вірне.

Використовуючи формули де Моргана довести, що коли F є σ-алгеброю, то з $A_i \in F$, $i=\overline{1,\infty}$, виходить $\bigcap_{i=1}^{\infty} A_i \in F$.

Найпростішими прикладами алгебр та σ-алгебр виявляються наступні.

1. Нехай Ω - простір елементарних подій. Тоді алгеброю та σ-алгеброю буде система множин $F=\{\Omega, \varnothing\}$. Ця "найхудіша" σ-алгебра мало цікавить нас з практичної точки зору, тому що містить події, відносно яких все відомо наперед (одна з них відбувається завжди, інша – ніколи не відбувається).

2. Нехай A – деяка підмножина множини Ω. Тоді $F=\{\Omega, \varnothing, A, \overline{A}\}$ є алгеброю та σ-алгеброю.

3. Ще один приклад алгебри та σ-алгебри - система всіх підмножин множини Ω. Як вже відмічалося, ця σ-алгебра використовується для скінченних Ω.

Розглянемо так звану σ-алгебру борелівських множин[4] для $\Omega=R^1=(-\infty,+\infty)$, яка часто використовується у різних імовірнісних побудовах.

Спочатку введемо наступне означення.

4 Названа ім'ям автора – Бореля Фелікса Едуарда Жюстена Еміля (1871 -1956), французького математика, члена Паризької АН.

Означення 3. Нехай А – деякий непорожній клас підмножин множини Ω. Тоді σ-алгебра σ (А) називається мінімальною σ-алгеброю, що містить клас А, якщо:

а) $A \subseteq \sigma(A)$;

б) для довільної σ-алгебри σ такої, що $A \subseteq \sigma$, має місце $\sigma(A) \subseteq \sigma$.

Має місце наступна теорема.

Теорема. Для довільного непорожнього класу множин А існує єдина мінімальна σ-алгебра σ (А), що містить клас А.

Доведення. Перевіримо, що в якості σ (А) можна взяти перетин усіх σ-алгебр, що містять клас А. По-перше, цей перетин непорожній тому, що існує хоча б одна σ-алгебра, що містить клас А (це σ-алгебра всіх підмножин множини Ω). По-друге, перетин довільного числа σ-алгебр є σ-алгеброю, тобто σ (А) - σ-алгебра. Нарешті, якщо σ - довільна σ - алгебра, що містить А, то $\sigma(A) \subseteq \sigma$ і таким чином, σ (А) є мінімальна σ - алгебра, що містить клас А.

Візьмемо тепер $\Omega = R^1 = (-\infty, +\infty)$.

Означення 4. Мінімальна σ-алгебра В, що містить клас А всіх інтервалів виду [а, в] (де а < в), називається σ-алгеброю борелівських множин в R^1. Зауважимо, що вона не співпадає з множиною всіх підмножин дійсної вісі, але містить у собі достатньо багатий клас множин, з якими ми маємо справу при вирішенні практичних задач.

- Вона містить у собі довільну одноточкову множину $\{a\}$. Дійсно,

$$\{a\} = \bigcap_{n=1}^{\infty} \left[a, a+\frac{1}{n}\right] \in \sigma(A) = B .$$

- Вона містить у собі множину натуральних чисел N та множину раціональних чисел Q (довести самостійно).

- Вона містить у собі довільні інтервали виду [а, в), (а, в], (а, в), а також всілякі скінченні або зліченні об'єднання та перетин зазначених інтервалів та відповідні доповнення.

Дійсно, наприклад, $[а, в) = [а, в] \setminus \{в\} \in B$ за властивостями σ-алгебри.

Розглянемо чотири класи скінченних інтервалів: інтервали виду [а, в], [а, в), (а, в], (а, в).

У зв'язку з тим, що з елементів одного з чотирьох вказаних інтервалів за допомогою зліченного числа операцій можна одержати довільний елемент іншого класу [Гих., Ск., Ядр.], σ-алгебри борелівських множин, побудовані на кожному з них співпадають. Тому можна дати, наприклад еквівалентне означення.

Означення 4. Мінімальна σ-алгебра B, що містить клас A всіх інтервалів виду [а, в), називається σ-алгеброю борелівських множин в R.

Ми побудували σ-алгебру борелівських множин в R^1. Аналогічно будується борелівська σ-алгебра в R^n.

Означення 5. Мінімальна σ-алгебра B_n, що містить всі паралелепіпеди виду $[a_1, в_1) \times ... \times [a_n, в_n)$, називається σ-алгеброю борелівських множин

в R^n.

Наприклад, в R^2 це мінімальна σ-алгебра, що містить всі прямокутники $[a_1, в_1) \times [a_2, в_2)$.

1.5 Ймовірнісні міри

Для різних класів множин імовірнісні міри вводяться так, щоб вони мали ті властивості, що характерні для частості.

Означення 6. Нехай F – деяка алгебра підмножин з Ω. Функція множини $P(\cdot)$, що визначена на F, зветься скінченно-адитивною імовірнісною мірою на F, якщо вона задовольняє таким вимогам:

- $P(A)\geq 0$ для кожної A з F (вимога невід'ємності);
- $P(\Omega)=1$ (вимога нормування);
- $P(A\cup B)=P(A)+P(B)$, якщо $A\in F$, $B\in F$, $A\cap B=\varnothing$ (вимога адитивності).

Розглянемо найпростіші наслідки цього означення.

1. Пропонуємо методом математичної індукції самостійно одержати узагальнення властивості 3) на суму довільного скінченного числа множин $A_1, A_2,..., A_m$ з F:

$$P\left(\bigcup_{i=1}^{m} A_i\right)=\sum_{i=1}^{m} P(A_i)$$, якщо $A_i \cap A_j = \varnothing$, $i\neq j$; $i, j=\overline{1,m}$.

2. З останньої рівності та з 1) виходить, що

$$0 \leq P(A) \leq 1.$$

3. Очевидно, що з $\varnothing=\overline{\Omega}\in F$ та з доведеної рівності виходить

$$P(\varnothing) = 0.$$

4. Якщо $A\in F$, $B\in F$ та $B\subseteq A$, то

$$P(A\backslash B) = P(A) - P(B).$$

Дійсно, у цьому випадку $(A\backslash B)\cup B=A$ та $(A\backslash B)$ і B несумісні. Звідси

$$P((A\backslash B)\cup B)= P(A\backslash B) + P(B) = P(A),$$

і властивість доведено.

5. З властивості 5) виходить, що коли $A\in F$, $B\in F$ та $B\subseteq A$, має міце

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

Означення 7. Функція множини $P(\cdot)$, що визначена на алгебрі F, називається зліченно-адитивною імовірнісною мірою на алгебрі F, якщо:

4. $P(A) \geq 0$ для кожної множини $A\in F$;
5. $P(\Omega) = 1$;

6. $P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P\left(A_i\right)$, якщо $A_i \in F$, i= $\overline{1,\infty}$;$A_i \cap A_j = \varnothing$, $i \neq j$, та $\overset{\infty}{\underset{i=1}{\text{¿}}} A_i$ $\in F$.

Таким чином, імовірнісна міра на алгебрі є зліченно-адитивною, якщо властивість адитивності поширюється на зліченні суми множин, що не перетинаються, коли ці суми належать алгебрі.

Деякі достатні умови того, що скінченно-адитивна міра на алгебрі F є зліченно-адитивною, дає наступна теорема [Гих., Ск., Ядр.].

<u>Теорема</u>. Нехай P(·) – скінченно-адитивна ймовірнісна міра на алгебрі F. Тоді наступні умови виявляються еквівалентними:

4. P(·) є зліченно-адитивна, тобто для $A_i \in F$, i= $\overline{1,\infty}$, таких, що $A_i \cap A_j = \varnothing$, $i \neq j$, $\overset{\infty}{\underset{i=1}{\text{¿}}} A_i$ $\in F$, має місце

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P\left(A_i\right);$$

5. Якщо $\{A_i\}$, i= $\overline{1,\infty}$ - довільна послідовність множин з F, що не спадає ($A_i \subseteq A_{i+1}$), та $\lim_{i\to\infty} = \overset{\infty}{\underset{i=1}{\text{¿}}} A_i$ $\in F$, то

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)=\lim_{i\to\infty} P\left(A_i\right);$$

6. Якщо $\{A_i\}$, i= $\overline{1,\infty}$ - довільна послідовність множин з F, що не зростає ($A_i \supseteq A_{i+1}$), та $\lim_{i\to\infty} = \overset{\infty}{\underset{i=1}{intersect}} A_i$ $\in F$, то

$$P\left(\overset{\infty}{\underset{i=1}{intersect}} A_i\right)=\lim_{i\to\infty} P\left(A_i\right);$$

7. 4) Якщо $\{A_i\}$, i= $\overline{1,\infty}$ - довільна послідовність множин з F, що не

зростає ($A_i \supseteq A_{i+1}$), та $\bigcap_{i=1}^{\infty} A_i = \varnothing$, то

$$\lim_{i \to \infty} P(A_i) = 0.$$

<u>Доведення.</u> Доведення проводиться за схемою: 1) ⇒ 2) ⇒ 3) ⇒ 4) ⇒ 1).

1. Перший крок (1) ⇒ 2)): нехай $\{A_i\}$, $i=\overline{1,\infty}$ - послідовність множин з F,

що не спадає, та $\bigcup_{i=1}^{\infty} A_i \in F$. Зведемо суму $\bigcup_{i=1}^{\infty} A_i$ до суми множин, що не перетинаються. Покладемо $A_o = \varnothing$, $B_i = A_i \setminus A_{i-1}$, $i=\overline{1,\infty}$, зауважимо, що за властивостями алгебри має місце $B_i \in F$. Далі,

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i \in F,$$

де $B_i \cap B_j = \varnothing$, $i \neq j$.

Тому (з урахуванням зліченної адитивності міри $P(\cdot)$) одержуємо

$$P(\lim_{n\to\infty} A_n) = P(\bigcup_{i=1}^{\infty} A_i) = P(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} [P(A_i) - P(A_{i-1})] = \lim_{n\to\infty} P(A_n)$$

1) Другий крок (2) ⇒ 3)): нехай $\{A_i\}$, $i=\overline{1,\infty}$ - послідовність множин з F, що не зростає та $\bigcap_{i=1}^{\infty} A_i \in F$. Тоді $\{A_i\}$ – послідовність множин, що не спадає, та

$$\overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \overline{A_i} \in F; \quad \overline{A_i} \in F, \; i = \overline{1,\infty}$$

,

за властивостями алгебри. Тоді з 2) маємо

$$P(\lim_{n\to\infty} A_n) = P(\bigcap_{i=1}^{\infty} A_i) = 1 - P(\overline{\bigcap_{i=1}^{\infty} A_i}) = 1 - P(\bigcup_{i=1}^{\infty} \overline{A_i}) = 1 - \lim_{n\to\infty} P(\overline{A_n}) = \lim_{n\to\infty} [1 - P(\overline{A_n})] = \lim_{n\to\infty} P(A_n)$$

.

2) Третій крок (3) $\Rightarrow$ 4)): нехай $\{A_i\}$, i= $\overline{1,\infty}$ - послідовність множин з F, що

не зростає та $\underset{i=1}{\overset{\infty}{intersect}} A_i$ $=\varnothing$.

Тривіально виходить з властивості 3)

$$P\left(\underset{i=1}{\overset{\infty}{intersect}} A_i\right)=P(\varnothing)=0=\lim_{n\leftarrow\infty} P(A_n)$$.

3) Четвертий крок (4) $\Rightarrow$ 1)): нехай $\{A_i\}$, i= $\overline{1,\infty}$ - послідовність

множин з F, що не перетинаються (A_i A_j $=\varnothing$, $i\neq j$), та $\underset{i=1}{\overset{\infty}{¿}} A_i$ $\in$F. Означимо B_n як

$$B_n=\underset{i=n+1}{\overset{\infty}{¿}} A_i$$.

Тоді $\{B_n\}$ – послідовність множин з F (це виходить з властивостей алгебри F); послідовність, що не зростає ($B_n \supseteq B_{n+1}$), та $\underset{n=1}{\overset{\infty}{intersect}} B_n$ $=\varnothing$. Останнє доведемо від супротивного. Нехай існує $\omega \in \underset{n=1}{\overset{\infty}{intersect}} B_n$. Тоді $\omega \in B_1 =$ $\underset{i=2}{\overset{\infty}{¿}} A_i$, і таким чином, існує к таке, що $\omega \in A_к$, та не належить $B_к =$ $\underset{i=k+1}{\overset{\infty}{¿}} A_k$, (тому що A_i - несумісні). Одержали протиріччя.

За припущенням $P\left(\underset{n=1}{\overset{\infty}{intersect}} B_n\right)=\lim P(B_n)=P(\varnothing)=0$. Треба довести, що

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i)$$.

Із скінченної адитивності виходить:

$$P(\overset{\infty}{\underset{i=1}{\cup}} A_i) = P(\overset{n}{\underset{i=1}{\cup}} A_i) + P(B_n) = \sum_{i=1}^{n} P(A_i) + P(B_n)$$.

Оскільки $\lim_{n\to\infty} P(B_n) = 0$, маємо:

$$P(\overset{\infty}{\underset{i=1}{\cup}} A_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{\infty} P(A_i)$$.

Теорему доведено.

<u>Означення 8.</u> Нехай F – деяка σ - алгебра підмножин з Ω. Функція множини P(·), що визначена на F, зветься ймовірністю, якщо вона володіє наступними властивостями:

1. P(A) ≥ 0 для кожної A ∈ F;

2. P(Ω) = 1;

3. якщо $\{A_i\}$ – послідовність (подій) $A_i \in F$, i= $\overline{1,\infty}$, така, що $A_i \cap A_j = \varnothing$,

i≠j, то $P(\overset{\infty}{\underset{i=1}{\cup}} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Зрозуміло, що ймовірність володіє доведеними властивостями скінченно-адитивної ймовірнісної міри.

Твердження, що задають властивості σ - алгебри та властивості ймовірності, складають систему аксіом теорії ймовірностей А.М.Колмогорова.

Трійка (Ω, F, P), де F є σ - алгеброю підмножин з Ω, а P(·) – ймовірнісна міра на F, зветься <u>ймовірнісним простором</u>.

1.6 <u>Приклади побудови ймовірнісних просторів</u>

1.6.1 <u>Простір із скінченною або зліченною множиною елементарних подій</u>

Нехай внаслідок імовірнісного експерименту (випробування) може відбутися скінченна або зліченна кількість елементарних подій ω_1, ω_2,...., ω_i,....

Таким чином, простір елементарних подій Ω є:

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

Випадковими подіями в цьому випадку будемо вважати довільні підмножини множини Ω. Таким чином, F у цій схемі – множина всіх підмножин з Ω.

Кожній елементарній події ω_i нехай відповідає деяка величина з$_i$, яку будемо називати ймовірністю елементарної події ω_i, якщо вона володіє наступними властивостями:

1) $p_i \geq 0$;

2) $\sum_{i=1}^{\infty} p_i = 1$.

Розглянемо довільну складну подію $A \in F$.

<u>Означення.</u> Ймовірністю складної події А будемо називати величину

$$P(A) = \sum_{\omega_i \in A} p_i ,$$

тобто суму ймовірностей всіх елементарних подій, що входять до події А.

Самостійно переконайтеся, що така означена ймовірність володіє усіма властивостями аксіоматики А.М.Колмогорова.

<u>Приклад</u>: Нехай деякий стрілець стріляє по мішені до першого попадання. Будемо вважати, що ймовірність того, що він улучає у мішень при одному пострілі, дорівнює 0,3; не улучає – 0,7.

Розглянемо простір елементарних подій

$$\Omega = \{\omega_1, \omega_2, ..., \omega_i, ...\omega_\infty\},$$

де ω_i – елементарна подія: стрілець перший раз улучає у мішень у i-му пострілі;

ω_∞ - стрілець ніколи не улучає у мішень.

Множиною всіх подій вважаємо довільні підмножини множини Ω.

У цій схемі доцільно кожній елементарній події ω_i приписати ймовірність

$P(\omega_i) = p_i = (0,7)^{i-1} \cdot 0,3$;

$p_\infty = 0$.

І, таким чином, всі $p_i \geq 0$. Перевіримо, що $\sum_{i=1}^{\infty} p_i$ =1.

Дійсно, $\sum_{i=1}^{\infty} p_i$ = $(0,7)^{i-1} \cdot 0,3 = 0,3/(1-0,7) = 1$.

Розглянемо складну подію А – стрільцю для попадання у мішень знадобиться не більш як три патрони. Очевидно, $A = \{\omega_1, \omega_2, \omega_3\}$.

Тоді $P(A)=0,3+0,7\cdot0,3+(0,7)^2\cdot0,3=0,657$.

1.6.2 Класичне означення ймовірності

Розглянемо окремий випадок схеми, що наведена раніш. А саме, нехай

$\Omega = \{\omega_1, \omega_2,...., \omega_m\}$ складається з скінченної кількості елементарних подій, та усі елементарні події однаково можливі, тобто жодній з них неможливо віддати перевагу до випробування, отже їх можна вважати рівноймовірними. Припишемо кожній ω_i ймовірність 1/m.

Очевидно, $\sum_{i=1}^{m} P(\omega_i)=1$.

Нехай до деякої складної події А належить к елементарних подій (або події А сприяють к елементарних подій). Відповідно до розглянутого раніш події А треба приписати ймовірність

$$P(A)=\sum_{\omega_i \in A} P(\omega_i)=\frac{k}{m}.$$

Таке означення ймовірності у схемі скінченного числа рівноможливих елементарних подій називається класичним означенням імовірності. З вивчення подібних схем починався розвиток теорії ймовірностей на прикладі задач, пов’язаних з азартними іграми.

Приклад. В книжковому магазині розігрується лотерея. Всього покупцю пропонується 10 лотерейних білетів. Відомо, що серед них 3 виграшних та 7 невиграшних білетів. Покупець навмання бере 5 білетів. Обчислити ймовірність того, що серед них буде 2 виграшних (подія А).

Елементарною подією в цьому випробуванні, буде взяти деяке сполучення з 10 білетів по 5. Ясно, що таких сполучень (елементарних подій) є $m = c_{10}^5$. Таким чином, ймовірність однієї елементарної події дорівнює $p_i = \frac{1}{c^5{}_{10}}$. Події А сприяє випадок, коли покупець з 7 невиграшних білетів візьме 3 та з трьох виграшних білетів візьме 2. Згідно з принципом множення в комбінаториці таких випадків (сприятливих до події А елементарних подій) є $k=c^3{}_7 * c^2{}_3$.

За класичним визначенням ймовірності маємо:

$$P(A)=\frac{c_7^3 \cdot c_3^2}{c_{10}^5}=\frac{10}{24}.$$

1.6.3 Приклад побудови ймовірнісного простору для випробування з незліченною множиною елементарних подій

Раніш розглядалися приклади випробувань, множина елементарних подій ,яких була скінченною або зліченною. У якості σ- алгебри F ми брали множину всіх підмножин простору елементарних подій Ω. Кожній елементарній події ми приписували ймовірність p_i так, щоб виконувалася умова нормування $\sum_{i=1}^{\infty} p_i = 1$. Але, як ми побачимо далі, такий підхід з ряду причин неможливо реалізувати для випробувань з незліченною множиною елементарних подій.

Розглянемо приклад, який ми взяли з книги [Гихм., Скор., Ядр.]

На інтервал [0,1) навмання кидається точка. Простір Ω=[0,1) має континуум елементарних подій. Слово “навмання” означає, що ніяка елементарна подія ω (точка з інтервалу [0,1)) не має переваг перед іншими елементарними подіями у сенсі можливості її наставання. Досі все схоже з класичною схемою рівноможливих елементарних подій.

Але, якщо спробувати приписати кожній елементарній події ω ненульову ймовірність, то ми одразу одержимо протиріччя. Дійсно, нехай $P(\omega) = c > 0$ для кожної $\omega \in [0,1)$. Візьмемо тільки зліченну множину раціональних чисел з цього інтервалу і просумуємо ймовірності попадання у відповідні точки. Ми одержимо, що ймовірність попадання у ці точки дорівнює нескінченності, тобто явно не виконується властивість ймовірності $P(\Omega) = 1$. Таким чином, оскільки усі ω “однаково можливі”, маємо $P(\omega) = 0$ для кожної $\omega \in [0,1)$.

Зауважимо, що рівність нулевій ймовірності події не означає її неможливість. При проведені випробування обов’язково настає якась ω з [0,1).

Тому у цьому випробуванні ненульові ймовірності приписуються не окремим точкам, а деяким їх множинам, наприклад, інтервалам, що належать [0,1). З точки зору рівні можливості інтервалам однакової довжини доцільно приписувати однакові ймовірності, а інтервал, який має більшу довжину ніж інший, повинен мати більшу ймовірність попадання в нього. З точки зору побудови σ- алгебри клас множин, яким приписуються ймовірності повинен бути замкнений відносно теоретико-множинних операцій.

Імовірнісний простір в цій схемі будується наступним чином. Розглянемо клас К всіх інтервалів $[a, b) \subseteq [0,1)$, а також скінченних об'єднань інтервалів $\overset{m}{\underset{i=1}{¿}}$ $[a_i, b_i) \subseteq [\,0,1)$, що не перетинаються. Клас К є алгебра, очевидно, що цій алгебрі належать також і певні зліченні суми інтервалів. Наприклад,

$\overset{\infty}{\underset{i=1}{¿}}$ $[a_i, b_i)$, де $a_{i+1} = b_i$. Нехай ß = σ(К)- мінімальна σ- алгебра, що містить клас К. Це σ- алгебра борелівських множин інтервалу [0,1). Імовірність Q(¿) попадання точки в інтервал [a,b),означимо як (b-a), а для об'єднання інтервалів

$\overset{m}{\underset{i=1}{¿}}$ $[a_i, b_i) \subseteq [0,1)$, що не перетинаються, ймовірністю будемо вважати

$\sum_{i=1}^{m}$ $(b_i - a_i)$ (тут m може бути не скінченним). Зліченно-адитивна ймовірнісна міра на К побудована.

Далі скористуємось теоремою, яку наведемо без доведення.

Теорема Каратеодорі [5] про продовження міри.

Нехай Q(·) – зліченно-адитивна ймовірнісна міра на деякій алгебрі К. Тоді існує єдина ймовірнісна міра P(·), що означена на мінімальній σ-алгебрі ß=σ(К), яка містить К, і є продовженням міри Q(·), тобто міра P(·) така, що P(A) = Q(A) для кожного A∈К.

Якщо повернутися до нашого прикладу, то можна помітити, що згідно з теоремою про продовження міри існує така єдина міра P(·) на ß=σ(К), що

5 Каратеодорі Костянтин (1873-1950)- німецький математик, професор Мюнхенського університету

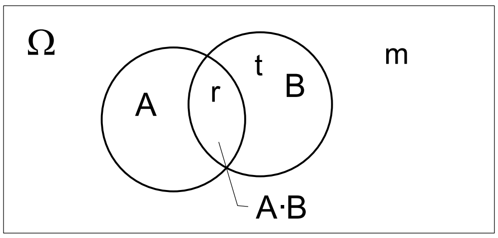
P([a,b)) = b-a. Випадковими подіями будемо вважати борелівські множини з [0,1), а для кожної події A $\in$ ß ймовірністю назвемо P(A). Імовірнісний простір (Ω,ß,P)побудовано.

Виникає питання, чому в якості σ- алгебри взята лише σ- алгебра борелівських множин з [0,1), а не система всіх підмножин множини [0,1). Без доведення зауважимо, що на останній не можливо побудувати таку імовірнісну міру P(·), щоб P([a,b))=b-aдля кожного [a,b) $\subset$ [0,1).

1.7Умовні ймовірності

Часто доводиться розглядати ймовірність наставання події A внаслідок випробування, якщо відомо що у цьому випробуванні відбулась подія B, що має додатню ймовірність. Цю умовну ймовірність позначають P(A/B). До того, як дати загальне формальне означення умовної ймовірності, розглянемо найпростіший приклад.

Нехай простір елементарних подій Ω складається з m рівноймовірних подій, подія B складається з елементарних подій, події AB сприяє r елементарних подій.



Зауважимо, що згідно класичному означенню ймовірності

P(B) = t/m, P(AB) = r/m.

Визначимо умовну ймовірність P(A/B). Відомо, що у цьому випробуванні подія B вже відбулася і, таким чином, множина усіх можливих у випробувані елементарних подій звузилась до t елементарних подій. Внаслідок рівно можливості, ймовірність кожної елементарної події, що належить B, тепер можна вважати рівною l/t. Події A сприяє тепер тільки r елементарних подій. Внаслідок рівно можливості природньо означити

$$P(A/B) = \frac{r}{t} = \frac{r/m}{t/m} = \frac{P(AB)}{P(B)} .$$

Розглянемо тепер більш загальну ситуацію, коли простір Ω не обов'язково складається з рівно можливих елементарних подій. Нехай проведено n випробувань і в n_B з них відбулася подія B, а в n_{AB} випробуваннях відбулася подія AB. Визначимо умовну частість наставання події A за умови, що відбулась подія B:

$$W_n(A/B) = \frac{n_{AB}}{n_B} = \frac{n_{AB}/n}{n_B/n} = \frac{W_n(AB)}{W_n(B)} .$$

Раніш вже відмічалося, що частість наставання події у певному розумінні збігається до ймовірності за кількості випробувань, що прямує до нескінченності.

Сказане дозволяє дати наступне загальне означення умовної ймовірності.

Означення. Нехай (Ω, F, P)- імовірнісний простір, $A \in F$, $B \in F$, $P(B) \neq 0$. Умовною ймовірністю події A за умови, що відбулася подія B, називається величина $P(A/B) = P(AB)/P(B)$.

При $P(B) = 0$ умовна ймовірність $P(A/B)$ не означена.

Можна переконатися, умовні ймовірності задовольняють наступним властивостям:

1) $P(A/B) \geq 0$;

2) $P(\Omega/B) = 1$;

3) $P(B/B) = 1$;

4) Якщо $\{A_i\}$- послідовність парно несумісних випадкових подій

$(A_i \cap A_j = \varnothing, i \neq j)$, то

$P(\bigcup_{i=1}^{\infty} A_i/B) = \sum_{i=1}^{\infty} P(A_i/B)$.

Властивості 1) – 3) перевірте самостійно.

Доведемо 4). Маємо:

$P(\bigcup_{i=1}^{\infty} A_i/B)=$

$$\frac{P((\bigcup_{i=1}^{\infty} A_i)\cap B)}{P(B)}=\frac{P(\bigcup_{i=1}^{\infty} A_i)\cap B))}{P(B)}=\frac{\sum_{i=1}^{\infty} P(A_i\cap B)}{P(B)}=\sum_{i=1}^{\infty}\frac{P(A_i\cap B)}{P(B)}=\sum_{i=1}^{\infty} P(A_i|B).$$

Якщо позначити через $F_в$ σ- алгебру усіх множин виду АВ, то з властивостей 1) – 4) виходить,що (В, $F_в$, Р(В)) – ймовірнісний простір.

З означення умовної ймовірності виходить

Р(АВ) = Р(В)Р(А/В) = Р(А)Р(В/А),

якщо Р(А)>0, Р(В)>0. Це співвідношення має назву <u>формули множення ймовірностей.</u>

Вона узагальнюється на випадок довільної кількості к множників. А саме нехай ми маємо систему випадкових подій $A_1, A_2 \ldots A_к$, і нехай

$P(A_1, A_2 \ldots A_{к-1})>0$.

Тоді $P(A_1, A_2 \ldots A_к) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_2A_1) \cdot \ldots \cdot P(A_к/A_1A_2 \ldots A_{к-1})$.

Перш за все зауважимо, що з $\bigcap_{i=1}^{k-1} A_i \subseteq \bigcap_{i=1}^{k-2} A_i \subseteq \ldots \subseteq A_1 \cap A_2 \subseteq A_1$,

та з $P(A_1, A_2 \ldots A_к)>0$ виходить, що

$P(\bigcap_{i=1}^{k-s} A_i)>0, s = \overline{1,k-1}$.

Отже усі умовні ймовірності $P(A_n/\bigcap_{i=1}^{n-1} A_i)$ означені.

Далі використовуємо метод математичної індукції. Раніш встановлено, що формула вірна при к=2. Нехай вона вірна для добутку (к-1) випадкових подій, тобто

$P(A_1, A_2...A_{к-1}) = P(A_1)\cdot P(A_2/A_1)\cdot P(A_3/ A_2A_1)\cdot...\cdot P(A_{к-1}/A_1 A_2...A_{к-2})$.

Позначимо $B= A_1, A_2...A_{к-1},$ так що $P(A_1, A_2...A_{к-1})=P(B)$.

Тоді $P(A_1, A_2...A_к) = P(A_к B) = P(B)\cdot P(A_к /B) =$

$P(A_1)P(A_2/A_1)...P(A_{к-1}/ A_1 A_2... A_{к-2})P(A_к/A_1 A_2...A_{к-1})$.

Таким чином, формула множення у довільному випадку доведена.

<u>Приклад</u>. На шістьох картках намальовані літери, з яких можна зіставити слово КАРЕТА. Картки перегортуються літерами до столу та добре перемішуються. Далі витягуються три картки по одній. Треба обчислити ймовірність того, що в порядку витягування ми одержимо слово РАК.

Введемо події:

Р - витягнути картку з літерою Р;

А – витягнути картку з літерою А;

К – витягнути картку з літерою К.

Нас цікавить ймовірність добутку цих подій РАК. За формулою множення ймовірностей маємо

$P(РАК)=P(Р)\cdot P(А/Р)\cdot P(К/РА)= \frac{1}{6}\cdot\frac{2}{5}\cdot\frac{1}{4}=\frac{1}{60}$.

1.8 <u>Незалежні події</u>

<u>Означення 1.</u> Нехай (Ω, F, P)- імовірнісний простір. Випадкові події A і B ($A\in F$, $B\in F$) називаються <u>незалежними</u>, якщо

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

Якщо ця рівність не виконується, то події називаються <u>залежними.</u>

Таке означення незалежних подій, на перший погляд, має чисто формальний характер. Якщо, одначе, відмовитись від розгляду випадкових подій нульової ймовірнісної міри, введене поняття незалежності має близьке до природних уявлень тлумачення. Доведемо наступне твердження.

Нехай P(B)>0. Випадкові події A і B незалежні тоді й тільки тоді, коли P(A/B)=P(A) (поява події B не змінює ймовірності події A).

Дійсно, якщо A і B незалежні та P(B)>0, то

P(A/B)= $\frac{P(A\cap B)}{P(B)}=\frac{P(A)P(B)}{P(B)}=P(A)$.

Якщо ж відомо, що P(A/B)= $\frac{P(A\cap B)}{P(B)}=P(A)$,

То $P(A\cap B) = P(A)P(B)$, а це означає, що A і B незалежні.

Таким чином, для подій ненульової ймовірнісної міри маємо еквівалентне означення незалежності: дві події A і B <u>незалежні</u> , коли умовна ймовірність наставання однієї з них, за умови, що друга відбулась, дорівнює безумовній імовірності першої події.

Мають місце наступні властивості незалежних подій: якщо A і B незалежні, то події A і $\overline{B}$, $\overline{A}$ і B, $\overline{A}$ і $\overline{B}$ також незалежні.

Доведемо, що A і $\overline{B}$ незалежні (решту довести самостійно). Зауважимо, що $A\overline{B} \cup AB = A$, $P(A\overline{B}) + P(AB) = P(A)$.

Тому $P(A\overline{B}) = P(A) - P(AB)$.

Якщо A і B незалежні, то

$P(A\overline{B}) = P(A) - P(A)P(B) = P(A)[1 – P(B)] = P(A)\,P(\overline{B})$.

Отже, A і $\overline{B}$ незалежні.

Довести самостійно, що у випадку, коли A і B незалежні та P(A)>0, P(B)>0, то A і B обов’язково сумісні.

<u>Означення 2.</u> Випадкові події $A_1, A_2, ..., A_n$, <u>незалежні у сукупності,</u> якщо для довільного k, $1 \le k \le n$, та для довільного набору індексів

$i_1, i_2, ..., i_k$, $1 \le i_1 < i_2 < ... < i_k \le n$,

$$P\left(\bigcap_{r=1}^{k} A_{i_r}\right) = \prod_{r=1}^{k} P\left(A_{i_r}\right).$$

Зокрема, якщо $A_1, A_2, ..., A_n$ незалежні у сукупності, то довільні дві події A_i і A_j, $i \ne j$, незалежні. Проте з парної незалежності незалежність у сукупності не випливає.

<u>Приклад</u> (С.Н. Бернштейна[6]). На площину кидають тетраедр, три грані якого пофарбовані відповідно у червоний, зелений, синій кольори, а на четверту грань нанесено усі три кольори.

Нехай подія Ч полягає у тому, що нижньою випадає грань,яка містить червоний колір, та нехай аналогічно означені події З і С. Так як кожний з трьох кольорів нанесено на дві грані, то

$$P(Ч) = P(З) = P(С) = 2/4 = 1/2 .$$

Далі

$P(Ч \cap З) = P(Ч \cap С) = p(З \cap С) = 1/4$,

І, отже події Ч, З, С парно незалежні. Але ці події не є незалежними у сукупності, тому що

$$P(Ч \cap З \cap С) = 1/4 \ne P(Ч)\,P(З)\,P(С) = 1/8.$$

Нехай для довільного набору вказаних індексів $i_1, i_2, ... i_к$ $P\left(\bigcap_{r=1}^{k} A_{i_r}\right) > 0$. Тоді для довільного $j \ne \overline{i_1, i_k}$, $1 \le j \le n$

$$P\left(A_j / A_{i_1} A_{i_2} \dots A_{i_k}\right) = P\left(A_j\right) . \quad (1)$$

Дійсно, з належності у сукупності $\{A_i\}$ маємо:

---

6 Бернштейн Сергій Натанович (1880 – 1968) – радянський математик, академік АН СРСР, академік АН УРСР

$$P(A_j / A_{i_1} A_{i_2} \cdots A_{i_k}) = \frac{P(A_j A_{i_1} A_{i2} \ldots A_{i_k})}{P(A_{i_1} A_{i_2} \ldots A_{i_k})} = \frac{P(A_j) \prod_{r=1}^{k} P(A_{i_r})}{\prod_{r=1}^{k} P(A_{i_r})} = P(A_j).$$

Навпаки, якщо для довільних подій з $\{A_i\}$ має місце (1), то A_i, i= $\overline{1,n}$, незалежні у сукупності (довести самостійно, використавши індукцію по k).

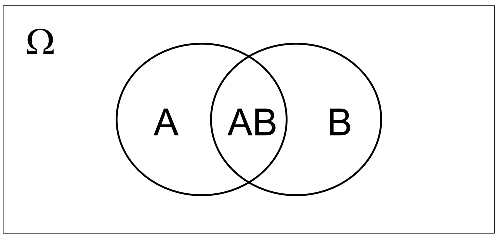
Таким чином для подій A_i, i= $\overline{1,n}$, таких, що для довільної послідовності індексів i_1, i_2, ... , ik $P\left(\bigcap_{r=1}^{k} A_{i_r}\right)$ >0, можна дати еквівалентне означення незалежності: події A_i, i= $\overline{1,n}$, <u>незалежні у сукупності</u>, якщо ймовірність наставання однієї (довільної) з них, за умови, що відбулась довільна сукупність інших подій з $\{A_i\}$, дорівнює безумовній імовірності.

1.9 <u>Формула додавання ймовірностей</u>

Згідно з третьою аксіомою теорії ймовірності імовірність суми несумісних подій дорівнює сумі ймовірностей доданків. Проте, коли доданки сумісні, згадане співвідношення, взагалі кажучи, не має місця.

Розглянемо дві події A і B. Доведемо, що:

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



Зауважимо, що $A \cup B$ можна зобразити як суму трьох несумісних доданків

$$A \cup B = (A \setminus AB) \cup AB \cup (B \setminus AB),$$

та $AB \subseteq A$, $AB \subseteq B$.

Звідси

P(A∪B)=P(A\AB)+P(AB)+P(B\AB)=P(A)-P(AB)+P(AB)+P(B)-P(AB)=

=P(A)+P(B)-P(AB).

Використовуючи метод математичної індукції, самостійно доведіть формулу додавання для довільної кількості доданків:

$$P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} P(A_iA_j)+\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{s=j+1}^{n} P(A_iA_jA_s)-\ldots+(-1)^{n-1}P(A_1A_2\ldots A_n).$$

<u>Приклад</u>: З колоди в 32 карти навмання витягується одна карта. Треба обчислити ймовірність того, що це буде дама або піка.

Вводимо події:

Д – витягнути даму;

П – витягнути піку.

Нас цікавить ймовірність суми Д∪П. За формулою додавання ймовірностей маємо:

$$P(Д\cup П)=P(Д)+P(П)-P(ДП)=\frac{4}{32}+\frac{8}{32}-\frac{1}{32}=\frac{11}{32}.$$

1.10 <u>Формула повної ймовірності</u>

Розглянемо систему к подій $B_1, B_2,\ldots, B_к$, які ще називаються гіпотезами.

Нехай $B_i \cap B_j=\varnothing$, $i\neq j$, тобто події системи парно несумісні, та нехай $P(B_i)>0$, $i=\overline{1,k}$. Якщо деяка подія A задовольняє умові $A=A(\bigcup_{i=1}^{k} B_i)$ (сума B_i накриває A), то має місце <u>формула повної ймовірності</u>

$$P(A)=\sum_{i=1}^{K} P(B_i)P(A/B_i).$$

Дійсно, якщо $B_1, B_2, ..., B_к$ несумісні, то несумісні $AB_1, AB_2, ..., AB_к$. Тоді з третьої аксіоми теорії ймовірностей та формули множення ймовірностей виходить

$$P(A) = P(A(\bigcup_{i=1}^{k} B_i)) = P(\bigcup_{i=1}^{k} AB_i) = \sum_{i=1}^{k} P(AB_i) = \sum_{i=1}^{k} P(B_i) P(A/B_i).$$

<u>Приклад:</u> Маємо урни трьох складів.

1) 5 урн складаються з 6 білих та 5 чорних куль.

2) 3 урни складаються з 10 білих та 1 чорної кулі.

3) 7 урн складаються з 1 білої та 10 чорних куль.

Навмання вибирається урна, а потім навмання з неї виймається куля.

Визначити ймовірність, що ця куля біла.

Вводимо гіпотези (вони несумісні і їх сума дорівнює Ω):

B_1 – вибрана урна першого складу;

B_2 – вибрана урна другого складу;

B_3 – вибрана урна третього складу.

Подія А – куля біла. Обчислюємо ймовірність гіпотез та ймовірність події А за умови, що мала місце та чи інша гіпотеза:

$P(B_1) = \frac{5}{15}$; $P(B_2) = \frac{3}{15}$; $P(B_3) = \frac{7}{15}$;

$P(A/B_1) = \frac{6}{11}$; $P(A/B_2) = \frac{10}{11}$; $P(A/B_3) = \frac{1}{11}$.

За формулою повної ймовірності:

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + P(B_3)P(A/B_3) =$$

$$\frac{5}{15} \cdot \frac{6}{11} + \frac{3}{15} \cdot \frac{10}{11} + \frac{7}{15} \cdot \frac{1}{11} = \frac{67}{165}.$$

1.11 <u>Формула Байєса</u>[7]

7 Байєс (Бейєс) Томас (1702-1761) – англійський математик, член Лондонського Королівського товариства.

Постановка задачі схожа на ту, що приводила до використання формули повної ймовірності, але є суттєві відмінності в постановці запитання. Докладніше, знову є система к подій B_1, B_2,..., $B_к$. Події несумісні, тобто $B_i \cap B_j = \varnothing$, $i \neq j$, та мають ненульові ймовірності $P(B_i)>0$, $i=\overline{1,k}$. Є подія А з ненульовою ймовірністю $P(A)>0$, яка задовольняє умові $A=A(\bigcup_{i=1}^{k} B_i)$. Відомо, що в результаті випробування подія А реалізувалась. Треба обчислити умовні ймовірності $P(B_i/A)$, $i=\overline{1,k}$. Ймовірність подій $P(B_i)$, $i=\overline{1,k}$, називаються <u>апріорними</u> (до випробування). Умовні ймовірності $P(B_i/A)$, $i=\overline{1,k}$, називаються <u>апостеріорними</u> (після випробування) ймовірностями подій

B_i, $i=\overline{1,k}$.

Відповідь на постановку задачі дає <u>формула Байєса</u>

$$P(B_i/A) = \frac{P(B_i)P(A/B_i)}{\sum_{j=1}^{k} P(B_j)P(A/B_j)}, \quad i=\overline{1,k}.$$

Доведення просте. За формулою множення ймовірностей:

$$P(AB_i)=P(A)\,P(B_i/A)=P(B_i)\,P(A/B_i).$$

Звідси

$$P(B_i/A)=\frac{P(B_i)P(A/B_i)}{P(A)},$$

а далі в знаменнику треба використати формулу повної ймовірності і одержати формулу Байєса.

<u>**Приклад**</u>: Умови задачі ті ж, що і в попередньому прикладі. Але тепер відомо, що після вибору урни і витягування з неї кулі вона виявилась білою (перша відмінність), тобто подія А реалізувалась в результаті

випробування. Треба обчислити ймовірності $P(B_i/A)$, $i=\overline{1,3}$, того, що була вибрана урна першого, другого, третього складу (друга відмінність).

Застосуємо формулу Байєса, використав вже підраховані ймовірності. Одержуємо:

$$P(B_1/A)=\frac{P(B_1)P(A/B_1)}{P(A)}=\frac{\frac{5}{15}\cdot\frac{6}{11}}{\frac{67}{165}}=\frac{30}{67};$$

$$P(B_2/A)=\frac{P(B_2)P(A/B_2)}{P(A)}=\frac{\frac{3}{15}\cdot\frac{10}{11}}{\frac{67}{165}}=\frac{30}{67};$$

$$P(B_3/A)=\frac{P(B_3)P(A/B_3)}{P(A)}=\frac{\frac{7}{15}\cdot\frac{1}{11}}{\frac{67}{165}}=\frac{7}{67}.$$

Результат може здатися дещо несподіваним через суттєву різницю білих куль у урнах різних складів. Але все стає зрозумілим, якщо врахувати кількість урн того чи іншого складу.

1.12 Композиція незалежних випробувань

1.12.1 Композиція двох випробувань

Розглядаються два випробування, кожне з яких має скінченну кількість елементарних подій.

Нехай (Ω_1, F_1, P_1) та (Ω_2, F_2, P_2) – ймовірнісні простори, що відповідають цим випробуванням, $\Omega_1=\{E_1, E_2,..., E_{m_1}\}$, $\Omega_2=\{Q_1, Q_2,..., Q_{m_2}\}$, F_k складається з усіх підмножин множини $\Omega_к$, а ймовірності складних подій $C_1\subseteq\Omega_1$ і $C_2\subseteq\Omega_2$ у випробуваннях

$$P_1(C_1)=\sum_{E_i\in C_1} P_1(E_i), P_2(C_2)=\sum_{Q_j\in C_2} P_2(Q_j)$$

задаються за допомогою ймовірностей елементарних подій $P_1(E_i)$, i= $\overline{1,m_1}$, та $P_2(Q_i)$, j= $\overline{1,m_2}$.

Зауважимо, що $P_1(\cdot)$ і $P_2(\cdot)$ – взагалі кажучи ,різні ймовірнісні міри.

Композицією двох випробувань зветься складне випробування, що полягає у проведенні першого та другого випробування, елементарними подіями якого є всілякі можливі пари виду $(E_i,\ Q_j)$, i= $\overline{1,m_1}$, j= $\overline{1,m_2}$. Таким чином, простір елементарних подій композицій двох випробувань є прямий добуток просторів Ω_1 і Ω_2 ,тобто

$$\Omega = \Omega_1 \text{ x } \Omega_2 .$$

Приклад: Перше випробування полягає у підкиданні симетричної монети, $\Omega_1=\{O, P\}$, друге випробування – підкидання грального кубика,

$\Omega_2=\{Q_1, Q_2,...,Q_6\}$. Простір елементарних подій композиції цих двох випробувань –

$\Omega=\{(O, Q_1),..., (O, Q_6), (P, Q_1),..., (P, Q_6)\}$.

Як складні події для композиції розглядаються усі можливі підмножини множини Ω, тобто усілякі теоретико-множинні суми пар виду $(E_i,\ Q_j)$ (та неможлива подія $\varnothing$). Таким чином Ω і F для композиції побудовані.

Для повного завдання ймовірнісного простору композиції необхідно ще означити ймовірнісну міру $P(\cdot)$ на F. Але, взагалі кажучи, тільки по $P_1(\cdot)$ і $P_2(\cdot)$ це зробити неможливо.

Розглянемо окремий випадок композиції.

Два випробування називаються незалежними, якщо для всіх пар ймовірність $P(\cdot)$ у композиційному просторі визначається співвідношенням:

$$P(E_i,\ Q_j)=P_1(E_i)P_2(Q_j).$$

Випробування незалежні, якщо неоднозначність результату кожного випробування визначена незв’язними між собою групами випадкових факторів, результат одного випробування не впливає на результат другого.

Ймовірність складної події С для композиції незалежних випробувань означимо як Р(С)= $\sum_{(E_i,Q_j)\in C} P(E_i,Q_j) = \sum_{(E_i,Q_j)\in C} P_1(E_i)P_2(Q_j)$.

Ймовірність Р, що так побудована, називається <u>прямим добутком</u> імовірностей P_1 та P_2 та позначається

$$P = P_1 \times P_2 .$$

Загальна структура незалежних подій у композиційному просторі (Ω, F, P) має наступний вигляд .

Розглянемо подію $A \in F$

$$A = \underset{l=1}{\overset{s_1}{\text{¿}}}\ \underset{j=1}{\overset{m_2}{\text{¿}}}(E_{i_l},Q_j) ,$$

Тобто внаслідок першого випробування відбувається одна з подій $E_{i_1}, E_{i_2}, \ldots, E_{i_{s_1}}$, $s_1 \leq m_1$ (подія $A_1 = \underset{l=1}{\overset{s_1}{\text{¿}}} E_{i_l} \text{¿} F_1$), а внаслідок другого випробування відбувається усе що завгодно, тобто одна з подій $Q_1, Q_2, \ldots, Q_{m_2}$ (вірогідна подія $\Omega_2 = \underset{j=1}{\overset{m_2}{\text{¿}}} Q_j$).

Нехай складна подія $B \in F \in$

$$B = \underset{i=1}{\overset{m_1}{\text{¿}}}\ \underset{t=1}{\overset{s_2}{\text{¿}}}(E_i,Q_{j_t}) ,$$

тобто внаслідок першого випробування відбувається усе що завгодно, тобто одна з подій E_1, E_2,..., E_{m_2} (вірогідна подія $\Omega_1 = \overset{m_1}{\underset{i=1}{¿}} E_i$), а внаслідок другого випробування відбувається одна з подій $Q_{j_1}, Q_{j_2}, \dots, Q_{j_{s_2}}$, $s_2 \le m_2$

(подія $B_2 = \overset{s_2}{\underset{t=1}{¿}} Q_{j_t} ¿ F_2$).

Можна коротко записати

$$A = A_1 \times \Omega_2,\ B = \Omega_1 \times B_2.$$

Переконаємось, що події A і B незалежні:

$$P(AB) = P(A) \cdot P(B).$$

Дійсно,

$$P(A) = \sum_{l=1}^{s_1} \sum_{j=1}^{m_2} P(E_{i_l}, Q_j) = \sum_{l=1}^{s_1} \sum_{j=1}^{m_2} P_1(E_{i_l}) P_2(Q_j) = \sum_{l=1}^{s_1} P_1(E_{i_l}) \sum_{j=1}^{m_2} P_2(Q_j) = \sum_{l=1}^{s_1} P_1(E_{i_l}) = P_1(A_1)$$

Аналогічно можна переконатись, що

$$P(B) = \sum_{l=1}^{s_2} P_2(Q_{j_t}) = P_2(B_2).$$

Очевидно,

$$AB = \sum_{l=1}^{s_1} \sum_{t=1}^{s_2} (E_{i_e}, Q_{j_t}) = A_1 ¿ B_2.$$

Тоді

$$P(AB) = \sum_{l=1}^{s_1} \sum_{t=1}^{s_2} P(E_{i_l}, Q_{j_t}) = \sum_{l=1}^{s_1} \sum_{t=1}^{s_2} P_1(E_{i_l}) P_2(Q_{j_t}) = \sum_{l=1}^{s_1} P_1(E_{i_l}) \sum_{t=1}^{s_2} P_2(Q_{j_t}) = P(A) \cdot P(B) = P_1(A_1) \cdot P_2(B_2)$$

,що і треба було довести.

1.12.2 Композиція n незалежних випробувань

Зовсім аналогічно будується схема композиції n незалежних випробувань.

Нехай здійснюється n випробувань із скінченими просторами елементарних подій

$$\Omega_i = \{ E_1^i, E_2^i, \ldots, E_{m_i}^i \}, \quad i=\overline{1,n}.$$

Складними подіями у кожному i-му випробуванні є всілякі підмножини множини Ω_i, що визначає F_i. Задані на кожному F_i ймовірнісні міри $P_i(\cdot)$. І, таким чином, маємо n імовірнісних просторів (Ω_i, F_i, P_i), $i=\overline{1,n}$.

Результат спільного проведення n випробувань вважаємо результатом проведення одного складного випробування (<u>композиції</u>).

Композиційний простір Ω будуємо як прямий добуток просторів Ω_i, $i=\overline{1,n}$,

$$\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n,$$

тобто у якості елементарних подій композиції беруться всілякі n-ки виду

$$(E_{j_1}^1, E_{j_2}^2, \ldots E_{j_n}^n); \quad j_1 = \overline{1, m_1}; \ldots; j_n = \overline{1, m_n}.$$

Складними подіями в композиційному просторі є теоретико-множинні суми вказаних n-ок.

Випробування називаються <u>незалежними</u>, якщо ймовірність кожної n-ки визначається співвідношенням

$$P(E_{j_1}^1, E_{j_2}^2, \ldots E_{j_n}^{\bar{n}}) = \prod_{i=1}^{n} P_i(E_{j_i}^i).$$

Ймовірності складних подій в композиційному просторі означаються звичним для дискретних просторів способом – як сума ймовірностей n-ок, що належать даній складній події.

Побудована таким чином імовірність P для подій композиційного простору Ω називається <u>прямим добутком імовірностей</u> P_i і позначається

$$P = P_1 \times P_2 \times \ldots \times P_n.$$

В побудованому просторі (Ω, F, P) виділимо клас подій A, що називаються прямокутниками. Нехай $A_i \in F_i$, $i=\overline{1,n}$. Прямокутник.

$$A = A_1 \times A_2 \times \ldots \times A_n$$

містить ті та й тільки ті n-ки E= $(E_{j_1}^1, E_{j_2}^2, \ldots, E_{j_n}^n)$, у яких $E_{j_i}^i ¿ A_i,\ i=\overline{1,n}$.

З означення ймовірності для композиції незалежних випробувань виходить, що ймовірність прямокутника A дорівнює:

$$P(A)=\sum_{E\in A} P(E)=\sum_{E_{j_1}^1\in A_1}\ldots\sum_{E_{j_n}^n\in A_n} P_1(E_{j_1}^1)\ldots P_n(E_{j_n}^n)=\sum_{E_{j_1}^1\in A_1} P_1(E_{j_1}^1)\ldots\sum_{E_{j_n}^n} P_n(E_{j_n}^n)=P_1(A_1)P_2(A_2)\ldots P_n(A_n).$$

Загальна структура незалежних подій в композиційному просторі задається прямокутниками виду

$$A_i' = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \ldots \times \Omega_n,$$

тобто подіями, які можна інтерпретувати наступним чином: в i-му випробуванні відбувається подія A_i, а в інших випробуваннях відбувається що завгодно.

З одержаного раніш співвідношення для ймовірності прямокутника маємо

$$P(A_i') = P_i(A_i).$$

Тому що $A = A_1 \times A_2 \times \ldots \times A_n = \bigcap_{i=1}^{n} A_i'$, то з того ж співвідношення виходить

$$P\left(\bigcap_{i=1}^{n} A_i'\right) = \prod_{i=1}^{n} P(A_i'),$$

тобто при різних i події A_i' незалежні.

Через те що у прямокутниках $A_i^{'}$ в якості A_i можуть фігурувати Ω_i, то ясно, що мова йде про незалежність у сукупності.

1.13 Елементарні факти з теорії інформації

Розглянемо два імовірнісних простори із скінченою кількістю

елементарних подій (кожний задамо таблицею, у якої у верхньому рядку вказані елементарні події, у нижньому – ймовірності цих подій):

$$\left\{E_1\ E_2\ E_3\ E_4 ¿\right\} ¿\{\} \qquad \left\{E_1^{'}\ E_2^{'}\ E_3^{'}\ E_4^{'} ¿\right\} ¿\{\}$$

Очевидно, що ступінь невизначеності результату випробування (до випробування) у першому ймовірнісному просторі вище, ніж у другому. Дійсно, до випробування у першому просторі ні одній з подій не можна віддати перевагу, а у другому майже напевно відбудеться подія $E_3^{'}$.

Виникає питання: чи можна якимось чином міряти ступінь невизначеності, що властива ймовірнісному простору. Першим, хто запропонував вираз для міри невизначеності, був К. Шеннон[8].

Нехай деякий імовірнісний простір задано таблицею

$$\left\{E_1\ E_2 \ldots E_s ¿\right\} ¿\{\}.$$

Для вимірювання ступеню невизначеності, що притаманна даному просторові, вводять функцію

$$H(\Omega) = -\sum_{i=1}^{s} p_i \log_2 p_i,$$

що зветься <u>ентропією</u>; при цьому приймається, що при $p_к=0$ $p_k \log_2 p_k = 0$.

8 Шеннон Клод Ельвуд (нар.1916) - американський математик і інженер, професор Масачусетського технологічного інституту, член Національної АН США.

Ентропії притаманні такі простіші властивості.

1) Якщо ймовірнісний простір не має невизначеності, тобто якесь p_i=1, а

інші дорівнюють нулеві, то ентропія дорівнює нулеві (очевидно).

2) Коли елементарні події рівноймовірні, тобто $p_i=\frac{1}{s},\ i=\overline{1,s}$, тоді ентропія приймає максимальне значення.

Для доведення треба вирішити задачу максимізації H(Ω) за умови

$\sum_{i=1}^{s} p_i=1$ Для цього використаємо метод невизначених множників Лагранжа[9**]: будемо шукати екстремум функції

$$F=-\sum_{i-1}^{s} p_i \log_2 p_i+\lambda\left(\sum_{i=1}^{s} p_i-1\right).$$

Диференцюємо цю функцію по p_i та прирівнюємо частинні похідні до нуля. Одержуємо систему рівнянь

$$-\log_2 p_i-\log_2 e+\lambda=0,\quad i=\overline{1,s};$$

або

$$\log_2 p_i=\lambda-\log_2 e,\quad i=\overline{1,s}.$$

Звідси виходить, що екстремум (перевірте, що це максимум) досягається при рівних між собою pi. З умови нормування $\sum_{i=1}^{s} p_i=1$ маємо

$$p_i=\frac{1}{s},\ i=\overline{1,s}.$$

За одиницю невизначеності приймається невизначеність імовірнісного простору з двома рівноймовірними елементарними подіями

9 Лагранж Жозеф Луї (1736-1813) – французький математик, член Берлінської АН, Паризької АН, Петербурзької АН.

$$\left\{ E_1\ E_2 ¿ \right\} ¿\{\}\ ,$$

$$H = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = \log_2 2 = 1\ .$$

Ця одиниця називається <u>біт</u> (від англійського binary digit – двійкова одиниця).

Невизначеність результату випробування до випробування автоматично визначає інформативність результату випробування після випробування. Тому у бітах також вимірюється інформативність результату.

Розглянемо два ймовірнісних простори:

$$\Omega_1 \left\{ E_1\ E_2 \ldots E_{s_1} ¿ \right\} ¿\{\}\ ,\ \Omega_2 \left\{ Q_1\ Q_2 \ldots Q_{s_2} ¿ \right\} ¿\{\}\ .$$

Проведемо композицію двох випробувань. Для композиційного простору таблиця має наступний вигляд:

$$\Omega_1 \text{ x } \Omega_2 \left\{ \left(E_i, Q_j \right) ¿ \right\} ¿\{\}\ ,\quad i = \overline{1, s_1}\ ;\quad j = \overline{1, s_2}\ .$$

Максимальна ентропія композиційного ймовірнісного простору досягається тоді, коли випробування незалежні (хоча б на основі якісного аналізу). Обчислимо ентропію композиційного простору для випадку незалежних випробувань:

$$H(\Omega_1 x \Omega_2) - \sum_{i=1}^{s_1}\sum_{j=1}^{s_2} p_i q_j \log_2 p_i q_j = -\sum_{i=1}^{s_1}\sum_{j=1}^{s_2} p_i q_j \log_2 p_i - \sum_{i=1}^{s_1}\sum_{j=1}^{s_2} p_i q_j \log_2 q_j =$$

$$= -\sum_{i=1}^{s_1} p_i \log_2 p_i - \sum_{j=1}^{s_2} q_j \log_2 q_j = H(\Omega_1) + H(\Omega_2).$$

Таким чином, для композиції двох випробувань маємо:

$$0 \le H(\Omega_1 \text{ x } \Omega_2) \le H(\Omega_1) + H(\Omega_2).$$

1.14 Біномний розподіл

Деякі n випробувань називається <u>схемою випробувань Бернуллі[10]</u>, якщо вони незалежні в кожному з них відбувається подія A з імовірністю p=P(A), або подія A не відбувається (відбувається подія $\bar{A}$) з імовірністю q=1- p. Поставимо задачу: обчислити ймовірність $P_n(m)$ того, що в результаті n проведених випробувань подія A відбудеться рівно m разів, 0 ≤ m ≤ n.

<u>**Зауваження.**</u> Хоча подія A у випробуванні може бути складною подією, для даної задачі це не важливо. Тому ми будемо вважати її елементарною, а простір елементарних подій в кожному і-му випробуванні таким, що містить лише дві елементарні події A і $\bar{A}$, тобто $\Omega_i\{A, \bar{A}\}$, $i=\overline{1,n}$.

Розглянемо композицію n незалежних і побудуємо композиційний простір елементарних подій. Загальний вигляд елементарної події цього простору – це n-ка ($A^{\alpha_1}, A^{\alpha_2}, \ldots, A^{\alpha_n}$), де $\alpha_i = \overline{0,1}$, $i=\overline{1,n}$; $A^1=A$; $A^0=\bar{A}$. Ймовірність такої елементарної події зважаючи на незалежність випробувань дорівнює:

$$p^{\sum_{i=1}^{n}\alpha_i} \cdot q^{n-\sum_{n=1}^{n}\alpha_i} .$$

Розглянемо в композиційному ймовірнісному просторі складну подію B: в n випробуваннях подія A відбулася m разів. Подія B складається з C_n^m елементарних подій, в кожну з яких символ A входить m разів, а символ

10 Бернуллі Якоб (1654-1705) – швейцарський математик, професор університету в м.Базелі.

$\bar{A}$ - (n - m) разів. Ймовірність кожної з цих елементарних подій однакова і дорівнює $p^m q^{n-m}$. Тоді за означенням шукана ймовірність події В (тобто $P_n(m)$) дорівнює

$$P_n(m) = C_n^m \, p^m q^{n-m}.$$

Одержані ймовірності при фіксованому n і різних $m=\overline{0,n}$ носять назву <u>біномного розподілу ймовірностей.</u>

До цього розподілу ми повернемось ще не раз.

<u>Приклад</u>:Гральний кубик кидається 4 рази. Обчислити ймовірність того, що "двійка" випаде 2 рази.

В даному випадку подія А – випадання "двійки" при одному киданні кубика. Очевидно, p=P(A)= $\frac{1}{6}, \quad q=\frac{5}{6}$. Шукана ймовірність $P_4(2)$ дорівнює: $P_4(2) = C_4^2 p^2 q^{4-2} = \frac{4!}{2!\,2!}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2 \approx 0,116$.

2. Теорія ймовірностей

2.1 Одновимірні випадкові величини

2.1.1 Означення випадкової величини. Функція розподілу

Перше, ніж давати формальне означення випадкової величини, на простому прикладі з'ясуємо ті властивості, які ми інтуїтивно вкладаємо у це поняття.

У якості випробування розглянемо підкидання грального кубика, і, таким чином, $\Omega=\{\omega_1, \omega_2,\ldots, \omega_6\}$, де ω_i – випадання грані з номером і, $i=\overline{1,6}$. Нехай виграш гравця дорівнює номеру цієї грані, що помножений на 10. Розмір виграшу – числова величина, що залежить від елементарної події, яка реалізується у даному випробуванні. Якщо виграш гравця позначити через φ, то $\varphi=\varphi(\omega)$, тобто деяка функція від елементарної події, а в нашому прикладі це $10 \cdot i$. Отже, випадкова величина є деяка числова скалярна функція, аргументом якої є елементарна подія з простору Ω. Таким чином, ми маємо справу з відображенням $\Omega \to R^1$.

Далі, нас звичайно буде цікавити питання, як вимірювати ймовірності, з якими наша величина приймає ті чи інші значення; ймовірності з якими вона перевищує деякі значення, або є менше цих значень, попадає у деякий інтервал і т.і. Але ймовірності означені лише для випадкових подій. Тому, якщо нас цікавить ймовірність попадання величини $\varphi(\omega)$ у деяку область B дійсної вісі, то необхідно, щоб $\{\omega: \varphi(\omega)\in B\}$ була випадковою подією, тобто належала σ - алгебрі підмножин з Ω, для яких означена ця ймовірність. Це накладає певні обмеження на вигляд функції $\varphi(\omega)$ та обмеження на клас множин на дійсній вісі, з якого береться B.

А зараз перейдемо до формального означення. Нехай маємо ймовірнісний простір (Ω, F, P).

<u>Означення 1.</u> Числова скалярна функція $\varphi=\varphi(\omega)$ зветься <u>вимірною відносно σ - алгебри F</u>, якщо для кожної дійсної x множина $\{\omega: \varphi(\omega)<x\}$ належить F.

<u>Означення 2.</u> <u>Випадковою величиною</u> зветься вимірна відносно F числова скалярна функція $\varphi(\omega)$, аргументом якої є елементарна подія $\omega\in\Omega$.

Якщо виконано випробування, внаслідок якого відбулася деяка елементарна подія $\omega\in\Omega$, то число $\varphi(\omega)$ зветься <u>реалізацією</u> випадкової величини у цьому випробуванні.

Таким чином, для випадкової величини $\varphi(\omega)$ означена ймовірність

$$F(x)=P(\omega: \varphi(\omega)<x)$$

для кожної дійсної $x \in R^1$.

Якщо розглядати x як змінну, що міняється від $-\infty$ до $+\infty$, то F(x) є функція змінної x; її звуть <u>функцією розподілу</u> випадкової величини $\varphi(\omega)$. За фіксованої x вона дорівнює ймовірності того, що випадкова величина попадає в область на дійсній вісі лівіше точки x.

2.1.2 <u>Деякі властивості функції розподілу</u>

1. Маємо дві випадкові події (припускаємо, що а<в):

$$A_1=\{\omega: \varphi(\omega) < в\} \in F;$$

$$A_2=\{\omega: \varphi(\omega) < a\} \in F.$$

З властивостей σ - алгебри виходить, що

$$A_3= A_1 \setminus A_2 =\{\omega: a \le \varphi(\omega) < в\} \in F,$$

тобто є також випадковою подією, а тому можна говорити про її ймовірність.

Зауважимо, що A_2 і A_3 несумісні, а також $A_1=A_2\cup A_3$. Тому за третьою аксіомою теорії ймовірностей.

$$P(A_1) = P(A_2) + P(A_3).$$

Але,

$$P(A_1) = F(в),$$

$$P(A_2) = F(a),$$

$$P(A_3) = P(\omega\colon a \le \varphi(\omega) < в).$$

Звідси

$$P(\omega\colon a \le \varphi(\omega) < в)= F(в) – F(a),$$

що дає формулу обчислення ймовірності попадання випадкової величини до інтервалу [a, в) за допомогою функції розподілу.

2. Тому що ймовірність невід'ємна, з останнього співвідношення виходить

$$F(a) \le F(в),\ a < в,$$

тобто F(x) – неспадна функція.

3. Доведемо, що F(x) неперервна зліва.

Це виходить з властивості неперервності ймовірності у нулі. Дійсно, тому що події $A_n=\{\omega\colon x-\frac{1}{n} \le \varphi(\omega) < x\}\downarrow\varnothing$, то $P(A_n)= F(x) – F(x-\frac{1}{n}) \xrightarrow[n\to\infty]{} 0$, тобто

$F(x-0)=F(x)$.

Зауваження. В літературі ви можете зустріти твердження про те, що функція розподілу неперервна справа. Одначе, зверніть увагу на те, що там інакше означена функція розподілу

$$F(x)= P(\omega\colon \varphi(\omega) \le x).$$

4. Переконаємось ще, що мають місце наступні властивості функції розподілу:

$$F(\infty)=\lim_{x\to\infty}F(x)=1\quad;$$

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

Зобразимо Ω у вигляді суми несумісних подій:

$$\Omega=\underset{n=-\infty}{\overset{\infty}{¿}}A_n\quad,$$

де $A_n=\{\omega\text{: } n-1 \le \varphi(\omega) < n\}$.

Тоді

$$1=P(\Omega)=\sum_{n=-\infty}^{\infty}P(A_n)=\lim_{N\to\infty}\sum_{n=-N+1}^{N}P(A_n)=\lim_{N\to\infty}\left[F(-N+1)-F(-N)+F(-N+2)-F(-N+1)+\ldots+¿¿¿+F(N)-F(N-1)\right]=\lim_{N\to\infty}\left[F(N)-F(-N)\right].¿$$
¿

Це може бути (враховуючи те, що F(x) – імовірність і неспадна функція) тільки тоді, коли $F(\infty)=\lim_{N\to\infty}F(N)=1,\ F(-\infty)=\lim_{N\to\infty}F(-N)=0$.

Підсумуємо доведені результати у вигляді теореми.

Теорема 1. Функція розподілу F(x) має такі властивості:

1. F(x) не спадає;

2. F(x) неперервна зліва;

3. $F(\infty)=1$;

4. $F(-\infty)=0$.

Приклад 1: Повернемось до випробування з підкиданням грального кубика, що розглянуте на початку 2.1.1. Простір елементарних подій в даному випадку $\Omega=\{\omega_1, \omega_2,..., \omega_6\}$, σ - алгебра подій F – сукупність всіх підмножин множини Ω. Імовірність елементарної випадкової події задається як

$$P(\omega_i)=\frac{1}{6}\ ,\quad i=\overline{1,6}\ .$$

Випадкова величина (виграш гравця) дорівнює:

$$\varphi(\omega) = 10 \cdot i,\ \omega=\omega_i,\quad i=\overline{1,6}\ .$$

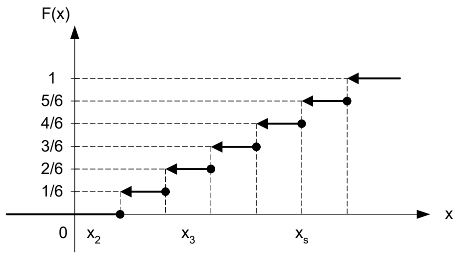
Зауважимо, що

$$\{\omega:\ \varphi(\omega) < x\} = \begin{cases} \emptyset, & x\le 10; \text{¿} \\ \{\omega_1,\dots,\omega_{i-1}\}, & x\le 10\cdot i,\ i=\overline{2,6}; \text{¿¿¿¿} \end{cases}$$

Таким чином, функція розподілу дорівнює:

$$F(x)= \begin{cases} 0, & x\le 10; \text{¿} \\ \dfrac{i-1}{6}, & x\le 10\cdot i,\ i=\overline{2,6}; \text{¿¿¿¿} \end{cases}$$

Графік цієї функції має вигляд



<u>Приклад 2:</u>Розглянемо кидання навмання точки на інтервал [0, 1) (див 1.5.3). У цьому випробуванні елементарною подією ω є число з інтервалу

[0, 1). Простір елементарних подій Ω=[0, 1), σ - алгебра подій F - це

σ - алгебра борелевих множин з інтервалу [0, 1). Імовірнісна міра задана для кожної події {ω: а ≤ ω < в} як Р(ω: а ≤ ω < в) = в - а (а≤в, 0≤а<1, 0≤в<1). Випадкову величину φ(ω) задамо як координату точки ω на осі, тобто φ(ω)=ω, ω∈[0, 1).

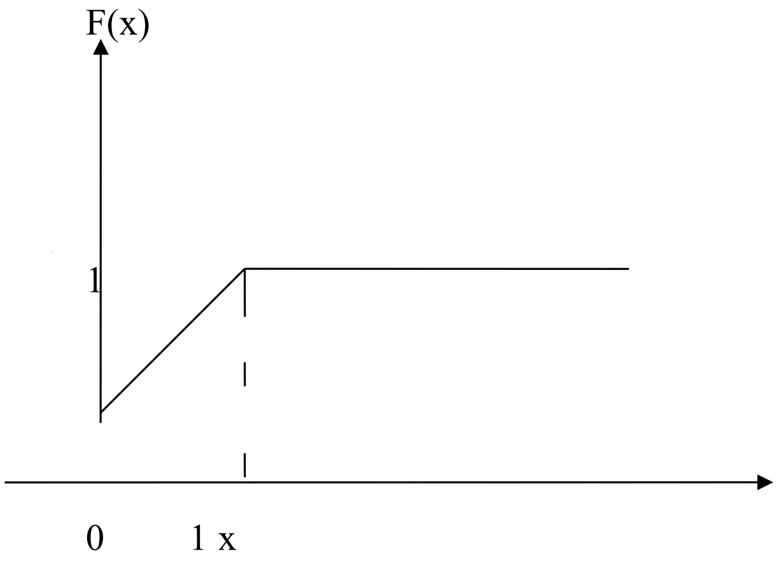
У цьому випадку

$$\{\omega:\ \varphi(\omega) < x\} = \begin{cases} \emptyset, & x\le 0; \text{¿¿¿¿} \\ \text{¿} \end{cases}$$

Функція розподілу випадкової величини φ(ω)має вигляд:

$$F(x)= \begin{cases} 0, \; x \le 0; ¿¿¿¿ \\ ¿ \end{cases}$$

Її графік



Це так званий рівномірний розподіл на інтервал [0, 1).

Тепер розглянемо питання: для яких $B \subset (-\infty, \infty)$ $\{\omega: \varphi(\omega) \in B\}$ є випадковою подією, тобто належить F.

З означення випадкової величини виходить, що таким чином множинами на осі є інтервали виду $B \subset (-\infty, x)$. Раніш також було доведено, що цю властивість мають інтервали виду $B=[a, в) \subset (-\infty, \infty)$. Виявляється, що цю властивість мають довільні борелєві множини на числовій осі.

<u>**Теорема 2.**</u> Нехай В – деяка довільна борелева множина на числовій осі. Задано імовірнісний простір (Ω, F, P), а також випадкову величину $\varphi(\omega)$ на ньому. Тоді множина

$$\{\omega: \varphi(\omega) \in B\} \in F, \qquad (1)$$

тобто є випадковою подією.

Доведення. Спочатку зауважимо, що клас σ множин B, для яких виконується (1), є деякою σ - алгеброю, тобто клас множин, замкнений відносно зліченної кількості теоретико – множинних операцій.

Дійсно, нехай

$$B_i : \{\omega: \varphi(\omega) \in B_i\} \in F, \quad i=\overline{1,\infty} \ .$$

Тоді очевидні рівності:

$$\{\omega:\phi(\omega)\in \bigcup_{i=1}^{\infty} B_i\} = \bigcup_{i=1}^{\infty} \{\omega:\phi(\omega)\in B_i\} \in F,$$

$$\{\omega:\phi(\omega)\in \bigcap_{i=1}^{\infty} B_i\} = \bigcap_{i=1}^{\infty} \{\omega:\phi(\omega)\in B_i\} \in F \ .$$

Зрозуміло також, що з (1) виходить

$$\{\omega:\phi(\omega)\in \overline{B}\} = \overline{\{\omega:\phi(\omega)\in B\}} \in F \ ,$$

та має місце

$$(\omega: -\infty \le \varphi(\omega) < \infty) = \Omega \in F.$$

Таким чином, маємо σ - алгебру множин B.

З того, що цій σ - алгебрі належать усі інтервали виду [а, в), виходить, що вона містить σ - алгебру борелевих множин на числовій осі. Отже твердження теореми має місце для довільної борелевої множини B.

З теореми 2 виходить, що для довільної борелевої множини B означена ймовірність

$$P_\varphi = P\{\omega: \varphi(\omega) \in B\}.$$

Зважаючи на це можна говорити про те, що відображення φ(ω) простору Ω у числову пряму $R^1=(-\infty, \infty)$ породжує новий імовірнісний простір $(R^1, B, P_\varphi(\cdot))$, де B - σ - алгебра борелевих множин на осі. У цьому просторі ймовірність події [а, в) визначається через функцію розподілу випадкової величини φ за формулою

$$P_\varphi ([а, в)) = F(в) - F(а),$$

а для подій $A=\underset{i=1}{\overset{n}{¿}}[a_i, в_i)$ з неперетинними інтервалами

$$P_\varphi(A)=\sum_{i=1}^{n}[F(в_i)-F(a_i)].$$

А. М. Колмогаров довів у певному розумінні зворотне твердження до теореми 1.

<u>Теорема 3</u> (Колмогорова). Якщо деяка функція F(x) володіє властивостями 1) –4), що фігурують у теоремі 1, тоді існує ймовірнісний простір (Ω, F, P) та випадкова величина φ(ω) на ньому така, що її функцією розподілу є F(x).

Не зупиняючись на деталях доведення, зазначимо лише, що за простір елементарних подій береться дійсна вісь $R^1=(-\infty, \infty)$, за σ - алгебру випадкових подій - σ - алгебра борелевих множин на числовій осі R^1.

Для скінчених сум неперетинних інтервалів

$$A=\underset{i=1}{\overset{n}{¿}}[a_i, в_i)$$

скінчено - адитивна ймовірнісна міра визначається співвідношенням

$$P(A)=\sum_{i=1}^{n}[F(в_i)-F(a_i)].$$

Доводиться, що вона зліченно – адитивна на алгебрі А інтервалів виду [а, в) та скінченних об'єднань цих інтервалів. Далі, використовуючи теорему Каратеодорі, вона продовжується на σ - алгебру борелевих множин числової осі R^1 . Як випадкова величина береться функція φ(ω)=ω (тобто тотожне відображення R^1 у себе), для якої функцією розподілу є F(x).

Наведені результати дають змогу стверджувати, що функція розподілу у певному розумінні є вичерпною характеристикою випадкової величини.

Якщо ви досліджуєте випадкову величину і вас цікавлять імовірності попадання її до деяких інтервалів на числовій осі, то не треба будувати

вихідний імовірнісний простір (Ω, F, P) та конкретний вигляд функції φ(ω). Можна вважати простором елементарних подій числову вісь R^1, випадковими подіями – борелеві множини на цій осі. Для обчислення шуканих імовірностей треба деяким чином оцінити функцію розподілу F(x) (пізніш буде показано як це робиться статистичними методами за результатами випробувань над випадковою величиною, тобто за її реалізаціями $x_1, x_2,..., x_n$) і по цій функції визначати ймовірності подій, що нас цікавлять.

Надалі випадкові величини будемо позначити великими літерами латинського алфавіту X, Y, Z,..., а для ймовірностей подій типу

P(ω: X(ω)∈B) часто будемо застосовувати скорочений запис P(X∈B). Так, для функції розподілу випадкової величини X будемо використовувати запис F(x)=P(X < x).

Доведемо ще деякі властивості функції розподілу, які нам знадобляться в майбутньому.

Нехай $\{x_n\}$ – деяка послідовність, що прямує до x справа, тобто $x_{n+1} < x_n$, $x_n \to x$, $x_n > x$.Позначимо F(x+0)= $\lim_{n\to\infty} F(x_n)$. Має місце

$$F(x + 0) = P(X \le x).$$

Дійсно, за таких умов

$$\{X \le x\} = \mathop{intersect}_{n=1}^{\infty} \{X < x_n\}$$

(послідовність множин {X ≤ x} не зростає). Тому, за властивістю неперервності імовірнісної міри, маємо

$$P(X \le x)=P(\lim_{N\to\infty} \mathop{intersect}_{n=1}^{N} \{X < x_n\} = P\left(\lim_{N\leftarrow\infty} \{X < x_N\}\right) = \lim_{N\to\infty} P(X < x_N) = F(x+0) .$$

Далі,

$$\{X \le x\}=\{ X < x\} \cup \{X = x\}.$$

Зауважимо, що доданки у правій частині останньої рівності несумісні. Тому

$$P(X \le x)=P(X < x) + P(X = x)$$

Звідси виходить, що

$$P(X = x)=P(X \le x)-P(X < x) = F(x + 0) - F(x).$$

У багатьох задачах теорії ймовірностей доводиться вивчати не тільки вихідну випадкову величину X, а й деяку функцію від неї g(X). Зрозуміло, що коли говорити про ймовірність попадання функції g(X) у деяку область на числовій осі, то необхідно щоб Y= g(X) сама була випадковою величиною. Це накладає деякі обмеження на клас функцій, що володіють цією властивістю.

Означення 3. Нехай g(X) – функція, що здійснює відображення $R^1 \to R^1$, та B - σ - алгебра борелевих множин у R^1. Функція g(X) зветься борелевою функцією, якщо вона вимірна відносно σ - алгебри B, тобто для кожного

$y \in R^1$ $\{x: g(x) < y\} \in B$.

Без доведення відзначимо, що клас борелевих функцій достатньо багатий: він містить, зокрема, усі неперервні та кусково-неперервні функції.

Теорема 4. Нехай g(x) – борелева функція на R^1, а X(ω) – випадкова величина, що задана на імовірнісному просторі (Ω, F, P). Тоді Y(ω)=g(X(ω)) – вимірна функція відносно σ - алгебри F, тобто є випадковою величиною.

Доведення. Через те, що функція g(x) борелева, то для кожного $y \in R^1$

$$B=\{x: g(x) < y\} \in B.$$

Отже, за теоремою 2 для кожного $y \in R^1$

$$\{\omega: Y(\omega)\} = g(X(\omega)) < y\} = \{\omega: X(\omega) \in B\} \in F,$$

і, таким чином, $Y(\omega)$ – випадкова величина.

2.1.3 Дискретні випадкові величини

Нехай (Ω, F, P) – імовірнісний простір.

Означення. Випадкова величина $X(\omega)$ називається дискретною, якщо внаслідок випробування вона може набути значення з скінченної або зліченної множини можливих значень.

Нехай дискретна випадкова величина $X(\omega)$ набуває значень $x_1, x_2, \ldots, x_s \ldots$. З того, що вона випадкова величина, та з того, що одноточкові множини на вісі є борелевими, виходить (згідно попереднім результатам), що означені ймовірності

$$p_i = P(\omega: X(\omega) = x_i).$$

Набір цих імовірностей називають розподілом випадкової величини $X(\omega)$.

Навпаки, якщо для деякої функції $X(\omega)$, що набуває зліченну множину значень, визначені для кожного x_i ймовірності $p_i = P(\omega: X(\omega) = x_i)$, то це означає, що $(\omega: X(\omega) = x_i) \in F$, а тому і для кожного дійсного x

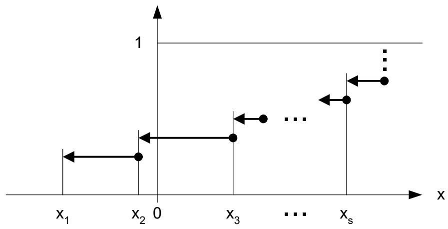
$$(\omega: X(\omega) < x) = \bigcup_{x_i < x} (\omega: X(\omega) = x_i) \in F,$$

тобто $X(\omega)$ вимірна відносно F та є випадковою величиною.

Функція розподілу випадкової величини має вигляд

$$F(x) = P(\omega: X(\omega) < x) = P\left(\bigcup_{x_i < x} \{\omega: X(\omega) = x_i\}\right) = \sum_{x_i < x} P(\omega: X(\omega) = x_i) = \sum_{x_i < x} p_i .$$

Це неспадна, ступінчаста, неперервна зліва функція, графік якої умовно можна зобразити наступним чином (припускаємо, що $x_1 < x_2 < \ldots < x_s \ldots$)



Одначе розподіл дискретної випадкової величини більш зручно характеризувати за допомогою таблиці

X~

Значення X(ω)	x_1	x_2	...	x_s	…
Ймовірність $P(X=x_i)$	p_1	p_2	...	p_s	...

Зауважимо, що $p_i \geq 0, \sum_{i=1}^{\infty} p_i = 1$.

<u>Приклад</u>. Випробування –композиція n незалежних випробувань, у кожному з яких відбувається подія A з імовірністю p, чи подія $\bar{A}$ з імовірністю q=1-p. Випадкова величина X – кількість разів, що подія A відбулась у n незалежних випробуваннях. Як вже відмічалось, розподіл величини X – біномний. Відповідна таблиця має вигляд

X~

0	1	2	...	m	...	n
q^n	$C_n^1 p q^{n-1}$	$C_n^2 p^2 q^{n-2}$	...	$C_n^m p^m q^{n-m}$	…	p^n

2.1.3.1 <u>Характеристики дискретних величин</u>

Нехай випадкова величина X набуває скінченну кількість різних значень $x_1, x_2,..., x_s$, та нехай $p_i = P(X = x_i)$ $i=\overline{1,s}$. Здійснемо n незалежних випробувань, у кожному з яких спостерігається величина X. Позначимо через $a_1, a_2,..., a_n$ реалізації випадкової величини у відповідних випробуваннях. Тоді середнє значення реалізацій можна наступним чином:

$$\frac{1}{n}\sum_{k=1}^{n} a_k=\sum_{i=1}^{s} x_i \frac{m_i}{n},$$

де m_i – частота появ події $\{X = x_i\}$ у n незалежних випробуваннях. Звідси

$$\frac{1}{n}\sum_{k=1}^{n} a_k=\sum_{i=1}^{s} x_i W_n(X=x_i),$$

де $W_n(X=x_i)$ – частість події $\{X=x_i\}$.

Для великої кількості випробувань частість прямує до ймовірності події. Замінюя частість на ймовірність, ми одержимо аналітичний аналог середнього значення реалізацій випадкової величини

$$MX=\sum_{i=1}^{s} x_i p_i,$$

який називається математичним сподіванням (або ймовірнісним середнім) випадкової величини X.

Для скінченої кількості s можливих значень випадкової величини проблем із збіжністю правої частини останнього співвідношення не виникає. Але це не так, якщо s нескінченно. Бажано, щоб результат не залежав від нумерації доданків у правій частині рівності. Цього можна домогтися, якщо припустити абсолютну збіжність ряду $\sum_{i=1}^{s} x_i p_i$, тобто $\sum_{i=1}^{s} |x_i| p_i < \infty$. Отже приходимо до загального означення.

Означення. Математичним сподіванням дискретної випадкової величини X, що набуває значень $x_1, x_2,..., x_s...$, з імовірностями $p_1, p_2,..., p_s...$, називається число

$$MX=\sum_{i=1} x_i p_i ,$$

за умови, що ряд праворуч збігається.

Математичне сподівання є як аналогом центру мас точкової механічної системи з координатами точок на осі $x_1, x_2,..., x_s$, та масами $p_1, p_2,..., p_s$.

Як центр мас:

$$MX=\frac{x_1 p_1+x_2 p_2+\ldots+x_s p_s}{p_1+p_2+\ldots+p_s} .$$

Але $p_1 + p_2 + ...+ p_s=1$, тобто

$$MX=\sum_{i=1}^{s} x_i p_i .$$

Зміст математичного сподівання міститься у наступному: це точка на числовій осі відносно якої групуються результати випробувань над дискретною величиною.

2.1.3.2 Властивості математичного сподівання

У подальшому детермінованою величиною (сталою) C будемо називати дискретну випадкову величину, що набуває єдине значення $x_1=C$ з імовірністю $p_1=1$.

1. Математичне сподівання сталої дорівнює цій сталій:

$$MC=C.$$

дійсно, за означенням

$$MC = x_1p_1 = C\cdot 1 = C.$$

Розглянемо найпростіші функції від дискретної випадкової величини X за умови, що існує її математичне сподівання.

2. Математичне сподівання величини СХ. Де С – стала, а Х дисктерна випадкова величина, дорівнює

$$MCX = CMX,$$

тобто стала виноситься з-під знаку математичного сподівання.

Таблиця для випадкової величини СХ має вигляд

$$CX \sim \left\{ Cx_1\ Cx_2 \dots Cx_s \text{¿} \right\} \text{¿} \{\} .$$

За означенням математичного сподівання

$$MCX = \sum_{i=1}^{s} Cx_i p_i = C \sum_{i=1}^{s} x_i p_i = CMX .$$

3. Математичне сподівання суми X+b, де Х дискретна випадкова величина, а b – стала, дорівнює

$$M(X+b) = MX + b.$$

Будуємо таблицю

$$X + b \sim \left\{ x_1 + b\ x_2 + b \dots x_s + b \text{¿} \right\} \text{¿} \{\} .$$

Маємо

$$M(X+b) = \sum_{i=1}^{s} (x_i + b) p_i = \sum_{i=1}^{s} x_i p_i + b \sum_{i=1}^{s} p_i = MX + b .$$

4. Самостійно довести, що

$$M(aX + b) = aMX + b,$$

якщо a та b сталі, X – дискретна випадкова величина.

5. Нехай випадкова величина У є функцією f(X) від дискретної випадкової величини Х. Для випадкової величини У=f (X) будуємо таблицю

$$У=f(X) \sim \left\{ f(x_1)\ f(x_2) \ldots f(x_s) ¿ \right\} ¿ \{\}.$$

Тут ми припускаємо, що усі $f(x_i)$ мають різні значення. Якщо це не так, об’єднуємо рівні функції, сумуючи відповідні ймовірності. Математичне сподівання випадкової величини Y дорівнює:

$$MY = \sum_{i=1}^{s} f(x_i)\, p_i$$

(для нескінченого s додатково тут і далі вимагаємо, щоб ряд у правій частині рівності абсолютно збігався).

Математичні сподівання деяких функцій від випадкової величини Х мають спеціальну назву.

<u>Початковим моментом к-го порядку</u> випадкової величини Х називається математичне сподівання випадкової величини $Y=X^к$:

$$v_k = MX^k = \sum_{i=1}^{s} x_i^k\, p_i .$$

Зрозуміло, що v_1 = МХ. Іноді позначають МХ=v (без індексу).

<u>Центрована випадкова величина</u> – це величина, що дорівнює $X^/=X-MX$.

Математичне сподівання $MX^/$ дорівнює нулеві. Дійсно, МХ – детермінована величина, і на підставі третьої властивості

$$MX^/ = M(X - MX) = MX – MX = 0.$$

<u>Центральним моментом k-го порядку</u> випадкової величини Х називається початковий момент k-го порядку випадкової величини $X^/$:

$$\mu_k = M(X')^k = \sum_{i=1}^{s} (x_i - MX)^k\, p_i .$$

<u>Дисперсією</u> випадкової величини Х називається її централний момент другого порядку:

$$DX = \mu_2 = \sum_{i=1}^{s} (x_i - v)^2 p_i \quad .$$

Дисперсія є мірою концентрації результатів конкретних випробувань над випадковою величиною.

Додатнє значення квадратного кореня з дисперсії має назву середнього квадратичного відхилення та позначається σ_x:

$$\sigma_x = + \sqrt{DX} \quad .$$

2.1.3.3 Властивості дисперсії

1) Чим менше дисперсія, тим більш тісно групуються результати конкретних випробувань навколо математичного сподівання.

Дійсно, нехай дисперсія мала, тоді малий і кожен доданок $(x_i - v)^2 p_i$ суми. Для x_i, яке значно відрізняється від математичного сподівання v, ймовірність p_i повинна бути малою величиною. Отже, велику ймовірність (а тому і частість у великій серії випробувань) можуть мати лише ті x_i, які мало відрізняються від математичного сподівання .

2) Коли дисперсія дорівнює нулеві, то X детермінована вличина з імовірністю одиниця.

Дійсно, нехай усі x_i різні та

$$DX = \sum_{i=1}^{s} (x_i - v)^2 p_i = 0 \quad .$$

Тому, що усі p_i не можуть дорівнювати нулеві (повинна виконуватись умова нормування), то ця рівність означає, що існує єдине $x_i = v$ і таке, що $p_i = 1$ (інших значень X може набувати лише з імовірністю нуль).

3) Дисперсія суми дискретної випадкової величини X та сталої C дорівнює дисперсії величини X:

$$D(X + C) = DX.$$

Дійсно,

$Y = X + C,$

$Y^{/} = Y – MY = X + C – M(X + C) = X – MX = X^{/},$

$DY = M\,(Y^{/})^2 = M(X^{/})^2 = DX.$

4) Стала виноситься з-під знаку дисперсії з квадратом:

$$D(CX) = C^2DX.$$

Дійсно,

$Y = CX,$

$Y^{/} = Y – MY = CX – M(CX) = CX – CMX = C(X - MX) = CX^{/},$

$DY = M(Y^{/})^2 = M(CX^{/})^2 = MC^2(X^{/})^2 = C^2M(X^{/})^2 = C^2DX.$

5) Має місце корисне співвідношення між дисперсією і початковими моментами першого та другого порядку:

$$DX = \nu_2 – \nu_1^2.$$

Дійсно,

$$DX=\sum_{i=1}^{s}(x_i-\nu)^2 p_i=\sum_{i=1}^{s} x_i^2 p_s-2\nu_1\sum_{i=1}^{s} x_i p_i+\nu_i^2\sum_{i=1}^{s} p_i=\nu_2-2\nu_1^2+\nu_1^2=\nu_2-\nu_1^2 .$$

2.1.3.4 Твірна функція

Досить часто виявляється зручним розглядати замість розподілу випадкової величини деякі його перетворення. У цьому курсі ми ознайомимось з перетвореннями, які мають назву характеристичних та твірних функцій.

Нехай дискретна випадкова величина X характеризується таблицею

$$X \sim \left\{ x_1\ x_2 \dots x_s ¿ \right\} ¿ \{\} .$$

Характеристичною функцією випадкової величини X називається комплексно – значна функція

$$\phi_x(t) = Me^{\sqrt{-1}Xt} = \sum_{i=1}^{s} e^{\sqrt{-1}x_i t} p_i\,,$$

де t – дійсна змінна.

З властивостями характеристичних функцій та їх застосуванням ми ознайомимось пізніш.

<u>Твірною функцією</u> (моментів) називається функція дійсної змінної t виду

$$m_x(t) = Me^{Xt} = \sum_{i=1}^{s} e^{x_i t} p_i\,.$$

Для нескінченного s ми будемо розглядати такі розподіли, для яких відповідні ряди збігаються для кожного t з деякого проміжку, що включає точку нуль.

2.1.3.5 <u>Властивості твірної функції</u>

1) Похідна k-го порядку твірної функції у точці t=0 дорівнює початковому моменту k-го порядку, тобто

$$\frac{d^k m_x(t)}{dt^k}\Big|_{t=0} = \nu_k\,.$$

Дійсно,

$$\frac{d^k\left(\sum e^{x_i t} p_i\right)}{dt^k}\Big|_{t=0} = \sum_{i=1}^{s} x_i^k e^{x_i t} p_i\Big|_{t=0} = \sum_{i=1}^{s} x_i^k p_i = \nu_k\,.$$

2) Розвинення твірної функції у ряд Маклорена [11] має вигляд:

$$m_x(t) = 1 + \frac{\nu_1}{1!}t + \frac{\nu_2}{2!}t^2 + \ldots + \frac{\nu_k}{k!}t^k + \ldots$$

(тут ми припускаємо, що для нашого розподілу твірна функція аналітична у нулі). Це очевидно виходить з першої властивості і з того, що

$$m_x(t) = m_x(0) + \frac{m_x'(0)}{1!}t + \frac{m_x''(0)}{2!}t^2 + \ldots + \frac{m_x^{(k)}}{k!}t^k + \ldots\,.$$

11 Маклорен Колін (1698-1746) – шотландський математик, професор, член Лондонського королівського товариства, учень і послідовник Ньютона.

Приклад: Розглянемо випадкову величину, що має біномний розподіл імовірностей

$$X \sim \{ i ¿ ¿ \}, \; i=\overline{0,n}, \; q=1-p \; .$$

Відшукаємо твірну функцію :

$$m_x(t)=\sum_{i=0}^{n} e^{it} C_n^i p^i q^{n-i} = \sum_{i=0}^{n} C_n^i (e^t p)^i q^{n-1} = (e^t p+q)^n \; .$$

Тут була використана формула біному Ньютона [12]:

$$(a+b)^n = a^n + C_n^1 b a^{n-1} + \ldots + C_n^i b^i a^{n-i} + \ldots + b^n \; .$$

Самостійно перевірити, що

$$MX = v_1 = \frac{dm_x(t)}{dt}\Big|_{t=0} = \frac{d(e^t p+q)^n}{dt}\Big|_{t=0} = np \; ,$$

$$v_2 = \frac{d^2 m_x(t)}{dt^2}\Big|_{t=0} = \frac{d^2(e^t p+q)^n}{dt^2}\Big|_{t=0} = n^2 p^2 + npq \; ,$$

$$DX = v_2 - v_1^2 = npq \; .$$

2.1.3.6 Перша модель розподілу Пуассона[13]

Нехай проведено необмежено велику серію випробувань. Внаслідок кожного випробування випадково з'являється точка на числовій осі. Припускаємо, що випадковий розподіл точок на числовій осі задовольняє наступним трьом властивостям (більш строго ми розглянемо цю задачу пізніш).

1) Стаціонарність. Імовірність того, що на відрізок фіксованої довжини

---

12 Ньютон Ісаак (1643-1727) –геніальний англійський фізик, астроном і математик, президент Лондонського королівського товариства, професор Кембриджського університету

13 [3]Пуассон Семєон Дені (1781-1840) –французький механік, фізик і математик, професор Сорбонни, член Паризької АН

попаде певна кількість точок визначається тільки довжиною цього відрізку та не залежить від розташування відрізку на числовій осі.

2) <u>Ординарність</u>. Імовірність того, що на достатньо малий відрізок

довжини Δx попаде точно одна точка, є величиною порядку $\vartheta(\Delta x)$. Імовірність того, що на цей відрізок попаде більш ніж одна точка, є нескінченно малою вищого порядку малості ніж Δx, тобто $\vartheta(\Delta x)$.

3) <u>Властивість відсутності післядії</u>. Ймовірність того, що на деякий

відрізок попаде певна кількість точок, не залежить від того, скільки точок попаде на відрізки, що не перетинаються з даним.

Обчислимо ймовірність того, що на відрізок довжини L попаде m точок.

Позначимо через X^1 випадкову величину, що дорівнює кількості точок, що попадають на відрізок одиничної довжини; відповідна таблиця має вигляд:

$$X^1 \sim \left\{ i \, ¿ \middle| ¿ \left\{ \right. \right\}, \; i=\overline{0,\infty}, \quad .$$

де $P_1(i)$ означає ймовірність попадання на відрізок одиничної довжини точно i точок.

Позначимо $MX^1 = \lambda$. За властивістю стаціонарності воно однаково для усіх відрізків одиничної довжини.

Розглянемо довільний відрізок довжини L та випадкову величину X^L – кількість точок, що попадає у цей відрізок. Використовуючи властивість математичного сподівання (яка буде доведена пізніш) – математичне сподівання суми випадкових величин дорівнює сумі їх математичних сподівань; покажіть самостійно, що $MX^1 = \lambda L$.

Розіб’ємо цей відрізок на n відрізків довжини

$$\Delta x = \frac{L}{n} .$$

Тоді з деякою похибкою, яка тим менш, чим більше n, можна вважати, що:

$$X^{\Delta x} \sim \{0 \quad 1¿\}¿\{\},$$,

де $P_{\Delta x}(0)$ – імовірність того, що на відрізок довжини Δx не попаде жодна точка;

$P_{\Delta x}(1)$ – ймовірність того, що на відрізок довжини Δx попаде одна точка.

Таким чином, для достатньо малого відрізку Δx ймовірність попадання до нього однієї точки $P_{\Delta x}(1) \approx \lambda \Delta x$, а ймовірність того, що не попаде жодна точка $P_{\Delta x}(0) \approx 1-\lambda \Delta x$ (точніш $P_{\Delta x}(1) \approx \lambda \Delta x + \vartheta(\Delta x)$, $P_{\Delta x}(0) \approx 1-\lambda \Delta x + \vartheta(\Delta x)$).

За нашими припущеннями на відрізок довжини L попаде m точок (наближено) тоді, коли до m малих відрізків попаде по одній точці. За 3-ю властивістю та за біномним розподілом

$$P_n(m) \approx C_n^m (\lambda \Delta x)^m (1-\lambda \Delta x)^{n-m} = \frac{n!}{m!(n-m)!} \left(\frac{\lambda L}{n}\right)^m \frac{\left(1-\frac{\lambda L}{n}\right)^n}{\left(1-\frac{\lambda L}{n}\right)^m} .$$

Точну ймовірність одержимо завдяки граничному переходу при $n \to \infty$:

$$\pi_m=\lim_{n\to\infty}P_n(m)=\lim_{n\to\infty}\frac{n!}{m!(n-m)!}\left(\frac{\lambda L}{n}\right)^m\frac{\left(1-\frac{\lambda L}{n}\right)^n}{\left(1-\frac{\lambda L}{n}\right)^m}=\lim_{n\to\infty}\frac{(n-m+1)\cdot ¿\cdot n}{n^m}\cdot\frac{a^m}{m!}\cdot\frac{\left(1-\frac{a}{n}\right)^n}{\left(1-\frac{a}{n}\right)^m}=$$

$$¿\frac{a^m}{m!}e^{-a}$$

де a=λL.

Самостійно перевірити, що $\sum_{m=0}^{\infty}\pi_m=1$.

Розподіл імовірностей

$$\pi_m=\frac{a^m}{m!}\cdot e^{-a},\ \ m=\overline{0,\infty},$$ де a>0

називається <u>розподілом Пуассона</u>.

Відшукаємо твірну функцію для випадкової величини, що має розподіл Пуассона:

$$m_x(t)=\sum_{m=0}^{\infty}e^{mt}\frac{a^m}{m!}e^{-a}=e^{-a}\sum_{m=0}^{\infty}\frac{(ae^t)^m}{m!}=e^{-a}e^{ae^t}=e^{a(e^t-1)}$$,

для скінченних t.

Скориставшись властивостями твірної функції самостійно переконайтеся, що ν_1=MX=a, $\nu_2=a^2+a$, DX=a.

2.1.3.7 <u>Друга модель розподілу Пуассона</u>

Розглядається звичайна схема випробувань Бернуллі, у якої n велике, а p – достатньо мале. Точна формула для ймовірності події A m разів у n випробуваннях має вигляд

$$P_n(m)=C_n^m p^m (1-p)^{n-m} \quad .$$

Для великих n за цією формулою обчислити $P_n(m)$ складно. Добре наближення у цьому випадку дає вираз

$$P_n(m)\approx\frac{a^m}{m!}\cdot e^{-a},$$

де a=np.

<u>Зауваження.</u> При практичних обчисленнях граничними результатами користуються наступним чином. Якщо деяка величина A_n збігається при $n\to\infty$ до величини A, то для достатньо великих n можна обчислювати A за наближеною формулою

$$A_n \approx A.$$

Питання про точність таких наближень залежить від швидкості збіжності A_n до A, яке ми залишим тут осторонь.

Повернемось до нашої задачі. Будемо вважати, що n прямує до нескінченності та p прямує до нуля злагоджено, а саме p= $\frac{a}{n}$, та $\lim\limits_{n\to\infty}\frac{a}{n}=0$, де a>0.

Тоді

$$\lim_{n\to\infty}P_n(m)=\lim_{n\to\infty}C_n^m p^m(1-p)^{n-m}=\lim_{n\to\infty}C_n^m\left(\frac{a}{n}\right)^m\left(1-\frac{a}{n}\right)^{n-m}=\lim_{n\to\infty}\frac{(n-m+1)\cdot ¿\cdot n}{n^m}\cdot\frac{a^m}{m!}\cdot\frac{\left(1-\frac{a}{n}\right)^n}{\left(1-\frac{a}{n}\right)^m}=$$

$$¿\frac{a^m}{m!}e^{-a}$$

і, таким чином, маємо наближену рівність (1).

2.1.4 <u>Неперервні випадкові величини</u>

Нехай *X* – випадкова величина з функцією розподілу *F*(*x*)=*P*(*X*<*x*). Розглянемо випробування, коли випадкова величина може приймати незліченну множину значень; наприклад, її можливі значення – усі точки числової осі або деякий її відрізок.

Припустимо, що $F(x)$ – неперервна функція. Тоді за властивістю функції розподілу ймовірність того, що випадкова величина X одержить довільне значення x дорівнює

$P(X = x) = F(x+0) - F(x) = 0$

Знову ми зустрілися з ситуацією, коли подія принципово може відбутися (не є неможливою), але має нульову ймовірність (згадаємо приклад з киданням навмання точки на інтервал [0;1)). Стандартним прийомом у цій ситуації є введення поняття щільності (як у механіці щільність маси тощо).

Означення. Говорять, що випадкова величина X має абсолютно неперервний розподіл, якщо існує невід'ємна функція $f(x)$ така, що для усіх x виконується рівність

$$F(x)=\int_{-\infty}^{x} f(u)\,du \qquad (1)$$

Функція $f(x)$ називається щільністю розподілу випадкової величини X. Випадкову величину, що володіє цією властивістю будемо називати неперервною.

Деякі властивості функції щільності.

1. Має місце (умова нормування):

$$\int_{-\infty}^{+\infty} f(x)\,dx=1 \qquad (2)$$

Дійсно,

$$\int_{-\infty}^{+\infty} f(x)\,dx=\lim_{y\to\infty}\int_{-\infty}^{y} f(x)\,dx=\lim_{y\to\infty}F(y)=F(\infty)=1$$

Зауваження. Властивість невід'ємності і властивість (2) є характеристичними для функції щільності розподілу. А саме: кожна невід'ємна функція, що володіє властивістю (2), є щільністю розподілу деякої випадкової величини. Це виходить з того, що коли $f(x)$ – така функція, то функція $F(x)$, означена рівністю (1), задовольняє усім умовам згадуваної раніше теореми Колмогорова. Тому ми надалі будемо часто оперувати з функціями, що володіють вказаними властивостями, як із щільностями розподілу деяких випадкових величин, не проводячи конструктивні побудови цих величин.

2. Якщо випадкова величина X є неперервною, то її функція розподілу $F(x)$ є неперервною функцією.

Це виходить з (1) та властивостей інтеграла із змінною верхньою межею.

3. Для неперервної випадкової величини має місце

$$P(a \le X < b) = \int_a^b f(x)\,dx \qquad (3)$$

Дійсно,

$$P(a \le X < b) = F(b) - F(a) = \int_{-\infty}^{b} f(x)\,dx - \int_{-\infty}^{a} f(x)\,dx = \int_a^b f(x)\,dx$$

Зауважимо, що з властивості 2 виходить

$$P(a \le X < b) = P(a \le X \le b) = P(a < X < b) = P(a < X \le b).$$

4. З курсу аналізу відомо, що коли має місце (1), то для майже всіх x (для всіх x, зокрема, можливих з множини нульової міри (довжини)) функція $F(x)$ диференційована і

$$f(x) = \frac{dF(x)}{dx} = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P(x \le X < x+\Delta x)}{\Delta x} \qquad (4)$$

Для таких x, зокрема, маємо

$$P(x \le X < x+\Delta x)=f(x)\Delta x+\vartheta(\Delta x)$$, (5)

тобто, ймовірність того, що в результаті випробування неперервна випадкова величина одержить значення на інтервалі достатньо малої довжини Δx , з точністю до $\vartheta(x)$ дорівнює $f(x)\Delta x$. На практиці, взагалі, використовуються такі розподіли неперервних випадкових величин, що (4), (5) має місце для усіх x за виключенням, можливо, однієї-двох точок.

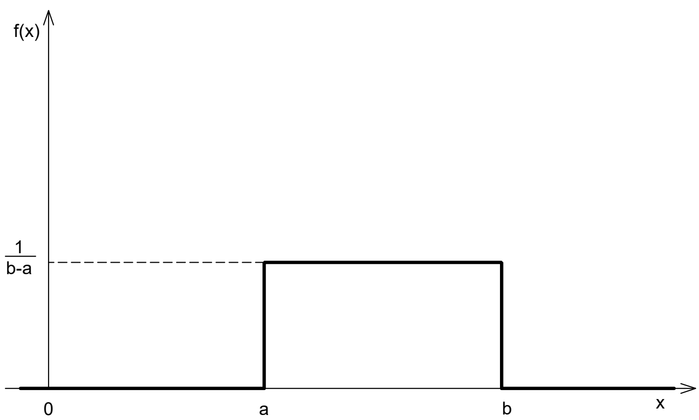
На закінчення зауважимо, що коли випадкова величина має щільність розподілу $f(x)$, то для довільної борелевої множини B має місце рівність

$$P(\omega : X(\omega) \in B)=\int_B f(x)\,dx$$

(без доведення). Останнє співвідношення можна було б узяти за означення щільності.

<u>Приклад 1.</u> Випадкова величина X має рівномірний розподіл на інтервалі (a, b), коли щільність розподілу $f(x)$ дорівнює

$$f(x)=¿(c,\, x\in(a,b),¿¿¿$$



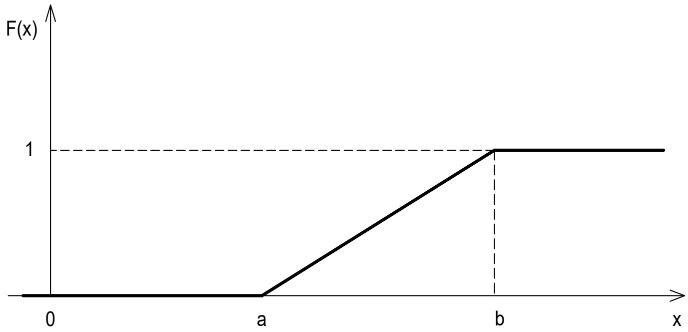
З умови нормування (2) маємо

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_a^b c\,dx = c(b-a) = 1$$

звідки $c = \dfrac{1}{b-a}$.

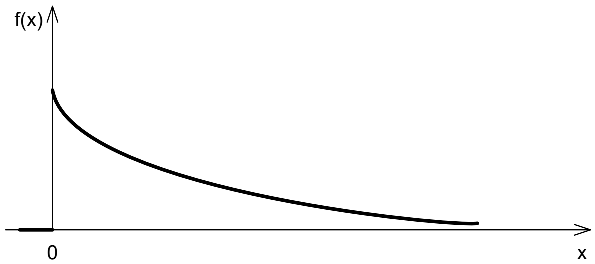
Відповідно, функція розподілу має вигляд

$$F(x) = \int_{-\infty}^{x} f(u)\,du = ¿ \begin{cases} 0, & x \le a ¿ \\ \dfrac{x-a}{b-a}, & a < x \le b ¿¿¿¿ \end{cases}$$



<u>Приклад 2.</u> Випадкова величина *X* має експоненційний (показниковий) розподіл, якщо щільність розподілу дорівнює

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



Очевидно, має місце

$$\int_{-\infty}^{+\infty} f(x)dx=\int_0^{+\infty}\lambda e^{-\lambda x}dx=\int_0^{+\infty} d\left(-e^{-\lambda x}\right)=-e^{-\lambda x}\Big|_0^{\infty}=1;$$

$$F(x)=\int_{-\infty}^{x} f(u)du=\left\{1-e^{-\lambda x},\ x>0\right.$$

Неперервна випадкова величина є математичною абстракцією і у чистому вигляді на практиці не зустрічається хоча б тому, що теоретично не може існувати вимірювальний пристрій, що обчислює значення цієї величини. Тому дослідник завжди має справу з дискретними випадковими величинами (результатами вимірювання). На практиці дискретну випадкову величину апроксимують неперервною, коли множина значень величини рівномірно і „щільно” заповнює деякий відрізок числової осі. Це зручно, бо ймовірнісні характеристики задаються не за допомогою сукупності чисел – ймовірностей елементарних подій, а за допомогою функції $F(x)$ або $f(x)$.

У теоретичному плані багато конструктивних результатів одержано за допомогою поняття неперервної випадкової величини.

2.1.5 Ймовірнісні характеристики неперервних випадкових величин

Нехай маємо неперервну випадкову величину Y, що є борелевою функцією від неперервної випадкової величини X, тобто $Y=\xi(X)$.

Означення. Математичним сподіванням випадкової величини Y називається число:

$$MY=\int_{-\infty}^{+\infty} \xi(x) f_X(x)dx,$$

де $f_X(x)$ – щільність розподілу випадкової величини X, за умови

$$\int_{-\infty}^{+\infty} \xi(x) f_X(x)dx<\infty.$$

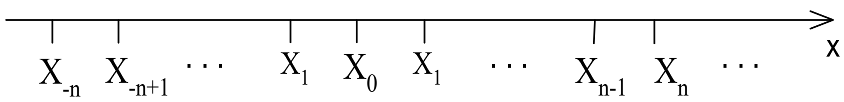
Надалі для простоти будемо вважати розподіли випадкових величин такими, що усі невласні інтеграли , що зустрічаються, існують, та не будемо з'ясовувати умови їх існування.

2.1.6 Обґрунтування запровадження такого означення

Апроксимуємо неперервну випадкову величину Y дискретною випадковою величиною Y*, що будується наступним чином.

Простором елементарних подій для випадкової величини X будемо всю числову вісь.

Розіб'ємо вісь на відрізки достатньо малої довжини Δx:



Якщо в результаті випробування випадкова величина X попадає у відрізок з початковою точкою x_j, то вважаємо, що випадкова величина Y^* одержує значення $\xi(x_j)$. Ймовірність того, що Y^* одержує значення $\xi(x_j)$, з точністю $\vartheta(\Delta x)$ дорівнює $f_X(x_j)\Delta x$.

Будуємо таблицю

$$Y^* \sim \left\{ \xi(x_j) ¿ \right\} ¿ \{\}$$

Тоді, за означенням математичного сподівання дискретної випадкової величини, маємо:

$$MY \approx MY^{¿} \approx \sum_j \xi(x_j) f_X(x_j) \Delta x .$$

При $n \to \infty,\ \Delta x \to 0$, суму замінюємо інтегралом:

$$MY = \int_{-\infty}^{+\infty} \xi(x) f_X(x) dx .$$

2.1.7 Моменти розподілу

Якщо $Y=X$, то маємо формулу для математичного сподівання випадкової величини X:

$$v_1 = MX = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx \quad .$$

Самостійно показати, що властивості математичного сподівання дискретної випадкової величини зберігаються і для неперервного випадка:

$$M(aX+b)=aMX+b.$$

Для неперервної випадкової величини початкові моменти, центральні моменти, дисперсія та середнє квадратичне відхилення означаються співвідношеннями:

$$v_k = \int_{-\infty}^{+\infty} x^k f_X(x)\, dx \quad ,$$

$$\mu_k = \int_{-\infty}^{+\infty} (x - v_1)^k f_X(x)\, dx \quad ,$$

$$\mu_2 = DX = \int_{-\infty}^{+\infty} (x - v_1)^2 f_X(x)\, dx \quad ,$$

$$\sigma_X = +\sqrt{DX} \quad .$$

Самостійно переконатися, що

$$D(CX) = C^2 DX \quad ,$$

$$D(C+X) = DX \quad ,$$

$$DX = v_2 - v_1^2 \quad .$$

Аналогічно дискретному випадку моментна твірна функція неперервної випадкової величини означається як

$$m_X(t) = Me^{Xt} = \int_{-\infty}^{+\infty} e^{xt} f_X(x)\, dx \quad .$$

У припущенні аналітичності функції $m_X(t)$ у нулі довести самостійно, що

$$\frac{d^k m_X(t)}{dt^k}\Big|_{t=0}=\nu_k,$$

$$m_X(t)=1+\frac{\nu_1}{1!}t+\frac{\nu_2}{2!}t^2+\ldots+\frac{\nu_k}{k!}t^k+\ldots.$$

<u>Приклад.</u> Треба обчислити ν_1, ν_2, DX та σ_X випадкової величини X, що має експоненційний розподіл з параметром $\lambda > 0$:

$$f(x)=¿\{\lambda e^{-\lambda x}, x\geq 0;¿¿¿¿$$

Зробимо це за допомогою моментної твірної функції. В даному випадку вона дорівнює

$$m_X(t)=\int_{-\infty}^{+\infty} e^{xt} f(x)\,dx=\int_{0}^{+\infty} \lambda e^{xt} e^{-\lambda x}\,dx=\frac{\lambda}{\lambda-t}.$$

Тоді

$$m'_X(t)=\frac{\lambda}{(\lambda-t)^2}, \quad m''_X(t)=\frac{2\lambda}{(\lambda-t)^3},$$

$$\nu_1=m'_X(t)\big|_{t=0}=\frac{1}{\lambda}, \quad \nu_2=m''_X(t)\big|_{t=0}=\frac{2}{\lambda^2},$$

$$DX=\nu_2-\nu_1^2=\frac{1}{\lambda^2}, \quad \sigma_X=\frac{1}{\lambda}.$$

Інтегруванням частинами самостійно переконайтеся, що ті ж самі результати можна одержати, якщо використовувати „прямі” формули:

$$\nu_1=\int_{-\infty}^{+\infty} x\, f(x)\,dx, \quad \nu_2=\int_{-\infty}^{+\infty} x^2 f(x)\,dx.$$

2.1.8 <u>Розподіл Гаусса[14] (нормальний розподіл)</u>

---

14 Гаусс Карл Фрідріх (1777-1855) – видатний німецький математик, астроном, фізик, геодезист.

Зараз ми ознайомимось з розподілом, який відіграє дуже важливу роль у теорії ймовірностей та математичній статистиці, теорії похибок і т.п.

Випадкова величина має <u>нормальний розподіл</u> (розподіл Гаусса), або називається нормально розподіленою, якщо її щільність розподілу має вигляд:

$$f_X(x)=n(x,a,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}},\ \ \sigma>0,\ \ x\in(-\infty,+\infty).$$

За означенням щільності, функція розподілу дорівнює:

$$F_X(x)=N(x,a,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(u-a)^2}{2\sigma^2}}du.$$

Бачимо, що $n(x,a,\sigma)$ - невід’ємна.

Перевіримо умову нормування (робимо заміну змінних $z=\frac{x-a}{\sigma}$, $dx=\sigma\ dz$):

$$\int_{-\infty}^{+\infty}f_X(x)dx=\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{z^2}{2}}\sigma\ dz=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{z^2}{2}}dz.$$

Перевіримо, що $\int_{-\infty}^{+\infty}e^{-\frac{z^2}{2}}dz=\sqrt{2\pi}$ (інтеграл Пуассона). Звідси буде виходити $\int_{-\infty}^{+\infty}f_X(x)dx=1$. Дійсно, $\int_{-\infty}^{+\infty}e^{-\frac{z^2}{2}}dz\int_{-\infty}^{+\infty}e^{-\frac{y^2}{2}}dy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{(z^2+y^2)}{2}}dzdy$.

Далі переходимо до полярних координат ($z=\rho\cos\phi$, $y=\rho\sin\phi$, $dzdy=\rho\ d\rho\ d\phi$). Одержуємо

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{(z^2+y^2)}{2}}dzdy=\int_{0}^{2\pi}\int_{0}^{\infty}\rho e^{-\frac{\rho^2}{2}}d\rho\ d\phi=\int_{0}^{2\pi}\int_{0}^{\infty}e^{-\frac{\rho^2}{2}}d\left(\frac{\rho^2}{2}\right)d\phi=2\pi.$$

Умову нормування доведено.

Знайдемо вираз для моментної твірної функції нормального розподілу (робиться заміна змінних $\frac{x-a}{\sigma}=z$, $x=\sigma z+a$, $dx=\sigma\, dz$ та $u=z-t\sigma$, $dz=du$.

$$m_X(t)=Me^{Xt}=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{xt-\frac{(x-a)^2}{2\sigma^2}}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{(\sigma z+a)t-z^2}dz=\frac{e^{at+\frac{\sigma^2t^2}{2}}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{\frac{-(z-\sigma t)^2}{2}}dz=$$
$$=\frac{e^{at+\frac{\sigma^2t^2}{2}}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{\frac{-y^2}{2}}dy=e^{at+\frac{\sigma^2t^2}{2}}.$$

Таким чином,

$$m_X(t)=e^{at+\frac{\sigma^2t^2}{2}}.$$

Відшукаємо деякі моменти випадкової величини з нормальним розподілом.

Перевірити самостійно, що

$$m_X'(t)|_{t=0}=a.$$

Це означає, що математичне сподівання величини X дорівнює a:

$$v_1=MX=a.$$

Далі (перевірити самостійно):

$$v_2=\frac{d^2m_X(t)}{dt^2}|_{t=0}=\sigma^2+a^2.$$

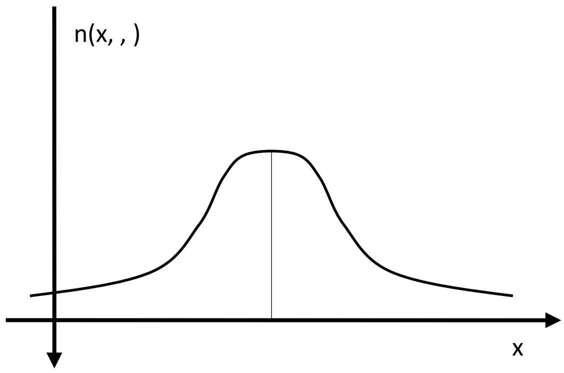
Звідси виходить, що дисперсія випадкової величини з нормальним розподілом дорівнює:

$$DX=v_2-{v_1}^2=\sigma^2,$$

а σ є середнє квадратичне відхилення. Тому надалі будемо писати

$$n(x,v,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-v)^2}{2\sigma^2}}.$$

Графік щільності розподілу $n(x,v,\sigma)$ має вигляд:



Непарні початкові моменти нормальної випадкової величини з нульовим математичним сподіванням $v=0$ дорівнюють нулеві:

$$v_{2n-1}=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^{2n-1}e^{\frac{-x^2}{2\sigma^2}}dx=0,\ n=1,\,2,\,3\ldots$$

тому що під знаком інтеграла – непарна функція, а інтервал інтегрування симетричний відносно нуля. Твірна функція для такої випадкової величини має вигляд

$$m_X(t)=e^{\frac{\sigma^2t^2}{2}}.$$

2.1.9 Функція Лапласа[15]

Функцією Лапласа нормованою називається функція виду

$$\Phi_0(z)=\frac{1}{\sqrt{2\pi}}\int_0^z e^{\frac{-u^2}{2}}du.$$

Властивості.

15Лаплас П’єр Симон (1749 - 1827) – французький математик, фізик і астроном, член Паризької АН.

- При z > 0 функція Лапласа є ймовірність попадання нормальної випадкової величини з параметрами $MX=0$, $DX=1$ в інтервал $[0,z)$.

- $\Phi_0(+\infty)=\frac{1}{2}$, $\Phi_0(-\infty)=-\frac{1}{2}$.

- $\Phi_0(z)$ - непарна функція: $\Phi_0(-z)=-\Phi_0(z)$.

Функція Лапласа табульована. Вона використовується для обчислення ймовірностей попадання нормальної випадкової величини до різних інтервалів.

Відшукаємо ймовірність того, що випадкова величина з щільністю $n(x,v,\sigma)$ попаде до інтервалу $[a,b)$:

$$P(a\le x\le b)=\frac{1}{\sigma\sqrt{2\pi}}\int_a^b e^{\frac{-(x-v)^2}{2\sigma^2}}dx=\left[\frac{x-v}{\sigma}=z,dx=\sigma\,dz\right]=\frac{1}{\sqrt{2\pi}}\int_{\frac{a-v}{\sigma}}^{\frac{b-v}{\sigma}} e^{-\frac{z^2}{2}}dz=$$

$$=\frac{1}{\sqrt{2\pi}}\int_0^{\frac{b-v}{\sigma}} e^{-\frac{z^2}{2}}dz-\frac{1}{\sqrt{2\pi}}\int_0^{\frac{a-v}{\sigma}} e^{-\frac{z^2}{2}}dz=$$

$$=\Phi_0\left(\frac{b-v}{\sigma}\right)-\Phi_0\left(\frac{a-v}{\sigma}\right).$$

2.1.10 Розподіл функції від випадкової величини

Нехай маємо неперервну випадкову величину *X* з щільністю розподілу $f_X(x)$.

Розглянемо випадкову величину *H*:

$$H = h(X) \; .$$

Задача полягає у тому, щоб по щільності розподілу $f_X(x)$ випадкової величини *X* визначити щільність розподілу $f_H(h)$ випадкової величини *H*.

Припустимо, що $f_X(x)$ неперервна, відображення *h=h(x)* взаємно однозначне і обернена функція *x=x(h)* неперервно диференційована.

Розглянемо у просторі елементарних подій для випадкової величини H відрізок $[h, h+dh]$ (тут $dh>0$).

Події – попаданню H у відрізок $[h, h+dh]$ завдяки однозначності $x(h)$ відповідає попадання випадкової величини X до відрізку $[x(h), x(h)+dx(h)]$ (це – одна й та ж подія). Зауважимо, що приріст $dx(h)$ може бути додатнім (якщо $x(h)$ зростає, тобто $\frac{dx(h)}{dh}>0$), а може бути і від'ємним (якщо $x(h)$ спадає, тобто $\frac{dx(h)}{dh}<0$).

Обчислити ймовірність попадання H до $[h, h+dh]$:

$$P(h \le H \le h+dh) = f_H(h)dh + o(dh). \quad (6)$$

Вона повинна дорівнювати ймовірності попадання X до $[x(h), x(h)+dx(h)]$. Маємо,

$$P(X\in[x(h),x(h)+dx(h)])=f_X(x(h))dx(h)+o(dx(h))=f_X(x(h))\frac{dx(h)}{dh}dh+o(dx(h))$$,

якщо $dx(h)>0$, та

$$P(X\in[x(h),x(h)+dx(h)])=-f_X(x(h))dx(h)+o(dx(h))=-f_X(x(h))\frac{dx(h)}{dh}dh+o(dx(h))$$

,

якщо $dx(h)<0$. Звідси

$$P(X\in[x(h),x(h)+dx(h)])=f_X(x(h))\left|\frac{dx(h)}{dh}\right|dh+o(dx(h)). \quad (7)$$

Порівнюючи (6) з (7), одержуємо

$$f_H(h)=f_X(x(h))\left|\frac{dx(h)}{dh}\right|. \quad (8)$$

Приклад. Нехай випадкова величина X має рівномірний розподіл на інтервалі $\left(-\frac{\pi}{2};\frac{\pi}{2}\right)$, тобто щільність розподілу дорівнює

$$f_X(x)=\begin{cases}\dfrac{1}{\frac{\pi}{2}-\left(-\frac{\pi}{2}\right)}=\dfrac{1}{\pi},\ x\in\left(-\frac{\pi}{2};\frac{\pi}{2}\right);\\ 0,\ x\notin\left(-\frac{\pi}{2};\frac{\pi}{2}\right).\end{cases}$$

Треба відшукати щільність розподілу випадкової величини $H=\sin X$.

Зауважимо, що функція $h=\sin x$ на інтервалі $\left(-\frac{\pi}{2};\frac{\pi}{2}\right)$ монотонно зростає, і тому на цьому інтервалі має обернену функцію $x=x(h)=\arcsin h$. Обчислимо її похідну:

$$\frac{dx(h)}{dh}=\frac{1}{\sqrt{1-h^2}}.$$

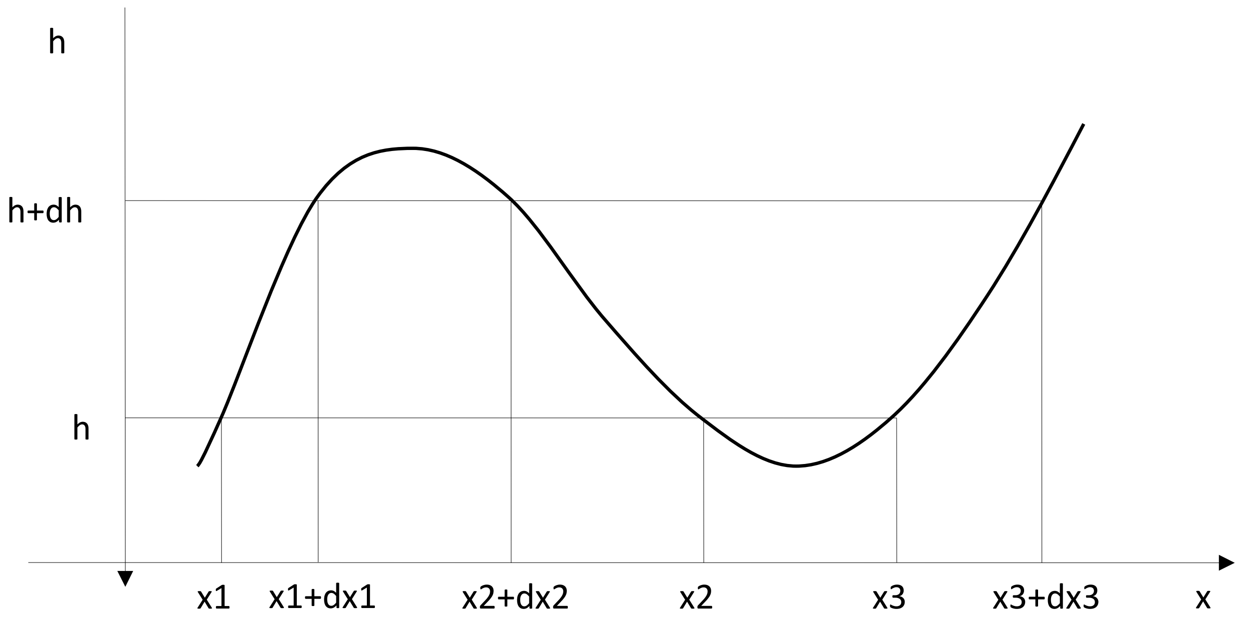
Далі,

$$f_X(x(h))=\begin{cases}\frac{1}{\pi},\ h\in(-1;1);\\ 0,\ h\notin(-1;1).\end{cases}$$

За формулою (8) маємо

$$f_H(h)=f_X(x(h))\left|\frac{dx(h)}{dh}\right|=\begin{cases}\frac{1}{\pi\sqrt{1-h^2}},\ h\in(-1;1);\\ 0,\ h\notin(-1;1).\end{cases}$$

Розглянемо узагальнення (8) на випадок, коли обернена функція до $h=h(x)$ неоднозначна, тобто якщо одному й тому ж $h(x)$ відповідає декілька різних значень x.



У цьому випадку аналізується ймовірність:

$$P(h\le H\le h+dh)=P\left(\bigcup_{x_i:h=h(x_i)} x_i\le X\le x_i+dx_i\right)=\sum_{x_i:h=h(x_i)} P(x_i\le X\le x_i+dx_i).$$

Проводячи побудови, аналогічні попередньому випадку, можна довести, що

$$f_H(h)=\sum_{x_i:h=h(x_i)} f_X(x_i(h))\left|\frac{dx_i(h)}{dh}\right|. \qquad (9)$$

Приклад. Нехай випадкова величина X має розподіл $f_X(x)=n(x,0,1)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\ -\infty<x<\infty.$ Треба обчислити щільність розподілу випадкової величини $H=X^2$.

Тут кожному $h\in(0,\infty)$ відповідає два значення x:

$$x_1(h)=\sqrt{h}\ ;\quad x_2(h)=-\sqrt{h}\ .$$

Похідні від цих обернених функцій дорівнюють:

$$x_1'(h)=\frac{1}{2\sqrt{h}}\ ;\quad x_2'(h)=-\frac{1}{2\sqrt{h}}.$$

Далі,

$$f_X(x_1(h))=\frac{1}{\sqrt{2\pi}}e^{-\frac{h}{2}}\ ,\quad f_X(x_2(h))=\frac{1}{\sqrt{2\pi}}e^{-\frac{h}{2}}.$$

За формулою (9) маємо для $h\in(0,\infty)$:

$$f_H(h)=f_X(x_1(h))\cdot|x_1'(h)|+f_X(x_2(h))\cdot|x_2'(h)|=\frac{1}{2\sqrt{2\pi h}}e^{-\frac{h}{2}}+\frac{1}{2\sqrt{2\pi h}}e^{-\frac{h}{2}}=\frac{1}{\sqrt{2\pi h}}e^{-\frac{h}{2}}.$$

Для $h\notin(0,\infty)$ щільність $f_H(h)=0$.

2.1.11 Нерівність Чебишова[16]

Для додавання багатьох фундаментальних результатів теорії ймовірностей використовуються нерівності, що пов'язані з випадковими величинами. Один з варіантів цих нерівностей ми розглянемо у цьому параграфі.

16Чебишов Пафнутій Львович (1821-1894) – видатний російський математик і механік, академік, член Петербурзької та декількох інших АН.

Нехай випадкова величина Z є неперервною та невід'ємною (тобто її щільність зосереджена на невід'ємній частині вісі z: $f(z)=0$ для z < 0 та $f(z)\geq 0$ для z ¿ 0). Нехай τ - деяке фіксоване число, що більше нуля.

Обчислимо математичне сподівання випадкової величини Z:

$$MZ=\int_{-\infty}^{\infty} zf(z)dz=\int_{0}^{\infty} zf(z)dz=\int_{0}^{\tau} zf(z)dz+\int_{\tau}^{\infty} zf(z)dz\geq\int_{\tau}^{\infty} zf(z)dz\geq\tau\int_{\tau}^{\infty} f(z)dz=\tau P(Z\geq\tau).$$

Звідси маємо:

$$P(Z\geq\tau)\leq\frac{MZ}{\tau}. \qquad (1)$$

Це перший варіант нерівності Чебишова.

Розглянемо випадкову величину $Z=|X-a|$, де X – деяка неперервна випадкова величина, a – деяке число. Тоді зауважимо, що $|X-a|\geq\tau$ та $(X-a)^2\geq\tau^2$ є дві однакові події, тобто $P(|X-a|\geq\tau)=P\left((X-a)^2\geq\tau^2\right)$.

Застосуємо вже одержану нерівність (1). Одержуємо:

$$P(|X-a|\geq\tau)=P\left((X-a)^2\geq\tau^2\right)\leq\frac{M(X-a)^2}{\tau^2}. \qquad (2)$$

Це другий варіант нерівності Чебишова.

У окремому випадку, коли $a=v=MX$, маємо

$$P(|X-v|\geq\tau)\leq\frac{DX}{\tau^2}. \qquad (3)$$

Цією нерівністю ми будемо користуватись неодноразово у подальшому.

А тепер обговоримо найпростіше застосування нерівності Чебишова. Нехай $\tau=t\sigma$. Тоді

$$P(|X-v|\geq t\sigma)\leq\frac{\sigma^2}{t^2\sigma^2}=\frac{1}{t^2}. \qquad (4)$$

Покладемо $t = 3$, тобто шукаємо оцінку ймовірності, що деяка випадкова величина за модулем перевищить математичне сподівання на 3σ (“правило трьох σ -м”). Бачимо, що у цьому випадку

$$P(|X-v|\geq 3\sigma)\leq \frac{1}{9}=0,1111. \qquad (5)$$

Зауважимо, що ці нерівності одержані за мінімальної інформації відносно випадкових величин (розподіл їх невідомий, можна переконатись, що вони вірні і для дискретних випадкових величин).

Зрозуміло, що більш точна інформація про випадкову величину дає більш точні оцінки ймовірності попадання її до деякого інтервалу.

Порівняємо, наприклад, “правило трьох σ -м” (5), якщо відомо, що випадкова величина X має нормальний розподіл з параметрами v та σ. За допомогою функції Лапласа обчислимо ймовірність

$$P(|X-v|<3\sigma)=P(v-3\sigma<X<v+3\sigma)=\Phi_0(3)-\Phi_0(-3)=2\Phi_0(3)=0,9973.$$

Ймовірність протилежної події дорівнює

$$P(|X-v|\geq 3\sigma)=0,0027$$

(порівняйте це з (5)).

Нехай X – деяка випадкова величина з $DX=0$, тоді $P(|X-v|\geq \tau)=0$ для довільного $\tau>0.$ Звідси виходить, що така випадкова величина з імовірністю один є константою, яка дорівнює v.

2.1.12 Деякі загальні властивості математичного сподівання випадкової величини

Розглянемо деякі властивості математичного сподівання випадкової величини, якими ми будемо користуватись надалі.

1. Якщо випадкова величина X невід’ємна, то $MX\geq 0$ (перевірте самостійно у дискретному та неперервному випадках).

2. Розглянемо індикатори подій $\{X \geq 0\}$ і $\{X<0\}$, тобто дискретні випадкові величини:

$$1_{(X\geq 0)}=\begin{cases}1, X\geq 0;\\ 0, X<0;\end{cases}$$

$$1_{(X<0)}=\begin{cases}1, X<0;\\ 0, X\geq 0;\end{cases}$$

Зрозуміло що $1_{(X\geq 0)}+1_{(X<0)}=1$. Тоді випадкову величину X можна подати у вигляді

$$X=X(1_{(X\geq 0)}+1_{(X<0)})=X\cdot 1_{(X\geq 0)}+X\cdot 1_{(X<0)}= X\cdot 1_{(X\geq 0)}-|X|\cdot 1_{(X<0)}$$

або

$$X=X^{+}-X^{-},$$

де $X^{+} = X\cdot 1_{(X\geq 0)}$, $X^{-} = |X|\cdot 1_{(X<0)}$

Зауважимо що X^{+} та X^{-} – невід’ємні випадкові величини.

Далі використаємо адитивну властивість математичного сподівання: математичне сподівання суми випадкових величин дорівнює сумі математичних сподівань доданків (доведення цього факту буде наведено пізніше). З (1) Маємо

$$MX = MX^{+}-MX^{-}, \tag{2}$$

Коли у правій частині (2) MX^{+} та MX^{-} не дорівнюють ∞ одночасно у цьому випадку кажуть, що MX<u>існує</u>. Коли одночасно $MX^{+} = MX^{-} = \infty$, то кажуть, що MX<u>не існує</u>. Коли $MX^{+} = \infty$, $MX^{-}< \infty$, то покладемо $MX = \infty$. Коли $MX^{+}< \infty$, $MX^{-} = \infty$, то покладемо $MX = -\infty$. У обох випадках кажуть, що MX<u>нескінченне</u>. Нарешті коли $MX^{+}< \infty$, $MX^{-}< \infty$, то кажуть, що MX<u>скінченне</u>.

Бачимо, що $|X|$ можна подати у вигляді $|X| = X^{+} + X^{-}$. З цього виходить, що

$$M|X| = MX^{+} + MX^{-},$$

і, таким чином, зі скінченності MX виходить скінченність $M|X|$.

Для моментів більш високих порядків теж зі скінченності MX^k виходить скінченність $M|X|^k$ (для цього треба лише помітити, що $|X^k| = |X|^k$).

3. Якщо $X \geq Y$, та MX і MY скінченні, то $MX \geq MY$.

Дійсно, $X - Y \geq 0$, а тому, за властивістю 1 та за властивістю адитивності, маємо

$$M(X - Y) = MX - MY \geq 0,$$

звідки і виходить потрібна нерівність.

4. Коли MX скінченне, то

$$|MX| \leq M|X|,$$

тобто модуль математичного сподівання випадкової величини не перевищує математичного сподівання модуля цієї величини.

Тривіально виходить з нерівності $-|X| \leq X \leq |X|$ та властивості 3, яка дає $-M|X| \leq MX \leq M|X|$.

5. Якщо скінченним є початковий момент порядку k випадкової величини $|X|$, то скінченним є і момент порядку s ($s \leq k$) цієї величини.

Доведемо це для неперервної випадкової величини X зі щільністю розподілу $f(x)$. За припущенням

$$M|X|^k = \int_{-\infty}^{\infty} |x|^k f(x)\,dx < \infty .$$

Очевидно, що коли $s \leq k$,

$$M|X|^k \geq \int_{|x|\geq 1} |x|^k f(x)\,dx \geq \int_{|x|\geq 1} |x|^s f(x)\,dx ,$$

тобто $\int_{|x|\geq 1} |x|^s f(x)\,dx < \infty$.

Далі зауважимо, що $\int_{|x|<1} |x|^s f(x)\,dx < \infty$. І, таким чином

$$M|X|^s = \int_{-\infty}^{\infty} |x|^s f(x)\,dx = \int_{|x|<1} |x|^s f(x)\,dx + \int_{|x|\ge 1} |x|^s f(x)\,dx < \infty$$

Випадок дискретної випадкової величини X розгляньте самостійно.

Висновок. Коли скінченним є момент порядку k деякої випадкової величини $|X|$, то скінченними є і моменти більш низьких порядків. Ясно і протилежне: якщо деякий момент величини $|X|$ є нескінченним, то нескінченними є усі моменти більш високих порядків (переконайтеся у цьому від супротивного).

2.2 Багатовимірні випадкові величини

2.2.1 Означення. Властивості функції розподілу

Розглянемо ймовірнісний простір $(\Omega, F, P(\cdot))$. Нехай на цьому просторі задано k функцій $\varphi_1(\omega), \varphi_2(\omega), \ldots, \varphi_k(\omega)$. Припустимо, що ці функції вимірні відносно σ-алгебри F, тобто для

$$\{\omega : \varphi_i(\omega) < x_i\} \in F, \; i = 1..k \qquad (1)$$

Таким чином, маємо k випадкових величин, що задані на одному просторі $(\Omega, F, P(\cdot))$.

Означення. Систему з k одновимірних випадкових величин $(\varphi_1(\omega), \varphi_2(\omega), \ldots, \varphi_k(\omega))$, що задані на одному імовірнісному просторі $(\Omega, F, P(\cdot))$, будемо називати k-вимірною випадковою величиною (випадковим вектором).

Для довільних $x_1, x_2, \ldots x_k$, що належить R_k, маємо

$$A_i = \{\omega : \varphi_i(\omega) < x_i\} \in F, \; i = 1..k,$$

тобто A_i є подіями. Тому, з властивостей σ-алгебри виходить, що

$$A = \bigcap_{i=1}^{k} A_i = \bigcap_{i=1}^{k} \{\omega : \phi_i(\omega) < x_i\} \in F,$$

тобто A є теж подією, для якої означена ймовірність $P(\cdot)$.

Взагалі можна довести, що для довільної борелевої множини B в R^k

$$\{ \omega: (\varphi_1(\omega), \varphi_2(\omega), \ldots, \varphi_k(\omega)) \in B\} \in F.$$

Цим ми будемо надалі користуватись без особливих пояснень.

Якщо розглядати x_i, $i=1..k$, як змінні, що змінюються від $-\infty$ до $+\infty$, то приходимо до наступного поняття.

Означення.<u>Функцією</u> розподілу k-вимірної випадкової величини називається функція k змінних

$$F(x_1, x_2, \ldots, x_k) = P\left(\bigcap_{i=1}^{k} \{\omega : \phi_i(\omega) < x_i\}\right) \qquad (2)$$

За традицією, замість (2) використовується запис

$$F(x_1,x_2,\ldots,x_k)=p(\varphi_1(\omega)<x_1,\ \varphi_2(\omega)<x_2,\ldots,\ \varphi_k(\omega)<x_k)$$

А, якщо позначити $X_i = \varphi_i(\omega)$, $i=1..k$, то будемо користуватись записом

$$F(x_1,x_2,\ldots,x_k)=p(X_1<x_1,\ X_2<x_2,\ldots,\ X_k<x_k) \qquad (3)$$

<u>Властивості функції розподілу k-вимірної випадкової величини.</u>

1. Для будь-якого $x=(x_1,x_2,\ldots, x_k) \in R^k$:

$$0 \le F(x_1,x_2,\ldots, x_k) \le 1.$$

Тривіально виходить з тог, що $F(x_1,x_2,\ldots, x_k)$ – імовірність.

2. Функція $F(x_1,x_2,\ldots, x_k)$ є неспадною функцією для будь-якого свого аргументу. Доводиться аналогічно одновимірному випадку.

3. Функція $F(x_1,x_2,\ldots, x_k)$ є неперервною зліва по будь-якій сукупності аргументів. Доводиться аналогічно одновимірному випадку.

4. Якщо деяка $x_i \to -\infty$, то $F(x_1,\ldots,x_i,\ldots, x_k) \to 0$. Доводиться аналогічно одновимірному випадку.

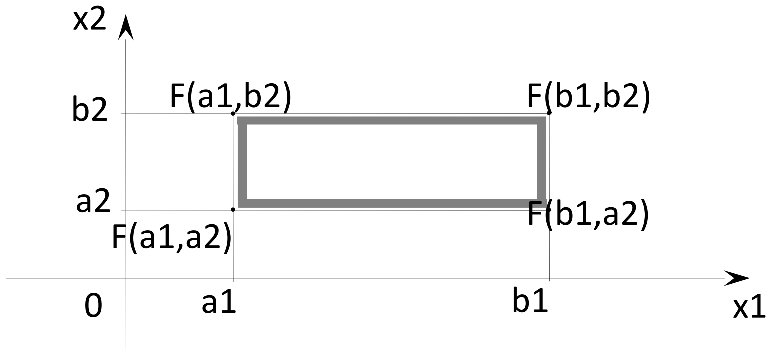
5. Якщо всі $x_i \to \infty$, $i=1..k$, то $F(x_1,\ldots,x_k) \to 1$. Доводиться аналогічно одновимірному випадку.

6. Розглянемо двовимірну випадкову величину

$$X(\omega)=(X_1(\omega), X_2(\omega)).$$

Через функцію розподілу $F(x_1,x_2)=P(X_1<x_1,\ X_2<x_2)$ обчислимо ймовірність попадання її до прямокутника $[a_1,b_1)\times[a_2,b_2)$, тобто

$$P(a_1\le X_1<b_1,\ a_2\le X_2<b_2).$$



За означенням функції розподілу виходить, що

$$P(a_1\le X_1<b_1,\ a_2\le X_2<b_2)=F(b_1,b_2)-F(a_1,b_2)-F(b_1,a_2)+F(a_1,a_2). \tag{4}$$

Виникає питання, які властивості функції $F(x_1,x_2)$ є характеристичними для функції розподілу (а саме, коли функція $F(x_1,x_2)$ є функцією розподілу деякої двовимірної випадкової величини). Без доведення відповімо, що кожна функція, що володіє властивостями 1.-5. та така, що для довільного прямокутника права частина (4) невід’ємна, є функцією розподілу деякої двовимірної випадкової величини (таким чином Теорема Колмогорова розповсюджується на багатовимірні випадкові величини). Функції розподілу компонент X_1 та X_2 виражаються через сумісну функцію розподілу $F(x_1,x_2)$ з очевидних співвідношень:

$$F_{X_1}(x_1)=P(X_1\text{¿}x_1)=P(X_1\text{¿}x_1,X_2\text{¿}\infty)=F(x_1,\infty);$$
$$F_{X_2}(x_2)=P(X_2\text{¿}x_2)=P(X_1\text{¿}\infty,X_2\text{¿}x_2)=F(\infty,x_2).$$

Вони називаються <u>одновимірними функціями розподілу</u> компонент.

В практичних задачах теорії ймовірностей, як правило, зустрічаються багатовимірні випадкові величини двох типів: дискретні, та неперервні.

2.2.2 Двовимірні дискретні випадкові величини

<u>Означення.</u>Двовимірна випадкова величина (X,Y) називається <u>дискретною</u>, якщо кожна компонента може приймати тільки скінченну або злічену множину значень, які позначимо x_i, $i=1..s$; y_j, $j=1..m$.

Випробування, що приводить до необхідності розглядати сумісно пару випадкових величин (X,Y) отже бути фізично одним випробуванням (одночасно вимірювання струму і напруги у мережі) , а також може бути композицією двох випробувань , кожне з яких породжує одновимірну дискретну випадкову величину .

Сумісний розподіл пари (X,Y) зручно задавати ймовірностями $p(x_i,y_j) = P(X = x_i, Y=y_j)$, тобто таблицею виду :

X Y	X_1	. . .	x_i	. . .	x_s
y_1	$P(x_1,y_1)$	. . .	$P(x_i,y_1)$	. . .	$P(x_s,y_1)$
⋮	⋮		⋮		⋮
y_j	$P(x_1,y_j)$	. . .	$P(x_i,y_j)$	. . .	$P(x_s,y_j)$
⋮	⋮		⋮		⋮
y_m	$P(x_1,y_m)$	. . .	$P(x_i,y_m)$	. . .	$P(x_s,y_m)$

Таблиця 1

Ймовірності $P(X = x_i, Y=y_j)$ мають зміст тому, що одно точкові множини (x_i, y_j) є борелевими в R.

Аналогічно одновимірному випадку будемо вважати елементарними подіями різні можливі пари :(x_i, y_j) , $i=\overline{1,s}$, $j=\overline{1,m}$.

В цьому просторі елементарних подій розглянемо складну подію А : внаслідок випробування над випадковою величиною (X,Y) випадкова величина X

приймає значення x_i , а випадкова величина Y – довільне значення , тобто

$$A = \{(x_i, y_1), (x_i, y_2), \ldots, (x_i, y_m)\}.$$

З того, що елементарні події, що складають A, є несумісними , виходить

$$p(X=x_i) = p(A) = \sum_{j=1}^{m} p(x_i, y_j) = p_X(x_i)$$

Запровадимо подію В : внаслідок випробування випадкова величина Y приймає значення y_j, а випадкова величина X – довільне значення , тобто

$$B = \{(x_1, y_j), (x_2, y_j), \ldots, (x_s, y_j)\}$$

Очевидно ,

$$P(Y=y) = p(B) = \sum_{i=1}^{s} p(x_i, y_j) = p_Y(y_j).$$

Розподіли ймовірностей $p_X(x_i)$, $i=\overline{1,s}$, а також $p_Y(y_j)$, $i=\overline{1,s}$, називаються **одновимірними** (маргінальними) розподілами випадкових величин X та Y відповідно.

Для компонент вектора (X, Y) можна говорити про моменти (початкові та центральні) різних порядків. Так, наприклад , математичне сподівання компоненти X визначається як

$$\nu_x = MX = \sum_{i=1}^{s} x_i p_X(x_i) \quad (1)$$

Дисперсію компоненти X визначаємо як

$$\sigma_X^2 = DX = \sum_{i=1}^{s} (x_i - \nu_X)^2 p_X(x_i) \quad (2)$$

Аналогічно, математичне сподівання ν_Y та дисперсія σ_y^2 компоненти Y визначається як

$$\nu_Y = MY = \sum_{j=1}^{m} y_j p_Y(y_j) \qquad (3)$$

$$\sigma_Y^2 = DY = \sum_{j=1}^{m} (y_j - \nu_Y)^2 p_Y(y_j) \qquad (4)$$

Але, на відміну від одновимірних випадкових величин, багатовимірні характеризуються окрім сумісних розподілів також умовними розподілами та відповідними умовними моментами випадкових величин.

Для розглянутих раніше подій A та B добуток дорівнює :

$$AB = (x_i, y_j).$$

Обчислимо умовну ймовірність :

$$p(B/A) = \frac{p(AB)}{p(A)} = \frac{p(x_i, y_j)}{p_X(x_i)} = p(Y = y_j / X = x_i) = p(y_j / x_i)$$

За умови $p_X(x_i) \neq 0$

Аналогічно ,

$$p(A/B) = \frac{p(AB)}{p(B)} = \frac{p(x_i, y_j)}{p_Y(y_j)} = p(X = x_i / Y = y_j) = p(x_i / y_j)$$

За умови $p_Y(y_j) \neq 0$

Переконаємось , що суми умовних ймовірностей дорівнюють одиниці. Дійсно,

$$\sum_{j=1}^{m} p(y_j / x_i) = \sum_{j=1}^{m} \frac{p(x_i, y_j)}{p_X(x_i)} = \frac{p_X(x_i)}{p_X(x_i)} = 1$$

Аналогічно ,

$$\sum_{i=1}^{s} p(x_i / y_j) = 1$$

Означення: Розподіли називаються **умовними розподілами** випадкових величин X та Y, відповідно.

Важливими поняттями в теорії ймовірностей є умовне математичне сподівання та умовна дисперсія.

Означення : Умовним математичним сподіванням випадкової величини Y , за умови , що випадкова величина X одержує значення x_i, називається сума:

$$M(Y/X = x_i) = \overline{y}(x_i) = \sum_{j=1}^{m} y_j\, p(y_j/x_i) .$$

Умовне математичне сподівання – це число (точка на осі y), відносно якого групуються результати випробувань над випадковою величиною Y , якщо в кожному з цих випробувань випадкова величина X приймає значення x_i.

Означення : **Умовною дисперсією** випадкової величини Y , за умови , що випадкова величина X приймає значення x_i, називається сума

$$D(Y/X = x_i) = \sigma^2_{y/x_i} = \sum_{j=1}^{m} \left(y_j - \overline{y}(x_i)\right)^2 p(y_j/x_i) \qquad (5)$$

Умовна дисперсія визначає ступінь концентрації результатів випробувань над величиною Y відносно умовного математичного сподівання , якщо відомо, що в кожному з цих випробувань випадкова величина X приймає значення x_i.

До цих понять ми ще повернемось після розгляду , так званих , неперервних багатовимірних випадкових величин.

Приклад : Розглянемо довільну випадкову величину (X,Y) , сумісний розподіл задано таблицею 2.

У

x_i / y_j	0	1	2	3	4	5	6	$P_Y(y_j)$
0	0,202	0,174	0,113	0,062	0,049	0,023	0,004	0,627
1	0	0,099	0,064	0,040	0,031	0,020	0,006	0,260
2	0	0	0,031	0,025	0,018	0,013	0,008	0,095
3	0	0	0	0,001	0,002	0,004	0,011	0,018
$p_X(x_i)$	0,202	0,273	0,208	0,128	0,100	0,060	0,029	1

останньому рядку та останньому стовпчику цієї таблиці обчислені одновимірні розподіли компонент *X* та *Y*, відповідно. Перевірте що одновимірні моменти компонент , що обчислюються за формулами (1) – (4), дорівнюють:

$$\nu_X = 1{,}947\,; \qquad \sigma_X^2 = 2{,}61$$

$$\nu_Y = 0{,}504\,; \qquad \sigma_Y^2 = 0{,}548$$

Умовний розподіл компоненти *Y*, тобто ймовірності $p(Y_j/x_i)$, представлено у таблиці 3.

Таблиця 3

x_i / y_j	0	1	2	3	4	5	6
0	1,000	0,637	0,543	0,484	0,490	0,383	0,138
1	0	0,363	0,308	0,313	0,310	0,333	0,207
2	0	0	0,149	0,195	0,180	0,217	0,276
3	0	0	0	0,008	0,020	0,067	0,379
$\overline{y}(x_i)$	0	0,363	0,606	0,729	0,730	0,968	1,896
σ^2_{y/x_i}	0	0,231	0,537	0,636	0,677	0,867	1,121

У останніх рядках цієї таблиці обчислені значення умовного математичного сподівання та умовної дисперсії випадкової величини *Y*.

2.2.3 Двовимірні неперервні випадові величини

Нехай знову маємо імовірнісний простір $(\Omega, F, p(.))$ та пару одновимірних випадкових величин $(X(\omega), Y(\omega))$ що задані на ньому. Але на цей раз розглянемо випадок, коли множина можливих значень цієї пари – незлічена

(це може бути вся площина *xy*, або деяка її область ненульової площі).

Означення: Говорять що двовимірна випадкова величина (X,Y) має абсолютно неперервний сумісний розподіл, якщо існує така невід'ємна функція f(x,y), що для довільних $(x, y) \in R^2$ її функція розподілу F(x,y) = $p(X<x, Y<y)$може бути надана у вигляді

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv \quad . (1)$$

Функція $f(x,y)$називається **сумісною щільністю розподілу** випадкової величини *(X,Y),* а саму випадкову величину, що володіє цією властивістю, будемо називати **неперервною**.

Зауважимо, що з (1) виходить, що функція розподілу неперервної двовимірної випадкової величини є неперервною функцією.

Надалі будемо розглядати випадок, коли щільність $f(x,y)$ є неперервною функцією майже усюди (за виключенням, можливо, множини точок в R^2 міри(площі) нуль). Тоді майже усюди щільність є другою мішаною частинною похідною функції розподілу

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x_1 \partial x_2} .$$

В точках неперервності функції щільності $f(x,y)$ маємо

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = \lim_{\substack{\Delta x -> \infty \\ \Delta y -> \infty}} \frac{[F(x+\Delta x, y+\Delta y) - F(x, y+\Delta y)] - [F(x+\Delta x, y) - F(x,y)]}{\Delta x \Delta y} =$$

$$= \lim_{\substack{\Delta x -> \infty \\ \Delta y -> \infty}} \frac{p(x \le X < x+\Delta x, y \le Y < y+\Delta y)}{\Delta x \Delta y}.$$

Звідси $p(x \le X < x+\Delta x, y \le Y < y+\Delta y) = f(x,y)\Delta x \Delta y + \sigma(\Delta x \Delta y)$, і ,таким чином , $f(x,y)\Delta x \Delta y$ дає (наближено) ймовірність попадання (X,Y)в достатньо малий окіл точки (x,y). Можна показати , що для довільної борелевої множини $G \in R^2$ означена ймовірність

$$p((X,Y) \in G) = \iint_G f(x,y) dxdy. \quad (2)$$

Мають місце такі основні властивості функції щільності:

4. $f(x,y) \ge 0$ (виходить з означення);

5. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) dxdy = F(\infty,\infty) = 1. \quad (3)$

<u>Приклад 1.</u> Двовимірна випадкова величина (X,Y) має щільність розподілу

$$f(x,y) = \frac{A}{\pi^2 (16+x^2)(25+y^2)}.$$

Треба визначити константу A та функцію розподілу $F(x,y)$.

Для визначення A використаємо умову нормування (3). Маємо:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{A dxdy}{\pi^2 (16+x^2)(25+y^2)} = \frac{A}{20\pi^2} arctg \frac{x}{4} \Big|_{-\infty}^{\infty} \cdot arctg \frac{y}{5} \Big|_{-\infty}^{\infty} =$$

$$= \frac{A}{20\pi^2}\left(\frac{\pi}{2}+\frac{\pi}{2}\right)\left(\frac{\pi}{2}+\frac{\pi}{2}\right) = \frac{A}{20} = 1.$$

Звідси $A = 20$.

Далі, за означенням (1):

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}\frac{20dudv}{(16+u^2)(25+v^2)}=\frac{20}{\pi^2}\int_{-\infty}^{x}\frac{du}{16+u^2}\int_{-\infty}^{y}\frac{dv}{25+v^2}=\frac{1}{\pi^2}arctg\frac{u}{4}\Big|_{-\infty}^{x}\cdot arctg\frac{v}{5}\Big|_{-\infty}^{y}=$$

$$=\frac{1}{\pi^2}(arctg\frac{x}{4}+\frac{\pi}{2})(arctg\frac{y}{5}+\frac{\pi}{2})=(\frac{1}{\pi}arctg\frac{x}{4}+\frac{1}{2})(\frac{1}{\pi}arctg\frac{y}{5}+\frac{1}{2}).$$

<u>Приклад 2.</u> Випадкова величина (X,Y) має щільність розподілу

$$f(x,y)=\begin{cases}c(R-\sqrt{x^2+y^2}), & x^2+y^2\le R^2;\\ 0, & x^2+y^2>R^2.\end{cases}$$

Треба визначити константу c та ймовірність попадання до круга радіуса $a<R$ з центром у початку координат.

З умови нормування (3) маємо (робимо заміну змінних $x=\rho\cos\phi$, $y=\rho\sin\phi$):

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=\iint_{x^6+y^2\le R^2}c(R-\sqrt{x^2+y^2})dxdy=c\int_0^{2\pi}\int_0^{R}(R-\rho)\rho d\rho d\phi=$$

$$=c\cdot 2\pi\cdot R^3(\frac{1}{2}-\frac{1}{3})=c\frac{\pi R^3}{3}=1.$$

Звідси $c=\frac{3}{\pi R^3}$.

Далі

$$P\big((X,Y)\in\{x^2+y^2<a^2\}\big)=\iint_{x^2+y^2<a^2}\frac{3}{\pi R^3}(R-\sqrt{x^2+y^2})dxdy=$$

$$=\frac{3}{\pi R^3}\int_0^{2\pi}\int_0^{a}(R-\rho)\rho d\rho d\phi=\frac{3a^2}{R^2}(1-\frac{2a}{3R}).$$

Знову розглянемо двовимірну випадкову величину (X,Y) із щільністю розподілу $f(x,y)$. Поставимо задачу: відшукати одновимірні (маргінальні) щільності розподілу $f_X(x)$, $f_Y(y)$ випадкових величин X та Y відповідно.

Для довільного $x \in R^1$ функція розподілу випадкової величини X дорівнює

$$F(x)=P(X<x)=P(X<x,Y<\infty)=\int_{-\infty}^{x}\left[\int_{-\infty}^{\infty} f(u,v)\,dv\right]du.$$

Звідси (за означенням функції щільності):

$$f_X(u)=\int_{-\infty}^{\infty} f(u,v)\,dv \qquad (4)$$

Аналогічно, для щільності розподілу випадкової величини Y маємо:

$$f_Y(v)=\int_{-\infty}^{\infty} f(u,v)\,du \qquad (5)$$

Висновок: для того, щоб по двовимірній функції щільності знайти одновимірну функцію щільності однієї з компонент, треба двовимірну функцію щільності проінтегрувати по осі, що відповідає другій випадковій величині.

Знаючи одновимірні функції щільності випадкових величин X та Y, можна визначити їх математичні сподівання та дисперсії за формулами:

$$v_X=MX=\int_{-\infty}^{\infty} x f_X(x)\,dx=\int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x,y)\,dydx;$$

$$v_Y=MY=\int_{-\infty}^{\infty} y f_Y(y)\,dy=\int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x,y)\,dxdy;$$

$$\sigma_{X^2}=DX=\int_{-\infty}^{\infty} (x-v_X)^2 f_X(x)\,dx=\int_{-\infty}^{\infty} (x-v_X)^2 \int_{-\infty}^{\infty} f(x,y)\,dydx;$$

$$\sigma_{Y^2}=DY=\int_{-\infty}^{\infty} (y-v_Y)^2 f_Y(y)\,dy=\int_{-\infty}^{\infty} (y-v_Y)^2 \int_{-\infty}^{\infty} f(x,y)\,dxdy;$$

По аналогії з дискретним випадком розглянемо умовні розподіли та умовні моменти для двовимірних випадкових величин.

Означення.<u>Умовна функція щільності</u> випадкової величини Y за умови, що випадкова величина X одержує значення x, визначається рівністю

$$f_{Y/X=x}(y/x)=f_{Y/x}(y/x)=\frac{f(x,y)}{f_X(x)} \qquad (6)$$

Умовна функція щільності випадкової величини X, за умови, що Y одержує значення y, визначається рівністю

$$f_{X/Y=y}(x/y)=f_{X/y}(x/y)=\frac{f(x,y)}{f_Y(y)} \qquad (7)$$

Доцільність такого означення в точках неперервності $f(x,y)$ можна пояснити таким чином. Розглянемо події:

$A=\{x\le X<x+\Delta x, Y<\infty\}$,

$B=\{x<\infty, y\le Y<y+\Delta y\}$,

$AB=\{x\le X<x+\Delta x, y\le Y<y+\Delta y\}$.

Імовірності цих подій дорівнюють:

$P(A)=f_X(x)\Delta x+o(\Delta x)$,

$P(B)=f_Y(y)\Delta y+o(\Delta y)$,

$P(AB)=f(x,y)\Delta x\Delta y+o(\Delta x\cdot\Delta y)$.

Умовна ймовірність $P(B/A)$ дорівнює

$$P(B/A)=P(y\le Y<y+\Delta y/x\le X<x+\Delta x)=$$

$$=\frac{f(x,y)\Delta x\Delta y+o(\Delta x\Delta y)}{f_X(x)\Delta x+o(\Delta x)}=\frac{f(x,y)}{f_X(x)}\Delta y+o(\Delta y).$$

Звідси стає зрозумілим запровадження означень (6) та (7).

Означення.<u>Умовним математичним сподіванням</u> випадкової величини Y за умови, що X одержує значення x, називається величина

$$M(Y/X=x)=\overline{y}(x)=\int_{-\infty}^{\infty} yf_{Y/x}(y/x)dy \qquad (8)$$

Функцію $\overline{y}(x)$ називають функцією регресії Y на X, а її графік – лінією регресії.

Умовне математичне сподівання X визначається рівністю

$$M(X/Y=y)=\overline{x}(y)=\int_{-\infty}^{\infty} x f_{X/y}(x/y)\,dx \qquad (9)$$

Умовною дисперсією величини Y за умови, що X одержує значення x, називається величина

$$D(Y/X=x)=\sigma_{Y/x}{}^2=\int_{-\infty}^{\infty}\left(y-\overline{y}(x)\right)^2 f_{Y/x}(y/x)\,dy \qquad (10)$$

Аналогічно:

$$D(X/Y=y)=\sigma_{X/y}{}^2=\int_{-\infty}^{\infty}\left(x-\overline{x}(y)\right)^2 f_{X/y}(x/y)\,dx \qquad (11)$$

В багатьох задачах теорії ймовірностей, наприклад, в задачах стохастичного прогнозування, доводиться розглядати умовне математичне сподівання і умовну дисперсію не як функції детермінованої змінної, а як функції випадкової величини.

Як функція від випадкової величини умова на математичне сподівання $\overline{y}(X)=M(Y/X)$ теж є випадковою величиною, і тому можна говорити про її математичне сподівання. Доведемо, що

$$M\,\overline{y}(X)=MM(Y/X)=MY\,.$$

Ця формула має назву формули повного математичного сподівання.

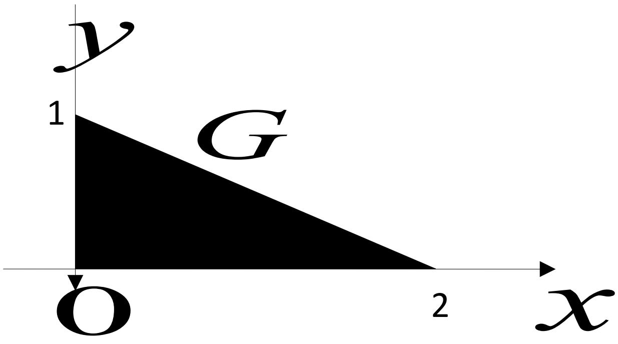
Дійсно, врахувавши (6), маємо:

$$M\,\overline{y}(X)=\int_{-\infty}^{\infty}\overline{y}(x)f_X(x)\,dx=\int_{-\infty}^{\infty}\int y f_{Y/x}(y/x)\,dy f_X(x)\,dx=$$

$$=\int_{-\infty}^{\infty} y\left[\int_{-\infty}^{\infty}\frac{f(x,y)}{f_X(x)}f_X(x)\,dx\right]dy=\int_{-\infty}^{\infty} y f_Y(y)\,dy=MY\,.$$

Пропонуємо самостійно довести формулу повного математичного сподівання для дискретної випадкової величини (X,Y).

<u>Приклад 3.</u> Нехай випадкова величина (X,Y) має рівномірний розподіл у трикутнику G з вершинами у точках (0; 0), (0; 1), (2; 0).



Це означає, що щільність розподілу (X,Y) дорівнює

$$f(x,y)=\begin{cases}\dfrac{1}{S_G}, & (x,y)\in G;\\ 0, & (x,y)\notin G,\end{cases}$$

де S_G – площа області G.

В нашому випадку $S_G=1$ і, таким чином,

$$f(x,y)=\begin{cases}1, & (x,y)\in G;\\ 0, & (x,y)\notin G.\end{cases}$$

Треба обчислити маргінальні щільності розподілу $f_X(x)$, $f_Y(y)$, умовні щільності $f_{X/y}(x/y)$, $f_{Y/x}(y/x)$, умовні математичні сподівання $\bar{y}(x)$, $\bar{x}(y)$ та умовні дисперсії $\sigma_{X/y}{}^2$, $\sigma_{Y/x}{}^2$.

Враховуючи те, що рівняння прямої, що проходить через точки (2; 0) та (0; 1), має вигляд $y=1-\dfrac{x}{2}$, одержуємо:

$$f_X(x)=\begin{cases}\displaystyle\int_0^{1-\frac{x}{2}} dy=1-\frac{x}{2}, & x\in[0;2];\\ 0 & x\notin[0;2].\end{cases}$$

$$f_Y(y)=\begin{cases}\int\limits_0^{2(1-y)} dx=2(1-y), & y\in[0;1];\\ 0 & y\notin[0;1].\end{cases}$$

Умовні щільності розподілу, які обчислюються за формулами (6) та (7), дорівнюють

$$f_{X/y}(x/y)=\begin{cases}\dfrac{1}{2(1-y)}, & x\in[0;2(1-y)], y\in[0,1];\\ 0 & x\notin[0;2(1-y)], y\in[0,1].\end{cases}$$

$$f_{Y/x}(y/x)=\begin{cases}\dfrac{2}{2-x}, & x\in[0;2], y\in[0,1-\dfrac{x}{2}];\\ 0 & x\in[0;2], y\notin[0,1-\dfrac{x}{2}].\end{cases}$$

При $y\notin[0;1]$ та $x\notin[0;2]$ умовні розподіли $f_{X/y}(x/y)$ та $f_{Y/x}(y/x)$ не означені (знаменники в (6) та (7) дорівнюють нулю).

Умовні математичні сподівання обчислюємо за формулами (8), (9):

$$\bar{x}(y)=\int\limits_0^{2(1-y)}\frac{xdx}{2(1-y)}=\frac{x^2}{4(1-y)}\Big|_0^{2(1-y)}=1-y;$$

$$\bar{y}(x)=\int\limits_0^{1-x/2}\frac{2ydy}{2-x}=\frac{y^2}{2-x}\Big|_0^{1-x/2}=\frac{2-x}{4}.$$

Умовні дисперсії обчислюємо за формулами (10), (11):

$$\sigma_{X/y^2}=\int\limits_0^{2(1-y)}\frac{(x-(1-y))^2dx}{2(1-y)}=\frac{(1-y)^2}{3};$$

$$\sigma_{Y/x^2}=\int\limits_0^{1-x/2}(y-\frac{2-x}{4})^2\frac{2dy}{2-x}=\frac{(2-x)^2}{48}.$$

2.2.4 Незалежні випадкові величини

Означення. Випадкові величини X та Y (або компоненти двовимірної випадкової величини (X,Y)), що задані на одному ймовірнісному просторі, називаються незалежними, якщо для довільних x, y має місце

$$F(x,y)=P(X<x,Y<y)=P(X<x)\cdot P(Y<y)=F_X(x)\cdot F_Y(y)\quad , \qquad (1)$$

тобто сумісна функція розподілу розпадається на добуток функцій розподілу компонент. В протилежному випадку X та Y <u>залежні</u>.

Бачимо, що випадкові величини X та Y незалежні, коли для довільних x,y незалежними є події $\{X<x\}$ та $\{Y<y\}$.

З означення виходить, що і всі події $\{a_1\le X<b_1\}$ та $\{a_2\le Y<b_2\}$ незалежні. Дійсно,

$$P(a_1\le X<b_1,a_2\le Y<b_2)=F(b_1,b_2)-F(a_1,b_2)-F(b_1,a_2)+F(a_1,a_2)=$$
$$=F_X(b_1)F_Y(b_2)-F_X(a_1)F_Y(b_2)-F_X(b_1)F_Y(a_2)+F_X(a_1)F_Y(a_2)=$$
$$=[F_X(b_1)-F_X(a_1)][F_Y(b_2)-F_Y(a_2)]=P(a_1\le X<b_1)\cdot P(a_2\le Y<b_2)\quad ,$$

тобто $\{a_1\le X<b_1\}$ та $\{a_2\le Y<b_2\}$ незалежні.

Ясно, що вірне й протилежне: якщо для довільних a_1,b_1,a_2,b_2 події $\{a_1\le X<b_1\}$ та $\{a_2\le Y<b_2\}$ незалежні, то X та Y – незалежні випадкові величини.

Можна довести і більш загальний результат: випадкові величини X та Y незалежні тоді і тільки тоді, коли незалежними є події $\{X\in G_1\}$ та $\{Y\in G_2\}$ для довільних борелевих множин G_1 та G_2 на відповідних осях.

Нехай X та Y – незалежні випадкові величини, а $\varphi_1(x)$, $\varphi_2(y)$ – борелеві функції. Тоді випадкові величини $\varphi_1(X)$, $\varphi_2(Y)$ теж незалежні.

Дійсно, позначимо $G_1 = ¿ ¿$. Це борелеві множини і тому

$$P(\phi_1(X)<z_1,\phi_2(Y)<z_2)=P(\{X\in G_1\}\cap\{Y\in G_2\})=$$
$$¿P(X\in G_1)P(Y\in G_2)=P(\phi_1(X)<z_1)P(\phi_2(Y)<z_2)$$

що і означає незалежність $\varphi_1(X)$, $\varphi_2(Y)$.

Розглянемо дискретний та неперервний випадок величини (X,Y).

Дискретний випадок

Нехай X та Y – дискретні випадкові величини, що приймають значення x_i, $i=1..s$ та y_j, $j=1..m$ відповідно. Сумісний розподіл їх задано ймовірностями

$$p(x_i,y_j)=p_X(x_i)\cdot p_Y(y_j) \quad (2)$$

для всіх і та j.

Дійсно, з урахуванням (2), сумісна функція розподілу їх може бути подана у вигляді

$$\begin{gathered} F(x,y)=P(X<x,Y<y)=\sum_{i:x_i<x}\sum_{j:y_j<y}p(x_i,y_j)= \\ =\sum_{i:x_i<x}\sum_{j:y_j<y}p_X(x_i)p_Y(y_j)=\sum_{i:x_i<x}p_X(x_i)\sum_{j:y_j<y}p_Y(y_j)= \\ =P(X<x)P(Y<y), \end{gathered}$$

а це, згідно з (1), і означає незалежність X та Y.

Навпаки, з незалежності X та Y виходить незалежність подій $\{X=x_i\}$ і $\{Y=y_j\}$ (одноточкові множини є борелевими), тобто має місце (2).

Переконайтеся самостійно, що умовні розподіли та умовні моменти цих випадкових величин співпадають з безумовними:

$$\begin{gathered} p(y_j/x_i)=p_Y(y_j),\, p(x_i/y_j)=p_X(x_i), \\ \overline{y}(x_i)=MY,\, \overline{x}(y_j)=MX, \\ \sigma^2_{Y/x_i}=\sigma^2_Y,\, \sigma^2_{X/y_j}=\sigma^2_X. \end{gathered}$$

Приклад 1. Підкидається гральний кубик два рази, X – кількість точок на верхній грані при першому підкиданні, Y – при другому. Двовимірна випадкова величина (X,Y) приймає 36 різних рівноможливих значень (x_i,y_j), де $x_i=\overline{1,6}$, $y_j=\overline{1,6}$. Тому

$$p(x_i,y_j)=\frac{1}{36},\, i,j=\overline{1,6}$$

.

З іншого боку, $p_X(x_i)=\frac{1}{6}, p_Y(y_j)=\frac{1}{6}$. І, таким чином, $p(x_i, y_j)=p_X(x_i)\cdot p_Y(y_j)$, тобто X та Y – незалежні.

<u>Приклад 2</u>. Підкидається монета. Розглянемо дві випадкові величини

$$X=\begin{cases}1, \text{ якщо зверху „орел”;}\\ 0, \text{ якщо зверху „решка”,}\end{cases}$$

$$Y=\begin{cases}1, \text{ якщо знизу „орел”;}\\ 0, \text{ якщо знизу „решка”.}\end{cases}$$

Таблиця сумісного розподілу $P(x_i,y_j)$ у даному випадку має вигляд

Y \ *X*	*0*	*1*
0	*0*	$\frac{1}{2}$
1	$\frac{1}{2}$	*0*

Але $p_X(0)=p_X(1)=p_Y(0)=p_Y(1)=\frac{1}{2}$, і тому $p(x_i, y_j)\neq p_X(x_i)\cdot p_Y(y_j)$ тобто X та Y – залежні.

<u>Неперервний випадок</u>

Нехай випадкова величина (X,Y) – неперервна з функцією розподілу $F(x,y)$ та щільності $f(x,y)$.

Випадкові величини X та Y незалежні тоді та й тільки тоді, коли

$$f(x,y)=f_X(x)\cdot f_Y(y), \quad (3)$$

тобто сумісна щільність розподілу розпадається на добуток щільностей компонент.

Дійсно, нехай *X* та *Y* – незалежні. Тоді

$$\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dudv = F(x,y) = F_X(x)\cdot F_Y(y) =$$
$$¿\int_{-\infty}^{x} f_X(u)\,du \int_{-\infty}^{y} f_Y(v)\,dv = \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(u) f_Y(v)\,dudv$$

для довільних *x* та *y*. Така рівність можлива за рівності підінтегральних функцій, тобто коли має місце (3). Навпаки, припускаючи (3), одержуємо (1) (для доведення використовуємо останні рівності у зворотному порядку).

Приклад 3. Випадкова величина (*X*,*Y*) має щільність розподілу

$$f(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}, -\infty < x < \infty, -\infty < y < \infty .$$

Тоді

$$f(x,y) = f_X(x) f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} ,$$

тобто *X* та *Y* – незалежні.

Переконатися самостійно, що і у випадку незалежних неперервних випадкових величин умовні розподіли та умовні моменти співпадають з безумовними.

Розглянемо багатовимірний випадок.

Маємо *n* випадкових величин $X_1,X_2,\ldots,X_n$, що задані на одному ймовірнісному просторі.

Означення. Випадкові величини $X_1,X_2,\ldots,X_n$, незалежні, якщо сумісна функція розподілу розпадається на добуток функцій розподілу цих випадкових величин, тобто для довільних $x_1,x_2,\ldots,x_n$ має місце

$$F(x_1,x_2,\ldots,x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) ,$$

або

$$P\left(X_1<x_1, X_2<x_2, \ldots, X_n<x_n\right)=P\left(X_1<x_1\right)\cdot P\left(X_2<x_2\right)\cdot\ldots\cdot P\left(X_n<x_n\right)$$

тобто події $\{X_1<x_1\},\ldots,\{X_n<x_n\}$ – незалежні.

Ясно, що йдеться про незалежність у сукупності. Дійсно, якщо покласти деякі x_i рівними $+\infty$ та використати $F_{X_i}(\infty)=P(X_i ¿ \infty)=1$, одержимо

$$P\left(X_{i_1}<x_{i_1}, X_{i_2} ¿ x_{i_2}, \ldots, X_{i_n} ¿ x_{i_n}\right)=$$
$$=P\left(X_{i_1}<x_{i_1}\right)\cdot P\left(X_{i_2} ¿ x_{i_2}\right)\cdot\ldots\cdot P\left(X_{i_n} ¿ x_{i_n}\right), 1\le k\le n.$$

На багатовимірні випадкові величини розповсюджуються результати, що одержані для двовимірних випадкових величин (X,Y).

А саме $X_i, i=\overline{1,n}$, є дискретні незалежні випадкові величини, тоді та й тільки тоді, коли

$$P\left(X_1=x_1, X_2=x_2, \ldots, X_n=x_n\right)=P\left(X_1=x_1\right)\cdot P\left(X_2=x_2\right)\cdot,\ldots,\cdot P\left(X_n=x_n\right)$$

для різних можливих $x_1,x_2,\ldots,x_n$.

Неперервні випадкові величини $X_i, i=\overline{1,n}$, незалежні тоді та й тільки тоді, коли сумісна щільність розподілу $f(x_1, x_2,\ldots, x_n)$ розпадається на добуток маргинальних щільностей $f_{X_i}(x_i)$:

$$f\left(x_1, x_2, \ldots, x_n\right)=\prod_{i=1}^{n} f_{X_i}\left(x_i\right)$$

для довільних можливих $x_1, x_2,\ldots, x_n$.

Випадкові величини $X_1,X_2,\ldots,X_n$ незалежні тоді і тільки тоді, коли для довільних борелевих множин $G_1, G_2,\ldots, G_n$

$$P\left(\mathop{intersect}_{i=1}^{n}\left\{X_i\in G_i\right\}\right)=\prod_{i=1}^{n} P\left(X_i\in G_i\right)$$

(без доведення).

Якщо $X_1, X_2, \ldots, X_n$ – незалежні випадкові величини, а $\phi_1(x_1), \phi_2(x_2), \ldots, \phi_n(x_n)$ – борелеві функції, то і випадкові величини $\phi_1(X_1), \phi_2(X_2), \ldots, \phi_n(X_n)$ також незалежні (доведення цього факту повторює доведення у двовимірному випадку).

2.2.5 Математичне сподівання функцій декількох випадкових величин

Розглянемо спочатку скалярну функцію $\varphi(X,Y)$ від двох випадкових величин X, Y, тобто деякі відображення $\varphi(x,y)$: $R^2 \rightarrow R^1$. Якщо функція $\varphi(x,y)$ борелева, то $Z=\varphi(X,Y)$ є випадковою величиною і можна говорити про її розподіл, моменти тощо.

Нехай (X,Y) – дискретна випадкова величина і заданий її сумісний розподіл $p(x_i, y_i), i=\overline{1,n}, j=\overline{1,m}$.

Означення.Математичним сподіванням функції $Z=\varphi(X,Y)$ називається величина

$$MZ = M\phi(X,Y) = \sum_{i=1}^{s} \sum_{j=1}^{m} \phi(x_i, y_j)\, p(x_i, y_j) \quad , (1)$$

за умови збіжності рядів, коли s,m нескінчені.

Бачимо, що це означення узгоджене з означенням математичного сподівання одновимірної дискретної випадкової величини тому, що сумуються всі можливі значення величин Z з вагами, що дорівнюють імовірностям, з якими приймаються ці значення.

Якщо (X,Y) – неперервна двовимірна випадкова величина із щільністю розподілу $f(x,y)$ та скалярна функція $z=\varphi(x,y)$ є борелевою, то приходимо до означення.

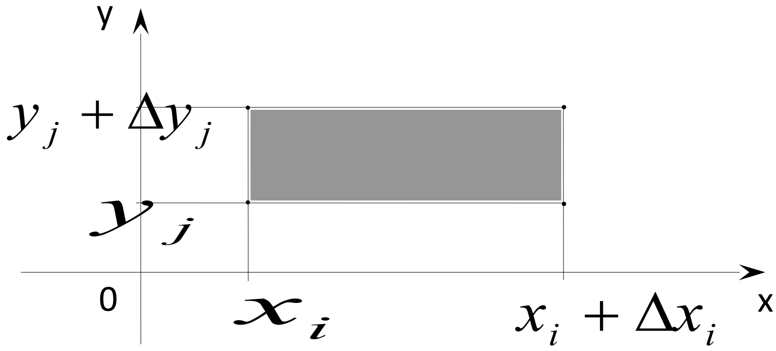
Означення.Математичним сподіванням функції $Z_1=\varphi(X,Y)$ називається величина

$$MZ = M\phi(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(x,y)f(x,y)dxdy \quad , (2)$$

якщо інтеграл у правій частині останньої рівності існує.

Мотивацію цього означення можна одержати з таких міркувань.

Розіб’ємо площину *x*,*y* на маленькі прямокутники зі сторонами, що визначені координатами $(x_i, x_i + \Delta x_i), (y_i, y_i + \Delta y_i)$.



Розглянемо дискретну випадкову величину $\phi^{¿}(X,Y)$, яка приймає значення $\phi^{¿}(x_i, y_i) = \phi(x_i, y_i)$ якщо неперервна випадкова величина (*X*,*Y*) попадає у вказаний прямокутник. Як зазначалось раніш, імовірність відповідної події (у точках неперервності *f*(*x*,*y*)) дорівнює

$$f(x_i, y_j)\Delta x_i \Delta y_j + o(\Delta x_i, \Delta y_j) \quad .$$

Математичне сподівання величини $\phi^{¿}(X,Y)$ слід шукати як

$$M\phi^{¿} = \sum_i \sum_j \left(\phi(x_i, y_j) f(x_i, y_j)\Delta x_i \Delta y_j + o(\Delta x_i, \Delta y_j)\right) \quad .$$

Звідси стає зрозумілою доцільність означення (2), якщо існує границя правої частини останньої рівності при $\max(\Delta x_i, \Delta y_j) \to 0$ (границя інтегральної суми).

Більш загальний випадок – це випадок, коли маємо n-вимірну випадкову величину (X_1, X_2,…, X_n) та скалярну борелеву функцію $\phi\left(X_1, X_2, \ldots, X_n\right)$.

Для дискретної випадкової величини (X_1, X_2,…, X_n) можливі значення позначимо $\left(x_{i_1}^1, x_{i_2}^2, \ldots, x_{i_n}^n\right)$, де $x_{i_j}^j, i_j = \overline{1, s_j}$, - можливі значення j-ої компоненти $X_j, j = \overline{1,n}$. Сумісний розподіл задається ймовірностями $p\left(x_{i_1}^1, x_{i_2}^2, \ldots, x_{i_n}^n\right) = P\left(X_1 = x_{i_1}^1, X_2 = x_{i_2}^2, \ldots, X_n = x_{i_n}^n\right)$.

<u>Означення.</u><u>Математичним сподіванням</u> функції φ дискретної випадкової величини (X_1, X_2,…, X_n) називається величина

$$M\phi\left(X_1, \ldots, X_n\right) = \sum_{i_1=1}^{s_1} \ldots \sum_{i_n=1}^{s_n} \phi\left(x_{i_1}^1, \ldots, x_{i_n}^n\right) f\left(x_{i_1}^1, \ldots, x_{i_n}^n\right), \quad (3)$$

якщо права частина (3) існує.

Для функції $\phi\left(X_1, \ldots, X_n\right)$ від неперервної випадкової величини $\left(X_1, \ldots, X_n\right)$ із щільністю розподілу $f\left(x_1, \ldots, x_n\right)$ маємо таке означення.

<u>Означення.</u> Математичним сподіванням функції ϕ неперервної випадкової величини $\left(X_1, \ldots, X_n\right)$ називається величина

$$M\phi\left(X_1, \ldots, X_n\right) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \phi\left(x_1, \ldots, x_n\right) f\left(x_1, \ldots, x_n\right) dx_1 \ldots dx_n, \quad (4)$$

якщо інтеграл у правій частині (4) існує.

Як ми вже домовилися, питання збіжності рядів та існування невластивих інтегралів обговорювати не будемо.

Далі розглянемо найпростіші функції випадкових величин: сумування та множення.

<u>Теорема.</u> Математичне сподівання суми випадкових величин дорівнює сумі їхніх математичних сподівань (за умови їх існування).

Обмежимось випадком двох доданків X та Y (узагальнення на випадок n доданків очевидно).

1. Нехай (X,Y) - дискретна випадкова величина. За означенням

$$M(X+Y)=\sum_{i=1}^{s}\sum_{j=1}^{m}(x_i+y_j)p(x_i,y_j)=\sum_{i=1}^{s}\sum_{j=1}^{m}x_i p(x_i,y_j)+\sum_{i=1}^{s}\sum_{j=1}^{m}y_j p(x_i,y_j)=$$

$$\text{¿}\sum_{i=1}^{s}x_i\sum_{j=1}^{m}p(x_i,y_j)+\sum_{j=1}^{m}y_j\sum_{i=1}^{s}p(x_i,y_j)=\sum_{i=1}^{s}x_i p_X(x_i)+\sum_{j=1}^{m}y_j p_Y(y_j)=MX+MY.$$

2. Нехай (X,Y) - неперервна випадкова величина зі щільністю розподілу $f(x,y)$. Тоді

$$M(X+Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)dxdy=\int_{-\infty}^{\infty}x\int_{-\infty}^{\infty}f(x,y)dydx+\int_{-\infty}^{\infty}y\int_{-\infty}^{\infty}f(x,y)dxdy=$$

$$\text{¿}\int_{-\infty}^{\infty}xf_X(x)dx+\int_{-\infty}^{\infty}yf_Y(y)dy=MX+MY.$$

Зауважимо, що ця теорема має місце для залежних і незалежних випадкових величин.

Розглянемо функцію множення.

Теорема. Математичне сподівання добутку незалежних випадкових величин дорівнює добутку їхніх математичних сподівань (за умови їх існування).

Знову обмежимось випадком двох співмножників.

1. Для дискретної випадкової величини (X,Y) маємо:

$$M(XY)=\sum_{i=1}^{s}\sum_{j=1}^{m}x_i y_j p(x_i,y_j)=\sum_{i=1}^{s}\sum_{j=1}^{m}x_i y_j p_X(x_i)p_Y(y_j)=\sum_{i=1}^{s}x_i p_X(x_i)\sum_{j=1}^{m}y_j p_Y(y_j)=MXMY.$$

2. У неперервному випадку

$$M(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)dxdy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_X(x)f_Y(y)dxdy=\int_{-\infty}^{\infty}xf_X(x)dx\int_{-\infty}^{\infty}yf_Y(y)dy=MXMY.$$

Випадок n співмножників пропонуємо дослідити самостійно.

Зробимо один важливий висновок: якщо математичне сподівання добутку випадкових величин не дорівнює добутку математичних сподівань співмножників, то випадкові величини - залежні.

Приклад. Розглянемо випадкові величини X та $Z=X+Y$, де X і Y – незалежні випадкові величини. Маємо

$$M(XZ)=M[X(X+Y)]=MX^2+MXMY.$$

І, таким чином, якщо $MX^2>0$, то X та Z – залежні.

Зауваження: якщо математичне сподівання добутку випадкових величин дорівнює добутку їхніх математичних сподівань, то це ще не означає їх незалежність.

2.2.6 Характеристики стохастичного зв'язку між випадковими величинами

Для характеризації зв'язку між випадковими величинами використовуються такі числові характеристики як коваріація та коефіцієнт корреляції.

Означення. Коваріацією випадкових величин X та Y називається величина

$$\operatorname{cov}(X,Y)=M[(X-MX)(Y-MY)].$$

Розглянемо деякі властивості коваріації.

1. Тривіальне виходить з означення, що

2. Аналогічно,

$$\operatorname{cov}(X,X)=DX.$$

3. Має місце формула

$$\operatorname{cov}(X,Y)=M(XY)-MXMY.$$

Дійсно (розкриваємо дужки та користуємось властивостями математичного сподівання):

$$\operatorname{cov}(X,Y)=M\left[(X-MX)(Y-MY)\right]=M(XY)-MXMY-MXMY+MXMY=M(XY)-MXMY .$$

4. Коваріація незалежних випадкових величин X та Y дорівнює нулеві. Дійсно, з незалежності X та Y виходить незалежність $(X-MX)$ та $(Y-MY)$ (перевірте самостійно). А тому

$$\operatorname{cov}(X,Y)=M\left[(X-MX)(Y-MY)\right]=M(X-MX)M(Y-MY)=0.$$

Зворотне твердження в загальному випадку невірне: з рівності нулеві коваріації ще не виходить незалежність випадкових величин.

<u>Приклад</u>. Випадкова величина *X* має нормальний розподіл з нульовим математичним сподіванням, тобто її щільність є $n(x,0,\sigma)$.

Нехай $Y=X^2$. Випадкова величина *X* та *Y* залежні.

Але,

$$\operatorname{cov}(X,Y)=M\left[X\left(X^2-MX^2\right)\right]=M\left(X^3-XMX^2\right)=MX^3-MXMX^2=0 ,$$

якщо згадати, що неперервні початкові моменти випадкової величини *X* дорівнюють нулеві.

Розглянемо нормовані випадкові величини *X* та *Y*, тобто величини $X^{¿}=\dfrac{X-MX}{\sigma_X}$, $Y^{¿}=\dfrac{Y-MY}{\sigma_Y}$. Їх математичне сподівання дорівнює нулеві, а дисперсія одиниці. Дійсно, наприклад,

$$M\left(\frac{X-MX}{\sigma_X}\right)=\frac{1}{\sigma_X}M(X-MX)=0 ,$$

$$D\left(\frac{X-MX}{\sigma_X}\right)=\frac{1}{\sigma_X^2}D(X-MX)=\frac{DX}{\sigma_X^2}=1.$$

Означення. Коефіцієнтом корреляції випадкових величин X та Y зветься коваріація величин $X^{¿}$ та $Y^{¿}$, тобто

$$\rho_{XY}=\operatorname{cov}\left(X^{¿},Y^{¿}\right)=M\left(X^{¿}Y^{¿}\right)=M\left[\frac{(X-MX)(Y-MY)}{\sigma_X\sigma_Y}\right]=\frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}.$$

Інакше, коефіцієнт корреляції є частка від ділення коваріації на добуток середніх квадратичних відхилень випадкових величин.

Мають місце такі властивості коефіцієнта корреляції (перші три доводяться тривіально).

1. $\rho_{XY}=\rho_{YX}$.

2. $\rho_{XX}=1$.

3. Для незалежних X та Y $\rho_{XY}=0$.

До того як перейти до інших властивостей доведемо дуже важливу формулу для дисперсії суми (різниці) випадкових величин X та Y:

$$D(X\pm Y)=DX+DY\pm 2\operatorname{cov}(X,Y).$$

Дійсно,

$$D(X\pm Y)=M\left[(X\pm Y)-M(X\pm Y)\right]^2=M\left[(X-MX)\pm(Y-MY)\right]^2=M(X-MX)^2+M(Y-MY)^2\pm$$
$$\pm 2M\left[(X-MX)(Y-MY)\right]=DX+DY\pm 2\operatorname{cov}(X,Y).$$

З цього виходить: для незалежних X та Y, або коли $\operatorname{cov}(X,Y)=0$, дисперсія суми (різниці) цих випадкових величин дорівнює сумі дисперсій.

Пропонуємо самостійно переконатися, що для суми n випадкових величин має місце формула

$$D\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}DX_i+\sum_{i\neq j}\operatorname{cov}\left(X_i,X_j\right)$$.

Для цього розкладіть вираз

$$\left(\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}MX_i\right)^2=\left[\sum_{i=1}^{n}\left(X_i-MX\right)\right]^2$$

на суму відповідних співмножників та візьміть від неї математичне сподівання.

Бачимо, що для незалежних X_i, $i=\overline{1,n}$, (або коли $\operatorname{cov}(X_i, X_j)=0,\ i \neq j$) виходить

$$D\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} DX_i .$$

А тепер повернемось до вивчення властивостей коефіцієнта корреляції.

4. $|\rho_{XY}| \leq 1.$

Для доведення цієї властивості зазначимо, що дисперсія не може бути від'ємною та

$$D\left(X^{¿} \pm Y^{¿}\right)=2 \pm 2\rho_{XY} \geq 0 . \quad (1)$$

Звідси маємо

$$-1 \leq \rho_{XY} \leq 1.$$

5. Якщо $Y=\alpha X+\beta$, $\alpha \neq 0$ (тобто X та Y пов'язані лінійно), то $|\rho_{XY}|=1.$

Дійсно, у цьому випадку $DY=\alpha^2 DX$, $\sigma_Y=|\alpha|\sigma_X$ і

$$\rho_{XY}=\frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y}=\frac{M\left[(X-MX)(Y-MY)\right]}{\sigma_X \sigma_Y}=\frac{M\left[(X-MX)(\alpha X+\beta-\alpha\, MX-\beta)\right]}{\sigma_X \sigma_Y}=$$

$$¿\frac{\alpha M(X-MX)^2}{\sigma_X |\alpha| \sigma_X}=\frac{\alpha}{|\alpha|}=\begin{cases}1,\ \alpha>0; \\ -1,\ \alpha<0.\end{cases}$$

6. Навпаки, коли $|\rho_{XY}|=1$, то існують $\alpha \neq 0$ і β такі, що з імовірністю одиниця

$$Y=\alpha X+\beta .$$

Нехай $\rho_{XY}=1.$ Тоді з (1) виходить

$$D\left(X^{¿}-Y^{¿}\right)=M\left(X^{¿}-Y^{¿}\right)=0 .$$

Випадкова величина $Z=\left(X^{¿}-Y^{¿}\right)^2$ - невід'ємна, тобто $P\left(Z\geq 0\right)=1$, а її математичне сподівання $MZ=0$. Це може бути тільки у тому випадку, коли $P\left(Z=0\right)=1$. Дійсно, наприклад, *Z* – дискретна невід'ємна випадкова величина з можливими різними значеннями z_i, $i=\overline{1,n}$. Тоді рівність

$$MZ=\sum_{i=1}^{n} z_i p_i=0$$

означає, що одне $z_i=0$, а $p_i=P\left(Z=z_i\right)=1$. Для інших $z_j\ \left(j\neq i\right)$ повинно бути $p_j=0$.

Далі, з $P\left(Z=0\right)=1$ виходить, що і $P\left(\left(X^{¿}-Y^{¿}\right)=0\right)=1$, тобто

$$P\left(\frac{X-MX}{\sigma_X}=\frac{Y-MY}{\sigma_Y}\right)=1.$$

Переходячи до позначень $v_X=MX$, $v_Y=MY$, маємо

$$P\left(Y=\frac{\sigma_Y}{\sigma_X}X+v_Y-\frac{\sigma_Y}{\sigma_X}v_X\right)=1.$$

Твердження доведено з $\alpha=\dfrac{\sigma_Y}{\sigma_X}$, $\beta=v_Y-\dfrac{\sigma_Y}{\sigma_X}v_X$.

Випадок $\rho_{XY}=-1$ пропонуємо розглянути самостійно. Для цього у якості *Z* візьміть $Z=\left(X^{¿}+Y^{¿}\right)^2$, скористайтеся (1) з плюсами, та доведіть

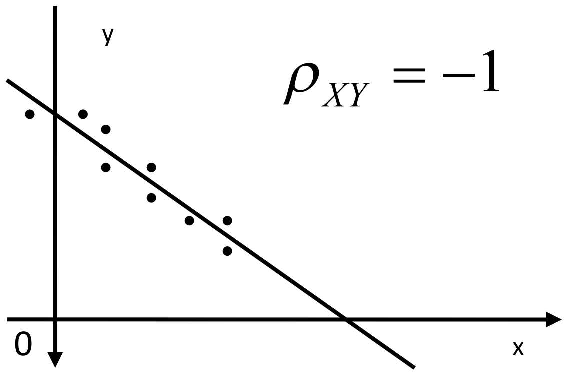
$$P\left(Y=-\frac{\sigma_Y}{\sigma_X}X+v_Y+\frac{\sigma_Y}{\sigma_X}v_X\right)=1.$$

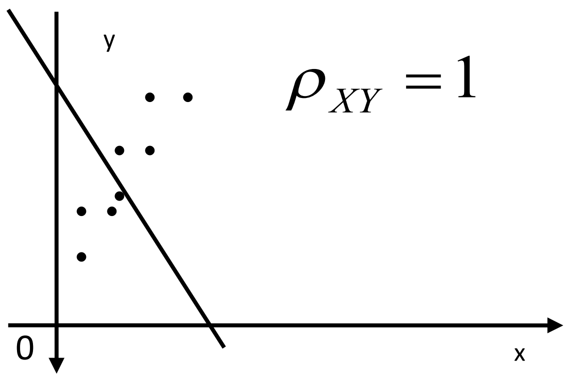
Якщо проаналізувати властивості коваріації та коефіцієнта корреляції, то можна зробити наступні корисні висновки.

1. Коли $\operatorname{cov}\left(X,Y\right)\neq 0$, а отже і $\rho_{XY}\neq 0$, то випадкові величини *X* та *Y* обов'язково залежні.

2. Рівність нулеві коваріації (або коефіцієнта корреляції) не означає незалежність випадкових величин. Це є необхідна умова незалежності, але не є достатньою. Неформально кажучи, величини з нульовою коваріацією є «підозрілими» на незалежність.

3. Близькість $|\rho_{XY}|$ до одиниці свідчить про те, що між випадковими величинами *X* та *Y* з великою ймовірністю існує лінійна залежність. Коли $\rho_{XY} \approx 1$, то це лінійно зростаюча залежність, коли $\rho_{XY} \approx -1$ це лінійно спадна залежність. За великої кількості випробувань реалізації точки (X, Y) повинні групуватися навколо деяких прямих $y = \alpha x + \beta$.





Означення. Коли $\rho_{XY}=0$, то випадкові величини X та Y називаються <u>некоррельованими</u>; у випадку $\rho_{XY}>0$ вони називаються <u>додатно коррельованими</u>; у випадку $\rho_{XY}<0$ - <u>від'ємно коррельованими</u>.

Розглянемо багатовимірну випадкову величину (X_1, X_2, … , X_n). Коваріації $\text{cov}(X_i, X_j)$ утворюють матрицю $B=\| \text{cov}\left(X_i, X_j\right) \|$, яка називається <u>матрицею коваріацій</u>. З властивостей коваріації виходить, що ця матриця – симетрична, тобто $B^T=B$. На головній діагоналі цієї матриці знаходяться дисперсії компонент. Якщо компоненти некоррельовані, тобто $\text{cov}(X_i, X_j)=0$, $i \neq j$ (а тим паче незалежні), то матриця B – діагональна.

Коефіцієнти кореляції $\rho_{X_i X_j}=\rho_{ij}$ теж утворюють матрицю $R=\|\rho_{ij}\|$, яку будемо називати матрицею коефіцієнтів <u>корреляції</u> або <u>корреляційною матрицею</u>. Вона теж симетрична, на головній її діагоналі знаходяться одиниці, поза головною діагоналлю – елементи (додатні, або від'ємні, або нульові), які за модулем не перевищують одиниці. У випадку некоррельованих компонент вона співпадає з одиничною матрицею.

<u>Приклад.</u> Дано матрицю коваріацій випадкових величин X_1, X_2, X_3:

$$B=\left\| \begin{matrix} 16 & -14 & 12 \\ -14 & 49 & -21 \\ 12 & -21 & 36 \end{matrix} \right\| .$$

Треба побудувати корреляційну матрицю R.

З того, що на головній діагоналі матриці B знаходяться дисперсії відповідних випадкових величин, одержуємо:

$$\sigma_{X_1}=4; \ \sigma_{X_2}=7; \ \sigma_{X_3}=6.$$

Тепер, для того, щоб обчислити, наприклад, елемент ρ_{23} матриці R, треба поділити cov(X_2, X_3)=-21 на добуток $\sigma_{X_2}\sigma_{X_3}=42$. Одержуємо $\rho_{23}=-\frac{21}{42}=-\frac{1}{2}$. Аналогічно обчислюються інші елементи матриці R. Остаточно,

$$R=\left\|\begin{matrix} 1 & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 1 \end{matrix}\right\|.$$

<u>Приклад.</u> Визначити матрицю коваріацій випадкових величин *X* та *Y*, сумісний розподіл яких задано щільністю:

$$f(x,y)=¿\{\cos x \cos y,\ \text{якщо } 0\le x\le\frac{\pi}{2} \text{ та } 0\le y\le\frac{\pi}{2};¿¿¿¿$$

Бачимо, що сумісна щільність розподілу *f(x,y)* розпадається на добуток щільностей компонент:

$$f_X(x)=¿\{\cos x,\ x\in\left[0,\frac{\pi}{2}\right];¿¿¿¿$$

$$f_Y(y)=¿\{\cos y,\ y\in\left[0,\frac{\pi}{2}\right];¿¿¿¿$$

З цього робимо висновок про незалежність *X* та *Y*, отож cov(*X*,*Y*)= cov(*Y*,*X*)=0.

Треба ще визначити cov(*X*,*X*)=*DX* та cov(*Y*,*Y*)=*DY*.

Обчислюємо *MX* (інтегруємо за частинами):

$$MX=\int_{-\infty}^{\infty} xf_X(x)dx=\int_0^{\frac{\pi}{2}} x\cos x dx=\int_0^{\frac{\pi}{2}} xd(\sin x)=x\sin x|_0^{\frac{\pi}{2}}-\int_0^{\frac{\pi}{2}} \sin x dx=\frac{\pi}{2}+\cos x|_0^{\frac{\pi}{2}}=\frac{\pi}{2}-1$$
.

Обчислюємо MX^2 (теж за частинами двічі):

$$MX^2=\int_{-\infty}^{\infty} x^2 f_X(x)dx=\int_0^{\frac{\pi}{2}} x^2\cos x dx=\int_0^{\frac{\pi}{2}} x^2 d(\sin x)=x^2\sin x|_0^{\frac{\pi}{2}}-2\int_0^{\frac{\pi}{2}} x\sin x dx=\frac{\pi^2}{4}+\int_0^{\frac{\pi}{2}} xd(\cos x)=\frac{\pi^2}{4}-2$$

.

Застосовуємо формулу $DX=MX^2-(MX)^2$ та одержуємо $DX=\pi$-3. Зрозуміло, що і $DY=\pi$-3. Остаточно, матриця коваріації має вигляд:

$$B=\begin{Vmatrix} \pi-3 & 0 \\ 0 & \pi-3 \end{Vmatrix}$$
.

2.2.7 Розподіл скалярної функції декількох випадкових величин

Раніше ми розглядали деякі числові характеристики функцій випадкових величин. Але в багатьох задачах теорії ймовірностей важливо знати не тільки «бідні тіні» випадкових величин, але й вичерпні їхні характеристики: функцію розподілу, щільність розподілу тощо. Випадок функції одного випадкового аргумента вже зустрічався в 2.1.10. Тут ми будемо цікавитися розподілом функцій декількох випадкових величин.

Знову основну ідею розв’язання цієї задачі спочатку доцільно простежити на прикладі скалярної функції $Z=\phi(X,Y)$ двох випадкових аргументів. Випадкову величину (X,Y) будемо вважати неперервною зі щільністю розподілу $f(x,y)$.

За означенням функції розподілу випадкової величини Z, вона дорівнює:

$$F_Z(z)=P(Z<z)=\iint_{\phi(x,y)<z} f(x,y)dxdy$$
. (1)

Обчислення цього інтегралу часто вирішує проблему знаходження функції розподілу *Z*. Щільність розподілу *Z* обчислюється як похідна $F'_Z(z)$ (для тих *z*, де вона існує), або зведенням інтегралу в правій частині (1) до вигляду:

$$F_Z(z)=\int_{-\infty}^{z} f_Z(u)du \quad .$$

Дуже важливою функцією є сумування, тобто функція $Z=X+Y$. Треба обчислити функцію розподілу $F_Z(z)$ та функцію щільності $f_Z(z)$, якщо відома сумісна щільність $f(x,y)$.

Бачимо, що

$$F_Z(z)=P(Z<z)=P(X+Y<z)=\iint_{x+y<z} f(x,y)dxdy=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-x} f(x,y)dy\right]dx=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-y} f(x,y)dx\right]dy$$

. (2)

В інтегралі $\int_{-\infty}^{z-x} f(x,y)dy$ робимо заміну змінних $y=u-x$ та одержуємо

$$\int_{-\infty}^{z} f(x,u-x)du \quad .$$

Далі в інтегралі $\int_{-\infty}^{\infty}\int_{-\infty}^{z} f(x,u-x)dudx$ міняємо порядок інтегрування і тоді виходить

$$F_Z(z)=\int_{-\infty}^{z}\left[\int_{-\infty}^{\infty} f(x,u-x)dx\right]du.$$

За означенням, функція щільності дорівнює

$$f_Z(z)=\int_{-\infty}^{\infty} f(x,z-x)dx. \quad (3)$$

Зовсім аналогічно можна показати, що

$$f_Z(z)=\int_{-\infty}^{\infty} f(z-y,y)dy \qquad (4)$$

(для цього зробіть відповідну заміну змінних в інтегралі $\int_{-\infty}^{z-y} f(x,y)dx$ і поміняйте порядок інтегрування в правій частині (2)).

Нехай *X* та *Y* – незалежні випадкові величини, тобто сумісна функція щільності розпадається на добуток маргинальних функцій щільностей

$$f(x,y)=f_X(x)f_Y(y)$$

Тоді (3), (4) можна записати у вигляді

$$f_Z(x)=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx=\int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy. \qquad (5)$$

Формули (5) носять назву «згортки» щільностей $f_X(x)$ та $f_Y(y)$. Визначення закону розподілу суми незалежних випадкових величин по законам розподілу доданків іноді називають ще <u>композицією</u> законів розподілу.

<u>**Приклад.**</u> Одержимо щільність розподілу суми двох незалежних випадкових величин *X* та *Y*, якщо кожна з них має експоненційний розподіл:

$$f_X(x)=¿\{\lambda_1 e^{-\lambda_1 x}, x\geq 0;¿¿¿ \qquad f_Y(y)=¿\{\lambda_2 e^{-\lambda_2 y}, y\geq 0;¿¿¿$$

Для $z\geq 0$ маємо згідно з (5):

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx=\int_0^z \lambda_1 e^{-\lambda_1 x}\lambda_2 e^{-\lambda_2(z-x)}dx=$$

$$=\lambda_1\lambda_2 e^{-\lambda_2 z}\int_0^z e^{-(\lambda_1-\lambda_2)x}dx=\frac{\lambda_1\lambda_2}{\lambda_1-\lambda_2}\left(e^{-\lambda_2 z}-e^{-\lambda_1 z}\right).$$

Для $z<0$ $f_Z(z)=0$.

Зауважимо, що формули (3), (4) можна одержати з (2) формальним диференціюванням інтегралів зі змінною верхньою межею.

Проілюструємо цей прийом на задачі знаходження розподілу частки двох випадкових величин $Z=\frac{X}{Y}$, де (X,Y) має сумісну щільність розподілу $f(x,y)$.

За означенням функції розподілу у цьому випадку маємо

$$F_Z(z)=\iint\limits_{\frac{x}{y}<z} f(x,y)dxdy=\int\limits_0^{\infty}\left[\int\limits_{-\infty}^{zy} f(x,y)dx\right]dy+\int\limits_{-\infty}^{0}\left[\int\limits_{zy}^{\infty} f(x,y)dx\right]dy.$$

Диференціювання по z дає (один інтеграл зі змінною верхньою межею, другий – зі змінною нижньою)

$$f_Z(z)=\int\limits_0^{\infty} y\, f(zy,y)dy-\int\limits_{-\infty}^{0} y\, f(zy,y)dy.$$

Пропонуємо самостійно одержати вираз для $f_Z(z)$ для цієї функції, коли X та Y незалежні.

Ми зупинилися на відображенні $R^2 \to R^1$. Більш загальні випадки функціонального зв'язку випадкових величин розглянемо далі при розв'язанні конкретних задач.

2.2.8 Двовимірний нормальний розподіл

Означення. Випадкова величина (X,Y) має (невироджений) двовимірний нормальний розподіл, якщо її функція щільності дорівнює:

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-v_1)^2}{\sigma_1^2}-2\rho\frac{(x-v_1)(y-v_2)}{\sigma_1\sigma_2}+\frac{(y-v_2)^2}{\sigma_2^2}\right]\right\},$$

(1)

де $\sigma_1>0$, $\sigma_2>0$, $|\rho|<1$, $-\infty<x<\infty$, $-\infty<y<\infty$.

Для того, щоб переконатися у тому, що (1) дійсно є щільністю розподілу деякої двовимірної випадкової величини (X,Y), а також з'ясувати ймовірнісний зміст констант v_1, v_2, σ_1, σ_2, ρ, проінтегруємо *f(x,y)* по змінній *x* від $-\infty$ до $+\infty$. Маємо

$$\int_{-\infty}^{\infty} f(x,y)dx = A\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-v_1)^2}{\sigma_1^2}-2\rho\frac{(x-v_1)(y-v_2)}{\sigma_1\sigma_2}\right]\right\}dx \quad , (2)$$

де $A=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\frac{(y-v_2)^2}{\sigma_2^2}\right\}$.

Далі в (2) вираз в квадратних дужках доповнюємо до повного квадрата

$$\frac{(x-v_1)^2}{\sigma_1^2}-2\rho\frac{(x-v_1)(y-v_2)}{\sigma_1\sigma_2}+\frac{\rho^2(y-v_2)^2}{\sigma_2^2}-\frac{\rho^2(y-v_2)^2}{\sigma_2^2}=\left[\frac{x-v_1}{\sigma_1}-\frac{\rho(y-v_2)}{\sigma_2}\right]^2-\frac{\rho^2(y-v_2)^2}{\sigma_2^2}.$$

Виносимо з інтеграла $B=\exp\left\{\frac{1}{2(1-\rho^2)}\frac{\rho^2(y-v_2)^2}{\sigma_2^2}\right\}$, одержуємо

$$\int_{-\infty}^{\infty} f(x,y)dx = AB\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x-v_1}{\sigma_1}-\frac{\rho(y-v_2)}{\sigma_2}\right]^2\right\}dx$$

Робимо заміну змінних

$$u=\frac{1}{\sqrt{1-\rho^2}}\left[\frac{x-v_1}{\sigma_1}-\frac{\rho(y-v_2)}{\sigma_2}\right]$$

та одержуємо (згадуємо інтеграл Лапласа $\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}du=\sqrt{2\pi}$)

$$\int_{-\infty}^{\infty} f(x,y)dx=\sigma_1\sqrt{1-\rho^2}AB\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}du=\frac{1}{\sigma_2\sqrt{2\pi}}e^{-\frac{(y-v_2)^2}{2\sigma_2^2}}.$$

Таким чином, $\int_{-\infty}^{\infty} f(x,y)dx$ є щільність нормального розподілу з математичним сподіванням v_2 та дисперсією σ_2^2.

Аналогічно,

$$\int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-v_1)^2}{2\sigma_1^2}}.$$

Далі тривіально виходить, що

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1.$$

Останнє свідчить про те, що (1) дійсно задає щільність розподілу деякої випадкової величини (*X,Y*) при прийнятих припущеннях відносно σ_1, σ_2, ρ. Окрім того можна зробити висновки про те, що компоненти *X* та *Y* мають нормальний розподіл з параметрами $v_X = v_1$, $v_Y = v_2$, $\sigma_X = \sigma_1$, $\sigma_Y = \sigma_2$. Неважко переконатися у тому (не повторюємо тривіальні, але дещо громіздкі викладі), що ρ в (1) можна інтерпретувати як коефіцієнт корреляції випадкових величин *X* та *Y*, тобто $\rho = \rho_{XY}$. Тому перепишемо (1) у вигляді:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}\exp\left\{-\frac{1}{2\left(1-\rho_{XY}^2\right)}\left[\frac{(x-v_X)^2}{\sigma_X^2} - 2\rho_{XY}\frac{(x-v_X)(y-v_Y)}{\sigma_X\sigma_Y} + \frac{(y-v_Y)^2}{\sigma_Y^2}\right]\right\}. \quad (3)$$

Бачимо, що коли *X* та *Y* некоррельовані ($\rho_{XY} = 0$) та випадкова величина (*X,Y*) має двовимірний нормальний розподіл, то *X* та *Y* незалежні ($f(x,y) = f_X(x) f_Y(y)$).

Зауваження. Зворотнє твердження не вірне: з нормальності розподілу (маргинальної) компонент *X* та *Y* та з $\rho_{XY} = 0$ не виходить незалежність *X* та *Y*.

В багатьох задачах важливо знати умовну щільність розподілу величини *Y* за умови, що випадкова величина *X* одержує значення *x*, тобто

$$f_{Y/x}(y/x) = \frac{f(x,y)}{f_X(x)}.$$

Пропонуємо читачам самостійно переконатися, що у випадку двовимірної нормальної величини

$$f_{Y/x}(y/x)=\frac{1}{\sigma_Y\sqrt{1-\rho_{XY}^2}\sqrt{2\pi}}e^{\frac{-\left[y-v_Y-\rho_{XY}\frac{\sigma_Y}{\sigma_X}(x-v_X)\right]}{2(1-\rho_{XY}^2)\sigma_Y^2}}.$$

Останнє означає, що цей умовний розподіл є нормальним, умовне математичне сподівання дорівнює $M(Y/X=x)=v_Y+\rho_{XY}\frac{\sigma_Y}{\sigma_X}(x-v_X)$, а умовна дисперсія $\sigma_{Y/x}^2=\sigma_Y^2(1-\rho_{XY}^2)$.

Знову повернемось до (3), та надамо цій рівності матричну форму. Введемо вектор-стовпець змінних $a=¿\left(x¿\right)¿¿¿ ¿$, вектор-стовпець математичних сподівань $v=¿\left(v_X¿\right)¿¿¿ ¿$. Матрицю коваріацій в даному випадку можна подати у вигляді:

$$B=\begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y\\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}.$$

Її визначник $|B|=\sigma_X^2\sigma_Y^2(1-\rho^2)>0$. Обернена до неї дорівнює

$$B^{-1}=\frac{1}{1-\rho^2}\begin{pmatrix}\frac{1}{\sigma_X^2} & -\frac{\rho}{\sigma_X\sigma_Y}\\ -\frac{\rho}{\sigma_X\sigma_Y} & \frac{1}{\sigma_Y^2}\end{pmatrix}$$

Неважко перевірити, що в матричній формі (3) має вигляд

$$f(x,y)=\frac{1}{2\pi|B|^{1/2}}e^{-\frac{1}{2}(a-v)^TB^{-1}(a-v)} \qquad (4)$$

Сказане дозволяє дати еквівалентне означення.

Означення. Двовимірна випадкова величина має невироджений нормальний розподіл, якщо її щільність можна подати у вигляді (4) з додатно означеною матрицею коваріацій B .

2.2.9 Багатовимірний нормальний розподіл

Розглянемо неперервну n -вимірну випадкову величину

$$X = \begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix}$$

(надалі нам буде зручно вважати її вектором-стовпцем). Вектор змінних, вектор математичних сподівань і матрицю коваріацій позначимо відповідно:

$$x = \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix}, \quad v_X = \begin{pmatrix} MX_1 \\ \dots \\ MX_n \end{pmatrix}, \quad B = \|\operatorname{cov}(X_i, X_j)\|$$

Означення. Випадкова величина X має невироджений n-вимірний нормальний розподіл, коли її щільність розподілу задається у вигляді

$$f(x_1, \dots, x_n) = n(x, v_X, B) = \frac{1}{(2\pi)^{n/2} |B|^{1/2}} e^{-\frac{1}{2}(x - v_X)^T B^{-1} (x - v_X)} \qquad (1)$$

з додатно означеною матрицею коваріацій B . Надалі слово „невироджений" в цьому контексті будемо опускати.

Нагадаємо, що симетрична матриця B називається додатно означеною, коли для довільного вектора $y = \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix} \neq \vec{0}$ виконується умова $y^T B y > 0$. З курсу алгебри відомо, що додатно означена матриця має додатні всі головні мінори, її визначник $|B|>0$, і, таким чином, існує обернена до неї матриця B^{-1} (яка теж є додатно означеною).

Виконуються такі властивості функції $n(x,v_X,B)$, які характеризують її як щільність розподілу.

1. Без доведення зазначимо, що має місце

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} n(x,v_X,B)\,dx_1\ldots dx_n=1$$

2. Невід'ємність $n(x,v_X,B)$ очевидна.

3. Якщо проінтегрувати $n(x,v_X,B)$ по нескінченним границях по всіх змінних, за виключенням $x_{i_1},x_{i_1},\ldots,x_{i_{1k}}$, то одержимо k-вимірний нормальний розподіл

$$f'(x_{i_1},x_{i_2},\ldots,x_{i_k})=n(x',v'_X,B')$$, де

$$x'=\begin{pmatrix}x_{i_1}\\ \ldots\\ x_{i_k}\end{pmatrix},\quad v'_X=\begin{pmatrix}MX_{i_1}\\ \ldots\\ MX_{i_k}\end{pmatrix},\quad B'=\|\operatorname{cov}(X_{i_l},X_{i_m})\|,\quad l,m=\overline{1,k}$$

Якщо X має n-вимірний нормальний розподіл та $\operatorname{cov}(X_i,X_j)=0$, для $i\neq j$, то випадкові величини $X_1,\ldots,X_n$ є незалежні.

Дійсно, в цьому випадку матриця коваріацій діагональна, на головній її діагоналі знаходяться дисперсії компонент вектора X, тобто

$$B=\begin{pmatrix}\sigma_{X_1^2} & 0 & \cdots & 0\\ 0 & \sigma_{X_2^2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_{X_n^2}\end{pmatrix}.$$

Визначник її дорівнює $|B|=\sigma_{X_1^2}\sigma_{X_2^2}\ldots\sigma_{X_n^2}$.

Обернена матриця B^{-1} має вигляд

$$B^{-1}=\begin{pmatrix} \frac{1}{\sigma_{X_1^2}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_{X_2^2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_{X_n^2}} \end{pmatrix}.$$

Якщо підставити все це в (1), то одержимо

$$f(x_1,\ldots,x_n)=\prod_{i=1}^{n}\frac{1}{\sigma_{X_i^2}\sqrt{2\pi}}e^{-\frac{(x_i-v_{X_i})^2}{2\sigma_{X_i}^2}},$$

що і говорить про незалежність.

Як і в двовимірному випадку зворотне твердження невірне: з нормальності маргинальних розподілів компонент та рівності нулеві коваріацій ще не виходить незалежність компонент.

2.2.10 Лінійні перетворення нормально розподілених випадкових величин

В багатьох задачах важливо знати розподіл вектора $Y=\begin{pmatrix} Y_1 \\ \ldots \\ Y_n \end{pmatrix}$, якщо відомо, що вектор $X=\begin{pmatrix} X_1 \\ \ldots \\ X_n \end{pmatrix}$ має n -вимірний нормальний розподіл та

$$Y=AX, \qquad (1)$$

де A – невироджена квадратна матриця вимірності $n\times n$.

Логіку міркувань при вирішенні цієї задачі продемонструємо на прикладі відображення $R^2 \to R^2$, яке може бути і нелінійним, а розподіл вектора X – не обов’язково нормальним.

Розглянемо відображення, що задається співвідношеннями

$$y_1 = y_1(x_1, x_2) \;;$$

$$y_2 = y_2(x_1, x_2) \;,$$

де функції $y_1(x_1,x_2)$, $y_2(x_1,x_2)$ такі, що існує однозначне обернене відображення

$$x_1 = x_1(y_1, y_2) \;;$$
$$x_2 = x_1(y_1, y_2) \;, \qquad (2)$$

а функції $x_1(y_1,y_2)$, $x_2(y_1,y_2)$ неперервно диференційовані по y_1, y_2.

Нас цікавить щільність розподілу $f_Y(y_1,y_2)$ випадкової величини $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, яка задана відображенням

$$Y_1 = y_1(X_1, X_2) \;,$$

$$Y_2 = y_2(X_1, X_2)$$

випадкового вектора $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, якщо відомо, що X має абсолютно неперервний розподіл зі щільністю $f_X(x_1,x_2)$.

Візьмемо довільну допустиму область S з площини $y_1 y_2$. Вона взаємно однозначно відображається в деяку область $A(S)$ площини $x_1 x_2$. Тоді

$$P(Y \in S) = \iint_S f_Y(y_1, y_2)\, dy_1 dy_2 = P(X \in A(S)) = \iint_{A(S)} f_X(x_1, x_2)\, dx_1 dx_2 \,.$$

В останньому інтегралі робимо заміну змінних (2) і одержуємо

$$\iint_S f_Y(y_1, y_2)\, dy_1 dy_2 = \iint_S f_X[x_1(y_1, y_2), x_2(y_1, y_2)]\, |J|\, dy_1 dy_2 \,,$$

$$J=\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

де – визначник Остроградського[*]-Якобі[**].

З того, що ми брали довільну область S виходить рівність підінтегральних функцій, тобто

$$f_Y(y_1,y_2)=f_X\left[x_1(y_1,y_2),x_2(y_1,y_2)\right]|J| , \qquad (3)$$

Що і дає розв'язок поставленої задачі.

__Приклад.__ Нехай вектор $X=\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ задає прямокутні координати випадкової точки на площині x_1x_2 і відомо, що він має двовимірний нормальний розподіл з параметрами $v_{X_1}=v_1$, $v_{X_2}=v_2$, $\sigma_{X_1}=\sigma_1$, $\sigma_{X_2}=\sigma_2$, $\rho_{X_1X_2}=\rho$.Треба обчислити щільність розподілу координат $Y=\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ цієї точки, якщо

$$\frac{X_1-v_1}{\sigma_1}=Y_1\cos Y_2, \quad \frac{X_2-v_2}{\sigma_2}=Y_2\sin Y_1,$$

де розподіл Y_1 зосереджено на інтервалі $[0;\infty)$ (тобто $P(Y_1\notin[0;\infty))=0$), а розподіл Y_2 — на інтервалі $[0;2\pi)$ (тобто $P(Y_1\notin[0;2\pi))=0$). Бачимо, що Y_1 , Y_2 — полярні координати точки з центрованими та нормованими прямокутними координатами X_1 , X_2 .

В даному випадку (2) має вигляд

$$x_1=\sigma_1 y_1\cos y_2+v_1 ,$$

$$x_2=\sigma_2 y_2\sin y_1+v_2 .$$

---

[*] Остроградський Михайло Васильович (1801–1862) — російський математик, один із засновників петербурзької математичної школи, член багатьох академій наук

[**] Якобі Карл Густав Якоб (1804–1851) — німецький математик, член багатьох академій наук

Якобіан J дорівнює

$$J=\begin{vmatrix}\sigma_1 \cos y_2 & -\sigma_1 y_1 \sin y_2 \\ \sigma_2 \sin y_2 & \sigma_2 y_1 \cos y_2\end{vmatrix}=\sigma_1 \sigma_2 y_1=|J|.$$

Ще раз нагадаємо щільність нормального розподілу

$$f_X(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-v_1)^2}{\sigma_1^2}+\frac{(x_2-v_2)^2}{\sigma_2^2}-2\rho\frac{(x_1-v_1)(x_2-v_2)}{\sigma_1\sigma_2}\right]\right).$$

Далі користуємось формулою (3) та одержуємо

$$f_X(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left[y_1^2\cos^2 y_2+y_1^2\sin^2 y_2-2\rho y_1^2\cos y_2 \sin y_2\right]\right)\sigma_1\sigma_2 y_1=$$

$$=\frac{y_1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{y_1^2}{2(1-\rho^2)}[1-\rho\sin 2y_2]\right),$$

за умови $0\le y_1<\infty$, $0\le y_2<2\pi$. Якщо ці умови не виконуються, то

$$f_Y(y_1,y_2)=0.$$

Сформулюємо узагальнення (3) на випадок відображення $R^n \to R^n$ (доведення аналогічне).

<u>Теорема 1.</u> Нехай задане відображення

$$y_1=y_1(x_1,\dots,x_n),$$
$$\dots\dots\dots\dots$$
$$y_n=y_n(x_1,\dots,x_n),$$

таке, що обернене відображення

$$x_1=x_1(y_1,\dots,y_n),$$
$$\dots\dots\dots\dots$$
$$x_n=x_n(y_1,\dots,y_n),$$

однозначне, а функції $x_i=x_i(y_1,\dots,y_n)$, $i=\overline{1,n}$, неперервно диференційовані по всіх аргументах.

Нехай випадкова величина $X=\begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix}$ — неперервна зі щільністю розподілу $f_X(x_1,\dots,x_n)$.

Тоді випадкова величина $Y=\begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}$, де $Y_i=y_i(X_1,\dots,X_n)$, $i=\overline{1,n}$ має щільність розподілу

$$f_Y(y_1,\dots,y_n)=f_X(x_1(y_1,\dots,y_n),\dots,x_n(y_1,\dots,y_n))|J|, \qquad (4)$$

з якобіаном

$$J=\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}.$$

Тепер повернемося до розв'язання задачі, що поставлена на початку параграфа.

<u>Теорема 2.</u> Нехай задано n-вимірну випадкову величину $X=\begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix}$, яка має нормальний розподіл $n(x,v_X,B)$, та довільну невироджену матрицю A вимірності $n\times n$. Тоді вектор $Y=AX$ має нормальний розподіл $n(y,Av_X,ABA^T)$, і, таким чином, вектор математичного сподівання $v_Y=Av_X$, а матриця коваріацій $B_Y=ABA^T$.

Доведення. Далі часто будемо використовувати відомі з курсу алгебри рівності:

$(A^{-1})^T=(A^T)^{-1}$, $(A\cdot B)^{-1}=B^{-1}\cdot A^{-1}$, $(A\cdot B)^T=A^T\cdot B^T$, $|AB|=|A||B|$,

$|A|=|A^T|$, $|A^{-1}|=\frac{1}{|A|}$.

В даному випадку ми маємо справу з відображенням $y=Ax$. Оскільки матриця A невироджена ($|A|\neq 0$), існує однозначне обернене відображення $x=A^{-1}y$.

$$x_1=a_{11}^{-1}y_1+a_{12}^{-1}y_2+\ldots+a_{1n}^{-1}y_n$$

$$x_n=a_{n1}^{-1}y_1+a_{n2}^{-1}y_2+\ldots+a_{nn}^{-1}y_n ,$$

де a_{ij}^{-1} - елементи оберненої матриці A^{-1}.

Бачимо що якобіан $I=|A^{-1}|=\frac{1}{|A|}$.

Скористаємось теоремою 1, формулою (4), та одержимо

$$f_Y(y_1,\ldots,y_n)=n\left(A^{-1}y,v_X,B\right)\operatorname{mod}I=\frac{1}{(2\pi)^{n/2}\operatorname{mod}|A|\sqrt{|B|}}e^{-\frac{1}{2}\left(A^{-1}y-v_X\right)^T B^{-1}\left(A^{-1}y-v_X\right)} .$$

Перетворимо константу перед експонентою. Для цього зобразимо $\frac{1}{\operatorname{mod}|A|}$ у вигляді:

$$\frac{1}{\operatorname{mod}|A|}=\sqrt{\frac{1}{|A|^2}}=\sqrt{\frac{|B|}{|A||B||A^T|}}=\frac{\sqrt{|B|}}{\sqrt{|ABA^T|}} .$$

І, таким чином згадувана константа дорівнює $\frac{1}{(2\pi)^{n/2}\sqrt{|ABA^T|}}$.

Перетворимо ступінь експоненти:

$$\left(A^{-1}y-v_X\right)^T B^{-1}\left(A^{-1}y-v_X\right)=\left(A^{-1}y-A^{-1}Av_X\right)^T B^{-1}\left(A^{-1}y-A^{-1}Av_X\right)=$$
$$¿\left[A^{-1}\left(y-Av_X\right)\right]^T B^{-1}A^{-1}\left(y-Av_X\right)=\left(y-Av_X\right)^T\left(A^T\right)^{-1}B^{-1}A^{-1}\left(y-Av_X\right)=$$
$$¿\left(y-Av_X\right)^T\left(ABA^T\right)^{-1}\left(y-Av_X\right)$$

Остаточно ,

$$f_Y\left(y_1,\ldots,y_n\right)=\frac{1}{(2\pi)^{n/2}\sqrt{|ABA^T|}}e^{-\frac{1}{2}\left(y-Av_X\right)^T\left(ABA^T\right)^{-1}\left(y-Av_X\right)}$$

Залишилося переконатися, що Av_X є математичне сподівання вектора *Y,*

а ABA^T - його матриця коваріацій (при цьому будемо вважати математичним сподіванням матриці з випадковими елементами матрицю їхніх математичних сподівань).

Переконаємося самостійно, що

MY=M(AX)=AMX=Av$_{X,}$

Тобто детермінована матриця з під знаку математичного сподівання виноситься .

Для цього AX запишіть у розгорнутій формі та скористайтеся властивістю математичного сподівання суми випадкових величин.

Доведіть також, що матриця коваріацій вектора X може бути подана у вигляді

$$B=\|\mathrm{cov}\left(X_i,X_j\right)\|=\|M\left[\left(X_i-MX_i\right)\left(X_j-MX_j\right)\right]\|=M\left(X-v_X\right)\left(X-v_X\right)^T A^T=$$
$$¿A\left[M\left(X-v_X\right)\left(X-v_X\right)^T\right]A^T=ABA^T.$$

Теорему повністю доведено .

Розглянемо тепер випадок лінійного перетворення *Y=AX*, де *X* такий самий вектор, що і у попередній теоремі, а матриця *A*-прямокутна вимірності

$m\times n, m<n.$ Тоді має компонент . Припустимо що ранг матриці дорівнює .

Теорема 3 Якщо вектор X має нормальний розподіл зі щільністю $n(x, v_X, B)$, то $Y=AX$ теж має нормальний розподіл зі щільністю $n(y, Av_X, ABA^T)$.

Доведення. Як і в попередньому випадку легко перевірити, що $MY = Av_X$, а матриця коваріацій $M(Y - Av_X)(Y - Av_X)^T = ABA^T$.

Розглянемо спочатку частковий випадок, коли матриця A має блочну структуру $A=(A_1A_2)$, де матриця A_1 вимірності $m \times m$ не вироджена.

Виведемо матрицю

$$C = \begin{pmatrix} A_1 & A_2 \\ 0 & E_{n-m} \end{pmatrix},$$

де E_{n-m}– одинична матриця вимірності $(n-m)\times(n-m)$. Матриця C не вироджена, тому що

$$|C| = |A_1||E_{n-m}| = |A_1| \neq 0.$$

За теоремою 2 вектор

$$Z = CX = \begin{pmatrix} A_1 & A_2 \\ 0 & E_{n-m} \end{pmatrix} \begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix} = \begin{pmatrix} Y_1 \\ \dots \\ Y_m \\ Y_{m+1} \\ \dots \\ Y_n \end{pmatrix}$$

має n – вимірний нормальний розподіл, при чому перші його компонент складають вектор Y. Але ми знаємо, що коли вектор має n – вимірний нормальний розподіл, то і довільна сукупність його компонент має нормальний розподіл. Тобто в цій ситуації твердження має місце .

Припустимо, що A_1 – вироджена. Але в матриці A повинен існувати мінор вимірності $(m \times m)$ відмінний від нуля (тому що ранг матриці дорівнює m). Переставленням змінюємо порядок стовпців в матриці A

(одночасно переставляючи однойменні компоненти вектора *X*) та збираємо в перших *m* стовпцях не вироджену матрицю. Зауважимо, що при цьому вектор *Y* не змінюється, а переставлення компонент вектора *X* не змінює його розподіл. В результаті отримаємо матрицю A', блочна структура якої $A' = \left(A_1' A_2' \right)$, де A_1' - не вироджена, та вектор X' з розподілом $n\left(x, v_X, B \right)$. Лінійне перетворення тепер має вигляд

$$Y = A' X' .$$

Тобто ми звели задачу до попереднього варіанту. Теорему доведено.

3. Граничні теореми теорії ймовірностей

3.1 Збіжність послідовності випадкових величин.

Розглянемо нескінчену послідовність випадкових величин $\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega), \ldots$, та випадкову величину $\xi(\omega)$, що задана на одному і тому ж самому імовірнісному просторі $(\Omega . F, p(.))$. Нас буде цікавити питання збіжності послідовності $\{\xi_n\}_{n=1}^{\infty}$ до випадкової величини ξ . Зверніть увагу на те, що мова не йде про збіжність послідовності чисел до деякого числа, а йдеться про збіжність послідовності функцій ($\xi_n(\omega)$ та $\xi(\omega)$ є функції елементарних подій, вимірні відносно σ – алгебри *F).* З курсу аналізу ви вже знаєте, що бувають різні види збіжності функцій, ми розглядаємо такі, що характерні для задач теорії ймовірностей.

Означення 1. Послідовність випадкових величин $\{\xi_n\}$ - збігається за ймовірністю до випадкової величини ξ , якщо для кожного $\varepsilon > 0$

$$\lim_{n \to \infty} p\left(\omega : |\xi_n(\omega) - \xi(\omega)| \ge \varepsilon\right) = 0 .$$

(1)

Збіжність за ймовірністю за звичай позначається $\xi_n \xrightarrow{p} \xi$.

Якщо перейти до протилежних подій, то замість (1) можна записати

$$\lim_{n \to \infty} p\left(\omega : |\xi_n(\omega) - \xi(\omega)| < \varepsilon\right) = 1 .$$

(2)

Збіжність за ймовірністю можна тлумачити так : зі збільшенням імовірність порушення нерівності $|\xi_n(\omega) - \xi(\omega)| < \varepsilon$ стає як завгодно малою.

<u>Означення 2.</u>послідовність випадкових величин $\{\xi_n\}$ збігається до випадкової величини в середньоквадратичному, якщо $M\xi_n^2<\infty$, $M\xi^2<\infty$ та

$$\lim_{n\to\infty} M\left[\xi_n(\omega)-\xi(\omega)\right]^2=0 \quad .(3)$$

Збіжність у середньоквадратичному часто позначають $\xi_n \overset{ck}{\to} \xi$.

Доведемо, що зі збіжності у середньоквадратичному виходить збіжність за ймовірністю. Використовуємо нерівність Чебишева, та одержуємо

$$p\left(\omega : |\xi_n(\omega)-\xi(\omega)| \geq \varepsilon\right) \leq \frac{M\left[\xi_n(\omega)-\xi(\omega)\right]^2}{\varepsilon^2} \quad .$$

Знаменник у правій частині малий, але більший нуля. Тому, коли чисельник прямує до нуля, то і вся права частина прямує до нуля, а, тим паче, ліва частина нерівності.

Ми довели що збіжність за ймовірністю більш слабка ніж збіжність у середньоквадратичному.

Можна було б розглянути збіжність функції $\xi_n(\omega)$ до $\xi(\omega)$ в усіх точках $\omega\in\Omega$ (так звану <u>збіжність усюди</u>) , але вона дуже рідко зустрічається на практиці. Більший інтерес викликає збіжність в усіх точках $\omega\in\Omega$, зокрема мабуть сукупності точок, імовірнісна міра якої дорівнює нулю.

Введемо подію $A=\left\{\omega : \lim_{n\to\infty} \xi_n(\omega)=\xi(\omega)\right\}$ (ξ_n збігається до ξ), яку за допомогою теоретико-множинних операцій можна визначити як

$$A=\underset{r=1}{\overset{\infty}{intersect}}\ \underset{m=1}{\overset{\infty}{¿}}\ \underset{n\geq m}{intersect}\left\{\omega : |\xi_n(\omega)-\xi(\omega)|<\frac{1}{r}\right\} \quad .(4)$$

Ця дещо, “загрозлива” на перший погляд формула стає більш прозорою, якщо згадати означення збіжності і прокоментувати її так: *A* складається з елементарних подій ω таких, має місце збіжність послідовності $\{\xi_n(\omega)\}$ до

$\xi(\omega)$, тобто для довільного $r=\overline{1,\infty}$ (дивись перший знак $\underset{r=1}{\overset{\infty}{intersect}}$)існує таке $m=\overline{1,\infty}$ (дивись другий знак $\underset{m=1}{\overset{\infty}{¿}}$) , що для кожного $n\geq m$ (третій знак $\underset{n\geq m}{intersect}$) має місце

$$|\xi_n(\omega)-\xi(\omega)|<\frac{1}{r}\ .$$

В тому, що $A\in F$ (тобто *A* є подією), можна переконатися, якщо згадати, що злічена кількість операцій сумування та множення елементів *F* не виводить результат з класу *F*.

<u>Означення 3</u>. Послідовність випадкових величин $\{\xi_n\}$ збігається <u>з імовірністю 1</u> (або <u>майже напевно)</u> до випадкової величини ξ , коли для кожного $r=\overline{1,\infty}$

$$p(A)=p\left(\underset{r=1}{\overset{\infty}{intersect}}\ \underset{m=1}{\overset{\infty}{¿}}\ \underset{n\geq m}{intersect}\left\{\omega:|\xi_n(\omega)-\xi(\omega)|<\frac{1}{r}\right\}\right)=1 \qquad .(5)$$

Збіжність з імовірністю 1(майже напевно) позначають $p(\xi_n\to\xi)=1$, або $\xi_n\xrightarrow{м.н.}\xi$.

Можна дати еквівалентне означення збіжності з імовірністю 1. Для цього одержимо наступний результат.

<u>Теорема.</u> Збіжність $p(\xi_n\to\xi)=1$ має місце тоді і тільки тоді, коли

$$\lim_{m\to\infty} p\left(\omega:\sup_{n\geq m}|\xi_n(\omega)-\xi(\omega)|\geq\frac{1}{r}\right)=0 \qquad .(6)$$

для кожного $r=\overline{1,\infty}$.

<u>Доведення</u>. Розглянемо протилежну до *A* подію (дивись(4)):

$$\overline{A}=\operatorname*{intersect}_{r=1}^{\infty}\ \mathop{\text{¿}}_{m=1}^{\infty}\ \operatorname*{intersect}_{n\ge m}\left\{\omega:\left|\xi_n(\omega)-\xi(\omega)\right|\ge\frac{1}{r}\right\}.$$

Зі збіжності з імовірністю 1 повинно бути $p\left(\overline{A}\right)=0$. Зафіксуємо *r* та позначимо:

$$\overline{A}_{r,m}=\left\{\omega:\sup_{n\ge m}\left|\xi_n(\omega)-\xi(\omega)\right|\ge\frac{1}{r}\right\}=\operatorname*{intersect}_{n\ge m}\left\{\omega:\left|\xi_n(\omega)-\xi(\omega)\right|\ge\frac{1}{r}\right\}.$$

Послідовність множин $\overline{A}_{r,m}$, $m=\overline{1,\infty}$, спадна тобто $\overline{A}_{r,m}\supseteq\overline{A}_{r,m+1}$, і тому за теоремою §1.4 для $\overline{A}_r=\operatorname*{intersect}_{m=1}^{\infty}\overline{A}_{r,m}$ маємо:

$$p\left(\overline{A}_r\right)=\lim_{m\to\infty}p\left(\overline{A}_{r,m}\right).$$

Далі міняємо *r* від 1 до ∞ ж бачимо що послідовність $\overline{A}_r$ зростає, тобто $\overline{A}_{r+1}\supseteq\overline{A}_r$, і тому за тією ж теоремою маємо для $\overline{A}=\mathop{\text{¿}}_{r=1}^{\infty}\overline{A}_r$:

$$p\left(\overline{A}\right)=\lim_{r\to\infty}p\left(\overline{A}_r\right)=\sup_r p\left(\overline{A}_r\right).$$

Але $p\left(\overline{A}\right)=\sup_r p\left(\overline{A}_r\right)=0$ тоді і тільки тоді, коли для кожного *r:*

$$p\left(\overline{A}_r\right)=\lim_{m\to\infty}p\left(\overline{A}_{r,m}\right)=\lim_{m\to\infty}p\left(\omega:\sup_{n\ge m}\left|\xi_n(\omega)-\xi(\omega)\right|\ge\frac{1}{r}\right)=0.$$

Теорему доведено.

З цього маємо еквівалентне означення.

<u>Означення 3′.</u> Послідовність випадкових величин $\{\xi_n\}$ збігається <u>з імовірністю 1</u> (або майже напевно) до випадкової величини ξ, коли для кожного $r=\overline{1,\infty}$

$$\lim_{m\to\infty} p\left(\omega : \sup_{n\ge m}\left|\xi_n(\omega)-\xi(\omega)\right|\ge\frac{1}{r}\right)=0$$
.

Аналізуючи (6) приходимо до висновку, що коли $p(\xi_n\to\xi)=1$, то $\lim_{n\to\infty} p\left(\omega : \left|\xi_n(\omega)-\xi(\omega)\right|\ge\frac{1}{r}\right)=0$, тобто зі збіжності з імовірністю 1 виходить збіжність за ймовірністю.

Нам необхідно розглянути ще один важливий вид збіжності: збіжність за розподілом.

<u>Означення 4</u>. Послідовність функцій розподілу $\{F_n(x)\}$ <u>слабко</u> збігається до функції $F(x)$, якщо $F_n(x)\to F(x)$ для кожного x, що є точкою неперервної функції $F(x)$. Якщо для послідовності випадкових величин $\{\xi_n\}$ послідовність відповідних функцій розподілу слабко збігається до функції розподілу

$F(x)$ випадкової величини ξ, то говорять що послідовність $\{\xi_n\}$ збігається до ξ <u>за розподілом</u>.

Збіжність за розподілом часто позначають $F_n(x)\Rightarrow F(x)$, або $\xi_n\Rightarrow\xi$.

Зі збіжності за розподілом виходить, що

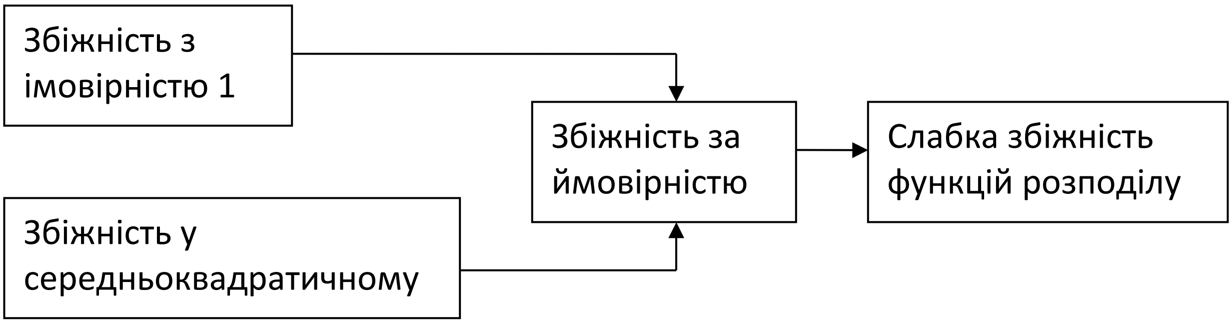
$$p(a\le\xi_n<b)\underset{n\to\infty}{\to} p(a\le\xi<b)$$
,

Коли $p(\xi=a)=p(\xi=b)=0$, тобто точки $a<b$ є точками неперервності функції розподілу $F(x)=p(\xi<x)$.

Можна довести, що зі збіжності за ймовірністю виходить збіжність $\{\xi_n\}$ до ξ

За розподілом.

Зв’язок між різними поняттями збіжності можна подати на схемі



3.2 <u>Теорема Бернуллі. Закон великих чисел</u>

Розглянемо схему Бернуллі: проводиться *n* незалежних випробувань, у кожному з яких подія A відбувається з імовірністю *p*, та не відбувається (відбувається подія $\bar{A}$) з імовірністю *q=1-p*. З кожним випробуванням пов’яжемо випадкову величину $X_i, i=\overline{1,n}$, таку що

$$X_i = \begin{cases} 1, & \text{з імовірністю } p; \\ 0, & \text{з імовірністю } 1-p. \end{cases}$$

Математичне сподівання, другий початковий момент, дисперсія цієї дискретної випадкової величини дорівнюють:

$$MX_i = 1\cdot p + 0\cdot(1-p) = p;$$
$$MX_i^2 = 1\cdot p + 0\cdot(1-p) = p;$$
$$DX_i = MX_i^2 - (MX_i)^2 = p - p^2 = p(1-p) = pq$$

Розглянемо частоту наставання випадкової події A в *n* незалежних випробуваннях, тобто випадкову величину $S_n = \sum_{i=1}^{n} X_i$. За властивостями математичного сподівання та дисперсії суми незалежних випадкових величин $X_i, i=\overline{1,n}$, маємо:

$$MS_n = M\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} MX_i = np;$$
$$DS_n = D\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} DX_i = npq.$$

Для частості наставання події A в n незалежних випробуваннях, тобто для випадкової величини $\frac{S_n}{n}$, маємо:

$$M\left(\frac{S_n}{n}\right) = p, D\left(\frac{S_n}{n}\right) = \frac{1}{n^2}npq = \frac{pq}{n}.$$

За нерівністю Чебишова для довільного $\varepsilon > 0$

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{pq}{n\varepsilon^2}, \quad (1)$$

і таким чином, $\lim\limits_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) = 0$, або $\frac{S_n}{n} \xrightarrow{P} p$.

Ми довели наступний дуже важливий результат.

Теорема Бернуллі. Частість наставання події в n незалежних випробуваннях збігається, за ймовірністю до ймовірності цієї події.

Бачимо, що властивість стійкості частості у великій кількості випробувань зберігається і в аксіоматичній теорії ймовірностей, що підтверджує вдалі введення такої аксіоматики А. М. Колмогоровим. Зауважимо, що цінність має і сама нерівність (1), яка дає уявлення про швидкість збіжності $\frac{S_n}{n}$ до p. Якщо задатися точністю ε та ймовірністю $\frac{pq}{n\varepsilon^2}$, то можна обчислити кількість n випробувань, які дозволяють визначити ймовірність події з заданою точністю та ймовірністю по частоті $\frac{S_n}{n}$.

Приклад. Імовірність деякої події визначається по частоті. Треба обчислити кількість незалежних випробувань, за якої з імовірністю не менше 0,99 одержимо оцінку невідомої ймовірності з точністю $\varepsilon=0,01$.

Використовуємо нерівність (1) (переходячи до ймовірності протилежної події). Одержуємо

$$P\left(\left|\frac{S_n}{n}-p\right|\geq 0,01\right)\leq\frac{pq}{n\varepsilon^2}\leq 0,01 .$$

Але p та q нам невідомі. Користуючись засобами аналізу переконайтеся, що $\max pq=\max p(1-p)$ досягається при $p=\frac{1}{2}$ і дорівнює $\frac{1}{4}$. Маємо

$$\frac{1}{4n\varepsilon^2}\leq 0,01 ,$$

звідки $n\geq\frac{1}{4\cdot 0,01\cdot\varepsilon^2}=250000$.

Розглянемо нескінченну послідовність незалежних випадкових величин X_1, X_2,…, X_n,…, які розподілені однаково та мають скінченну дисперсію $\sigma^2=DX_i<\infty, i=\overline{1,\infty}$.

Теорема.(Слабкий закон великих чисел). Середнє арифметичне цих випадкових величин

$$\overline{x_n}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

збігається за ймовірністю до математичного сподівання однієї з них, тобто $\overline{x_n} \vec{p}\, MX_i$.

<u>Доведення.</u> Маємо

$$M\,\overline{x_n} = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} MX_i = MX_i = \nu\,;$$

$$D\,\overline{x_n} = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{\sigma^2}{n}.$$

Знову використовуємо нерівність Чебишова та одержуємо:

$$P\left(|\bar{x}_n - \nu| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} \quad . (2)$$

Бачимо, що при довільному $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(|\bar{x}_n - \nu| \geq \varepsilon\right) = 0 \quad .$$

Запам'ятаємо цей важливий результат: для того, щоб достатньо точно оцінити математичне сподівання деякої випадкової величини *X*, треба взяти над нею велику кількість незалежних випробувань та взяти середнє арифметичне результатів цих випробувань $X_1, X_2, \ldots, X_n$.

3.3<u>Посилений закон великих чисел</u>

Раніш ми довели закон великих чисел у найпростішій формі: збіжність середнього арифметичного до математичного сподівання за ймовірністю. Тепер наша задача одержати більш сильний результат – збіжність з ймовірністю 1 та й ще в ситуації, коли випадкові величини, що сумуються, мають різні розподіли.

Перш за все установимо важливу для подальшого нерівність.

Теорема 1 (Нерівність Колмогорова). Нехай X_1, X_2,…, X_n – незалежні випадкові величини, у яких $MX_i = v_i = 0$ та $DX_i = \sigma_i^2 < \infty, i = \overline{1,n}$.

Покладемо $S_m = \sum_{i=1}^{m} X_i, m = \overline{1,n}$. Тоді для довільного $\varepsilon > 0$

$$P\left(\max_{1 \le m \le n} |S_m| \ge \varepsilon\right) \le \frac{1}{\varepsilon^2} DS_n .$$

Доведення. Розглянемо події

$$A_m = \{\omega : |S_1| < \varepsilon, \dots, |S_{m-1}| < \varepsilon, |S_m| < \varepsilon,\}, m = \overline{1,n} ;$$

тобто A_m є подія, яка полягає у тому, що суми випадкових величин X_i до X_{m-1} включно за модулем менше ε, а $\sum_{i=1}^{m} X_i$ за модулем не менше ε.

Події A_m несумісні, сума цих подій дорівнює

$$\bigcup_{m=1}^{n} A_m = \left\{\max_{1 \le m \le n} |S_m| \ge \varepsilon\right\} ,$$

тобто події, ймовірність якої нам треба оцінити.

Введемо індикатори подій A_m, тобто дискретні випадкові величини $1_m, m = \overline{1,n}$, такі, що приймають два значення:

$$1_m = \begin{cases} 1, & \text{якщо } \omega \in A_m ; \\ 0, & \text{якщо } \omega \notin A_m ; \end{cases}$$

Очевидно, $M 1_m = P(A_m)$.

Нехай $m < n$. Випадкові величини X_{m+1},…, X_n не залежать від X_1,…, X_m та величин 1_m. Незалежними є і функції $\phi_1(X_{m+1}, \dots, X_n)$ та $\phi_2(X_1, \dots, X_m, 1_m)$. Тому, за властивістю математичного сподівання добутку незалежних випадкових величин, та враховуючи $MS_m = 0$, маємо

$$M\left(S_n^2\cdot 1_m\right)=M\left\{\left[S_m+\left(S_n-S_m\right)\right]^2\cdot 1_m\right\}=M\left[S_m^2\cdot 1_m\right]+2M\left[S_m\left(S_n-S_m\right)\cdot 1_m\right]+$$
$$+M\left[\left(S_n-S_m\right)^2\cdot 1_m\right]=M\left[S_m^2\cdot 1_m\right]+2M\left[S_n-S_m\right]\cdot M\left[S_m\cdot 1_m\right]+$$
$$+M\left[S_n-S_m\right]^2 M\,1_m=M\left[S_m^2\cdot 1_m\right]+M\left[S_n-S_m\right]^2 P\left(A_m\right)\geq M\left[S_m^2\cdot 1_m\right]$$

Далі зауважимо, що випадкова $\sum_{m=1}^{n} 1$, може приймати тільки два значення: нуль або одиниця (серед 1_m тільки одна може дорівнювати одиниці, тому що події A_m несумісні, або усі 1_m дорівнюють нулеві). І, таким чином,

$S_n^2 \geq S_n^2 \sum_{m=1}^{n} 1_m$, звідки (з урахуванням (3)):

$$MS_n^2 \geq M\left[S_n^2 \sum_{m=1}^{n} 1_m\right] = \sum_{m=1}^{n} M\left[S_n^2 \cdot 1_m\right] \geq \sum_{m=1}^{n} M\left[S_m^2 \cdot 1_m\right]. \quad (4)$$

Очевидна нерівність $S_m^2 \cdot 1_m \geq \varepsilon^2 \cdot 1_m$, і тому $M\left[S_m^2 \cdot 1_m\right] \geq \varepsilon^2 M 1_m = \varepsilon^2 P\left(A_m\right)$. (5)

Тепер з (4) одержуємо (завдяки (2) і (5)):

$$DS_n = MS_n^2 \geq \varepsilon^2 \sum_{m=1}^{n} P\left(A_m\right) = \varepsilon^2 P\left(\bigcup_{m=1}^{n} A_m\right) = \varepsilon^2 P\left(\max_{1\leq m\leq n}\left|S_m\right| \geq \varepsilon\right).$$

З останньої нерівності виходить (1). Теорему доведено.

Вище ми припускали, що $MX_i = 0, i = \overline{1,n}.$ Якщо це не так, тобто існують $MX_i = \nu_i < \infty, i = \overline{1,n}$, то заміною $X_i - \nu_i$ на X_i можна звести цю ситуацію до попередньої та одержати замість (1) (пропонуємо переконатись в цьому самостійно) нерівність

$$P\left(\max_{1\leq m\leq n}\left|S_m - MS_m\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} DS_n. \quad (6)$$

<u>Теорема 2 (посилений закон великих чисел).</u> Нехай $X_1,\ldots,X_n,\ldots,$ нескінченна послідовність незалежних випадкових величин з $MX_n = \nu_n < \infty, DX_n = \sigma_n^2 < \infty, n = \overline{1,\infty}.$ Нехай виконується умова

$$\sum_{n=1}^{\infty}\frac{\sigma_n^2}{n^2}<\infty. \qquad (7)$$

Тоді

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\nu_i\right)=0\right)=1, \qquad (8)$$

тобто різниця $\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}\nu_i$ збігається до нуля з імовірністю один.

<u>Доведення.</u> Знову заміною $X_n-\nu_n$ на X_n зводимо задачу до дослідження послідовності випадкових величин $\{X_n\}$ з нульовими математичними сподіваннями. Згадуємо еквівалентне означення збіжності з імовірністю 1 та переконаємося, що для довільного $\varepsilon>0$ має місце

$$\lim_{m\to\infty}P\left(\sup_{n\ge m}\left|\frac{S_n}{n}\right|\ge\varepsilon\right)=0,$$

де $S_n=\sum_{i=1}^{n}X_i.$

Вводимо події

$$B_k=\left\{\max_{2^{k-1}\le n\le 2^k}\left|\frac{S_n}{n}\right|\ge\varepsilon\right\},\ k=\overline{1,\infty}.$$

Використовуємо нерівність Колмогорова (1) та одержуємо

$$P\left(B_k\right)\le P\left(\max_{2^{k-1}\le n\le 2^k}\left|S_n\right|\ge\varepsilon\cdot 2^{k-1}\right)\le\frac{DS_{2^k}}{\varepsilon^2\cdot 2^{2(k-1)}}=4\varepsilon^{-2}\cdot 2^{-2k}\sum_{n\le 2^k}\sigma_n^2.$$

Знаходимо оцінку для суми (користуємось (7) та міняємо порядок сумування):

$$\sum_{k=1}^{\infty}P\left(B_k\right)\le 4\varepsilon^{-2}\sum_{k=1}^{\infty}2^{-2k}\sum_{n\le 2^k}\sigma_n^2=4\varepsilon^{-2}\sum_{n=1}^{\infty}\sigma_n^2\sum_{2^{-2k}\le n^{-2}}2^{-2k}\le 4\varepsilon^{-2}\sum_{n=1}^{\infty}\frac{\sigma_n^2}{n^2}\frac{1}{1-1/4}<\infty.$$

Бачимо, що послідовність $\sum_{k=1}^{m} P(B_k)$ не спадна та обмежена, тобто вона збігається, а тому $\lim_{m\to\infty} \sum_{k=m}^{\infty} P(B_k) = 0.$ Звідки виходить, що

$$P_{2^{m-1}} = P\left(\sup_{n \ge 2^{m-1}} \left| \frac{S_n}{n} \right| \ge \varepsilon \right) = P\left(\bigcup_{k=m}^{\infty} B_k \right) \le \sum_{k=m}^{\infty} P(B_k) \underset{m\to\infty}{\to} 0,$$

а тому і

$$P_m = P\left(\sup_{n \ge m} \left| \frac{S_n}{n} \right| \ge \varepsilon \right) \underset{m\to\infty}{\to} 0.$$

Теорему доведено.

Наслідок (теорема Бореля[17]). Частість настававння події в n незалежних випробуваннях $\frac{S_n}{n}$ збігається з ймовірністю 1 до ймовірності p цієї події

$$P\left(\frac{S_n}{n} \to p \right) = 1$$

(дивись позначення у теоремі Бернуллі з 3.2).

Пропонуємо читачам самостійно переконатися, що умови теореми 2 тут виконуються.

17Борель Фелікс Едуард Жюстен Еміль (1871–1956) – французький математик, член Паризької АН.