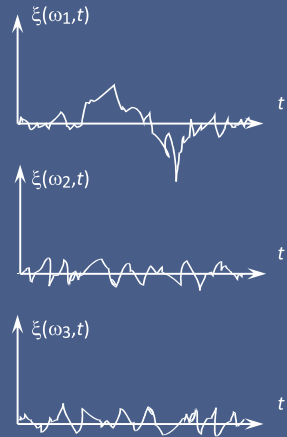
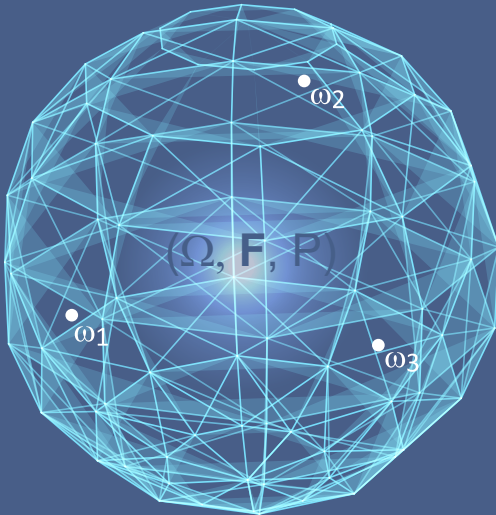


В.О. Омельченко

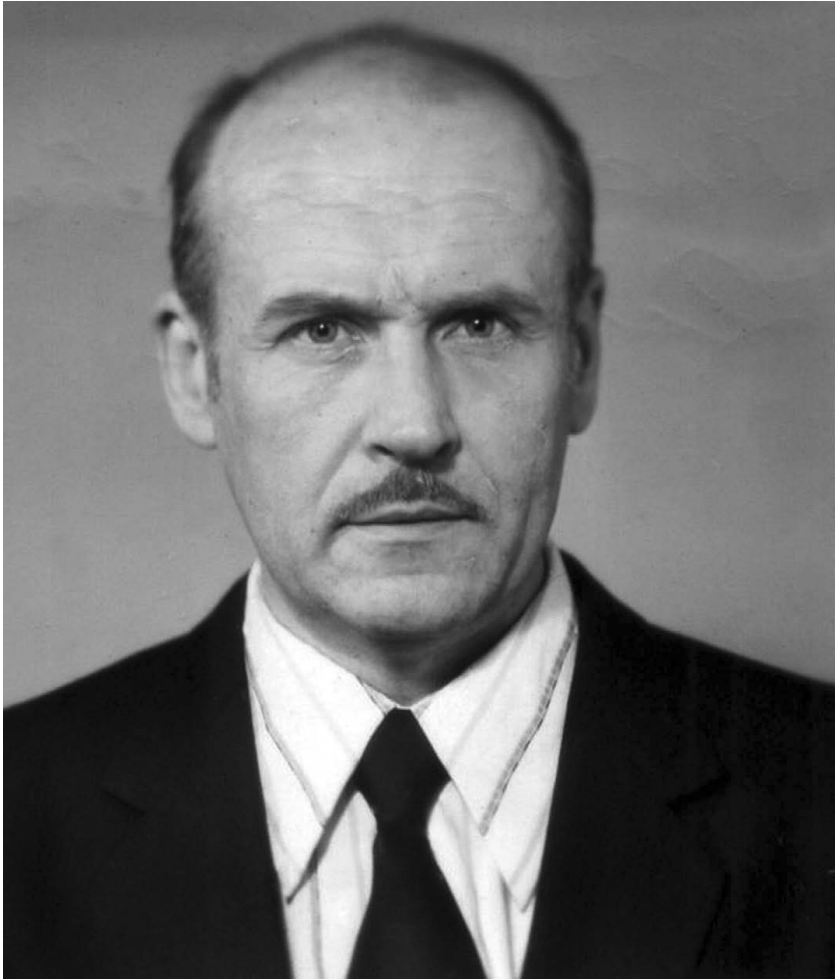
ТЕОРІЯ ЙМОВІРНОСТЕЙ У ЕЛЕКТРОЗВ'ЯЗКУ



$$I(X;Y) = MJ(X,Y) = \iint_{-\infty}^{\infty} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dx dy$$

В.О. Омельченко

Теорія ймовірностей у електровз'язку



Віктор Олександрович Омельченко у молодому віці

В.О. Омельченко

ТЕОРІЯ ЙМОВІРНОСТЕЙ У ЕЛЕКТРОЗВ'ЯЗКУ

**Класична теорія ймовірностей, випадкові процеси,
теорія інформації, математична статистика
(теорія статистичних рішень)**

Харків
«Контраст»
2024

УДК 621.391:537.86:519

О57

Омельченко В.О.

О57 Теорія ймовірностей у електрозв'язку. Класична теорія ймовірностей, випадкові процеси, теорія інформації, математична статистика (теорія статистичних рішень). – Харків: «Контраст», 2024. – 272 с.
ISBN 978-617-7405-74-9

На основі теорії множин, міри, інтеграла за мірою і функціональних просторів розглядаються принципові результати класичної теорії ймовірностей, теорії випадкових процесів, теорії інформації, теорії статистичних рішень (як основних задач математичної статистики) – з позицій потреб теорії електрозв'язку. Мотивація теорії, наведені міркування, фізичний зміст основних понять та результатів теорії ілюструються прикладами з теорії електрозв'язку.

Розрахована на студентів, аспірантів та викладачів вузів у галузі зв'язку; може бути корисною фахівцям у галузі радіотехніки, радіофізики, а також там, де використовується обробка сигналів.

УДК 621.391:537.86:519

Рецензент:

А.С. Мазманішвілі,

доктор фізико-математичних наук, професор (Національний науковий центр «Харківський фізико-технічний інститут»)

ISBN 978-617-7405-74-9

© В.О. Омельченко, 2024

ПЕРЕДМОВА

Існує різноманітна література з теорії ймовірностей та її гілок — для інженерів, прикладників-математиків, «чистих» математиків тощо. Однак серед широкого кола книг цього профілю нема достатньо повного висвітлення питань теорії у напрямку радіотехніки та зв'язку. Особливо це стосується сучасних спеціальностей у галузі телекомунікацій та мереж зв'язку.

Книга ставить ціль дати теоретико-ймовірнісний та статистичний фундамент теорії електричного зв'язку, яка сама по собі на Заході розглядається як самостійна гілка математики. Вона відрізняється від відомих книг з теорії ймовірностей та її гілок відбором матеріалу, способом викладання її положень і особливо — наповненням фізичним змістом абстрактних теоретичних співвідношень з галузі електровз'язку.

У першому розділі стисло викладається математичний апарат, на якому будуватиметься сучасна теорія ймовірностей та її гілки. Другий розділ присвячено поняттям та результатам класичної теорії ймовірностей. У третьому розділі викладаються випадкові процеси. Четвертий розділ присвячено поняттям та результатам теорії інформації. П'ятий розділ присвячено деяким задачам математичної статистики — теорії статистичних рішень.

У розділах паралельно з викладанням математичних понять приводяться необхідні відомості з теорії електричного зв'язку, на основі яких дається відповідна мотивація постановок математичних задач та трактовок результатів їх вирішення.

Книга заснована на спецкурсі «Спеціальні розділи математики», який три роки читався автором у Харківському національному університеті радіоелектроніки студентам спеціальності у галузі зв'язку. Книга створювалась на основі студентських конспектів лекцій. Подяки заслужують наукові співробітники автора і студенти кафедри «Мережі зв'язку», які приймали участь у створенні цієї книги. Особливої подяки заслужують студентки цієї кафедри Аліна Кучерява та Наталя Біловус, а також студентка кафедри «Прикладна математика» Людмила Омельченко за суттєву допомогу у роботах по комп'ютерному набору та упорядкуванню тексту. Слід також відмітити Шаталова А.П. — спонсора автобіографічної книги автора («Два с половиною года в игровом бизнесе или можно ли выиграть миллион»), підтримка видання якої врешті решт дозволила без затримки видати і цю книгу.

1. МАТЕМАТИЧНИЙ АПАРАТ СУЧАСНОЇ ТЕОРІЇ ЙМОВІРНОСТЕЙ ТА ЇЇ ГІЛОК

У сучасній радіотехніці та зв'язку багато задач не розв'язуються у рамках класичної вищої математики, які викладаються згідно до програм цього курсу. Для їх постановок, розв'язання і навіть для пояснення результатів необхідно використовувати більш широкі математичні поняття, ніж ті, що вивчаються у технічному вузі у курсі вищої математики. Ідеальним варіантом було б мати освіту професійного математика, яку дає механіко-математичний факультет університетів. Однак, як відомо, неможливо «обійняти необіймане».

Все ж таки існує можливість студентам, які навчаються у технічному вузі, після звичайного курсу вищої математики у відповідних спецкурсах на рівні визначень та понять, дати уявлення з більш глибоких та сучасних гілок математики, які базуються на теорії множин, вимірних функцій та функціональному аналізі.

Більшість з них є широким узагальненням понять та результатів, відомих з розділів курсу вищої математики. З другого боку, саме такий математичний апарат виявляється адекватним багатьом задачам у напрямку радіотехніки та зв'язку.

У цьому розділі розглядається необхідний математичний апарат – це відомості з теорії множин, міри, інтеграла за мірою, вимірних функцій, функціональних просторів, без яких неможливо викласти сучасні результати теорії ймовірностей та її гілок.

1.1. Елементи теорії множин

Основи теорії множин були закладені у кінці XIX – на початку XX століть. Теорія множин учинила великий вплив на наступний розвиток усієї математики. Через теорію функцій та функціональний аналіз її поняття проникли у теорію електров'язку.

1.1.1. Поняття множини

Поняття множин первинно у цій теорії. Тому дати визначення множини через більш широкі поняття неможливо. Можливо лише дати пояснення на прикладах (множина студентів в аудиторії, множина чисел на дійсній осі, множина атомів у сонячній системі тощо). Кажуть, що множина – це багато чогось, що розглядається як єдине ціле. Множини позначаються великими літерами, її елементи – малими.

Якщо $a \in A$ є елементом множини A , це записують символом $a \in A$, а якщо $a \notin A$ не є елементом множини A , — символом $a \notin A$.

Якщо кожний елемент множини A є елементом множини B , то говорять, що множина A — частина (підмножина) B . Це записують так: $A \subset B$. Говорять також, що A міститься в B . Будь-яка множина містить порожню, тобто завжди $\emptyset \subset A$. Коли $A \subset B$ і $B \subset A$, множини A і B збігаються, тобто дорівнюють одна одній, що записують символом $A = B$. Ці множини еквівалентні, але не навпаки (далі ми дамо означення, що множини еквівалентні ($A \sim B$), якщо між їх елементами можна встановити взаємно—однозначну відповідність).

1.1.2. Теоретико-множинні операції

У класі множин уводять ряд теоретико-множинних операцій над множинами, основними з яких є об'єднання, різниця, переріз.

Об'єднанням двох множин A і B називають множину, котра містить усі елементи A , усі елементи B і не містить жодних інших елементів. Об'єднання позначають так: $A \cup B$.

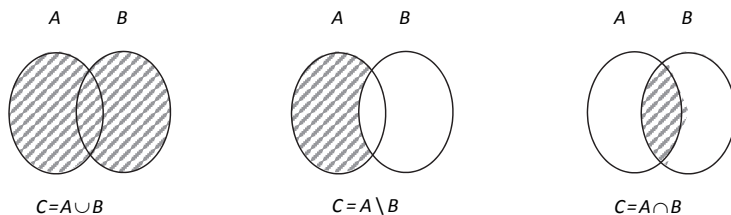


Рис. 1.1. — Теоретико-множинні операції

Різницею двох множин A і B називають множину, яка складається з тих елементів множини A , що не є елементами множини B . Різницю позначають символом $A \setminus B$.

Перерізом двох множин A і B називають множину її спільних елементів. Операцію перерізу позначають так: $A \cap B$. Є також інші операції.

Симетричною різницею двох множин A і B називають суму різниць $A \setminus B$ та $B \setminus A$. Її позначають символом $A \Delta B$. Отже, за означенням:

$$A \Delta B = (A \setminus B) \cup (B \setminus A).$$

Із цього означення видно, що назва операції «симетрична різниця» не зовсім вдале, бо, по суті, результатом її є об'єднання тих частин множин A та B , що мають пустий переріз. До речі, об'єднання пустих

множин називають сумою і позначають $A+B$. У загальному випадку, симетрична різниця близька до поняття операції суми за модулем 2, коли береться об'єднання двох множин, але елементи, що зустрічаються двічі – викидаються. [Колмогоров і Фомін, ст.21, Колмогоров, ст.11]

Доповнення множини A до основної множини S називають різницю $S \setminus A$. Позначають: $\bar{A} = S \Delta A$.

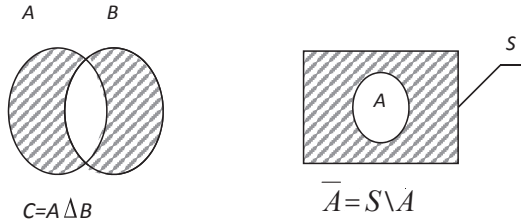


Рис. 1.2. – Теоретико-множинні операції

Використовуючи введені теоретико-множинні операції, означають монотонні послідовності множин $A_1, A_2, \dots, A_n, \dots$ та їх границі.

Якщо для послідовності можна побудувати включення:

$$A_1 \subset A_2 \subset \dots \subset A_n \subset \dots, \text{ або:}$$

$$A_1 \supset A_2 \supset \dots \supset A_n \supset \dots,$$

її називають монотонною. У першому разі це *неспадна* послідовність ($A_n \uparrow$), у другому – *незростаюча* ($A_n \downarrow$).

Кожна монотонна послідовність має границю. Нижню та верхню границі визначають відповідно рівностями:

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

$$\overline{\lim}_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

Існують методи конструювання по заданих множинах нових множин. Один із них – це побудова добутку множин.

Прямим добутком $A \times B$ множин A і B називається сукупність усіх пар елементів вигляду (a, b) , де $a \in A$, $b \in B$.

Наведене означення може бути узагальнене на довільне скінченне, або, навіть, нескінченне число множин. Так, під прямим добутком

$$\prod_{i \in I} A_i, \quad I = \{1, 2, \dots, n\},$$

де для n може припускатися і нескінченність, розуміється сукупність усіх елементів вигляду (a_1, a_2, \dots, a_n) , $a_i \in A_i$.

1.1.3. Еквівалентні множини

Множини називають *еквівалентними*, якщо між ними можна встановити взаємно-однозначну відповідність.

Встановлення такої відповідності є одним із способів перерахування «кількості» елементів у множині. Існують скінченні та нескінченні множини.

Скінченні множини мають скінченну кількість елементів. Множина називається *нескінченною*, якщо при викиданні будь-якого числа її елементів у неї залишиться хоча би один елемент. Для нескінченних множин встановлення взаємно-однозначної відповідності даної множини та деякої стандартної множини – це єдиний спосіб знаходження *кількості елементів*.

Нескінченна множина має характеристичну особливість: існує така її частина, що ця частина еквівалентна усій множині. Тобто деяка частина нескінченної множини у певному розумінні дорівнює (а точніше – еквівалентна) цілому. Ця властивість знаходиться у різкому протиріччі з відомою властивістю скінченної множини, що частина не дорівнює цілому.

Наприклад, візьмемо нескінченну множину A натуральних чисел, та B – множину всіх парних чисел:

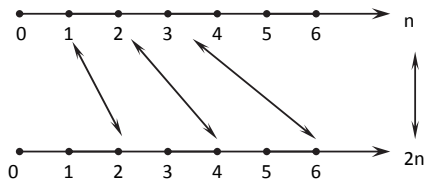


Рис.1.3. – Множина $A = \{1, 2, 3, \dots, n, \dots\}$ еквівалентна множині $B = \{2, 4, 6, \dots, 2n, \dots\}$, яка є частиною A

Якщо кожному $n \in A$ зіставити $2n \in B$, буде встановлена взаємно-однозначна відповідність, тобто одержимо, що ці множини еквіва-

лентні ($A \sim B$). Але B міститься у A , тобто $B \in$ частиною A . Множина A еквівалентна деякій своїй частині. (рис. 1.3) Ця властивість нескінченної множини може бути прийнята за її означення.

Розглядають зчисленні та незчисленні множини.

Множина називається *зчисленною*, якщо вона еквівалентна множині натуральних чисел: $N = \{1, 2, 3, \dots, n, \dots\}$.

Множину, еквівалентну відрізьку $[0, 1]$, називають *множиною потужності континууму*. Вона незчисленна. (Термін *потужність* у теорії множин еквівалентний терміну *число елементів* для скінченних множин. Підкреслимо, що термін *потужність* у інших розділах книги буде використано також у іншому розумінні).

Тут та у подальшому тексті термін «відрізок» вживається для позначення «інтервалу у широкому розумінні» — довільної зв'язаної множини на прямій. При цьому не уточнюють, входять чи не входять кінцеві точки до розглядуваної множини. В разі потреби уточнюють, належить чи не належить кінцева точка до розглядуваної множини і позначають цей факт відповідно квадратною дужкою чи то круглою, вживаючи також терміни: «*сегмент*», «*інтервал*». При встановленні взаємно-однозначної відповідності це має суттєве значення, але при визначенні потужності нескінченної множини скінченну кількість окремих точок можна не враховувати.

Приклади зчисленних множин. Сума скінченного числа зчисленних множин є зчисленною множиною. Сума зчисленної множини зчисленних множин є множиною зчисленна. Так, множина усіх цілих та раціональних чисел зчисленна; множина усіх цілих та раціональних точок площини зчисленна (раціональною точкою площини називають точку, абсциса та ордината якої раціональні).

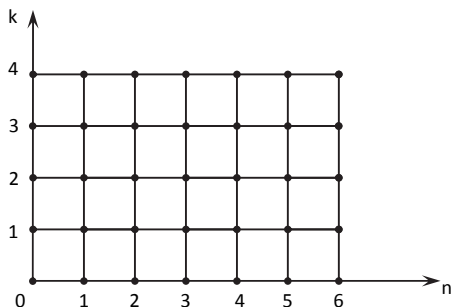


Рис. 1.4. — Цілі точки площини (n, k) — зчисленна множина

Множина усіх цілих та раціональних точок n -вимірного простору також зчисленна.

Приклади незчислених множин. Існування незчислених множин доводиться теоремою Кантора: множина, еквівалентна відріzkу $(0,1)$, – незчисленна.

Доводиться, що сума скінченного числа множин потужності континууму є множина потужності континууму; сума зчисленного числа множин потужності континууму є множина потужності континууму. Так, множина точок усїєї дійсної осі є множина потужності континууму.

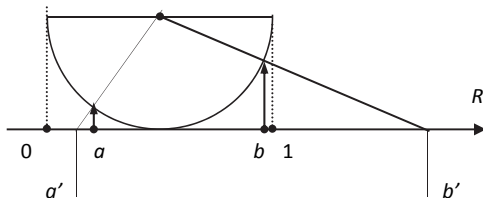


Рис. 1.5. – Встановлення взаємно-однозначної відповідності точок відріzkу $[0,1]$ з точками дійсної осі R : $(a \leftrightarrow a')$; $(b \leftrightarrow b')$

Множина точок одиничного квадрату на площині та множина усіх точок площини також є множинами потужності континууму; множина усіх точок n -вимірного простору є множина потужності континууму.

Між зчисленною множиною та множиною потужності континууму існує певний зв'язок. Якщо взяти зчисленну множину та побудувати клас усіх її підмножин, то він буде мати потужність континууму. Якщо аналогічну операцію виконати для множини потужності континууму одержимо множину, потужність якої більше, ніж континуум. Цікаво підкреслити, що множина усіх дійсних функцій на інтервалі $(0,1)$ має саме таку потужність – більш, ніж потужність континууму.

1.1.4. Класи множин

Для подальшого нам треба ввести крім окремих множин ще *класи множин* (множину множин).

Класом множин називають множину, елементами якої є будь-які множини. Особливо важливими є класи, замкнені відносно тих чи інших теоретико-множинних операцій.

Важливим класом є *кільце множин*. Це непустий клас, замкнений відносно операцій об'єднання і різниці. При цьому кільце виявляється замкненим також відносно операції перерізу, тому що цю операцію можна виразити через операції різниці:

$$A \cap B = A \setminus (A \setminus B).$$

Операції об'єднання та різниці називають основними для кільця. Якщо цей клас замкнений відносно зчисленних операцій об'єднання. Його називають σ -кільцем.

У класі множин вводять поняття *одиниця класу*. Кажуть, що E – одиниця класу множин, якщо для довільної множини A із цього класу: $A \cap E = A$, тобто E – це *максимальна* множина класу, яка включає в себе елементи всіх інших множин.

σ -алгеброю називається σ -кільце з одиницею класу. Тобто σ -алгебра замкнена відносно операції різниці, зчисленних операцій об'єднання. Виявляється, що вона також замкнена відносно зчисленних операцій перерізу.

*Наведемо приклад алгебри. Класи множин будемо позначати великими напівжирними літерами: **A**, **B**, **F** або *A*, *B*, *F* тощо. Нехай маємо множину A , яка складається із трьох елементів:*

$$A = \{a_1, a_2, a_3\}.$$

Побудуємо клас, у який входять усі підмножини цієї множини.

Разом з пустою множиною клас має $n=2^3=8$ елементів:

$$\mathbf{A} = \{\emptyset, \{a_1\}, \{a_2\}, \{a_3\}, \{a_1, a_2\}, \{a_1, a_3\}, \{a_2, a_3\}, \{a_1, a_2, a_3\}\}.$$

Клас \mathbf{A} є алгеброю, тому що операції об'єднання та різниці не виводять за межі цього класу. Наприклад:

$$\{a_1\} \cup \{a_2\} = \{a_1, a_2\} \in \mathbf{A};$$

$$\{a_2, a_3\} \setminus \{a_2\} = \{a_3\} \in \mathbf{A}.$$

Автоматично, операція перерізу також не виводить за межі класу, так:

$$\{a_1, a_3\} \cap \{a_2\} = \emptyset \in \mathbf{A}.$$

Підкреслимо, що літерами a_1, a_2, a_3 позначаються елементи множини; ці самі літери у фігурних дужках вже позначають відповідні множини – у даному разі одноелементні множини: $\{a_1\}$, $\{a_2\}$, $\{a_3\}$ тощо.

1.2. Міра, вимірні функції, інтеграл за мірою

1.2.1. Міра

Міра – це функція множин, яка задовольняє певні умови.

Функція множин – це функція, яка визначена на класі множин тих чи інших елементів і набуває значень із деякого лінійного простору. Звичайно це множина дійсних чи комплексних чисел.

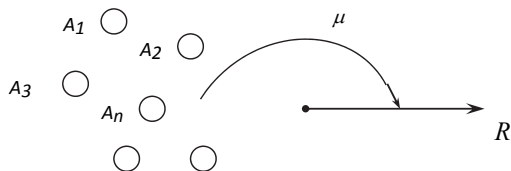


Рис. 1.6. – Функція множин

Функцію множин $\mu(A)$, задану на деякому класі \mathcal{C} множин, називають мірою, якщо вона невід’ємна, адитивна і монотонна, тобто:

а) $\mu(A) \geq 0$;

б) $\mu(A \cup B) = \mu(A) + \mu(B)$, $A \cap B = \emptyset$, $A, B \in \mathcal{C}$; (1.1)

в) $\mu(A) \geq \mu(B)$, якщо $A \supset B$.

Міра називається *зчисленно-адитивною* (або *σ -адитивною*), якщо:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i), \quad A_i \cap A_j = \emptyset, \quad A_i \in \mathcal{C}.$$

Міра – це узагальнене поняття довжини відрізка, площі плоскої фігури, об’єму просторової фігури, тощо. Зокрема для двовимірного випадку міру вводять так. Спочатку міру t задають на прямокутниках як площу. Потім її поширюють на елементарні множини, тобто множини, що їх можна подати у вигляді об’єднання скінченного числа попарно неперерізних прямокутників, і вводять вимірну множину – множину, яку як завгодно точно можна наблизити елементарними множинами. Далі вводять зовнішню міру як нижню грань за всіма можливими покриттями множини A – скінченними чи зчисленими системами прямокутників. Здобуту функцію t , яку розглядають лише на вимірних множинах, називають *мірою Лебега*.

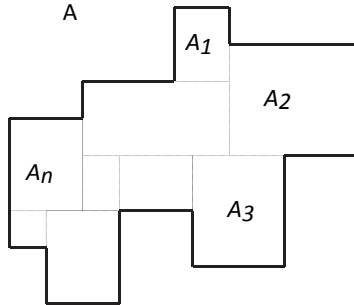


Рис. 1.7. – Введення міри m елементарної множини A як суми площ $S(A_i)$ попарно неперерізних прямокутників $A_i, i=\overline{1, n}$; $m(A) = \sum_{i=1}^n S(A_i)$

Окрім міри Лебега застосовують також *міру Стілтєса*. При цьому вважають, що задано деяку неспадну неперервну зліва функцію $\varphi(x)$ у деякому проміжку на прямій. Тоді, завдавши функцію $m([x_1, x_2]) = \varphi(x_2) - \varphi(x_1)$, згідно зі згадану процедурою можна побудувати міру. Сукупність множин, вимірних щодо такої міри, буде σ -кільцем. Міри Стілтєса визначаються вибором функцій $\varphi(x)$. Якщо $\varphi(x) = x$, дістаємо міру Лебега.

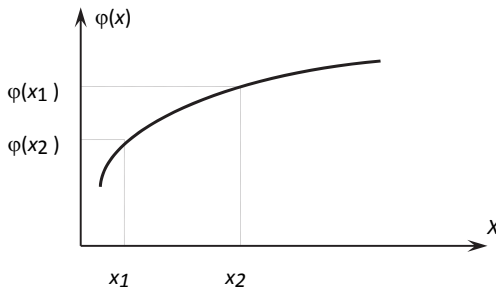


Рис. 1.8. – Завдання міри Стілтєса: $m([x_1, x_2]) = \varphi(x_2) - \varphi(x_1)$

Поняття міри використовується для побудування теорії інтегрування вимірних функцій, клас яких ширший за клас неперервних функцій. У теорії ймовірностей це поняття є фундаментальним – там використовується *ймовірнісна міра*.

1.2.2. Вимірні функції

У загальній теорії функцій функцію розглядають як деяке відображення. Якщо кожному елементу x множини X поставлено у відповідність елемент y множини Y , то маємо відображення X в Y .

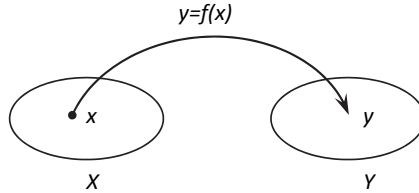


Рис.1.9. — Означення функції $y=f(x)$ як відображення множини X в Y

Елемент y у відображенні $y=f(x)$ називається *образом* елемента x , а елемент x — *прообразом* елемента y .

Якщо задано множину $A \subset Y$, то сукупність усіх тих елементів x , для яких $f(x)$ належить A , називається *повним прообразом* множини A . Його позначають $f^{-1}(A)$: $\{x: f(x) \in A\} = f^{-1}(A)$.

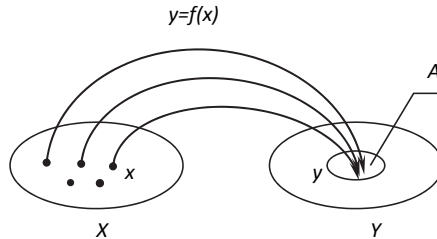


Рис.1.10. — Ілюстрація повного прообразу множини A як сукупності елементів x , для яких $y=f(x)$ належить A : $f^{-1}(A) = \{x: f(x) \in A\}$

Нехай на множині X задано σ -алгебру G_X , а на множині Y — σ -алгебру G_Y . Кажуть, що множина вимірна, якщо вона належить σ -алгебрі (доводиться, що це означення еквівалентне попередньому, даному у пункті «міра»).

Функцію $f(x)$ називають *вимірною*, якщо повний прообраз $f^{-1}(A)$, ($A \in G_Y$) множини A належить G_X . Інакше кажучи, функція $f(x)$ вимірна, якщо повний прообраз вимірної множини є вимірною множиною.

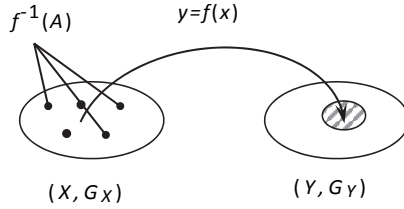


Рис.1.11. – Ілюстрація вимірної функції $y=f(x)$, де $A \in G_Y$, $f^{-1}(A) \in G_X$

Говорять, що задано деякий вимірний простір, коли задано пару об'єктів (X, \mathbf{B}) , де X – деяка довільна множина, \mathbf{B} – деяка σ -алгебра підмножин із X .

Важливим поняттям у теорії вимірних функцій є поняття їх еквівалентності.

Функції f і g , задані на одній і тій самій вимірній множині X , називаються еквівалентними, якщо вони різняться лише на множині міри нуль:

$$\mu \{x: f(x) \neq g(x)\} = 0. \quad (1.2)$$

Так, узагальнена функція $\delta(x)$, яка у інженерних роботах інтерпретується як функція

$$\delta(x) = \begin{cases} \infty, & x=0; \\ 0, & x \neq 0; \end{cases}$$

у теорії вимірних функцій еквівалентна функції, яка тотожно дорівнює нулю: $0(x) \equiv 0$ – це

$$\mu \{x: \delta(x) \neq 0(x)\} = 0. \quad (1.3)$$

У теорії вимірних функцій вводять принципово нові типи збіжностей послідовностей функцій порівнянно з відомими з області диференціального та інтегрального числення.

Послідовність вимірних функцій $\{f_n(x)\}$ збігається майже скрізь до $f(x)$, якщо множина тих x , для яких $\lim_{n \rightarrow \infty} f_n(x) \neq f(x)$, має міру нуль:

$$\mu \{x: \lim_{n \rightarrow \infty} |f_n(x) - f(x)| \neq 0\} = 0. \quad (1.4)$$

Послідовність вимірних функцій $\{f_n(x)\}$ збігається за мірою до функції $f(x)$, якщо для будь-якого $\varepsilon > 0$ границя мір множин, для яких $|f_n(x) - f(x)| \geq \varepsilon$, дорівнює нулю:

$$\lim_{n \rightarrow \infty} \mu \{x: |f_n(x) - f(x)| \geq \varepsilon\} = 0. \quad (1.5)$$

Зі збіжності майже скрізь ($f_n \xrightarrow{M.CKP.} f$) випливає збіжність за мірою ($f_n \xrightarrow{\mu} f$).

Введені вимірні функції утворюють клас, який ширше класу неперервних функцій. Цей клас замкнений не тільки відносно арифметичних операцій, але й операцій граничного переходу, тобто границя послідовності вимірних функцій є функція вимірна. (Відомо, що клас неперервних функцій не замкнений відносно операції граничного переходу, він тільки замкнений відносно операції рівномірного граничного переходу). Однак, у деякому розумінні, вимірні функції досить близькі до неперервних. Так, згідно з теоремою Лузіна для будь-якої числової вимірної функції $f(x)$, яка не приймає значень ∞ , для довільного $\varepsilon > 0$ можна вказати неперервну функцію $f_\varepsilon(x)$, таку, що міра тієї множини, на якій $f(x) \neq f_\varepsilon(x)$, менше ε .

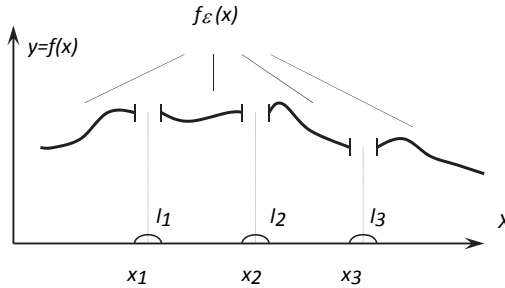


Рис. 1.12. – Ілюстрація теореми Лузіна щодо заміни вимірної функції $f(x)$

неперервною $f_\varepsilon(x)$ на множині, міра якої менше ε : $\sum_i l_i \leq \varepsilon$

1.2.3. Інтеграл за мірою

Поняття міри та вимірної функції використовують для побудови інтеграла за мірою. Основна ідея побудови інтеграла Лебега полягає в тому, що точки x_i у інтегральній сумі групуються не за ознаками близькості їх у множині X (як при побудові інтеграла Рімана), а за ознаками близькості відповідних значень функції.

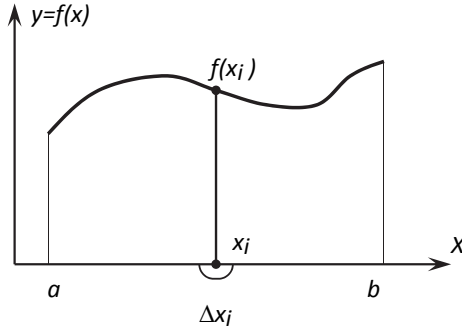


Рис.1.13. — Побудова інтегральної суми у інтегралі Рімана: $\sum_{i=1}^n f(x_i)\Delta x_i$

Побудова інтеграла Лебега:

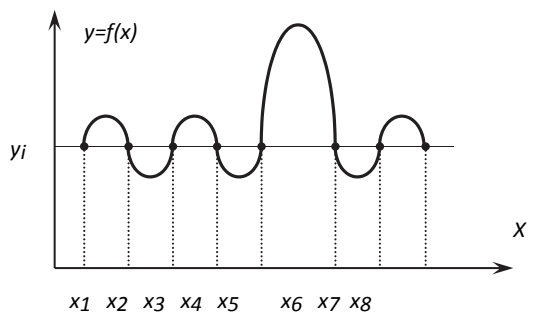


Рис.1.14. — Побудова інтеграла Лебега: $y_i = f(x_1), y_i = f(x_2), \dots, y_i = f(x_8)$,
 тобто $\{x_1, x_2, \dots, x_8\}$ — множина, образ якої є y_i ;
 тобто це його повний прообраз

Інтеграл вводять спочатку для простої функції $f(x)$, яка набуває не більш як зчисленну множину різних значень:

$$y_1, y_2, \dots, y_n; (y_i \neq y_j)$$

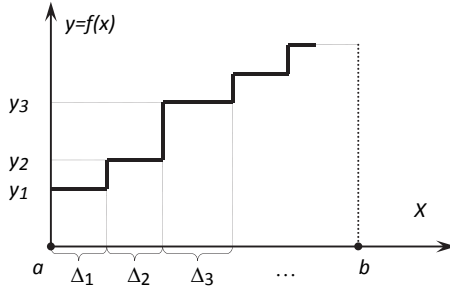


Рис.1.15. — $f(x) = \{y_1, y_2, y_3, \dots\}$ — проста функція, $\mu\{f^{-1}(y_i)\} = \mu(\Delta_i)$;

У даному випадку $\mu(\Delta_i) = \Delta_i$, де $\Delta_i = b_i - a_i$; тобто

$$\int_{[a,b]} f(x) d\mu = \sum_i y_i \mu(\Delta_i)$$

Тоді за означенням:

$$\int_A f(x) d\mu = \sum_n y_n \mu(A_n), \quad (1.6)$$

де A — вимірна множина, а

$$A_n = \{x : x \in A, f(x) = y_n\} -$$

підмножина A , образ якої є y_n .

Функцію f називають *інтегрованою на множині A за мірою μ* , якщо існує послідовність простих інтегрованих за мірою функцій $\{f_n\}$, які рівномірно збігаються на A до f . Границю

$$\lim_{n \rightarrow \infty} \int_A f_n(x) d\mu = \int_A f(x) d\mu \quad (1.7)$$

називають *інтегралом функції f на множині A* . Окремим випадком інтеграла за мірою є інтеграл Стілтєса, в якому міра $m([x_1, x_2])$ задається приростом неспадної функції $\Delta\varphi = \varphi(x_2) - \varphi(x_1)$.

Поняття інтеграла Лебега узагальнює поняття інтеграла Рімана. Якщо для деякої функції інтеграл Рімана існує, то він збігається з інтегралом Лебега. Інтеграл Лебега-Стілтєса узагальнює відповідне поняття інтеграла Рімана-Стілтєса.

1.3. Функціональні простори

Простір — це множина елементів, в якій певним чином визначено поняття збіжності послідовності елементів.

Елементи множин у математиці не мають ніяких фізичних ознак – вони абстрактні. Тому, розглядаючи використання ідей функціональних просторів у електров'язку, можна буде надавати елементам цих просторів різне розуміння – детерміновані сигнали, випадкові величини, випадкові сигнали тощо. Але це ми зробимо далі у відповідних підрозділах.

Розглянемо три типи функціональних просторів – *лінійний метричний простір, нормований простір, гільбертів простір*.

1.3.1. Лінійні метричні простори

Множину X елементів x, y, z, \dots називають *лінійним*, або *векторним простором*, якщо вона задовольняє такі умови:

1. Для двох її елементів $x, y \in X$ однозначно визначено третій елемент $z \in X$, який називають їх сумою і позначають $x + y$, причому

- а) $x + y = y + x$ – комутативність;
- б) $(x + y) + z = x + (y + z)$ – асоціативність;

в) у X існує такий елемент 0 , що $x + 0 = x$ для всіх $x \in X$, (тобто існує нульовий елемент);

г) для кожного $x \in X$ існує такий елемент $-x \in X$ («мінус ікс»), що $x + (-x) = 0$ (тобто існує протилежний елемент).

2. Для кожного числа α і будь-якого елемента $x \in X$ визначено елемент $\alpha x \in X$ (добуток елемента x на число α), причому:

- а) $\alpha(\beta x) = \alpha\beta x$;
- б) $1 \cdot x = x$;
- в) $(\alpha + \beta)x = \alpha x + \beta x$;
- г) $\alpha(x + y) = \alpha x + \alpha y$.

Елементи лінійного простору називають також *векторами*. Вектори x_1, \dots, x_n називають *лінійно незалежними*, якщо рівність

$$\sum_{j=1}^n \alpha_j x_j = 0$$

виконується тоді і тільки тоді, коли всі $\alpha_j = 0$. Їх називають також *базисними векторами простору*.

Вектори x_1, \dots, x_n називають *лінійно залежними*, якщо принаймні один із них є лінійною комбінацією решти:

$$x_k = \sum_{\substack{j=1 \\ j \neq k}}^n \alpha_j x_j.$$

Метричним простором називають абстрактну множину, для довільних елементів x і y якої визначено функцію $\rho(x,y)$ – так звану *відстань*, що задовольняє умови:

а) $\rho(x,y) \geq 0$, причому $\rho(x,y) = 0$ тоді й тільки тоді, коли $x = y$ (невід'ємність відстані);

б) $\rho(x,y) = \rho(y,x)$ – симетрія відстані;

в) $\rho(x,z) \leq \rho(x,y) + \rho(y,z)$ – нерівність трикутника для відстані.

У практичних ситуаціях звичайно розглядають лінійні метричні простори, елементи яких можна додавати та множити на числа.

Впроваджену відстань з вказаними властивостями називають *метрикою*. Властивості відстані поміж абстрактними елементами ті ж самі, що й властивості відстані поміж точками фізичного тривимірного простору. Це дає можливість надати геометричну інтерпретацію абстрактним елементам, а далі – тим фізичним об'єктам у відповідності з фізичним розумінням цих абстрактних елементів, яке ми будемо далі їм надавати.

У метричному просторі елементи x, y геометрично можна трактувати як точки або вектори, що йдуть з початку координат до цих точок; можна розглядати відстань поміж цими точками.

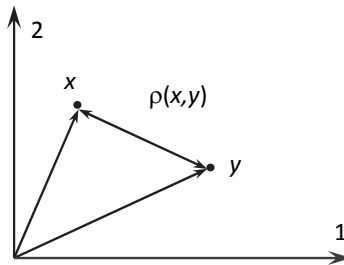


Рис. 1.16. – Геометричне зображення елементів метричного простору

Приклади метричних просторів

1. Простір $L_2(S)$.

Елементами множини X є комплекснозначні функції $x(S)$, задані на деякій множині S з мірою μ на σ -алгебрі її підмножин. Вважають,

що функції $x(S)$ є функціями з інтегрованим у розумінні Лебега квадратом, тобто

$$\int_S |x(S)|^2 d\mu < \infty.$$

Фізично (у електровз'язку, наприклад), $L_2(S)$ – це простір детермінованих сигналів зі скінченною енергією. Відстань вводиться так:

$$\rho(x, y) = \sqrt{\int_S |x(S) - y(S)|^2 d\mu}. \quad (1.9)$$

2. Простір l_2 .

Множиною X є множина послідовностей чисел:

$$x_1 = (x_{11}, x_{12}, \dots, x_{1j}, \dots),$$

$$x_2 = (x_{21}, x_{22}, \dots, x_{2j}, \dots),$$

які задовольняють умову

$$\sum_{j=1}^{\infty} |x_{sj}|^2 < \infty.$$

Відстань визначається формулою:

$$\rho(x_1, x_2) = \sqrt{\sum_{j=1}^{\infty} |x_{1j} - x_{2j}|^2}, \quad (1.10)$$

яка також, як можна довести, задовольняє умови метрики. У електровз'язку такий простір фізично трактується як простір дискретизованих детермінованих сигналів.

У частковому випадку, коли

$$x_1 = (x_{11}, \dots, x_{1n}),$$

$$x_2 = (x_{21}, \dots, x_{2n})$$

маємо для дійсних елементів

$$\rho(x_1, x_2) = \sqrt{\sum_{j=1}^n [x_{1j} - x_{2j}]^2}, \quad (1.10a)$$

що визначає відстань у просторі R^n . Це n -вимірний підпростір простору l_2 дійсних елементів. З іншого боку R^n є прямим узагальненням тривимірного евклідового простору. Покладаючи $n=3$, одержуємо характерну для нього відстань між точками.

Простір R^n має назву n -вимірний евклідовий простор. Розглядають як простір R^n дійсних, так і комплексних елементів, для яких формула (1.10а) набуває вигляд:

$$\rho(x_1, x_2) = \sqrt{\sum_{j=1}^n |x_{1j} - x_{2j}|^2}. \quad (1.10б)$$

У наведених прикладах $L_2(S)$ та l_2 існують зчисленні множини лінійно незалежних елементів, які називаються *базисами*. Такі простори називають *сепарабельними*. Якщо у просторі не існує зчисленного базису, його називають *несепарабельним*.

Простір $L_2(S)$ може бути також несепарабельним. У останньому випадку відстань уводиться інакше.

3. Простір $B_2(S)$.

Множина X є множиною комплекснозначних функцій $x(S)$, заданих на числовій осі $(-\infty, +\infty)$. Введена міра $\mu(S)$. Функції задовольняють умову:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(S)|^2 d\mu(S) < \infty.$$

Відстань задається формулою:

$$\rho(x, y) = \sqrt{\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(S) - y(S)|^2 d\mu(S)}. \quad (1.11)$$

Це простір $B_2(S)$ з метрикою Бора – Безиковича. Він несепарабельний. Іноді він теж позначається як $L_2(S)$.

1.3.2. Нормовані простори

Функціональний простір називають *нормованим*, якщо він лінійний і для кожного елемента $x \in X$ уведено функцію – норму, яка задовольняє такі умови:

а) $\|x\| > 0$, причому $\|x\| = 0$ тоді й тільки тоді, коли $x = 0$;

б) $\|\lambda x\| = |\lambda| \cdot \|x\|$;

в) $\|x + y\| \leq \|x\| + \|y\|$ (нерівність трикутника для норми).

Норма узагальнює поняття довжини вектора. Усякий нормований простір стає метричним, якщо ввести відстань як:

$$\rho(x, y) = \|x - y\|.$$

Збіжність елементів $x_1, x_2, \dots, x_n, \dots$, $\|x_n - x\| \xrightarrow{n \rightarrow \infty} 0$ називають *сильною*, або *збіжністю за нормою*. При цьому $\rho(x_n, x) \xrightarrow{n \rightarrow \infty} 0$.

Елементи нормованого простору геометрично розглядають як точки або вектори; можна обчислювати їх «довжину» – норму, а також визначати відстань між цими елементами.

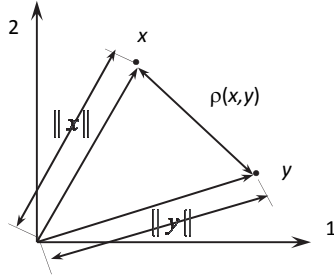


Рис.1.17. – Геометричне зображення елементів нормованого простору

Приклади нормованих просторів

1. Простір $L_2(S)$.

У цьому просторі норму визначають так:

$$\|x\| = \sqrt{\int_S |x(S)|^2 d\mu}. \quad (1.12)$$

Тоді відстань вводиться як:

$$\rho(x, y) = \|x - y\| = \sqrt{\int_S |x(S) - y(S)|^2 d\mu},$$

що збігається з уведеною раніше в $L_2(S)$ формулою (1.9).

2. Простір l_2 .

Тут норму задають формулою:

$$\|x_S\| = \sqrt{\sum_{j=1}^{\infty} |x_{Sj}|^2}, \quad (1.13)$$

тобто відстань

$$\rho(x_1, x_2) = \|x_1 - x_2\| = \sqrt{\sum_{j=1}^{\infty} |x_{1j} - x_{2j}|^2},$$

що збігається з формулою (1.10).

3. Простір B_2 .

Норму задають формулою:

$$\|x\| = \sqrt{\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(S)|^2 d\mu(S)}, \quad (1.14)$$

звідси:

$$\rho(x, y) = \|x - y\| = \sqrt{\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(S) - y(S)|^2 d\mu(S)},$$

що збігається з формулою (1.11).

Простір B_2 у електрозв'язку трактують як простір детермінованих сигналів, заданих на всій осі, які мають скінченну середню потужність.

1.3.3. Гільбертів простір

Нескінченновимірний комплексний лінійний простір H називають гільбертовим, якщо кожній парі елементів x і y із H поставлено у відповідність комплексне число (x, y) – так званий скалярний добуток, що задовольняє умови:

а) $(x, x) \geq 0$, причому $(x, x) = 0$ тоді й тільки тоді, коли $x = 0$;

б) $(x, y) = \overline{(y, x)}$;

в) $(x_1 + x_2, y) = (x_1, y) + (x_2, y)$;

г) $(\lambda x, y) = \lambda(x, y)$.

Властивості в) і г) називають *лінійністю* скалярного добутку.

Розглядають сепарабельні та несепарабельні гільбертові простори. У сепарабельному випадку у означенні гільбертова простора додатково покладається, що існує зчисленний базис.

Усякий гільбертів простір стає *нормованим*, якщо норму задати як

$$\|x\| = \sqrt{(x, x)},$$

і *метричним*, коли:

$$\rho(x, y) = \|x - y\| = \sqrt{(x - y, x - y)}.$$

Нульові елементи гільбертового простору називають ортогональними, якщо їх скалярний добуток дорівнює нулю: $(x, y) = 0$.

Якщо простір H дійсний, окрім того вводять кут між векторами $\langle x, y \rangle$, косинус якого визначається через скалярний добуток:

$$\cos \langle x, y \rangle = \frac{(x, y)}{\|x\| \cdot \|y\|}.$$

Елементи гільбертова простору можна геометрично трактувати як точки або вектори. Можна обчислювати їх норму, відстань між елементами і, окрім цього, знаходити для дійсних елементів кут $\varphi = \langle x, y \rangle$ між векторами.

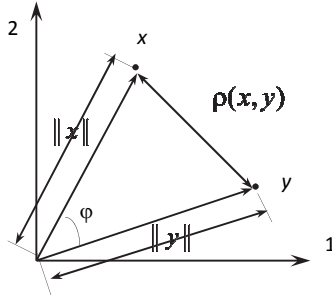


Рис.1.18. – Геометричне зображення елементів гільбертового простору.

У гільбертовому просторі H вводять узагальнений ряд Фур'є для елементів $x \in H$ за довільною повною ортонормованою системою

$$\{\psi_k\}_{k=1}^{\infty}, \psi_k \in H:$$

$$x = \sum_{k=1}^{\infty} C_k \psi_k, \quad (1.16)$$

де $C_k = (x, \psi_k)$ – (1.17)

узагальнений коефіцієнт Фур'є елемента x за ортонормованим базисом $\{\psi_k\}_{k=1}^{\infty}$, $(\psi_k, \psi_n) = \delta_{kn}$;

$$\delta_{kn} = \begin{cases} 1, & k=n; \\ 0, & k \neq n. \end{cases} \text{ – символ Кронекера.}$$

Збіжність елементів гільбертового простору $\|x_n - x\| \rightarrow 0$ називають збіжністю у середньоквадратичному. З неї, як із збіжності «в середньому» для простору Банаха L_1 , випливає збіжність за мірою:

$$\left. \begin{array}{l} x_n \xrightarrow{\text{сеп.}} x \\ x_n \xrightarrow{\text{м.скр.}} x \end{array} \right\} \Rightarrow x_n \xrightarrow{\mu} x. \quad (1.18)$$

У загальному випадку зі збіжності в середньоквадратичному не впливає збіжність майже скрізь, і навпаки.

Приклади гільбертових просторів

1. Простір $L_2(S)$.

Скалярний добуток, норма і відстань визначаються так:

$$(x, y) = \int_S x(S) \overline{y(S)} d\mu;$$

$$\|x\| = \sqrt{(x, x)} = \sqrt{\int_S |x(S)|^2 d\mu}; \quad (1.19)$$

$$\rho(x, y) = \|x - y\| = \sqrt{\int_S |x(S) - y(S)|^2 d\mu}.$$

Простір позначають L_2 , $L_2(S)$, $L_2(S, \mu)$, або найповніше $L_2(S, \mathfrak{F}, \mu)$, де \mathfrak{F} – відповідна σ -алгебра підмножин S .

Узагальнений ряд Фур'є у просторі L_2 :

$$x(S) = \sum_{k=1}^{\infty} C_k \psi_k(S),$$

де $C_k = (x, \psi_k) = \int_S x(S) \overline{\psi_k(S)} d\mu;$ (1.20)

$$(\psi_k, \psi_n) = \int_S \psi_k(S) \overline{\psi_n(S)} d\mu = \delta_{kn}.$$

2. Простір l_2 .

Скалярний добуток, норма і відстань визначаються відповідно за формулами:

$$(x_1, x_2) = \sum_{k=1}^{\infty} x_{1k} \overline{x_{2k}};$$

$$\|x\| = \sqrt{(x, x)} = \sqrt{\sum_{j=1}^{\infty} |x_j|^2}; \quad (1.21)$$

$$\rho(x_1, x_2) = \|x_1 - x_2\| = \sqrt{\sum_{j=1}^{\infty} |x_{1j} - x_{2j}|^2}.$$

Узагальнений ряд Фур'є у просторі l_2 має вигляд:

$$\begin{aligned}
 x &= \sum_{k=1}^{\infty} C_k \Psi_k ; \\
 C_k &= (x, \Psi_k) = \sum_{j=1}^{\infty} x_j \overline{\Psi_{kj}} ; \\
 (\Psi_k, \Psi_n) &= \sum_{j=1}^{\infty} \Psi_{kj} \overline{\Psi_{nj}} = \delta_{kn} .
 \end{aligned} \tag{1.22}$$

3. Простір B_2 .

Тут скалярний добуток, норму і відстань уводять так:

$$\begin{aligned}
 (x, y) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(S) y(S) d\mu ; \\
 \|x\| &= \sqrt{(x, x)} = \sqrt{\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(S)|^2 d\mu} ; \\
 \rho(x, y) &= \|x - y\| = \sqrt{\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(S) - y(S)|^2 d\mu} .
 \end{aligned} \tag{1.23}$$

Розглянемо тепер приклади більш складних гільбертових просторів, так званих *гільбертових просторів над гільбертовими просторами*. Вони бувають сепарабельні й несепарабельні.

4. Простір $L_2(S, L_2(F))$.

Множина X є множиною функцій $x(s, f)$, заданих на S , які приймають значення із гільбертова простору $L_2(F)$. Ці функції задовольняють умову:

$$\int_S \int_F |x(s, f)|^2 d\mu_1(s) d\mu_2(f) < \infty .$$

Скалярний добуток, норма і відстань уводяться відповідно за формулами:

$$\begin{aligned}
 (x, y) &= \int_S \int_F x(s, f) \overline{y(s, f)} d\mu_1(s) d\mu_2(f) ; \\
 \|x\| &= \sqrt{\int_S \int_F |x(s, f)|^2 d\mu_1(s) d\mu_2(f)} ;
 \end{aligned} \tag{1.24}$$

$$\rho(x, y) = \sqrt{\int_S \int_F |x(s, f) - y(s, f)|^2 d\mu_1(s) d\mu_2(f)}.$$

Такий простір називають гільбертовим простором над гільбертовим простором. Простір $L_2(F)$ називають основним. Коли основний простір сепарабельний простір $L_2(S, L_2(F))$ теж сепарабельний.

У теорії електрозв'язку такий простір фізично трактують як простір випадкових сигналів зі скінченною енергією.

5. Простір $L_2(S, B_2(F))$

У цьому випадку множина X є множиною відображень (функцій), заданих на S , які приймають значення із несепарабельного гільбертова простору $B_2(F)$. Функції задовольняють умову:

$$\int_F \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(s, f)|^2 d\mu_1(s) d\mu_2(f) < \infty.$$

Скалярний добуток, норма і відстань уводяться так:

$$(x, y) = \int_F \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(s, f) \overline{y(s, f)} d\mu_1(s) d\mu_2(f);$$

$$\|x\| = \sqrt{\int_F \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(s, f)|^2 d\mu_1(s) d\mu_2(f)}; \quad (1.25)$$

$$\rho(x, y) = \sqrt{\int_F \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(s, f) - y(s, f)|^2 d\mu_1(s) d\mu_2(f)}.$$

Цей простір теж називають гільбертовим простором над гільбертовим простором; на відміну від попереднього прикладу він несепарабельний (тому що основний простір B_2 несепарабельний).

У теорії електрозв'язку такий простір фізично трактують як простір випадкових сигналів зі скінченною середньою потужністю.

1.4. Аксиоматизація А.М. Колмогорова теорії ймовірностей

Говорять, що експеримент поставлено (або він відбувся без участі експериментатора) завжди, коли реалізується певний комплекс натурних умов, який спричинює появу одного з наслідків такого експерименту. Експеримент називають статистичним, якщо цей комплекс умов приводить до непередбачуваного наслідку з наперед (апріорі)

відомої сукупності всіх наслідків цього експерименту. Хоча сама поява того чи іншого результату є випадковою в звичайному розумінні цього слова, у застосовуваному математичному апараті, який описує експеримент, усі об'єкти не є випадковими.

В аксіоматизованій А.М. Колмогоровим теорії ймовірностей подано теоретико-множинну інтерпретацію об'єктів і притаманних їм операцій та співвідношень. (таб.1.1).

Таблиця 1.1.

Відповідність між термінами теорії ймовірностей та теорії вимірних функцій і просторів з мірою

<i>Терміни теорії вимірних функцій і просторів з мірою.</i>	<i>Терміни теорії ймовірностей.</i>
1.	2.
Простір з нормованою мірою $(\Omega, \mathcal{F}, \mu)$	Ймовірнісний простір (Ω, \mathcal{F}, P)
Точка ω , яка належить простору Ω .	Елементарний наслідок (подія) ω .
Вимірні множини A, B, \dots	Події A, B, \dots
Простір Ω .	Вірогідна подія.
Порожня множина \emptyset	Неможлива подія.
$A \subset B$ – множина A міститься в B .	Із події A випливає B .
$A \cup B$ – об'єднання множин A і B .	Подія, яка полягає в тому що відбудеться A або B .
$A \cap B$ – переріз множин A і B .	Подія, яка полягає в тому, що відбудеться і A , і B .
$A \setminus B$ – різниця множин A і B .	Подія, яка полягає в тому, що відбудеться A , але не відбудеться B .
$\bar{A} = \Omega \setminus A$ – доповнення до множини A .	\bar{A} – подія, протилежна A .
$A \cap B = \emptyset$ – множини A і B не мають спільних точок.	Події A і B несумісні.
Нормована міра m .	Імовірність P (імовірнісна міра).
Майже скрізь.	Майже напевне (з імовірністю 1)
Вимірна функція.	Випадкова величина.
Інтеграл за мірою $\int (\circ) d\mu$.	Математичне сподівання M .

Відповідно до табл.1.1. математичною формалізацією статистичного експерименту є ймовірнісний простір (Ω, \mathcal{F}, P) . Випадкова величина – це вимірна функція, що відображає множину елементарних наслідків Ω на дійсну вісь. Математичне сподівання випадкової величини – це інтеграл вимірної функції за мірою. Загалом, будь-які ситуації, розглядувані в теорії ймовірностей, можна виразити в термінах просторів з мірою та вимірних функцій.

Величезним значенням для розвитку теорії ймовірностей має той факт, доказаний А.М. Колмогоровим, що вся теорія може бути побудована лише на п'яти дуже простих аксіомах, а якщо система \mathfrak{T} множин кінцева, то навіть на чотирьох аксіомах. Наведемо їх:

Нехай Ω – множина елементів ω – елементарних подій;

F – множина підмножин із Ω , випадкових подій.

Тоді може бути введена система аксіом, що не мають між собою протиріч:

1. F є алгебра множин;
2. Кожній множині A з F поставлене у відповідність невід'ємне дійсне число $P(A)$. Це число зветься імовірністю події A ;
3. $P(\Omega)=1$;
4. Якщо A та B не перетинаються $A \cap B = \emptyset$, то $P(A+B) = P(A) + P(B)$;
5. Для кінцевих F аксіома 5 витікає із аксіом 1-4: для неспадної послідовності $A_1 \supseteq A_2 \supseteq \dots \supseteq A_n \supseteq \dots$, подій із F такої, що $\bigcap_n A_n = \emptyset$, має місце рівність $\lim_n A_n = 0$.

Скориставшись теоретико-ймовірнісною термінологією, співвідношення між різними типами збіжностей, наведені у підрозділах 1.2.2 і 1.1.3, можна виразити мовою теорії ймовірностей:

$$\begin{array}{ccc}
 \xi_n \xrightarrow{сер.} \xi & \Downarrow & \\
 & \xi_n \xrightarrow{P} \xi & \\
 \xi_n \xrightarrow[\substack{м.н. \\ P=1}]{} \xi & \Uparrow &
 \end{array} \tag{1.26}$$

де скорочення *сер.*, *м.н.* – ($P=1$), P означають збіжність у середньому, майже напевне (з імовірністю 1), та за імовірністю.

Висновки

1. Сучасна теорія ймовірностей базується на теорії множин, міри, вимірних функцій та функціональному аналізі.

2. Основними для потреб теорії ймовірностей у теорії множин є поняття множини, теоретико-множинні операції (об'єднання, різниця, переріз), класи множин, замкнених відносно теоретико-множинних операцій (σ -кільце, σ -алгебра).

3. У теорії вимірних функцій найважливішими є поняття міри (яке узагальнює поняття довжини відрізка, площі плоскої фігури, об'єму просторової фігури), поняття інтеграла за мірою, зокрема інтеграла Лебега (який узагальнює поняття інтеграла Рімана), а також поняття вимірної функції (клас яких розширює клас неперервних функцій). Ці поняття розширюють відомі з попередніх розділів математики й збігаються з ними у більш простих випадках.

4. У функціональному аналізі найбільш важливими є поняття та конструкція функціональних просторів – лінійного метричного, нормованого та гільбертова, які дають можливість узагальнити геометричні поняття «точка», «вектор», «довжина вектора», «відстань між точками», «кут між векторами» на елементи абстрактних множин.

5. У аксіоматизованій Колмогоровим теорії ймовірностей встановлюється відповідність понять теорії ймовірностей та теорії множин, вимірних функцій і функціональних просторів з мірою.

Запитання, задачі й вправи

1. Дайте означення основних теоретико-множинних операцій – об'єднання, різниці, перерізу множин.

а) Довести: із $A \subset B$ випливає, що $A \cap B = A$ та $A \cup B = B$;

б) перевірити: $A \Delta B = (A \cap \bar{B}) \cup (B \cap \bar{A})$;

в) довести: $A \Delta A = \emptyset$ та $A \Delta \emptyset = A$;

г) визначити й зобразити на рисунках множини $A \cup B$, $A \cap B$, $A \setminus B$, $B \setminus A$, якщо $A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$, $B = \{(x, y) \in \mathbb{R}^2 : xy \geq 0\}$.

2. Що таке кільце і алгебра, σ -кільце і σ -алгебра?

а) Перелічити різні класи типу кільця і алгебри, які можна побудувати з підмножин триелементної множини $X = \{a, b, c\}$ і порожньої множини \emptyset .

3. Які множини називають зчисленими та незчисленими? Наведіть приклади.

а) Довести, що коли відстань між двома точками множини E на прямій більша від одиниці, то множина E скінченна та зчисленна.

4. Дайте означення міри. Що таке міра Лебега і як вона будується? Що таке вимірна функція? Назвіть достатній критерій вимірності.

а) Показати, що міра порожньої множини дорівнює нулю: $(\mu(\emptyset)=0)$.

б) Показати, що функція $\mu[(a,b))=b-a$ є мірою на класі одновимірних пів інтервалів.

в) Довести, що функція $|f|$ вимірна, якщо вимірна f .

г) Нехай функція f вимірна й не перетворюється на нуль. Довести, що $1/f$ є вимірною.

5. Що таке інтеграл за мірою? Як будується інтеграл Лебега?

6. Що таке лінійний метричний простір? Наведіть приклади.

а) Чи є множина дійсних чисел метричним простором, якщо відстань між елементами цієї множини визначено так: $\rho(x,y)=\sqrt{|y-x|}$?

7. Що таке нормований простір? Наведіть приклади.

а) Показати, що умова лінійної незалежності системи з N ортогональних на $[0,T]$ функцій $\varphi_i(t)$ не виконується, якщо в системі знайдеться принаймні одна функція з нульовою енергією.

8. Що таке гільбертів простір? Наведіть приклади.

а) Показати, що для скалярного добутку у гільбертовому просторі виконується співвідношення: $(x,\lambda y)=\bar{\lambda}(x,y)$.

б) Визначити скалярний добуток сигналів: $s_1(t)=a_1 \cos 2\pi f_1(t)$ і $s_2(t)=a_2 \cos 2\pi f_2(t)$, заданих на інтервалі T , за умови $f_1 + f_2 > f_1 - f_2 = \Delta f$. Знайти норми сигналів $s_1(t)$ і $s_2(t)$.

9. Перевірити виконання аксіом, наведених А.М. Колмогоровим для прикладу $\Omega=\omega$, $\mathfrak{F}=\{\omega,\emptyset\}$, $P(\Omega)=1$, $P(\emptyset)=0$.

2. КЛАСИЧНА ТЕОРІЯ ЙМОВІРНСТЕЙ

Теорія ймовірностей має справу з «випадковістю». Але вона вивчає не будь-яку випадковість, теорія має справу з випадковостями, які характеризуються стохастичною стійкістю – при багатократному повторенні у них проявляються відповідні закономірності.

Важливим поняттям теорії ймовірностей є статистичний експеримент. У сучасній теорії його математична модель будується на основі теорії множин та функціональних просторів з мірою, основні поняття яких викладені у попередньому розділі. У цьому розділі на такій математичній основі з використанням ідеї аксіоматизації теорії ймовірностей за Колмогоровим розглядаються випадкові події, випадкові величини та вектори, охоплюються питання закону великих чисел та граничних теорем теорії ймовірностей [5,6].

Вказані питання є класичними для теорії ймовірностей. Звичайно при їх викладанні даються ілюстративні приклади з різних галузей практики. На відміну від відомої літератури у цій книзі мотивація теорії, тлумачення понять та вкладання фізичного розуміння у результати теорії береться з галузі теорії електров'язку [3].

2.1. Випадкові події

2.1.1. Імовірнісний простір

Розглядається абстрактна множина Ω – вибірковий простір.

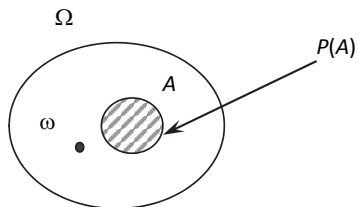


Рис. 2.1. – Вибірковий простір: Ω – абстрактна множина; ω – елементарна подія; A – подія; $P(A)$ – ймовірність події A

Його елемент $\omega \in \Omega$ – це елементарна подія, або елементарний наслідок.

Підмножина A множини Ω – це подія. Розглядається клас подій – σ -алгебра \mathcal{F} . Цьому класові разом з подією A належить протилежна подія $\bar{A} = \Omega \setminus A$, неможлива подія – \emptyset , а також вірогідна подія Ω . Згідно

з означенням σ -алгебри для яких завгодно події з класу F можна виконувати операції об'єднання (\cup), знаходження їх різниці (\setminus) та перетину (\cap) – результати завжди належатимуть класові; більш того, операції об'єднання та перетину можна виконувати зчисленну кількість разів. Нарешті, для множин з класу F вводиться міра P – функція множин, яка є невід'ємною, монотонною, σ -адитивною функцією; додатково вимагається, щоб вона була нормованою, тобто:

- а) $P(A) \geq 0$;
- б) $P(A) \leq P(B)$, якщо $A \subset B$;
- в) $P(A \cup B) = P(A) + P(B)$, якщо події несумісні, тобто $A \cap B = \emptyset$;
- г) $P(\Omega) = 1$.

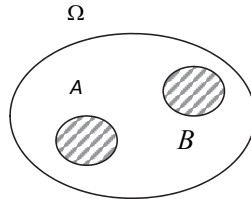


Рис. 2.2. – Вибірковий простір: несумісні події: $A \cap B = \emptyset$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i), A_i \cap A_j = \emptyset. \quad (2.1)$$

«Трійка» (Ω, F, P) згідно з аксіоматизацією за Колмогоровим теорії ймовірностей є математичною моделлю стохастичного експерименту.

Принципово неможливо розв'язувати будь-які задачі, пов'язані зі статистичним експериментом, якщо явно чи неявно не введена його математична модель – (Ω, F, P) .

Ця «трійка» носить назву ймовірнісного простору.

Приклад 1:



Рис 2.3. – Джерело двійкових повідомлень

Джерело двійкових повідомлень формує послідовність одиниць та нулів. Однократний експеримент полягає у тому, що формується (передається) один із двох символів – «1» або «0», які випадково вибираються. Необхідно побудувати модель цього експерименту.

Згідно з особливостями задачі вибіркового простір складається з двох елементарних подій – «1» або «0»:

$$\Omega = \{\langle 0 \rangle, \langle 1 \rangle\}. \quad (2.2)$$

Алгебра подій у даному випадку має 4 елементи:

$$F = \{\emptyset, \{\langle 0 \rangle\}, \{\langle 1 \rangle\}, \{\langle 0 \rangle, \langle 1 \rangle\}\}, \quad (2.3)$$

тобто сюди входить неможлива подія \emptyset , 2 елементарні множини $\{\langle 0 \rangle\}$ і $\{\langle 1 \rangle\}$ як події, та одиниця класу – увесь вибіркового простір $\Omega = \{\langle 0 \rangle, \langle 1 \rangle\}$, як вірогідна подія.

Для повного завдання моделі однократного експерименту необхідно задати ймовірності передання одиниці – $P(1)$ та нуля – $P(0)$. Тоді з властивості (2.1) ймовірності можна знайти ймовірність інших множин алгебри F . Так, маємо

$$P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset),$$

тобто $P(\emptyset) = 0$; а також

$$P(\langle 0 \rangle \cup \langle 1 \rangle) = P(\Omega) = P(0) + P(1) = 1. \quad (2.4)$$

Таким чином, побудовано ймовірнісний простір (Ω, F, P) , де Ω – вибіркового простір (2.2), F – алгебра (2.3), ймовірнісна міра на F : $\{P(\emptyset), P(0), P(1), P(\Omega)\}$ задається відповідно з (2.4).

Приклад 2:

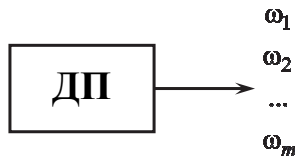


Рис 2.4. – Джерело m -кових повідомлень

Джерело повідомлень у цьому випадку формує (передає) послідовність, кожний елемент якої випадково вибирається із m символів. Необхідно побудувати модель однократного експерименту, який полягає у тому що формується один з m можливих символів.

У цьому випадку вибіркового простір складається із m елементарних подій ω_i :

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

Алгебра F має 2^m елементів, куди входять усі підмножини множини Ω , неможлива подія \emptyset та вірогідна подія, тобто входить Ω як одиниця класу. Наприклад, коли $m=3$, маємо:

$$F = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_3\}\}.$$

Для побудови ймовірносної міри на класі F необхідно задати ймовірності передання кожного з m можливих символів — $P(\omega_j)$, $j=\overline{1, m}$. Тоді, аналогічно попередньому прикладу, використовуючи властивості ймовірності, можна обчислити ймовірності усіх інших подій з алгебри F .

«Трійка» (Ω, F, P) — ймовірнісний простір — повністю описує даний статистичний експеримент.

Наведені приклади показують, що у розглянутих випадках джерела дискретних повідомлень для завдання ймовірнісного простору достатньо задати множину елементарних подій ω_j , $j=\overline{1, m}$ та їх

ймовірності $P(\omega_j)$, $j=\overline{1, m}$. У теорії електровз'язку цю пару $\{\omega_j, P_j\}_{j=1}$ називають ансамблем дискретних повідомлень. Його завдання еквівалентне завданню ймовірнісного простору, який розглядається у теорії ймовірностей.

2.1.2. Модель незалежних випробувань

Незалежними називаються випробування, якщо результат першого випробування не впливає на результат другого; взагалі, результат попереднього випробування не впливає на результат наступного.

Приклад 3:

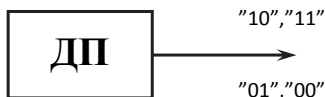


Рис. 2.5.— Джерело двійкових повідомлень при дворовому випробуванні

Джерело двійкових повідомлень послідовно два рази випадково вибирає повідомлення «0» чи «1». При цьому формується одне з повідомлень «00», «01», «10», «11». Тобто при першому випробуванні фор-

мується «0» чи «1», а при другому – незалежно від результату першого теж передається «0» чи «1».

Для формалізації цього більш складного статистичного експерименту необхідно побудувати вибірковий простір незалежних випробувань, ввести на ньому алгебру (у загальному випадку – σ -алгебру подій), а також задати ймовірнісну міру на цьому класі подій. Стисло кажучи – необхідно задати модель цього статистичного експерименту у вигляді відповідного ймовірнісного простору.

Дамо відповідні означення.

Прямим добутком просторів елементарних подій $\Omega_1 \times \Omega_2$ називається множина упорядкованих пар:

$$(\omega_k^{(1)}, \omega_j^{(2)}), \omega_k^{(1)} \in \Omega_1, \omega_j^{(2)} \in \Omega_2 : \Omega_1 \times \Omega_2 = \{(\omega_k^{(1)}, \omega_j^{(2)})\}. \quad (2.5)$$

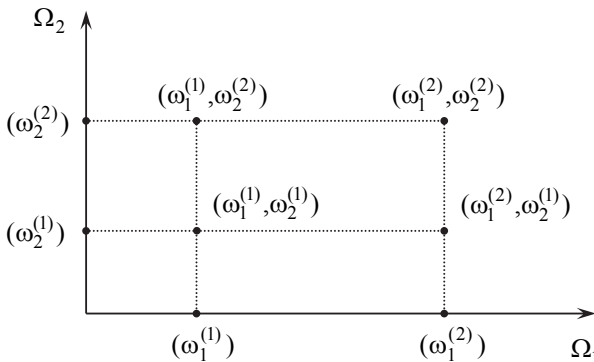


Рис. 2.6. – Прямий добуток просторів елементарних подій $\Omega_1 \times \Omega_2$

Сукупність всіх підмножин множини $\Omega_1 \times \Omega_2$, неможлива подія \emptyset та вірогідна подія $\{\Omega_1 \times \Omega_2\}$ – одиниця класу – дає, так званий, прямий добуток алгебр $F_1 \times F_2$.

Прямим добутком міри називається міра, яка задається формулою

$$P(\omega_k^{(1)}, \omega_j^{(2)}) = P(\omega_k^{(1)})P(\omega_j^{(2)}). \quad (2.6)$$

Таким чином вводиться модель незалежних випробувань – ймовірнісний простір $(\Omega_1 \times \Omega_2, F_1 \times F_2, P_1 \times P_2)$.

У межах цієї моделі за означенням вводиться поняття незалежних подій: події називаються *незалежними*, якщо ймовірність їх добутку дорівнює добутку ймовірностей окремих подій:

$$P(AB) = P(A \cap B) = P(A)P(B). \quad (2.7)$$

Введена модель двох незалежних випробувань безпосередньо узагальнюється на випадок скінченного числа незалежних випробувань.

Аналогічно в дискретному випадку з використанням відповідного математичного апарату вводиться і у загальному випадку прямий добуток просторів з мірою.

Приклад 4:

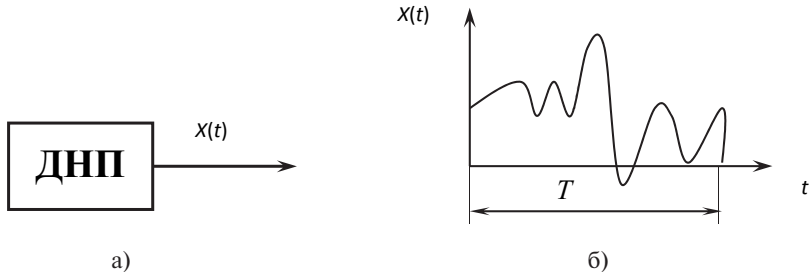


Рис 2.7. – а – Джерело неперервних повідомлень, б – реалізація повідомлень

Даний приклад показує, що у теорії електрозв'язку такі випадки достатньо типові. Тут показано джерело неперервних повідомлень (типу мовних сигналів), які розглядаються на інтервалі $t \in T$. Елементарна подія – функція часу $x(t)$. Простір елементарних подій $\Omega = \{x(t)\}$ – це, наприклад, простір R^T усіх дійсних функцій, заданих на інтервалі T . Ймовірнісний простір треба задавати на основі такого функціонального простору. Далі у загальному випадку для узагальнення описаної моделі треба відповідним чином ввести прямий добуток σ -алгебр та мір.

2.1.3. Залежні події

Події називаються *залежними*, якщо ймовірність їх добутку не дорівнює добутку ймовірностей окремих подій:

$$P(AB) \neq P(A)P(B). \quad (2.8)$$

Для обчислення ймовірності добутку подій у загальному випадку вводиться умовна ймовірність.

Розглянемо подію A та деяку фіксовану подію H (гіпотезу).

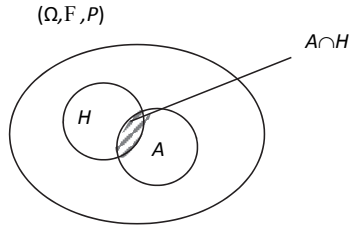


Рис. 2.8. – Звуження міри: $m(A) = P(AH)$

Звуженням (частиною) міри називається міра перерізу H і A : $\mu(A) = P(AH)$. Це, дійсно, міра, тому що виконуються властивості міри:

- а) $P(AH) \geq 0$;
- б) $P(AH) \leq P(BH)$, якщо $A \subset B$;
- в) $P\left(\bigcup_i A_i H\right) = \sum_i P(A_i H)$;
- г) $P(\Omega H) = P(H) \neq 1$.

Однак, як покаже властивість (г), міра ненормована.

Умовною ймовірністю події по відношенню до H називається нормована звужена міра:

$$P(A|H) = \frac{\mu(A)}{P(H)} = \frac{P(AH)}{P(H)}. \quad (2.9)$$

Якщо розглядати подію H за умови, що вона вже трапилась, з формули (2.9) можна обчислити ймовірність добутку подій у загальному випадку:

$$P(AH) = P(H)P(A|H). \quad (2.10)$$

Тобто ймовірність добутку подій у загальному випадку, коли вони залежні, дорівнює добутку $P(H)$ на умовну ймовірність другої події (A) по відношенню до першої (H).

Формула повної ймовірності. Нехай простір елементарних подій Ω розбито на скінченну або зчисленну множину вимірних множин $\{H_i\}_{i=1}^{\infty}$, які парами не перетинаються: $H_i \cap H_k = \emptyset$.

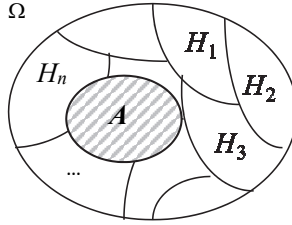


Рис. 2.9. – Повна група подій $\{H_i\}$

Кажуть, що $\{H_i\}$ утворює *повну групу подій*.

Повна група подій характеризується тим, що кожне випробування відбувається у одному і тільки одному H_i . Події H_i називаються *гіпотезами*.

За вищезгаданим малюнком знайдемо $P(A)$ через умовні ймовірності при кожній гіпотезі:

$$P(A) = P(A \cap \Omega) = P\left(A \cap \left[\bigcup_i H_i\right]\right) = P\left(\bigcup_i [A \cap H_i]\right) = P\left(\bigcup_i AH_i\right) = \sum_i P(H_i)P(A|H_i). \quad (2.11)$$

Формула $P(A) = \sum_i P(H_i)P(A|H_i)$ називається *формулою повної ймовірності*.

Формула Байєса. Нехай $\{H_i\}$ утворює повну групу подій. Задана подія A . Обчислимо ймовірність H_i за умови, що A – трапилось, тобто $P(H_i|A)$.

Згідно з означенням умовної ймовірності маємо:

$$P(H_i|A) = \frac{P(AH_i)}{P(A)}. \quad (2.12)$$

Враховуючи, що $P(AH_i) = P(H_i)P(A|H_i)$, а також формулу повної ймовірності (2.11), тобто що $P(A) = \sum_i P(H_i)P(A|H_i)$, із співвідношення (2.12) одержуємо

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_i P(H_i)P(A|H_i)}. \quad (2.13)$$

Ця формула носить назву формули Байєса. Її треба розуміти у статистичному сенсі, тобто, що випробування повторюються. Тоді $P(H_i)$ слід трактувати, як априорну ймовірність, а $P(H_i|A)$ – апостеріорну

ймовірність. Формула Байєса дає можливість перерахувати ймовірність з використанням експериментальних даних.

Формула Байєса має величезне значення у теорії електров'язку при розробці алгоритмів оптимального приймання сигналів.

2.2. Випадкові величини

Згідно з аксіоматизацією за Колмогоровим теорії ймовірностей випадкова величина позначається як вимірна функція. У технічній літературі, у тому разі і у теорії електров'язку випадкову величину раніше визначали як величину, яка приймає випадкові значення. Це вірно, але під таким означенням відсутній математичний апарат, який дає можливість розв'язувати задачі, пов'язані з випадковими величинами. Коли ж випадкова величина означається як вимірна функція, відразу під це визначення закладається глибоко розвинений математичний апарат теорії функцій та просторів з мірою. Базуючись на поняттях та властивостях математичних об'єктів з цих гілок математики, які описані у першому розділі, розглядаються способи повного завдання випадкових величин та їх числові характеристики.

2.2.1. Означення випадкової величини та можливості її повного описування

Випадковою величиною називається вимірне відображення ймовірнісного простору (Ω, \mathcal{F}, P) у дійсну вісь R^1 .

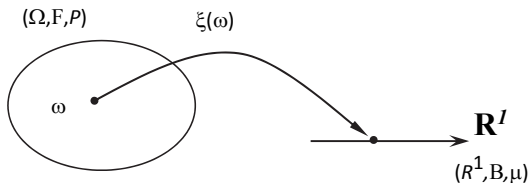


Рис. 2.10. – Випадкова величина

Випадкові величини позначаються $\xi(\omega), h(\omega), \xi, h, \dots$ або великими літерами латинської абетки X, Y, \dots

Згідно з означенням вимірного відображення крім дійсної осі R^1 необхідно ввести σ -алгебру подій. Вона задається як клас так званих *борелевських множин* \mathcal{B} : це напівінтервали типу $a < x \leq b$, їх суми, різниці, перетини та результати граничних переходів. Нарешті, треба задати міру на класі подій \mathcal{B} . Вона вводиться як індукована міра, тобто

міра $\mu(A)$, де $A \in \mathbf{B}$ – будь-яка подія з класу \mathbf{B} , визначається як ймовірність P повного прообразу множини A :

$$\mu(A) = P\{\xi^{-1}(A)\} = P\{\omega: \xi(\omega) \in A\}. \quad (2.14)$$

У теорії вимірних функцій доводиться, що замість довільної множини A у (2.14) можна взяти нескінченний напівінтервал $(-\infty, x]$; це не зменшить загальності означення. Відповідна індукована міра має назву *функції розподілу випадкової величини*:

$$F(x) = P\{\xi^{-1}(-\infty, x]\} = P\{\omega: \xi(\omega) \in (-\infty, x]\} = P\{\xi(\omega) \leq x\}. \quad (2.15)$$

Згідно з означенням (2.15) кажуть, що функція розподілу $F(x)$ – це ймовірність того, що випадкова величина не перевищує заданої величини x : ця ймовірність попадання ξ у напівінтервал $(-\infty, x]$.

Випадкова величина повністю описується своєю функцією розподілу. Вживається також інший спосіб її описання – характеристична функція (х.ф.). За означенням маємо:

$$f(u) = \int_{-\infty}^{\infty} \exp[j(u, x)] dF(x), \quad (2.16)$$

тобто це – перетворення Фур'є-Стілтьєса функції розподілу $F(x)$.

Як відомо з теорії функцій, функція $F(x)$ у загальному випадку може бути розкладена на три складові

$$F(x) = F_{ан}(x) + F_{стр}(x) + F_{синг}(x),$$

де $F_{ан}(x)$ – абсолютно неперервна функція,

$F_{стр}(x)$ – функція стрибків,

$F_{синг}(x)$ – сингулярна функція, яка є неперервною функцією, але її похідна майже скрізь дорівнює нулеві.

Якщо $F(x) = F_{ан}(x)$, можна ввести її похідну

$$\frac{dF(x)}{dx} = p(x), \quad (2.17)$$

яка називається *щільністю ймовірності*. Тоді х.ф. може бути зображена у вигляді перетворення Фур'є щільності ймовірності:

$$f(u) = \int_{-\infty}^{\infty} p(x) e^{jux} dx. \quad (2.16a)$$

Відповідно, існує зворотне перетворення Фур'є:

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(u) e^{-jux} du, \quad (2.18)$$

яке дає можливість обчислювати щільність ймовірності випадкової величини, якщо відома її характеристична функція.

Формули (2.15) і (2.16) дають «на різних мовах» – «мовах» функцій розподілу та характеристичних функцій – повне описування випадкової величини у загальному випадку.

Існують більш прості частинні випадки.

Випадкова величина називається *простою*, якщо вона приймає скінченну множину значень: $\xi(\Omega) = \{a_1, a_2, \dots, a_n\}$.

Випадкова величина називається *дискретною*, якщо вона приймає не більш, ніж зчисленну множину значень: $\xi(\Omega) = \{a_1, a_2, \dots, a_n, \dots\}$. У загальному випадку, коли вона приймає незчисленну множину значень, іноді кажуть про *неперервну випадкову величину*.

Формули (2.15) і (2.16) дають можливість описати які завгодно випадкові величини; вони можуть бути використані і у відокремлених часткових випадках. Однак у цьому разі можливі більш прості описування випадкових величин.

Приклади:

Неперервна випадкова величина з абсолютно неперервною функцією розподілу. Типовий вигляд функції розподілу і щільності ймовірностей цієї випадкової величини ілюструє цей малюнок.

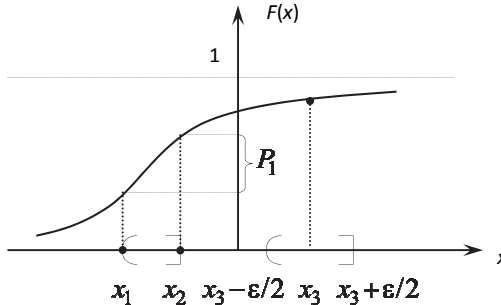


Рис. 2.11. – Функція розподілу

Функція розподілу має такі основні властивості:

- а) $F(-\infty) = 0$;
- б) $F(\infty) = 1$;
- в) $F(x)$ – неперервна справа;

$$\text{г) } F(x_2) - F(x_1) = P\{\xi \in (x_1, x_2]\} = P;$$

д) $P\{\xi = x\} = 0$, якщо x – точка неперервності функції $F(x)$.

Доведемо властивість г): ймовірність попадання випадкової величини на інтервал $(x_1, x_2]$ дорівнює різниці $F(x_2) - F(x_1)$. Справді, згідно з означенням функції розподілу, маємо

$$P\{\xi \in (x_1, x_2]\} = P\{\xi \in (-\infty, x_2]\} - P\{\xi \in (-\infty, x_1]\} = F(x_2) - F(x_1).$$

Доведемо також властивість д): ймовірність того, що $P\{\xi = x\} = 0$, якщо x – точка неперервності $F(x)$. Дійсно, візьмемо ε -окіл точки x_3 , знайдемо ймовірність попадання у цей інтервал (як це зроблено вище) та виконаємо граничний перехід, коли $\varepsilon \rightarrow 0$:

$$\lim_{\varepsilon \rightarrow 0} P\left\{\xi \in \left(x_3 - \frac{\varepsilon}{2}, x_3 + \frac{\varepsilon}{2}\right]\right\} = \lim_{\varepsilon \rightarrow 0} \left[F\left(x_3 + \frac{\varepsilon}{2}\right) - F\left(x_3 - \frac{\varepsilon}{2}\right) \right] = F(x_3) - F(x_3) = 0.$$

Знайдемо ще раз ймовірність попадання на інтервал $(x_1, x_2]$, але використовуючи замість $F(x)$ її похідну – щільність імовірності. Маємо:

$$P\{\xi \in (x_1, x_2]\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} p(x) dx,$$

тобто шукана ймовірність обчислюється як відповідна площа:

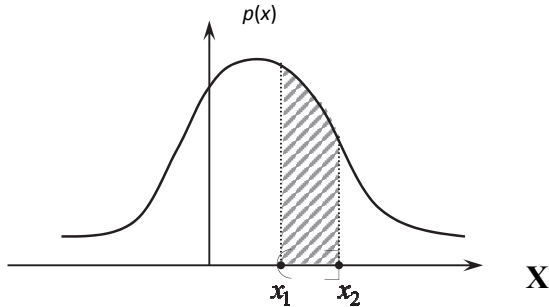


Рис. 2.12. – Щільність імовірності

Проста та дискретна випадкова величина. У цих випадках функція розподілу є функцією стрибків $F(x) = F_{\text{cmp}}(x)$:

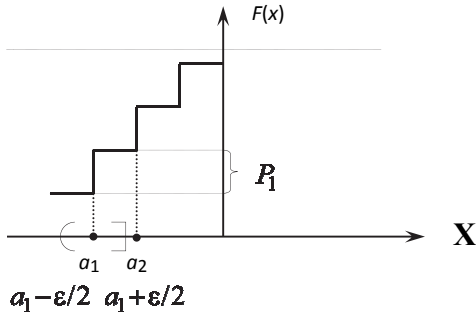


Рис. 2.13. – Функція розподілу простої та дискретної випадкової величини

Загальні властивості функції розподілу залишаються як і у попередньому випадку, але на функції $F_{cmp}(x)$ з'являються стрибки та ділянки постійності. Їх властивості такі:

- а) $P\{\xi \in (a_1, a_2)\} = 0$;
- б) $P\{\xi = a_1\} = P_1$.

Тобто, ймовірність попадання випадкової величини на ділянку постійності (без точки стрибка) дорівнює нулю; ймовірність прийняти конкретне значення (якщо це точка стрибка) дорівнює величині стрибка – функції $F_{cmp}(x)$ у цій точці. Доведемо останнє:

$$\lim_{\varepsilon \rightarrow 0} P\left\{\xi \in \left(a_1 - \frac{\varepsilon}{2}, a_1 + \frac{\varepsilon}{2}\right)\right\} = \lim_{\varepsilon \rightarrow 0} \left[F\left(a_1 + \frac{\varepsilon}{2}\right) - F\left(a_1 - \frac{\varepsilon}{2}\right) \right] = P_1.$$

Ці властивості функції розподілу дають можливість замість $F_{cmp}(x)$ використовувати інший спосіб завдання таких випадкових величин – задати точки стрибків та відповідні ймовірності: $\{a_1, P_1; a_2, P_2; \dots; a_n, P_n; \dots\}$. У теорії електров'язку таке завдання випадкової величини має назву *ансамблю* – це значення, які приймає випадкова величина та ймовірності прийняття цих значень.

2.2.2. Числові характеристики випадкових величин

При розв'язанні прикладних задач замість повного описування випадкової величини у ряду випадків достатньо знайти її окремі ха-

рактики. На математичній мові – це моменти (початкові та центральні) та кумулянти. Серед моментів особливо часто використовують початковий момент першого порядку – математичне сподівання випадкової величини, та центральний момент другого порядку – її дисперсію. Дамо відповідні означення та проаналізуємо властивості цих основних числових величин.

Означення моментів та кумулянтів випадкової величини. Згідно з аксіоматизацією за Колмогоровим теорії ймовірностей математичне сподівання випадкової величини є інтеграл за мірою від вимірної функції:

$$M\xi = M\xi(\omega) = \int_{\Omega} \xi(\omega) dP. \quad (2.19)$$

Враховуючи означення випадкової величини, математичне сподівання може бути також обчислене так:

$$M\xi = \int_{-\infty}^{\infty} x dF(x), \quad (2.20)$$

а для абсолютно неперервних функцій розподілу, для яких може бути введена щільність ймовірності, також так:

$$M\xi = \int_{-\infty}^{\infty} xp(x) dx. \quad (2.20a)$$

Початковим моментом k -го порядку називають величину

$$m_k = M[\xi]^k = \int_{\Omega} [\xi(\omega)]^k dP = \int_{-\infty}^{\infty} x^k dF = \int_{-\infty}^{\infty} x^k p(x) dx, \quad k=1,2,\dots \quad (2.21)$$

де скорочення а.н. – тільки для абсолютно неперервних функцій розподілу.

Центральним моментом k -го порядку називають величину

$$m_{0k} = M[\xi - m_1]^k = \int_{\Omega} [\xi(\omega) - m_1]^k dP = \int_{-\infty}^{\infty} [x - m_1]^k dF(x) = \int_{-\infty}^{\infty} [x - m_1]^k p(x) dx. \quad (2.22)$$

Серед цих моментів особливо значення мають

$$m_1 = M\xi = \int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} xp(x) dx \quad (2.23)$$

та

$$m_{02} = D\xi = \int_{-\infty}^{\infty} [x - m_1]^2 dF(x) = \int_{-\infty}^{\infty} [x - m_1]^2 p(x) dx \quad (2.24)$$

— відповідно математичне сподівання та дисперсія випадкової величини.

Моменти випадкової величини можуть бути обчислені також по характеристичній функції:

$$m_k = j^{-k} \left[\frac{d}{du^k} f(u) \right]_{u=0}, \quad k=1,2,\dots \quad (2.25)$$

Відомий також зворотній зв'язок: маючи нескінченну послідовність моментів, можна знайти характеристичну функцію:

$$f(u) = 1 + \sum_{k=1}^{\infty} \frac{m_k}{k!} (ju)^k. \quad (2.26)$$

Крім моментів випадкової величини використовують також кумулянти (або семі інваріанти): це коефіцієнти розкладу логарифму характеристичної функції у степеневий ряд:

$$\ln f(u) = \sum_{k=1}^{\infty} \frac{\aleph_k}{k!} (ju)^k. \quad (2.27)$$

Між моментами та кумулянтами існує взаємозв'язок. Так, наприклад,

$$\begin{aligned} m_1 &= \aleph_1; \\ m_2 &= \aleph_2 + \aleph_1^2; \\ m_3 &= \aleph_3 + 3\aleph_1\aleph_2 + \aleph_1^3; \\ m_4 &= \aleph_4 + 3\aleph_2^2 + 4\aleph_1\aleph_3 + 6\aleph_1^2\aleph_2 + \aleph_1^4 \end{aligned} \quad (2.28)$$

тощо.

Приклади:

1. Обчислення середнього значення випадкової величини. Нехай

$$\xi(\Omega) = \{a_1, a_2, \dots, a_n\}.$$

$$\text{Тоді} \quad M\xi = \int_{\Omega} \xi(\omega) dP = \sum_{k=1}^n a_k P\{\xi = a_k\}. \quad (2.29)$$

Розглянемо частковий випадок рівномірного розподілу, тобто покладаємо

$$P\{\xi = a_k\} = \frac{1}{n}.$$

Тоді

$$M\xi = \sum_{k=1}^n a_k \cdot \frac{1}{n} = \frac{\sum_{k=1}^n a_k}{n}. \quad (2.30)$$

Це формула середньоарифметичного.

У загальному випадку, позначаючи $P\{\xi = a_k\} = P_k$, з (2.29) одержуємо

$$M\xi = \sum_{k=1}^n a_k P_k. \quad (2.29a)$$

Результат (2.29a) обчислення математичного сподівання простої випадкової величини з врахуванням розуміння формули (2.30) трактують як середнє значення випадкової величини, де усереднення виконується з врахуванням «вагових коефіцієнтів» P_k при її відповідних значеннях a_k .

2. Обчислення середнього значення дискретної випадкової величини.

Тут $\xi(\Omega) = \{a_1, a_2, \dots, a_n, \dots\}$, тоді

$$M\xi = \int_{\Omega} \xi(\omega) dP = \sum_{k=1}^{\infty} a_k P\{\xi = a_k\} = \sum_{k=1}^{\infty} a_k P_k. \quad (2.31)$$

Формула (2.31) безпосередньо узагальнює відповідний результат для простої випадкової величини.

Властивості математичного сподівання випадкової величини.

Властивості математичного сподівання випадкової величини визначаються її означенням як інтеграла та властивостями інтеграла. Однак замість того, щоб раз за разом «доводити» ці властивості, слід знати їх та користуватись ними.

Властивості математичного сподівання:

а) $M\{a\xi\} = aM\xi$;

б) $M\{\xi_1 + \xi_2\} = M\xi_1 + M\xi_2$;

в) $M\eta = M\varphi(\xi) = \int_{-\infty}^{\infty} \varphi(\xi) dP_{\xi}(x)$, де $\eta = \varphi(\xi)$;

г) $M\{\xi_1\xi_2\}=M\xi_1M\xi_2$, якщо випадкові величини ξ_1 та ξ_2 незалежні. (2.32)

Випадкові величини називаються *незалежними*, якщо всі події, які пов'язані з ними є незалежними.

Властивості дисперсії випадкової величини. Із означення (2.24) маємо:

$$D\xi = M[\xi - m_1]^2, \quad (2.33)$$

або

$$D\xi = M[\xi^2 - 2\xi m_1 + m_1^2] = M\xi^2 - m_1^2. \quad (2.33a)$$

а) $D\xi \geq 0$ – ця властивість безпосередньо витікає із означення.

Нехай $D\xi = 0$. Якщо

$$D\xi = M[\xi - m_1]^2 = \int_{\Omega} [\xi(\omega) - m_1]^2 dP = 0,$$

одержуємо, що майже всюди $\xi(\omega) = m_1$, тобто, з ймовірністю 1 випадкова величина дорівнює математичному сподіванню: $\xi = const$. Інакше кажучи, дисперсія випадкової величини може дорівнювати нулю тільки у постійної (з ймовірністю 1) функції.

б) $D(a\xi) = a^2 D\xi$.

Дійсно,

$$\begin{aligned} D(a\xi) &= \int_{\Omega} [a\xi(\omega) - m_1]^2 dP = a^2 \int_{\Omega} [\xi(\omega)]^2 dP - a^2 m_1^2 = a^2 \left[\int_{\Omega} [\xi(\omega)]^2 dP - m_1^2 \right] = \\ &= a^2 \int_{\Omega} [\xi(\omega) - m_1]^2 dP = a^2 D\xi. \end{aligned}$$

в) $D(\xi_1 + \xi_2) = D\xi_1 + D\xi_2$, якщо випадкові величини ξ_1 та ξ_2 незалежні.

Дійсно, якщо випадкові величини незалежні, маємо:

$$\begin{aligned} D(\xi_1 + \xi_2) &= M[(\xi_1 + \xi_2)^2] - M^2(\xi_1 + \xi_2) = M(\xi_1^2 + 2\xi_1\xi_2 + \xi_2^2) - (M^2\xi_1 + 2M\xi_1\xi_2 + M^2\xi_2) = \\ &= [M\xi_1^2 - M^2\xi_1] + [M\xi_2^2 - M^2\xi_2] + 2[M(\xi_1\xi_2) - M\xi_1M\xi_2] = D\xi_1 + D\xi_2 + 2[M(\xi_1\xi_2) - M\xi_1M\xi_2]. \end{aligned}$$

Тому, якщо випадкові величини незалежні, тобто $M(\xi_1\xi_2) = M\xi_1M\xi_2$, звідси одержуємо шукану рівність.

Приклад (на використанні моментів):

Узагальнена нерівність Чебишева. Нехай $g(x)$ – довільна парна зростаюча функція. Часто покладають $g(x)=|x|^2$. Необхідно довести, що

$$P(|\xi| \geq a) \leq \frac{Mg(\xi)}{g(a)}. \quad (2.34)$$

Маємо:

$$Mg(\xi) = \int_{-\infty}^{\infty} g(x) dP_{\xi}(x) = \int_{|x| < a} g(x) dP_{\xi}(x) + \int_{|x| \geq a} g(x) dP_{\xi}(x) \geq 0 + g(a) \int_{|x| \geq a} dP_{\xi} = g(a) P(|\xi| \geq a).$$

Звідси одержуємо (2.34), а коли $g(x)=|x|^2$, маємо

$$P(|\xi| \geq a) \leq \frac{M\xi^2}{a^2}. \quad (2.34a)$$

Це – класична нерівність Чебишева.

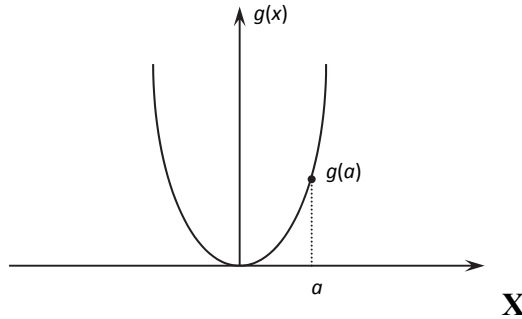


Рис. 2.14. – Ілюстрація до доведення нерівності Чебишева

2.3. Класичні закони розподілу ймовірностей

Випадкова величина повністю описується своїм законом розподілу. Тому особливу увагу приділяють вивченню деяких класичних законів розподілу, які часто зустрічаються у прикладних задачах.

2.3.1. Біноміальний закон розподілу

У пункті (2.1.2) ми розглядали модель незалежних випробувань. Там увага приділялась введенню за означенням ймовірності добутку

подій. Зараз ми вже знайомі з поняттям випадкової величини і маємо можливість розглядати цю модель з позицій знання випадкової величини і вивчати відповідний закон розподілу ймовірностей.

Припускається, що n разів повторюється в однакових умовах випробування. Ми слідкуємо за «успіхом», тобто за деякою подією. Задача полягає у знаходженні ймовірностей того, що «успіх» досягається точно k разів.

Цю схему незалежних випробувань називають схемою Бернуллі.

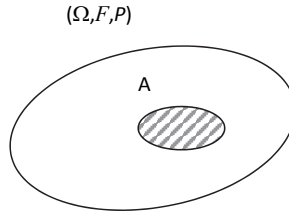


Рис. 2.15. — Ілюстрація до схеми Бернуллі: Ω — простір елементарних подій для одного випробування; A — успіх у просторі Ω , Ω^n — простір елементарних подій для n випробувань;

$$\left\{ \begin{array}{l} \overline{A \times \bar{A} \times A \times \bar{A} \times \dots \times A} \\ k \text{ разів } A \\ n-k \text{ разів } \bar{A} \end{array} \right\} - \text{подія у просторі } \Omega^n; \Omega^n = \underbrace{\Omega \times \Omega \times \dots \times \Omega}_n.$$

Задамо ймовірнісний простір одного випробування — (Ω, F, P) . Згідно зі схемою незалежних випробувань маємо ймовірнісний простір n випробувань — (Ω^n, F^n, P^n) , де

$$\begin{aligned} \Omega^n &= \underbrace{\Omega \times \Omega \times \dots \times \Omega}_n - \text{прямий добуток } n \text{ копій } \Omega, \\ F^n &= \underbrace{F \times F \times \dots \times F}_n - \text{прямий добуток відповідних алгебр}; \\ P^n &= \underbrace{P \times P \times \dots \times P}_n - \text{прямий добуток мір}. \end{aligned} \tag{2.35}$$

Випадкова величина — число виникнень успіху A . Це проста випадкова величина:

$$\xi(\Omega^n) = \{0, 1, 2, \dots, n\}. \tag{2.36}$$

Таким чином, практично, задача полягає у тому, щоб знайти ймовірність кожної з подій $P_k = P\{\xi = k\}$, інакше кажучи – ймовірнісну міру на просторі (2.36). У просторові Ω^n нові події мають вигляд типу $\underbrace{A \times \bar{A} \times A \times \bar{A} \times \dots \times A}_n$.

Позначимо:

$P(A) = p$ – ймовірність виникнення «успіху» у одному випробуванні (у просторі Ω);

$P(\bar{A}) = 1 - p = q$ – ймовірність «неуспіху» у одному випробуванні (у просторі Ω).

Тоді одержуємо (у просторі Ω^n):

$$P_0 = P\{\xi = 0\} = P\left\{\underbrace{\bar{A} \times \bar{A} \times \dots \times \bar{A}}_n\right\} = [P(\bar{A})]^n = q^n,$$

а також –

$$P_1 = P\{\xi = 1\} = P\left\{A \times \underbrace{\bar{A} \times \dots \times \bar{A}}_{n-1} \cup \bar{A} \times A \times \underbrace{\bar{A} \times \dots \times \bar{A}}_{n-2} \cup \dots \cup \bar{A} \times \dots \times \bar{A} \times A\right\} = npq^{n-1}. \quad (2.37)$$

Розглядається n несумісних подій у просторі Ω^n , у кожній з яких є один «успіх» та $n-1$ «неуспіх»; усі «успіхи» та «неуспіхи» незалежні.

У загальному випадку, коли «успіх» досягається k разів, а «неуспіх» – $n-k$ разів, маємо

$$P_k = P\{\xi = k\} = P\left\{\bigcup \underbrace{A \times \bar{A} \times \dots \times A \times \bar{A}}_{\substack{k \text{ разів } A \\ n-k \text{ разів } \bar{A}}}\right\} = \sum P\left\{\underbrace{A \times \bar{A} \times \dots \times A \times \bar{A}}_{\substack{k \text{ разів } A \\ n-k \text{ разів } \bar{A}}}\right\} = C_n^k p^k q^{n-k}.$$

Тут враховано, що у просторі Ω^n розглядається сукупність несумісних подій, у яких k разів реалізується «успіх» і $n-k$ разів «неуспіх»; таких несумісних подій усього

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

Таким чином, доведена формула

$$P\{\xi=k\}=C_n^k p^k q^{n-k} \quad (2.38)$$

— це закон Бернуллі, або біноміальний закон розподілу. На мові ймовірнісних мір — це міра, яка задана на дійсній осі (на дискретному просторі):

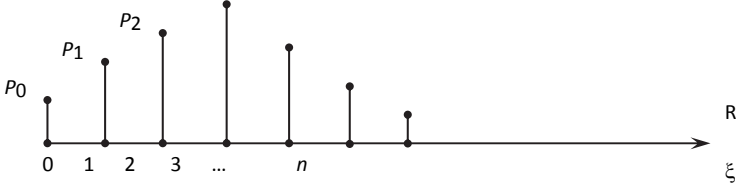


Рис. 2.16. — Біноміальний закон розподілу: $P_k = P\{\xi=k\}$

Зауважимо, що (2.38) є елементами розкладу біному Ньютона $(p+q)^n$; розподіл залежить від двох параметрів — n і p .

Приклади:

1. Характеристики біноміального закону розподілу ймовірностей.

а) Перевірка нормованості міри:

$$P\{\xi(\Omega)\} = \sum_{k=0}^n P_k = \sum_{k=0}^n C_n^k p^k q^{n-k} = (p+q)^n = 1.$$

б) Перевірка одновершинності закону.

Знайдемо

$$\frac{P_{k+1}}{P_k} = \frac{C_n^{k+1} p^{k+1} q^{n-k-1}}{C_n^k p^k q^{n-k}} = \frac{p(n-k)}{q(k+1)}.$$

Нехай

$$\frac{p}{q} \cdot \frac{n-k}{k+1} \geq 1,$$

тобто $pn - pk \geq qk + q$. Ця умова виконується для $k \leq -q + pn$, тобто розподіл ймовірностей дійсно одновершинний.

в) Середнє значення (математичне сподівання):

$$M\xi = \sum x_k P\{\xi=x_k\} = \sum_{k=0}^n k C_n^k p^k q^{n-k} = np.$$

г) Дисперсія:

$$D\xi = M\xi^2 - [M\xi]^2 = \sum_{k=0}^n k^2 P\{\xi=k\} - (np)^2 = \sum_{k=0}^n k^2 C_n^k p^k q^{n-k} - n^2 p^2 = npq.$$

2. Класичне визначення ймовірності.

Обчислення ймовірності подій при доведенні біноміального закону розподілу ілюструє особливості класичного означення ймовірності. Вона витікає із припущення рівноможливості як об'єктивної властивості вивчаємих явищ, що ґрунтується на реальній симетрії. Саме так випроваджується за означенням прямий добуток просторів з мірою.

Згідно з класичним означенням ймовірністю випадкової події A називається відношення числа несумісних рівноможливих елементарних подій, із яких складається A до числа всіх можливих елементарних подій.

Природно, що ймовірності подій у цьому випадку підраховуються на основі комбінаторики.

3. Ймовірність k -кратної помилки у n -значному коді при передаванні по дискретному каналу зв'язку без пам'яті.

По дискретному каналу зв'язку передається n -значний код, наприклад, двійковий. Канал без пам'яті, тобто помилки у всіх елементах кодової групи, незалежні. Необхідно знайти ймовірність k -кратної помилки, коли відома ймовірність помилки одного символу — p .

Ймовірність помилки у одному символі та ймовірність його правильного приймання утворюють повну групу подій $p+q=1$. Помилки та правильне приймання усіх n елементів кодової групи незалежні, тобто ймовірність k -кратної помилки конкретної конфігурації та $n-k$ правильного приймання дорівнює $p^k q^{n-k}$. Нарешті, кількість усіх можливих варіантів k помилок та $n-k$ правильно прийнятих символів, дорівнює C_n^k . Тобто, враховуючи несумісність подібних подій, маємо біноміальний закон розподілу k -кратних помилок:

$$P_{ном}(k) = C_n^k p^k q^{n-k}. \quad (2.39)$$

2.3.2. Закон розподілу Пуассона

Повернемось до попереднього прикладу. Введемо відносну частоту успіху у схемі Бернуллі: $\mu_n = \frac{\xi_n}{n}$, де ξ_n – число виникаючих «успіхів» у серії з n випробувань.

Тоді маємо такі значення μ_n : $0, \frac{1}{n}, \frac{2}{n}, \dots, 1$. У цьому випадку середнє значення і дисперсія у схемі Бернуллі дорівнюють:

$$\begin{aligned} M\mu_n &= M\left(\frac{\xi_n}{n}\right) = p; \\ D\mu_n &= D\left(\frac{\xi_n}{n}\right) = \frac{pq}{n}, \end{aligned} \quad (2.40)$$

тобто середнє значення відносної частоти події дорівнює її ймовірності, а дисперсія спадає із зростанням n . Відомо, що чим довша серія випробувань, тим менше відхилення частоти «успіху» від ймовірності, тобто у деякому розумінні тут є відповідна збіжність (але ці питання ми розглянемо пізніше).

Такі навідні міркування пояснюють нову постановку задачі. У серії n випробувань цікавимося числом «успіхів» ξ_n :

$$P\{\xi_n = k\} = C_n^k p^k q^{n-k}, k=0,1,\dots,n.$$

Вважаємо, що середнє число «успіхів» при $n \rightarrow \infty$ залишається незмінним:

$$M\xi_n = np = \lambda. \quad (2.41)$$

Для цього покладаємо

$$p = \frac{\lambda}{n} \rightarrow 0, \quad (2.42)$$

тобто розглядаємо дуже рідкі події у дуже довгій серії.

Необхідно знайти граничну ймовірність успіху.

При вказаних умовах, як можна довести, ця границя існує:

$$\lim_{n \rightarrow \infty} P(\xi_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (2.43)$$

Граничні ймовірності визначають деяку ймовірнісну міру на дискретному просторі – на зчисленній множині точок $k=0,1,2,\dots,n,\dots$ (рис.2.17):

$$p(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}, \lambda > 0. \quad (2.44)$$

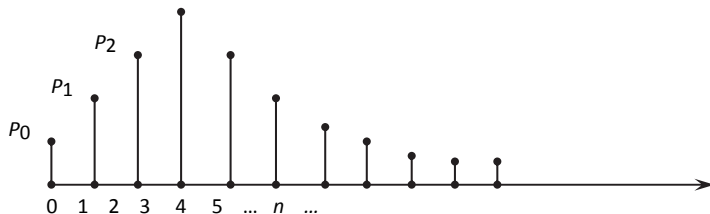


Рис. 2.17. — Закон Пуассона: $P_k = P\{\xi_n = k\}$

Цей результат по суті є граничною теоремою. Вказаний закон розподілу ймовірностей носить назву *закону Пуассона*. Це — закон рідких подій.

Приклади:

1. Обчислення характеристик закону Пуассона.

а) Перевірка нормованості міри:

$$\sum_{k=0}^{\infty} p(k; \lambda) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

б) Середнє значення випадкової величини, яка розподілена за законом Пуассона:

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

в) Дисперсія випадкової величини ξ з врахуванням (2.42):

$$D\xi = \lim_{n \rightarrow \infty} D\mu_n = \lim_{n \rightarrow \infty} npq = \lim_{n \rightarrow \infty} \lambda \left(1 - \frac{\lambda}{n}\right) = \lambda;$$

або безпосередньо

$$D\xi = \sum_{k=0}^{\infty} (k - \lambda)^2 p(k; \lambda) = \lambda.$$

2. Течія викликів (у телефонії). Течія викликів вважається незалежною, однорідною та ординарною.

а) Однорідність:

Ймовірність виникнення довільного числа викликів на ділянці тривалістю Δt сек. залежить від її величини, а не від положення на осі часу t .

б) Незалежність:

Ймовірність виникнення викликів на інтервалі Δt_2 не залежить від ймовірності їх виникнення на інтервалі Δt_1 ($\Delta t_1 \cap \Delta t_2 = \emptyset$).

в) Ординарність:

Нехай Δt – мале. Позначимо ймовірність виникнення k викликів за Δt сек. через $P_k(\Delta t)$. Тоді за означенням ординарності:

$$P_1(\Delta t) = \lambda \Delta t + o(\Delta t),$$

де λ – середнє число викликів за t сек.;

$o(\Delta t)$ – нескінченно мала величина вищого степеня мализни порівняно з Δt . Тобто ординарність означає, що на як завгодно малому інтервалі часу існує принаймі один виклик.

Можна довести, що за цих умов ймовірність одержати k викликів за Δt сек. визначається законом Пуассона

$$P_k(\Delta t) = \frac{(\lambda \Delta t)^k}{k!} e^{-\lambda \Delta t}.$$

2.3.3. Гаусів (нормальний) закон розподілу

Гаусовим (або нормальним) розподілом $N(a, \sigma^2)$ з параметрами a і σ^2 називається закон зі щільністю ймовірності

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (2.45)$$

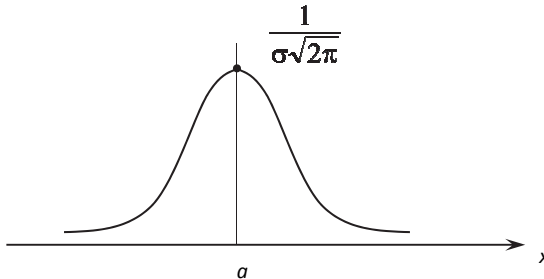


Рис 2.18. – Гаусів розподіл: a – середнє значення; σ^2 – дисперсія

Нормованість функції (2.45) легко перевіряється, якщо покласти $y = \frac{x-a}{\sigma}$ і використати табличний матеріал $\int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy = \sqrt{2\pi}$. Безпосередньо можна знайти, що середнє значення дорівнює

$$M\xi = \int_{-\infty}^{\infty} xp(x)dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right] dx = a;$$

відповідно дисперсія –

$$D\xi = \int_{-\infty}^{\infty} (x-a)^2 p(x)dx = \int_{-\infty}^{\infty} (x-a)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right] dx = \sigma^2.$$

Якщо $\sigma \ll 1$ вершина розподілу «уходить» далеко вгору, а ширина мала (тому що площа завжди дорівнює одиниці); випадкова величина тут має мале розсіяння.

У випадку $\sigma \gg 1$ ширина графіка значна; випадкова величина має велике розсіяння.

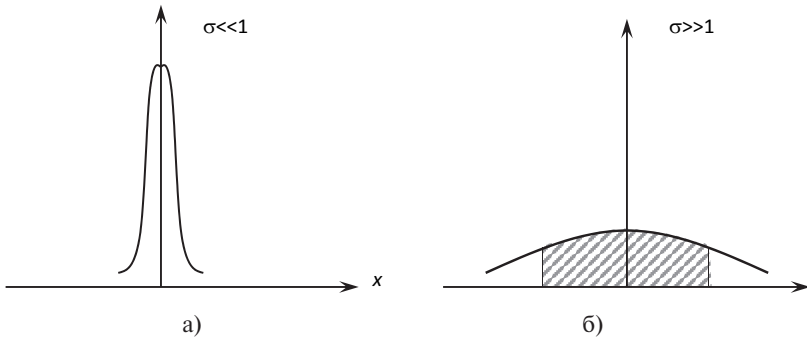


Рис. 2.19. – Гаусів розподіл при різних σ

При цьому у значному інтервалі розподіл ймовірності практично рівномірний.

Функція розподілу гаусова закону має вигляд

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-a)^2}{2\sigma^2}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{u^2}{2}} du = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right), \quad (2.46)$$

де покладено

$$u = \frac{z-a}{\sigma};$$

$$\Phi(x) = \frac{1}{2\pi_0} \int_{-\infty}^x e^{-\frac{u^2}{2}} du - \text{функція Лапласа};$$

$$\int_{-\infty}^0 \frac{1}{2\pi} e^{-\frac{u^2}{2}} du = \frac{1}{2}.$$

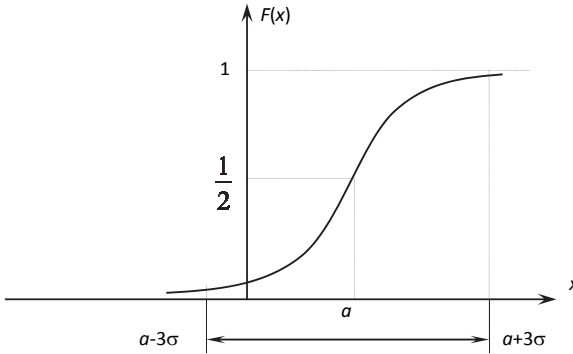


Рис. 2.20. – Функція розподілу гаусова закону

Характерна така властивість гаусова розподілу: ймовірність попадання випадкової величини на інтервал $a-3\sigma \leq x \leq a+3\sigma$ приблизно дорівнює одиниці, тобто $F(a-3\sigma) \approx 0$; $F(a+3\sigma) \approx 1$. У технічній літературі цю властивість називають «законом трьох сигма».

Гаусів розподіл грає виключно важливу роль ще й тому, що він є результатом ряду граничних теорем.

Так, розглянемо ще раз схему Бернуллі. Згідно з (2.40) відносна нормована і центрована частота «успіху»

$$\hat{\mu}_n = \frac{\mu_n - M\mu_n}{\sqrt{D\mu_n}}$$

має вигляд:

$$\hat{\mu}_n = \frac{\mu_n - p}{\sqrt{\frac{pq}{n}}} = \frac{\xi_n - p}{\sqrt{\frac{pq}{n}}} = \frac{\xi_n - np}{\sqrt{npq}}, \xi_n = 0, 1, \dots, n; \quad (2.47)$$

вона приймає такі значення:

$$-\frac{np}{\sqrt{npq}}, \frac{1-np}{\sqrt{npq}}, \frac{2-np}{\sqrt{npq}}, \dots, \frac{n-np}{\sqrt{npq}}.$$

Для випадкової величини $\hat{\mu}_n$ (яку стисло називають *нормованою частотою «успіху»*) справджується *Теорема Муавра-Лапласа*. Нормована частота «успіху» у схемі Бернуллі при $n \rightarrow \infty$ прагне по розподілу до нормального закону з параметрами $m=0$, $\sigma=1$:

$$F_{\hat{\mu}_n}(x) \rightarrow F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz, \quad (2.48)$$

або

$$P\{a \leq \hat{\mu}_n < b\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz, \quad -\infty < a < \infty, \quad (2.48a)$$

де $F_{\hat{\mu}_n}(x)$ – функція розподілу $\hat{\mu}_n$ із (2.47); збіжність розуміється, як збіжність у точках неперервності функцій розподілу; збіжність рівномірна відносно a і b .

На відміну від попередньої граничної теореми (п. 2.3.2), яка характеризувала дискретну випадкову величину, у даному разі при граничному переході одержується неперервна випадкова величина з гаусовим розподілом.

Абсолютна більшість класичних задач теорії електров'язку, пов'язаних з неперервними сигналами, розглядалась у припущенні гаусова закону розподілу сигналів та завад.

2.3.4. Рівномірний та інші закони розподілу

Рівномірним називається закон розподілу, щільність ймовірності якого задовольняє умову

$$p(x) = \begin{cases} c, & x \in (a, b], \\ 0, & x \notin (a, b], \end{cases} \quad (2.49)$$

$$\text{де } \int_{-\infty}^{\infty} p(x) dx = \int_a^b c dx = c(b-a), \text{ тобто } c = \frac{1}{b-a}.$$

За межами інтервалу $(a, b]$ випадкова величина практично не приймає значень; точніше, ймовірність попадання за межі цього

інтервалу нульова. У межах цього інтервалу ймовірність попадання на однаковий інтервал та ж сама.

Функція розподілу, яка відповідає щільності ймовірності (2.49), має вигляд

$$F(x) = \int_{-\infty}^x p(y) dy = \begin{cases} 0, & x \leq a; \\ \frac{x-a}{b-a}, & a < x \leq b; \\ 1, & x > b. \end{cases} \quad (2.50)$$

Графіки функцій $p(x)$ та $F(x)$ наведені нижче

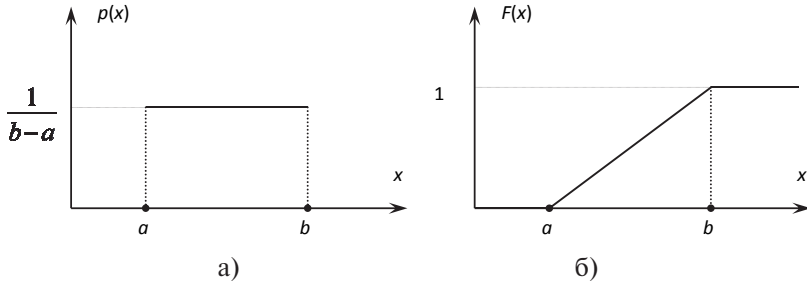


Рис. 2.21. – Рівномірний закон розподілу:
а) щільність ймовірності; б) функція розподілу

Випадкова величина має експоненціальний розподіл з параметром λ , якщо її функція розподілу дорівнює

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0; \\ 0, & x \leq 0. \end{cases} \quad (2.51)$$

Щільність ймовірності такої випадкової величини дорівнює

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0; \\ 0, & x \leq 0. \end{cases} \quad (2.52)$$

Розподіл Вейбулла-Гнеденко:

$$F(x) = \begin{cases} 1 - e^{-\lambda x^\alpha}, & x > 0; \\ 0, & x \leq 0. \end{cases} \quad (2.53)$$

де $\lambda > 0$, $\alpha > 0$. У частковому випадку при $\alpha = 1$ розподіл (2.53) перетворюється у показниковий (експоненціальний) (2.51).

Цей закон широко використовується у теорії надійності.

Закон розподілу випадкової величини називається *розподілом Коші*, якщо функція розподілу і щільність ймовірності відповідно задаються формулами:

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} \frac{x - \mu}{a}; \quad (2.54)$$

$$p(x) = \frac{1}{\pi} \frac{a}{(x - \mu)^2 + a^2}, \quad a > 0. \quad (2.55)$$

2.4. Випадкові вектори

Випадковий вектор згідно з аксіоматизацією за Колмогоровим теорії ймовірностей розглядається як відповідне вимірне відображення. На ідейному рівні це дає аналогічні методи її повного описування. Нижче дається означення випадкового вектора, розглядаються методи його описування та приводяться приклади багатовимірних законів розподілу.

2.4.1. Означення випадкового вектора та можливості його повного описування

Випадковим вектором називається вимірне відображення ймовірнісного простору (Ω, \mathcal{F}, P) у n -вимірний евклідов простір (R^n, \mathcal{B}^n) з σ -алгеброю борелівських множин \mathcal{B}^n .

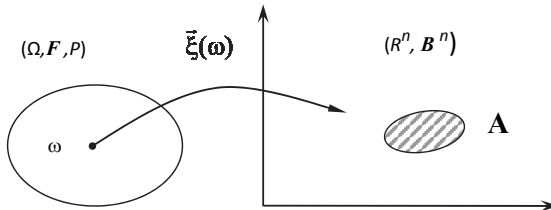


Рис. 2.22. — Випадковий вектор

Випадковий вектор називають також багатовимірною випадковою величиною.

Клас \mathcal{B}^n утворюється з n -вимірних напівінтервалів, їх об'єднань, різниць, перетинів та результатів граничних переходів. На цьому класі задається індукована міра, аналогічно тому, як це було показано для випадкових величин. Повторимо цю процедуру, використовуючи відповідні позначення.

Беремо довільну вимірну множину A як це показано на малюнку. Це означає, що вона належить класові $\mathbf{B}^n : A \in \mathbf{B}^n$. Індукована міра $P^n(A)$ визначається як міра P повного прообразу A :

$$P^n(A) = P\{\xi^{-1}(A)\} = P\{\omega : \xi(\omega) \in A\}. \quad (2.56)$$

У загальній теорії функцій доводиться, що без обмеження загальності замість довільної множини $A \in \mathbf{B}^n$ можна взяти деякі стандартні, у тому разі, напівнескінченні інтервали:

$$F_n(x_1, x_2, \dots, x_n) = P\{\xi_1(\omega) \leq x_1, \xi_2(\omega) \leq x_2, \dots, \xi_n(\omega) \leq x_n\}, \quad (2.56a)$$

де $\xi_1, \xi_2, \dots, \xi_n$ – координати вектора $\vec{\xi}_n$.

Індукована міра (2.56a) має назву *n-вимірної функції розподілу*.

Якщо функція $F_n(\vec{x})$ абсолютно неперервна, вводиться *n-вимірна щільність ймовірності*

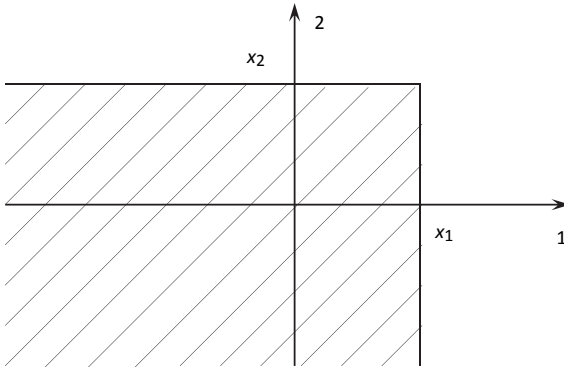


Рис. 2.23. – Напівнескінченний інтервал (двовимірний випадок)

$$p_n(\vec{x}) = \frac{\partial^n F_n(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n}. \quad (2.57)$$

Ця функція має такі основні властивості:

- а) $p_n(\vec{x}) \geq 0$;
- б) $\int_{R^n} p_n(\vec{x}) d\vec{x} = 1$.

Випадковий вектор може бути описаний також характеристичною функцією

$$f_n(\bar{u}) = M \exp[j(\bar{u}, \bar{x})] = \int_{-\infty}^{\infty} \exp[j(\bar{u}, \bar{x})] dF(\bar{x}), \quad (2.58)$$

де $(\bar{u}, \bar{x}) = \sum_{k=1}^n u_k x_k$ – скалярний добуток у просторі R^n .

Співвідношення (2.58) – це *перетворення Фур’є-Стільтьєса* функції розподілу $F_n(\bar{x})$. Якщо існує щільність ймовірності (2.57), формула (2.58) приймає звичайний вигляд n -вимірного перетворення Фур’є:

$$f_n(\bar{u}) = \int_{-\infty}^{\infty} \exp[j(\bar{u}, \bar{x})] p_n(\bar{x}) d\bar{x}. \quad (2.58a)$$

Існує зворотне перетворення Фур’є:

$$p_n(\bar{x}) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \exp[-j(\bar{u}, \bar{x})] f_n(\bar{u}) d\bar{u}. \quad (2.58b)$$

Таким чином, функція розподілу, ймовірності та характеристична функція дають повне описування n -вимірного вектора, як сукупності n випадкових величин (його координат):

$$\vec{\xi}(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega)). \quad (2.59)$$

Характерна особливість такого описування – це можливість дістати звідси відповідне описування будь-якого вектора меншої вимірності, який утворюється сукупністю окремих випадкових величин з (2.59).

Так, щоб одержати функцію розподілу $F_{n-1}(\bar{x})$ з $F_n(\bar{x})$, достатньо у відповідній координат поставити $+\infty$. Дійсно, маємо

$$F_n(x_1, x_2, \dots, x_n) = P\{\xi_1 \leq x_1; \xi_2 \leq x_2; \dots; \xi_n \leq x_n\}.$$

Тоді, замінюючи x_1 на $+\infty$, дістаємо

$$F_n(\bar{x}) = P\{\xi_1 \leq \infty; \xi_2 \leq x_2; \dots; \xi_n \leq x_n\} = P\{\xi_2 \leq x_2; \dots; \xi_n \leq x_n\} = F_{n-1}(\bar{x}),$$

тобто одержуємо функцію розподілу $n-1$ – вимірного вектора.

Якщо існує щільність ймовірності, можна записати

$$F_n(\bar{x}) = \int_{-\infty}^{\bar{x}} p_n(\bar{x}) d\bar{x} = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} p_n(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n.$$

Тоді вірні рівності

$$F_{n-1}(x_2, \dots, x_n) = F_n(\infty, x_2, \dots, x_n) = \int_{-\infty}^{\infty} d\bar{x} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} p_n(x_1, x_2, \dots, x_n) dx_2 \dots dx_n.$$

тобто приходимо до такого результату:

$$p_{n-1}(x_2, \dots, x_n) = \int_{-\infty}^{\infty} p_n(x_1, \dots, x_n) dx_1$$

– якщо n -вимірну щільність ймовірності проінтегрувати по одній координаті, то одержується $n-1$ - вимірна щільність ймовірності $n-1$ - вимірного вектора.

Аналогічно можна з $F_n(\bar{x})$ та $p_n(\bar{x})$ дістати функцію розподілу та щільність ймовірності $n-k$ -вимірного вектора, який одержується як відповідна сукупність координат n -вимірного вектора (2.586). Для цього необхідно у функції $F_n(\bar{x})$ замість відповідних координат x_j підставити $+\infty$, а функцію $p_n(\bar{x})$ проінтегрувати по всій вісі по відповідних координатах.

Приклади:

1. Випадковий вектор з незалежними координатами: n - вимірний випадковий вектор називається *вектором з незалежними координатами*, якщо його функція розподілу (або щільності ймовірності) розкладається на n співмножників, незалежних один від одного. Нагадаємо, що означення – це необхідна і достатня умова.

Проілюструємо цю тезу. Так, хай компоненти вектора $\bar{\xi}(\omega) = \{\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega)\}$ незалежні. Тоді функція розподілу

$$F_n(x_1, x_2, \dots, x_n) = P\{\xi_1 \leq x_1; \xi_2 \leq x_2; \dots; \xi_n \leq x_n\}$$

характеризує незалежні події $\xi_1 \leq x_1, \xi_2 \leq x_2, \dots, \xi_n \leq x_n$ і може бути зображена у вигляді

$$F_n(x_1, x_2, \dots, x_n) = P\{\xi_1 \leq x_1\} P\{\xi_2 \leq x_2\} \dots P\{\xi_n \leq x_n\}$$

або

$$F_{\bar{\xi}}(x_1, x_2, \dots, x_n) = F_{\xi_1}(x_1) F_{\xi_2}(x_2) \dots F_{\xi_n}(x_n).$$

Таким чином, n -вимірна функція розподілу вектора з незалежними координатами дійсно дорівнює добутку n одновимірних функцій розподілу його координат.

Аналогічно справджується також зворотне твердження.

Зауважимо, що у загальному випадку, як доведено нижче, n -вимірна функція розподілу (або щільність ймовірності) дає можливість обчислити всі відповідні функції меншої вимірності. У тому числі є можливість обчислити всі одновимірні функції. Однак знання всіх одновимірних функцій розподілу не дає можливості знайти у загальному випадку відповідну n -вимірну функцію розподілу. Винятком є вектор з незалежними координатами: знання всіх одновимірних функцій розподілу дає можливість обчислити n -вимірну функцію розподілу цього вектора.

2. Перетворення випадкових векторів.

Нехай у області $A \subset R^n$ задані неперервно диференційовані функції

$$\begin{aligned} y_1 &= g_1(x_1, \dots, x_n); \\ &\dots \\ y_n &= g_n(x_1, \dots, x_n), \end{aligned} \tag{2.60}$$

які однозначно розв'язувані відносно x_1, \dots, x_n :

$$\begin{aligned} x_1 &= g_1^{-1}(y_1, \dots, y_n); \\ &\dots \\ x_n &= g_n^{-1}(y_1, \dots, y_n), \end{aligned} \tag{2.61}$$

Припускається, що якобіан $\mathfrak{J} = \left| \frac{\partial x_i}{\partial y_j} \right| \neq 0$.

Тоді справджується твердження: якщо випадковий вектор $\bar{\xi} = (\xi_1, \dots, \xi_n)$ має щільність ймовірності $p_{\bar{\xi}} = (x_1, \dots, x_n)$, то випадковий вектор $\bar{\eta} = (\eta_1, \dots, \eta_n)$, де $\eta_i = g_i(\xi_1, \dots, \xi_n)$, має щільність ймовірності, яка обчислюється за формулою

$$p_{\bar{\eta}}(y_1, \dots, y_n) = p_{\bar{\xi}}(g_1^{-1}(y_1, \dots, y_n), \dots, g_n^{-1}(y_1, \dots, y_n)) |\mathfrak{J}|. \tag{2.62}$$

Дійсно, нехай B – довільна борелівська множина ($B \subset A$), а B' – її образ при перетворенні (2.60). Використовуючи правило заміни змінних у кратних інтегралах, маємо $B \subset B^n$

$$\begin{aligned} P\{\bar{\xi} \in B\} &= \int_B p_{\bar{\xi}}(x_1, \dots, x_n) dx_1 \dots dx_n = \int_{B'} p_{\bar{\xi}}(g_1^{-1}(y_1, \dots, y_n), \dots, g_n^{-1}(y_1, \dots, y_n)) |\mathfrak{J}| dy_1 \dots dy_n = \\ &= P\{\bar{\eta} \in B'\}. \end{aligned}$$

Тобто має місце рівність (2.62).

2.4.2. Багатовимірні закони розподілу

1. Вектор $\vec{\xi}$ називається *рівномірно розподіленим у область D* якщо у нього існує щільність ймовірності, яка дорівнює константі у цій області та нулю за її межами:

$$p(\vec{x}) = \begin{cases} C, \vec{x} \in D; \\ 0, \vec{x} \notin D. \end{cases} \quad (2.63)$$

Константа C обчислюється із умови нормованості міри:

$$\int_{R^n} p(\vec{x}) d\vec{x} = \int_D p(\vec{x}) d\vec{x} = CmD = 1,$$

де mD – міра Лебега області D . Тобто $C = \frac{1}{mD}$.

2. Нормальний вектор з незалежними координатами.

Нехай $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ – вектор n -вимірному простору. Тоді вектор $\vec{a} = (a_1, a_2, \dots, a_n)$, де $a_i = M\xi_i$, $i = \overline{1, n}$, називається *вектором математичного сподівання* (або просто математичним сподіванням). Відомі також дисперсії координат $D\xi_i = \sigma_i^2$; тобто розподіли координат –

$$p_{\xi_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i - a_i)^2}{2\sigma_i^2}}. \quad (2.64)$$

Згідно з означенням вектора з незалежними координатами та (2.64) одержуємо

$$p(\vec{x}) = \prod_{i=1}^n p_{\xi_i}(x_i) = \frac{1}{(2\pi)^{n/2} \sigma_1 \dots \sigma_n} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - a_i}{\sigma_i} \right)^2 \right\}. \quad (2.65)$$

Це щільність ймовірного n -вимірному нормального вектора з незалежними координатами – найпростіший вигляд нормального закону розподілу.

2. Багатовимірний нормальний розподіл.

Випадковий вектор називається *нормальним*, якщо довільна лінійна комбінація його координат є нормальною випадковою величиною.

Введемо спочатку означення моменту випадкового вектора.

Нехай k_1, k_2, \dots, k_n – довільні цілі невід’ємні числа. *Моментом вектора* $\vec{\xi} = (\xi_1, \dots, \xi_n)$ порядку $k = k_1 + k_2 + \dots + k_n$ називається величина

$$M \xi_1^{k_1} \xi_2^{k_2} \dots \xi_n^{k_n} = \int \dots \int x_1^{k_1} x_2^{k_2} \dots x_n^{k_n} dF_n(x_1, \dots, x_n).$$

Якщо існують моменти другого порядку, вводять кореляційну матрицю вектора $\vec{\xi}$:

$$R = [R_{ij}], \quad (2.66)$$

де $R_{ij} = M(\xi_i - m_i)(\xi_j - m_j) = \int \dots \int (x_i - m_i)(x_j - m_j) dF_n(x_1, \dots, x_n)$;

$m_i = M\xi_i = \int \dots \int x_i dF_n(x_1, \dots, x_n)$;

$1 \leq i \leq n, 1 \leq j \leq n$.

Якщо використати нормальний вектор $\vec{\xi}$ з незалежними координатами та сформувати новий вектор $\vec{\eta} = (\eta_1, \dots, \eta_n)^{tr}$ за формулою $\vec{\xi} = C\vec{\eta}$ (або $\vec{\eta} = C^{tr}\vec{\xi}$), де C – довільна ортогональна матриця (тобто $CC^{tr} = I$), можна одержати розподіл $\vec{\eta}$ у вигляді

$$p_{\vec{\eta}}(\vec{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |R|} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{a})^{tr} R^{-1} (\vec{x} - \vec{a}) \right\}, \quad (2.67)$$

де \vec{x} – вектор стовпець;

$\vec{x} = (x_1, x_2, \dots, x_n)^{tr}$;

tr – символ операції транспонування матриці;

$\vec{a} = (a_1, a_2, \dots, a_n)^{tr}$, $\vec{a} = M\vec{\eta}$;

R – кореляційна матриця вектора $\vec{\eta}$;

$|R|$ – визначник матриці R ;

R^{-1} – матриця, зворотня матриці R .

Випадковий вектор $\vec{\eta}$ зі щільністю ймовірності (2.67) називається *n-вимірним нормальним випадковим вектором*, а його щільність – *n-вимірною нормальною щільністю*.

Співвідношення (2.67) може бути записано також з використанням символу скалярного добутку у просторові R^n у вигляді

$$p_{\vec{\eta}}(\vec{x}) = \frac{1}{(2\pi)^{n/2} |R|^{1/2}} \exp \left\{ -\frac{1}{2} (R^{-1}(\vec{x} - \vec{m}), (\vec{x} - \vec{m})) \right\}. \quad (2.67a)$$

Використовуючи означення характеристичної функції, можна нормальний вектор охарактеризувати також на мові характеристичних функцій співвідношенням

$$f_{\vec{\eta}}(\vec{u}) = \exp\left\{j(\vec{m}, \vec{u}) - \frac{1}{2}(R\vec{x}, \vec{x})\right\}. \quad (2.68)$$

Наведемо деякі властивості нормального вектора:

1. якщо $\vec{\xi}$ – нормальний вектор, то його лінійне перетворення – теж нормальний вектор;

2. якщо $\vec{\xi}$ – n -вимірний нормальний вектор, а B – матриця порядку $m \times n$ рангу m ($m \leq n$), то випадковий вектор $\vec{\eta} = B\vec{\xi}$ має m – вимірний нормальний розподіл;

3. якщо $\vec{\xi}$ – n -вимірний вектор, кореляційна матриця якого має ранг $m < n$, розподіл ймовірностей зосереджений у m -вимірному підпросторі (гіперплощині L) n -вимірного простору R^n . Гіперплощина L простору R^n може бути зображена у вигляді

$$L = \vec{a} + BR^n,$$

тобто L є сукупність усіх векторів $y \in R^n$ вигляду $\vec{y} = \vec{a} + B\vec{x}$. У цьому підпросторі існує щільність розподілу

$$p(\vec{y}) = \frac{1}{(2\pi)^{m/2} |B|^{1/2}} \exp\left\{-\frac{1}{2}(B^{-1}(\vec{y}-\vec{a}), (\vec{y}-\vec{a}))\right\},$$

де $|B|$ – визначник матриці, яка задає оператор B у підпросторі $R^m = BR^n$, а B^{-1} є оператор у цьому підпросторі, зворотній до B .

Приклади:

1. Двовимірний нормальний розподіл. Розглянемо частковий випадок багатовимірного нормального розподілу, коли $n=2$. Нехай $D\xi_1 = \sigma_1^2$, $D\xi_2 = \sigma_2^2$. Кореляційна матриця вектора $\vec{\xi} = (\xi_1, \xi_2)$ має вигляд

$$R = \begin{bmatrix} \sigma_1^2 & R_{12} \\ R_{12} & \sigma_2^2 \end{bmatrix}. \quad (2.69)$$

Введемо коефіцієнт кореляції випадкових величин ξ_1, ξ_2 :

$$r = \frac{R_{12}}{\sqrt{D\xi_1 D\xi_2}}. \quad (2.70)$$

Тоді (2.69) можна зобразити у вигляді

$$R = \begin{vmatrix} \sigma_1^2 & \sigma_1 \sigma_2 r \\ \sigma_1 \sigma_2 r & \sigma_2^2 \end{vmatrix}. \quad (2.69a)$$

У цих позначеннях, матриця, зворотня матриці R , має вигляд

$$R^{-1} = \begin{vmatrix} \frac{1}{\sigma_1^2(1-r^2)} & -\frac{r}{\sigma_1 \sigma_2(1-r^2)} \\ -\frac{r}{\sigma_1 \sigma_2(1-r^2)} & \frac{1}{\sigma_2^2(1-r^2)} \end{vmatrix}; \quad (2.71)$$

визначник цієї матриці $-\frac{1}{[\sigma_1^2 \sigma_2^2 (1-r^2)]}$.

Таким чином, у двовимірному випадку нормальний закон розподілу має щільність

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp \left\{ -\frac{1}{2(1-r^2)} \left[\left(\frac{x_1 - a_1}{\sigma_1} \right)^2 - 2r \left(\frac{x_1 - a_1}{\sigma_1} \right) \left(\frac{x_2 - a_2}{\sigma_2} \right) + \left(\frac{x_2 - a_2}{\sigma_2} \right)^2 \right] \right\}. \quad (2.72)$$

де $a_1 = M\xi_1$, $a_2 = M\xi_2$.

У частковому випадку $r=0$, тобто $R_{12}=0$, одержуємо $p(x_1, x_2)$ у вигляді добутку щільностей ξ_1 і ξ_2 , що повертає нас до розподілу (2.65), коли $n=2$.

Розглянемо геометричну інтерпретацію двовимірного розподілу.

Функція (2.72) визначає деяку поверхню, задану на площині (x, y) .

Поверхню $p(x_1, x_2)$ можна досліджувати лініями рівня. Так, задамо

$$p(x_1, x_2) = C > 0.$$

Тоді з (2.72) одержуємо

$$\left(\frac{x_1 - a_1}{\sigma_1} \right)^2 - 2r \left(\frac{x_1 - a_1}{\sigma_1} \right) \left(\frac{x_2 - a_2}{\sigma_2} \right) + \left(\frac{x_2 - a_2}{\sigma_2} \right)^2 = C_1. \quad (2.73)$$

Це сім'я еліпсів з центром у точці (a_1, a_2) площини. Якщо $\sigma_1 = \sigma_2$ еліпси перетворюються у кола з центром у (a_1, a_2) .

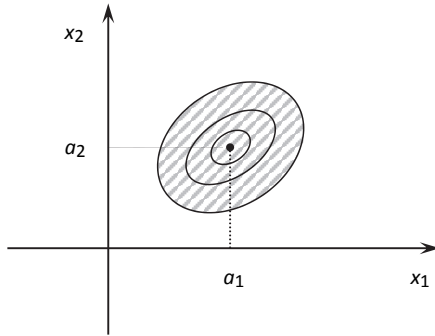


Рис. 2.24. – Двовимірний нормальний розподіл : еліпси рівної ймовірності

2. Властивості коефіцієнта кореляції.

Властивості r :

а) $|r| \leq 1$.

Запишемо $|r|$ згідно з (2.70) і (2.66) у вигляді

$$|M(\xi_i - m_i)(\xi_j - m_j)| \leq \sqrt{D\xi_i D\xi_j},$$

або

$$\left| \int_{R^2} (x_1 - m_1)(x_2 - m_2) dP_{\xi} \right| \leq \sqrt{\int_{R^2} (x_1 - m_1)^2 dP_{\xi}} \sqrt{\int_{R^2} (x_2 - m_2)^2 dP_{\xi}}.$$

Це є нерівність Буняковського-Шварца

$$|\int f_1(x)f_2(x)d\mu| \leq \sqrt{\int [f_1(x)]^2 d\mu} \sqrt{\int [f_2(x)]^2 d\mu},$$

що й доводить властивість (а).

б) Якщо $r = -1$, випадкові величини лінійно залежні.

Маємо:

$$D(\xi_1 + \xi_2) = \sigma_1^2 + \sigma_2^2 + 2r\sigma_1\sigma_2.$$

Введемо

$$\tilde{\xi}_1 = \frac{\xi_1 - m_1}{\sigma_1}, \quad \tilde{\xi}_2 = \frac{\xi_2 - m_2}{\sigma_2}.$$

Тоді

$$D(\tilde{\xi}_1 + \tilde{\xi}_2) = 1 + 1 + 2r = 2(1+r).$$

Тому при $r = -1$ одержуємо $D(\tilde{\xi}_1 + \tilde{\xi}_2) = 0$. Тобто з ймовірністю одиниця $\tilde{\xi}_1 + \tilde{\xi}_2 = \text{const}$, що означає наявність лінійної залежності випадкових величин ξ_1 і ξ_2 .

в) Якщо випадкові величини незалежні, то $r = 0$.

Дійсно:

$$r = \frac{M(\xi_1 - m_1)(\xi_2 - m_2)}{\sqrt{D\xi_1}\sqrt{D\xi_2}} = \frac{M(\xi_1 - m_1)M(\xi_2 - m_2)}{\sqrt{D\xi_1}\sqrt{D\xi_2}} = 0. \quad (2.74)$$

Кажуть, що випадкові величини некорельовані, якщо їх коефіцієнт кореляції дорівнює нулю.

Із (2.74) витікає, що, якщо випадкові величини незалежні, вони некорельовані. Зворотнє твердження у загальному випадку невірне.

г) Некорельованість координат нормального вектора.

Згідно з (2.72) двовимірний нормальний вектор має щільність

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left[\left(\frac{x_1-m_1}{\sigma_1}\right)^2 - 2r\left(\frac{x_1-m_1}{\sigma_1}\right)\left(\frac{x_2-m_2}{\sigma_2}\right) + \left(\frac{x_2-m_2}{\sigma_2}\right)^2\right]\right\},$$

яка при $r = 0$ приводиться до вигляду

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-m_1}{\sigma_1}\right)^2 + \left(\frac{x_2-m_2}{\sigma_2}\right)^2\right]\right\} = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{x_1-m_1}{\sigma_1}\right)^2\right\} \times \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{x_2-m_2}{\sigma_2}\right)^2\right\} = p(x_1)p(x_2),$$

тобто щільність ймовірностей двовимірного вектора дорівнює добутку щільностей його координат. Звідси витікає, що у нормальному випадку із некорельованості координат витікає їх незалежність.

2.5. Збіжності у теорії ймовірностей. Граничні теореми

Ми вже зустрічались з тим, що у схемі Бернуллі незалежних випробувань, коли число випробувань прагне до нескінченності, у різних умовах існують різні границі. У одних умовах проста випадкова величина (яка характеризує скінченне число випробувань) прагне до дискретної випадкової величини з законом розподілу Пуассона,

а в інших умовах вона прагне до неперервної гаусової випадкової величини. У цьому підрозділі ми побачимо, що нічого дивного тут нема – обидва закони належать широкому класу законів з деякими загальними властивостями. Однак наведені приклади підкреслюють, по-перше, важливість граничних переходів самих по собі, а по-друге, важливість умов, за яких розглядаються відповідні граничні теореми.

Нижче розглядаються деякі класичні питання теорії ймовірностей, пов'язані з граничними переходами: закон великих чисел, класична гранична теорема та більш сучасні результати – граничні теореми у класі безмежно подільних законів розподілу.

2.5.1. Типи збіжностей випадкових величин

Розглянемо послідовність величин $\xi_1, \xi_2, \dots, \xi_n, \dots$, які задані на ймовірнісному просторі (Ω, F, P) .

Згідно з аксіоматизацією за Колмогоровим теорії ймовірностей, типи збіжностей послідовностей випадкових величин у відповідних позначеннях повторюють відповідні типи збіжностей вимірних функцій. Наведемо їх.

1. Послідовність випадкових величин $\{\xi_n\}$ збігається майже напевне (або з ймовірністю 1) до ξ , якщо

$$P\left\{\omega: \lim_{n \rightarrow \infty} |\xi_n(\omega) - \xi(\omega)| \neq 0\right\} = 0. \quad (2.75)$$

Ця збіжність позначається так:

$$\xi_n \xrightarrow{M.H} \xi \text{ або } \xi_n \xrightarrow{P=1} \xi. \quad (2.76)$$

2. Послідовність випадкових величин $\{\xi_n\}$ збігається за ймовірністю до ξ , якщо

$$\lim_{n \rightarrow \infty} P\{\omega: |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\} = 0. \quad (2.77)$$

Фізично це означає, що ймовірність великого відхилення прагне до нуля для будь-якого ε .

Збіжність позначається так:

$$\xi_n \xrightarrow{P} \xi. \quad (2.78)$$

У теорії функцій доводиться, що зі збіжності майже напевне витекає збіжність за ймовірністю:

$$\xi_n \xrightarrow{M.H} \xi \Rightarrow \xi_n \xrightarrow{P} \xi.$$

3. Послідовність випадкових величин $\{\xi_n\}$ збігається до ξ у середньоквадратичному, якщо

$$\lim_{n \rightarrow \infty} M(\xi - \xi_n)^2 = 0. \quad (2.79)$$

Використовуючи нерівність Чебишева (2.34a) одержуємо

$$0 \leq P\{|\xi - \xi_n| \geq \varepsilon\} \leq \frac{M(\xi - \xi_n)^2}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Тобто із збіжності майже напевне витікає збіжність у середньоквадратичному.

Збіжність позначається так:

$$\xi_n \xrightarrow{\text{сеп.}} \xi. \quad (2.80)$$

Відомо також (це теорема Ріса), що якщо послідовність збігається по ймовірності, то з неї можна виділити підпослідовність, яка збігається майже напевне. Відомо також, що для обмежених величин із збіжності за ймовірністю витікає збіжність у середньоквадратичному.

Співвідношення між цими типами збіжності можна подати так:

$$\left. \begin{array}{l} \xi_n \xrightarrow{\text{сеп.}} \xi \\ \xi_n \xrightarrow{\text{м.н.}} \xi \end{array} \right\} \Rightarrow \begin{array}{l} \xi_n \xrightarrow{P} \xi \\ (\xi_n - \text{обмежені}) \\ \xi_{n_k} \xrightarrow{\text{м.н.}} \xi \\ \downarrow \\ \xi_n \xrightarrow{\text{сеп.}} \xi \end{array} \quad (2.81)$$

Тут $\{\xi_{n_k}\}$ – підпослідовність послідовності $\{\xi_n\}$.

4. Збіжність за розподілом.

Крім типів збіжностей, які в інших позначеннях повторюють відповідні типи збіжностей послідовностей вимірних функцій, у теорії ймовірностей використовуються також інші типи збіжності, у тому разі – збіжність за розподілом. Ця збіжність характерна саме для теорії ймовірностей.

Дійсно, у теорії ймовірностей суттєвим поняттям є функція розподілу $F_\xi(x)$ випадкової величини $\xi(\omega)$. Вивчаються ті властивості $\xi(\omega)$, які виражаються через функцію розподілу. Тому тут випадкові величини, які мають однакову функцію розподілу, розглядаються як тотожні.

Послідовність випадкових величин $\{\xi_n\}$ збігається за розподілом до ξ , якщо послідовність їх функцій розподілу $\{F_{\xi_n}(x)\}$ збігається до $F_\xi(x)$ у точках неперервності:

$$\lim_{n \rightarrow \infty} F_{\xi_n}(x) = F_\xi(x). \quad (2.82)$$

Цю збіжність називають також слабкою збіжністю функцій розподілу.

Збіжність позначається так:

$$\xi_n \xrightarrow{\text{позн.}} \xi \text{ або } F_n(x) \xrightarrow{\text{сл}} F(x). \quad (2.83)$$

Доводиться, що із збіжності за ймовірністю витікає слабка збіжність функцій розподілу:

$$\xi_n \xrightarrow{P} \xi \Rightarrow \xi_k \xrightarrow{\text{позн.}} \xi (F_n(x) \xrightarrow{\text{сл}} F(x)). \quad (2.84)$$

2.5.2. Закон великих чисел

У теорії ймовірностей одним із центральних положень є закон великих чисел.

З точки зору, як теорії, так і практики відомо, що подія з ймовірністю, близької до одиниці, майже обов'язково наступить. Тоді кажуть про практичну вірогідність події.

Якщо подія має нульову ймовірність, кажуть, що подія практично неможлива. Але експериментально багатократно доведено, що події, ймовірності яких дуже близькі до нульових, все ж таки трапляються, якщо число випробувань (в кожному з яких вона може відбутися з тою самою ймовірністю) дуже велике.

Ці закономірності і вивчає закон великих чисел.

Розглянемо конкретні постановки задач.

Закон великих чисел у формі Бернуллі. Розглянемо схему Бернуллі незалежних випробувань. Позначимо: n – довжина серії випробувань; $p = P(A)$ – ймовірність «успіху»; ξ_n – число «успіхів» у серії, причому $\xi_n = 0, 1, 2, \dots, n$, $\mu_n = \frac{\xi_n}{n}$ – відносна частота «успіху»; причому

$$\mu_n = 0, \frac{1}{n}, \frac{2}{n}, \dots, 1.$$

Справджується таке твердження: Відносна частота «успіху» у схемі Бернуллі прагне за ймовірністю до ймовірності «успіху»:

$$\frac{\xi_n}{n} \xrightarrow{P} p, \text{ або } P\left(\left|\frac{\xi_n}{n} - p\right| \geq \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0. \quad (2.85)$$

Дійсно, використовуючи формулу Чебишева, маємо

$$P\left(\left|\frac{\xi_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{M\left(\frac{\xi_n}{n} - p\right)^2}{\varepsilon^2} = \frac{D\left(\frac{\xi_n}{n}\right)}{\varepsilon^2} = \frac{pq}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Тут $q=1-p$; $D\left(\frac{\xi_n}{n}\right) = \frac{pq}{n}$ (див. (2.40)).

Закон великих чисел у формі Чебишева. Нехай розглядається послідовність незалежних випадкових величин ξ_1, ξ_2, \dots . Припускається, що їх математичні сподівання однакові $M\xi_n = a$, а дисперсії обмежені – $D\xi_n < A$.

Справджується таке твердження: Послідовність середньоарифметичних $S_n = \frac{1}{n} \sum_{k=1}^n \xi_k$ прагне за ймовірністю до числа a :

$$S_n = \frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow{P} a, \text{ або } P(|S_n - a| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} 0. \quad (2.86)$$

Дійсно, як і вище, маємо

$$P(|S_n - a| \geq \varepsilon) \leq \frac{M(S_n - a)^2}{\varepsilon^2} = \frac{D(S_n)}{\varepsilon^2} = \frac{D\left(\sum_{k=1}^n \xi_k\right)}{n^2 \varepsilon^2} = \frac{\sum_{k=1}^n D\xi_k}{n^2 \varepsilon^2} \leq \frac{A}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Тут також враховано, що $D(S_n) = \frac{D(\sum \xi_k)}{n^2}$; ξ_n – незалежні, тобто

$D(\sum \xi_k) = \sum D\xi_k$; дисперсії ξ_n обмежені – $D\xi_k < A$.

Перевіримо, що a – середнє S_n . Дійсно:

$$M(S_n) = \frac{1}{n} \sum_{k=1}^n M\xi_k = a.$$

Статистична інтерпретація закону така: якщо ξ_n – результат n -го вимірювання величини a , то середньоарифметичне S_n приблизно дорівнює a ($S_n \approx a$).

Це твердження є узагальненням закону великих чисел у формі Бернуллі. Дійсно, якщо ξ_k – число «успіхів» у кожному випробуванні,

$$S_n = \frac{\sum_{k=1}^n \xi_k}{n} = \frac{\xi_n}{n}.$$

Підсилений закон великих чисел (у формі Бореля). Скажімо, що послідовність випадкових величин $\xi_1, \xi_2, \dots, \xi_n, \dots$ підпорядковується підсиленому закону великих чисел, якщо для будь-яких $\varepsilon > 0$ і $\eta > 0$ можна вказати таке n_0 , що для довільних s і n , що задовольняють нерівностям $n_0 \leq n \leq n_0 + s$, ймовірність нерівності

$$\max_{n_0 \leq n \leq n_0 + s} \left| \frac{\sum_{k=1}^n \xi_k}{n} - \frac{\sum_{k=1}^n M\xi_k}{n} \right| < \varepsilon \quad (2.87)$$

більша, ніж $1 - \eta$.

Це означення охоплює умови закону великих чисел у формі Чебишева та Бернуллі. Так у першому випадку

$$S_n = \frac{\sum_{k=1}^n \xi_k}{n},$$

$$a = \frac{\sum_{k=1}^n M\xi_k}{n},$$

тому що умовою теореми Чебишева є $M\xi_k = a$. Таким чином, маємо

$$\max_{n_0 \leq n \leq n_0 + s} |S_n - a| < \varepsilon. \quad (2.87a)$$

У другому випадку, коли $S_n = \frac{\xi_n}{n}$, безпосередньо одержуємо

$$\max_{n_0 \leq n \leq n_0 + s} \left| \frac{\xi_n}{n} - p \right| < \varepsilon. \quad (2.87б)$$

Справджується таке твердження (Борель):

Відносна частота «успіху» у схемі Бернуллі прагне майже напевне до ймовірності «успіху»: $\frac{\xi_n}{n} \xrightarrow{M.H.} p$, або

$$P\left(\left|\frac{\xi_n}{n} - p\right| \neq 0\right) \xrightarrow{n \rightarrow \infty} 0. \quad (2.88)$$

Закон великих чисел у формі Бернуллі згідно з (2.85) не стверджує, що у одному випробуванні при $n \rightarrow \infty$ $\frac{\xi_n}{n} \xrightarrow{n \rightarrow \infty} p$. Підсилений закон у

формі Бореля стверджує, що $\frac{\xi_n}{n} \xrightarrow{n \rightarrow \infty} p$ майже напевне, тобто усюди, за

виключенням точок, майже неймовірних. Інакше кажучи, для довільних $\varepsilon > 0$ і $\eta > 0$ можна вказати таке n_0 , що для будь-якого s ймовірність нерівності

$$\left| \frac{\xi_n}{n} - p \right| < \varepsilon$$

для всіх n , що задовольняють $n_0 \leq n \leq n_0 + s$, більша $1 - \eta$. Тобто тут

$$P\left(\frac{\xi_n}{n} \xrightarrow{n \rightarrow \infty} p\right) = 1;$$

закон великих чисел у формі Бернуллі дає лише

$$P\left(\left|\frac{\xi_n}{n} - p\right| < \varepsilon\right) \xrightarrow{n \rightarrow \infty} 1;$$

що витікає з (2.85), якщо взяти протилежну подію.

Наведемо ще узагальнення закону великих чисел:

1. Якщо дисперсії випадкових величин ξ_n обмежені тою самою постійною A , то послідовність взаємно незалежних випадкових величин $\xi_1, \xi_2, \dots, \xi_n, \dots$ підпорядковується підсиленому закону великих чисел.

2. Існування математичного сподівання є необхідною й достатньою умовою застосування підсиленого закону великих чисел до по-

слідовності однаково розподілених і взаємно незалежних випадкових величин.

Закон великих чисел дає теоретичне обґрунтування принципової можливості обчислення ймовірності по частоті «успіху» з якою завгодно високою точністю при достатньо великому числі спостережень.

Взагалі, як вже підкреслювалось, закон великих чисел є одним із основних положень теорії ймовірностей. У теорії ймовірностей і у реальній дійсності можна вказувати, які події є практично неможливими, а які є практично вірогідними. Однак, подія, яка має додатню ймовірність (як би мала ця ймовірність не була), може трапитись. При цьому, якщо число випробувань, при кожному з яких подія може трапитись з однаковою ймовірністю, дуже велико, то ймовірність хоча би його однократного виникнення може стати як завгодно близько до одиниці. Ось це і є суть закону великих чисел.

2.5.3. Класична гранична теорема

Ще раз повернемося до схеми Бернуллі незалежних випробувань в умовах збіжності до нормального закону. Придамо інтегральній теоремі Муавра-Лапласа іншу форму, яка дасть можливість розглянути проблеми дослідження у області класичної граничної теорема.

Позначимо через ξ_k число виникнення події A у k -тому випробуванні. Тоді число виникнення події A у n послідовних випробуваннях дорівнює $\xi_n = \sum_{k=1}^n \xi_k$; як було показано у (п. 2.3.1) $M \sum_{k=1}^n \xi_k = np$,

$D \sum_{k=1}^n \xi_k = npq$. Використовуючи ці данні, запишемо твердження інте-

гральної теорема Муавра-Лапласа у вигляді

$$P \left\{ a \leq \frac{\sum_{k=1}^n (\xi_k - M\xi_k)}{\sqrt{\sum_{k=1}^n D\xi_k}} < b \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz. \quad (2.89)$$

Це співвідношення означає, що розглядається сума відхилень незалежних випадкових величин від їх математичних сподівань. Сума нормована на корінь із суми дисперсій її доданків. Знаходиться ймовірність попадання цієї нормованої суми у інтервал $[a, b)$. Стверджу-

ється, що при збільшенні числа доданків цієї суми до нескінченності, шукана ймовірність рівномірно, відносно a і b , прагне до величини

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz = \Phi(b) - \Phi(a).$$

У області дослідження класичної граничної теореми ставилось питання: як взагалі пов'язані властивості доданків ξ_k з існуванням збіжності у співвідношенні (2.89). У класичних роботах було доведено, що на доданки необхідно накласти лише достатньо загальні обмеження – окремі доданки у деякому розумінні повинні незначно впливати на суму. Ця постановка задачі безпосередньо відбиває умови багатьох прикладних досліджень, де на досліджуване явище впливає велике число незалежно діючих випадкових факторів, кожен з яких дуже мало впливає на досліджуване явище.

Для математичного дослідження цієї проблеми замість вивчення суми дуже великого (але скінченного) числа доданків переходять до вивчення послідовності сум, де число членів раз від разу збільшується. Шукається гранична функція розподілу для послідовності функцій розподілу сум.

Таким чином, розглядається послідовність незалежних випадкових величин $\xi_1, \xi_2, \dots, \xi_n, \dots$, які мають математичні сподівання $M\xi_k = a_k$ та дисперсії $D\xi_k = b_k$. Позначаємо $B_n^2 = \sum_{k=1}^n b_k^2 = D \sum_{k=1}^n \xi_k$. За-

дача полягає у тому, що необхідно знайти умови, які накладаються на ξ_k , щоб розподіл суми

$$\frac{1}{B_n^2} \sum_{k=1}^n (\xi_k - a_k) \quad (2.90)$$

збігався до нормального закону. Ліндеберг показав, що умова

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{k=1}^n \int_{|x-a_k| > \epsilon B_n} (x-a_k)^2 dF_k(x) = 0 \quad (2.91)$$

є достатньою для цієї збіжності; пізніше Феллер довів, що вона є також і необхідною. Тут $\lambda > 0$, $F_k(x)$ – функція розподілу ξ_k . Можна довести, що умова (2.91) є по суті вимогою рівномірної малізми доданків $\frac{1}{B_n^2}(\xi_k - a)$ у сумі (2.90).

Справджується таке твердження:

Теорема Ліндеберга. Якщо послідовність взаємно незалежних випадкових величин $\xi_1, \xi_2, \dots, \xi_n \dots$ при довільнім постійним $\lambda > 0$ задовольняє умову

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{k=1}^n \int_{|x-a_k| > \tau B_n} (x-a_k)^2 dF_k(x) = 0,$$

то при $n \rightarrow \infty$ рівномірно відносно $x \in (-\infty, \infty)$

$$P \left\{ \frac{1}{B_n} \sum_{k=1}^n (\xi_k - a_k) < x \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz.$$

З цієї теореми витікає ряд наслідків, які до неї розглядалися як часткові розв'язки центральної граничної задачі.

Наслідок 2.1 Якщо незалежні випадкові величини $\xi_1, \xi_2, \dots, \xi_n \dots$ однаково розподілені і мають скінченну відмінну від нуля дисперсію, то при $n \rightarrow \infty$ рівномірно по x

$$P \left\{ \frac{1}{B_n} \sum_{k=1}^n (\xi_k - M\xi_k) < x \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz.$$

Наслідок 2.2 (теорема Ляпунова) Якщо для послідовності взаємно незалежних випадкових величин $\xi_1, \xi_2, \dots, \xi_n \dots$ можна підібрати таке додатне число $\delta > 0$, що при $n \rightarrow \infty$

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^n M |\xi_k - a_k|^{2+\delta} \rightarrow 0,$$

то при $n \rightarrow \infty$ рівномірно по x

$$P \left\{ \frac{1}{B_n} \sum_{k=1}^n (\xi_k - a_k) < x \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz.$$

Достатньо довгий час умова Ліндеберга вважалася достатньою умовою нормалізації і різні відомі теореми, пов'язані з класичною проблемою нормалізації, було цікаво розглядати з позицій виконання цієї умови. Однак, коли Феллер довів, що ця умова є і необхідною умовою, стало ясно, що будь яка теорема, пов'язана з нормалізацією обов'язково є частинним випадком (наслідком) теореми Ліндеберга-Феллера.

2.5.4. Сучасні граничні теореми

До тридцятих років 20-го століття під центральною задачею теорії ймовірностей розуміли відшукування умов нормалізації, тобто знаходження умов, за яких функція розподілу сум незалежних випадкових величин збігається до нормального закону, навіть дотепер у ряді книг навчального характеру збереглася ця точка зору. Наприклад, часто вживають термін «центральна гранична теорема теорії ймовірностей», що підкреслює саме таку точку зору. У тридцяті роки були знайдені необхідні і достатні умови нормалізації і тим самим завершена класична проблематика. Вона почала розглядатися у більш широкому розумінні – досліджувався цілий клас законів розподілу, які можуть бути типовими для сум незалежних випадкових величин. Так прийшли до класу, так званих, безмежно подільних законів, одним із яких є нормальний закон. Таким чином центральна проблема теорії ймовірностей, як вона розглядалась до 30-х років, стала рядовим частинним випадком загальної проблеми вивчення класу нескінченно подільних розподілів. Цей період почався в 30-ті роки і продовжувався далі.

Безмежно подільні закони розподілу. Функція розподілу $F(x)$ та її характеристична функція $f(u)$ називаються безмежно подільними, якщо при кожному додатному n на фіксованому ймовірнісному просторі (Ω, F, P) існує n незалежних і однаково розподілених випадкових величин $G_{n_k}(\omega)$, $k=1, n$, для яких функція розподілу суми

$$G_n(\omega) = \sum_{k=1}^n G_{n_k}(\omega), \quad (2.92)$$

збігається до $F(x)$, або при кожному n існує така однакова для всіх n_k характеристична функція $f^{n_k}(u)$, що

$$f(n) = f^{(n)}(u) = \left[f^{(n_k)}(u) \right]^n. \quad (2.93)$$

Термін «безмежно подільний» відносять також до випадкової величини. Це виправдано тим, що вона для будь-якого n подається у вигляді суми n незалежних випадкових величин

$$\xi(\omega) = \xi_1(\omega) + \xi_2(\omega) + \dots + \xi_n(\omega), \quad (2.94)$$

які мають однаковий закон розподілу. Така властивість ширше відома для гаусової випадкової величини.

До безмежно подільних законів розподілу, крім нормального, належать пуассонів, гамма-розподіл, логарифмічно нормальний, Лапласа і такі інші.

Стосовно послідовності сум незалежних випадкових величин (2.92) Хінчин довів, що сім'я всіх законів розподілу для сум $G_{n_k}(\omega)$ збігається з класом усіх безмежно подільних розподілів.

Для цього класу розподілів природно описувати випадкову величину $G_n(\omega)$ характеристичною функцією, яка у довільному випадку незалежних доданків дорівнює добутку їх характеристичних функцій

$$f^n(u) = \prod_{k=1}^n f^{n_k}(u).$$

На мові функцій розподілу ця випадкова величина описується композицією відповідних функцій розподілу

$$F^n(x) = F^{n_1}(x) * F^{n_2}(x) * \dots * F^{n_n}(x).$$

Канонічні зображення безмежно подільних законів. Характеристична функція безмежно подільних випадкових величин може бути зображена у деяких канонічних формах – Леві-Хінчина, Леві та Колмогорова. Наведемо дані канонічної форми Колмогорова, яка буде використана далі.

Справджується таке твердження для випадкових величин зі скінченною дисперсією.

Теорема 2.1 (А.М. Колмогорова). Для того, щоб функція розподілу $F(x)$ зі скінченною дисперсією була безмежно подільною, необхідно та достатньо, щоб логарифм її характеристичної функції мав вигляд

$$\ln f(u) = jmu + \int_{-\infty}^{+\infty} \frac{e^{jux} - 1 - jux}{x^2} dK(x), \quad (2.95)$$

де $K(x)$ визначається однозначно по заданій характеристичній функції $f(u)$ у всіх точках неперервності,

$$K(x) = \frac{1}{2\pi} \lim_{\nu \rightarrow \infty} \lim_{\tau \rightarrow \infty} \int_{-\tau}^{+\tau} \frac{e^{-j\nu y} - e^{-j\nu x}}{j\nu} \frac{d^2}{du^2} \ln f(u) du. \quad (2.96)$$

Приймається $K(-\infty) = 0$.

Наведемо приклади.

Випадкова величина має нормальний розподіл $N(a, \sigma^2)$ з дисперсією σ^2 та математичним сподіванням a .

Враховуючи, що

$$M\xi = (-j) \frac{d}{du} \ln f(u) \Big|_{u=0};$$

$$D\xi = \left[\frac{d^2}{du^2} \ln f(u) \right]_{u=0},$$

із (2.95) можна знайти

$$D\xi = \int dK(x) = K(+\infty).$$

Тоді для нормального закону одержуємо

$$m = a; \quad K(x) = \begin{cases} 0, & x < 0; \\ \sigma^2, & x > 0. \end{cases} \quad (2.97)$$

Дійсно, ця функція і постійна m приводять до даного закону, тому що

$$\int_{-\infty}^{\infty} (e^{jux} - 1 - jux) \frac{1}{x^2} dK(x) = \lim_{\tau \rightarrow 0} \frac{e^{j\mu\tau} - 1 - j\mu\tau}{\tau^2} [K(+0) - K(-0)] = -\frac{u^2 \sigma^2}{2},$$

а завдяки єдиності канонічного зображення (2.95) інші функції $K(x)$ не можуть дати нормального закону.

Випадкова величина має Пуассонів закон розподілу з характеристичною функцією

$$f(u) = e^{\lambda(e^{jua} - 1) + jbu}.$$

Аналогічно можна одержати

$$K(x) = \begin{cases} 0, & x < a; \\ a^2 \lambda, & x > a, \end{cases} \quad (2.98)$$

$$m = b + a\lambda.$$

Слід підкреслити, що безмежно подільний закон є або композицією скінченного числа законів Пуассона і нормального закону, або границею послідовності таких законів, коли розглядається рівномірна збіжність. Тобто ці два закони є основними «цеглинками», із яких складається будь-який закон із класу нескінченно подільних розподілів. Дуже важливою характеристикою цього класу є його замкненість відносно лінійності операцій.

Граничні теореми для безмежно подільних законів. Граничні теореми цього класу зручно формулювати, користуючись метрикою Леві $L(.,.)$ як відстанню між функціями розподілу в просторі розподілів:

$$L(G, F) = \inf \{h: F(x-h) - h \leq G(x) \leq F(x+h) + h, -\infty < x < \infty\}. \quad (2.99)$$

Тобто ми розглядатимемо збіжність за розподілом, але на відміну від слабкої збіжності (п.2.5.1) цю збіжність називають сильною.

Сформулюємо в термінах метрики Леві три з можливих граничних теорем для послідовності сум незалежних випадкових величин. Вони розглядаються в умовах, коли доданки послідовностей сум (2.92) при будь яких $n=1,2,\dots$ задовольняють умову рівномірної нескінченної мализни, або умову граничної нехтуваності

$$\lim_{n \rightarrow \infty} \max_k P\{G_{n_k}(\omega) \geq \varepsilon\} = 0 \quad (2.100)$$

для кожного $\varepsilon > 0$. Це означає, що при $n \rightarrow \infty$ роль кожного доданка в до граничній сумі для

$$G(\omega) = \lim_{n \rightarrow \infty} \sum_{k=1}^n G_{n_k}(\omega) \quad (2.101)$$

стає як завгодно малою.

Теорема 2.2 (Ліндеберга-Феллера). Нехай доданки в $G_n(\omega)$ є гранично нехтуваними, мають нульові математичні сподівання та скінченні дисперсії, причому $M[G_{n_k}(\omega)]^2 = 1$.

Тоді співвідношення $L(F^n, \Phi) \rightarrow 0$, (Φ – стандартний нормальний закон) еквівалентне умові Ліндеберга:

$$\Lambda_n(\delta) = \sum_{k: |x| > \delta} \int x^2 dF^{(n_k)}(x) \rightarrow 0$$

при кожному $n \rightarrow \infty$ і кожному $\delta > 0$.

Нагадаємо, як вже підкреслювалось у (п. 2.5.3), що теорема Ліндеберга-Феллера дає необхідну і достатню умову збіжності $F^n(x)$ до нормального закону.

Позначимо Π закон Пуассона з характеристичною функцією

$$f(u; 0; \lambda) = \exp[\lambda(e^{ju} - 1 - ju)].$$

Теорема 2.3. Нехай доданки в $G_n(\omega)$ є гранично нехтуваними, мають нульові математичні сподівання та скінченні дисперсії, причому

$$M[G_{n_k}(\omega)]^2 = \lambda > 1.$$

Тоді співвідношення

$$L(F^{(n)}, \Pi) \rightarrow 0$$

еквівалентне співвідношенню

$$\sum_k \int_{|x-1|>\delta} x^2 dF^{(n_k)}(x) \rightarrow 0$$

при кожному $n \rightarrow \infty$ і кожному $\delta > 0$.

Теорема 2.4. Нехай доданки в $G_n(\omega)$ є гранично нехтуваними та мають нульові математичні сподівання, а дисперсії сум рівномірно обмежені.

Тоді співвідношення

$$L(F^{(n)}, G) \rightarrow 0$$

(G -безмежно подільний закон, зображення логарифма характеристичної функції якого за формулою Колмогорова визначається функцією $K(x)$) еквівалентне умові:

$$K^{(n)}(x) = \sum_{k=-\infty}^x y^2 dF^{n_k}(x) \rightarrow K(x).$$

Ці теореми демонструють розширення класичних уявлень щодо збіжності сум незалежних випадкових величин до нормального розподілу. Насправді вони мають збіжність до деякого закону з класу безмежно подільних законів. Більш того, умова граничної нехтуваності, яка є загальною для всіх наведених граничних теорем, у технічній літературі навіть розглядають як достатню умову нормалізації — збіжності до нормального закону. Це у ряді прикладних задач теорії електров'язку може привести до принципових помилок.

Висновки

1. Для теорії електров'язку основними у класичній теорії ймовірностей є поняття випадкової події, випадкової величини, випадкового вектора, які врешті решт входять у означення випадкового сигналу (процесу).

2. Означення і властивості випадкових подій базуються на аксіоматизації Колмогорова теорії ймовірностей. Важливим є модель не-

залежних випробувань, означення умовної ймовірності як нормованої звуженої міри, формули повної ймовірності та Байєса.

3. Випадкова величина означається як вимірне відображення імовірнісного простора у дійсну вісь. Відповідно функція розподілу вводиться як частинний випадок індукованої міри; щільність імовірностей — як похідна функції розподілу (коли така похідна існує) та характеристична функція — як перетворення Фур'є функції розподілу або щільності імовірностей; числові характеристики (моменти, кумулянти) дають деяку неповну «інформацію» щодо випадкової величини.

4. Важливою для теоретичних досліджень та практики є схема Бернуллі незалежних випробувань. Така схема визначає просту випадкову величину з біноміальним законом розподілу. Існують два діаметрально протилежних приклада граничних переходів на базі схеми Бернуллі. В умовах, коли розглядаються дуже рідкі події у дуже довгій серії, одержується дискретна випадкова величина, яка підпорядкована закону Пуассона. У другому випадку, коли розглядається нормована частина успіху, у разі граничного переходу одержується неперервна випадкова величина з гаусовим законом розподілу.

5. Випадковий вектор означається як вимірне відображення імовірнісного простору у n -вимірний евклідів простір. Його називають також векторною випадковою величиною — це сукупність n скалярних випадкових величин. У ідейному плані випадковий вектор описується аналогічно випадковій величині — n -вимірною функцією розподілу, n -вимірною щільністю ймовірностей, n -вимірною характеристичною функцією. Використовуються також числові характеристики вектора, перш за все, — математичне сподівання та кореляційна матриця.

6. Класичними проблемами теорії ймовірностей є закон великих чисел та умови нормалізації — умови збіжності сум послідовностей незалежних випадкових величин до нормального закону.

Закон великих чисел у підсиленому розумінні має такий зміст: теоретично та фізично можна стверджувати, які події є практично вірогідними. Практично вірогідні події відбуваються з імовірністю близькою до одиниці. Але, якщо ймовірність події як завгодно мала, вона може трапитись. Більш того, як стверджує закон великих чисел, якщо число випробувань дуже велике, ймовірність хоча би одного виникнення такої події може стати як завгодно близькою до одиниці.

На протязі довгого часу центральною проблемою класичної теорії ймовірностей було відшукування найбільш загальних умов, за яких

функції розподілу сум незалежних випадкових величин збігаються до нормального закону. Достатньо загальні умови дає теорема Ляпунова. Необхідну і достатню умову дає теорема Ліндеберга-Феллера. Умова Ліндеберга по суті дає сучасну вимогу рівномірної мализни доданків у розглядуваній сумі незалежних випадкових величин, нормованих відповідним чином.

7. Сучасна точка зору на центральну задачу теорії ймовірностей полягає у тому, що ця задача є частинним випадком більш загальної проблеми знаходження умов, які накладаються на доданки суми незалежних величин, коли існують будь-які границі для послідовностей функції розподілу таких сум.

Клас таких граничних законів збігається з класом безмежно подільних законів. Нормальний закон є один із представників такого класу. Другим представником є закон Пуассона. Більш того, будь-який безмежно подільний закон є або композицією скінченного числа законів Пуассона і нормального закону, або границею послідовностей таких законів. Клас безмежно подільних законів замкнений відносно лінійних операцій.

Запитання, задачі й вправи

1. Дайте означення стохастичного експерименту.
2. Що є математичною моделлю стохастичного експерименту згідно з аксіоматизацією Колмогорова теорії ймовірностей?
3. Як вводиться модель незалежних випробувань?
4. Що таке умовна ймовірність події?
5. Довести формули повної ймовірності та Байєса.
6. Що таке випадкова величина і як вона повністю описується?
7. Як обчислюються моменти та кумулянти випадкової величини?
8. Які властивості характерні для математичного сподівання та дисперсії випадкової величини?
9. Охарактеризуйте схему Бернуллі незалежних випробувань і біноміальний закон розподілу.
10. Охарактеризуйте закон розподілу Пуассона як границю розподілів у схемі Бернуллі для дуже рідких подій і дуже довгої серії.
11. Охарактеризуйте нормальний закон у межах теореми Муавра-Лапласа як граничний для нормованої частоти успіху у схемі Бернуллі.
12. Дайте означення випадкового вектора та охарактеризуйте можливості його повного описування.

13. Наведіть приклади законів розподілу випадкових векторів.
14. Охарактеризуйте основні типи збіжностей у теорії ймовірностей.
15. У чому полягає закон великих чисел?
16. Сформулюйте закон великих чисел у формі Бернуллі, Чебишева, Бореля.
17. У чому полягає класична гранична теорема теорії ймовірностей?
18. Охарактеризуйте суть теореми Ляпунова і Ліндеберга.
19. Що таке безмежно подільний закон розподілу?
20. Що таке умова рівномірної нескінченної мализни, або умова граничної нехтуваності?
21. У чому суть сучасних граничних теорем теорії ймовірностей?
22. Охарактеризуйте основні граничні теореми для безмежно подільних законів розподілу.
23. Задана умовна ймовірність $P(B|A)$, причому $P(A) \neq 0$.
 - а – Знайти значення $P(A|A)$, $P(\emptyset|A)$, $P(B|A)$, якщо $B \supseteq A$.
 - б – Довести властивість $P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A)$.
 - в – За умови незалежності подій A і B довести незалежність подій A і \overline{A} , \overline{A} і B , A і \overline{B} .
24. По каналу зв'язку передається п'ять повідомлень. Кожне з них (незалежно від решти) спотворюється з імовірністю 0,2. Знайти імовірність таких подій:

$A = \{\text{всі повідомлення передано без спотворень}\};$
 $B = \{\text{всі повідомлення будуть спотворені}\}.$
25. По каналу зв'язку передаються два сигнали (події A_1 і A_2) з імовірностями $P(A_1) = 2P(A_2)$. Канал зв'язку спотворює перший сигнал (переводить його в другий) з імовірністю 0,1, а другий (переводить його в перший) – з імовірністю 0,2. Якою буде ймовірність одержання першого сигналу?
26. Дана гаусова випадкова величина з параметрами a і σ^2 . Обчислити її характеристичну функцію.
27. Випадкова величина ξ має рівномірний розподіл на $[a, b]$. Знайдіть математичне сподівання $M\xi$ та дисперсію $D\xi$.
28. Дано випадковий вектор $[\xi_1(\omega), \xi_2(\omega)]$ із функцією розподілу $F(x_1, x_2)$. Знайти функцію розподілу $F_n(x)$ випадкової величини $\eta(\omega) = \max\{\xi_1(\omega), \xi_2(\omega)\}$.

29. Знайти функцію розподілу випадкового вектора $\vec{\xi}=(\xi_1, \dots, \xi_n)$, який має рівномірний розподіл у паралелепіпеді $a_i \leq \xi_i \leq b_i$, $1 \leq i \leq n$.
30. Довести, що потік викликів у телефонії за умови незалежності, однорідності та ординарності має закон розподілу Пуассона.
31. Нехай $\xi_1, \xi_2, \dots, \xi_n$ – незалежні випадкові величини, які мають гаусів розподіл з однаковими параметрами a і σ^2 . Знайти функцію розподілу величини

$$\chi^2 = \frac{1}{\sigma^2} \sum_{k=1}^n (\xi_k - a)^2.$$

32. Дана послідовність випадкових величин $\{\xi_k\}$, для яких $D\xi_k < c$, $R_{ij} \rightarrow 0$ при $|i - j| \rightarrow \infty$. Довести, що до даної послідовності можна застосувати закон великих чисел.
33. Дана випадкова величина зі щільністю ймовірності

$$p(x) = \begin{cases} 0, & x \leq 0; \\ \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x > 0, \end{cases}$$

де $\alpha > 0$, $\beta > 0$. Довести, що ця випадкова величина безмежно подільна. Обчислити функцію $K(x)$ та параметр m у канонічному зображенні логарифма характеристичної функції.

3. ВИПАДКОВІ ПРОЦЕСИ

Випадкові процеси мають особливо велике значення у теорії електров'язку, тому що вони описують сигнали, тобто фізичні процеси, які застосовуються для передання повідомлень. З точки зору електров'язку вивчаються фізичні явища через описуючі їх випадкові функції – моделі сигналів. Тільки ті характеристики фізичного процесу, які є у наявності у моделі, і можна вивчати по моделі. Модель сигналу вибирається, з одного боку, з точки зору її повноти та достатності для розв'язання тих чи інших задач електров'язку; з другого боку, вона повинна бути достатньо простою, придатною для практичного застосування.

Вимога, яка висувається до випадкових функцій як до моделей фізичних процесів, приводить до необхідності їх різнобічного вивчення. Широке коло різноманітних задач електров'язку додатково вимагає вивчення різних типів випадкових процесів.

У цьому розділі випадкові процеси вивчаються як єдині математичні об'єкти з властивим для них математичним описуванням; розглядаються можливості їх більш-менш однобічного вивчення на рівні кореляційної теорії (але достатньої для розв'язання деякого кола задач практики); приводиться декілька замкнених класів випадкових процесів, які відносяться до класичних, але й достатньо сучасних процесів, які дають можливість розв'язати нові задачі, висуванні практикою; підкреслюються також перспективні напрямки у розвитку цих моделей.

3.1. Основні означення та підходи до вивчення випадкових процесів

Теорія випадкових процесів – це гілка теорії ймовірностей, яка, у свою чергу, з математичної точки зору може розглядатися як гілка загальної теорії функцій, міри та функціонального аналізу [2,7,9].

3.1.1. Означення випадкових функцій

Випадкова функція – це параметрична сім'я

$$\Xi_{\Theta}(\omega) = \{\xi_{\nu}(\omega), \nu \in \Theta\} \quad (3.1)$$

випадкових величин, визначених на імовірнісному просторі $(\Omega, \mathcal{F}, \mathcal{P})$, яка задовольняє деякі умови.

Випадковим процесом називають випадкову функцію $\Xi_{\Theta}(\omega)$, якщо абстрактний параметр ν тлумачиться як час t з інтервалу T :

$$\Xi_T(\omega) = \{\xi(\omega, t), t \in T\}. \quad (3.1a)$$

Коли T – зчисленна множина чисел, говорять про часову послідовність, процес з дискретним часом тощо:

$$\Xi_Z(\omega) = \{\xi(\omega, k), k \in Z\}, \quad (3.1б)$$

де $Z = \{0, \pm 1, \pm 2, \dots\}$.

Якщо Θ – множина багатовимірних точок Q (точок скінченновимірного евклідового простору), сім'ю $\Xi_\Theta(\omega)$ називають *випадковим полем*:

$$\Xi_Q(\omega) = \{\xi(\omega, \vec{q}), \vec{q} \in Q\}. \quad (3.1в)$$

Зокрема, коли $\vec{q} = (\vec{r}, t), \vec{r} = (x, y, z)$ – точки тривимірного евклідового простору, маємо *просторово-часове поле*:

$$\Xi_Q(\omega) = \{\xi(\omega, t, x, y, z), t \in T, x \in X, y \in Y, z \in Z\}. \quad (3.1г)$$

Сім'ї $\Xi_\Theta(\omega)$ можуть бути компонентами деякого випадкового вектора

$$\vec{\Xi}_{\Theta, S}(\omega) = (\Xi_{\Theta_1}(\omega), \Xi_{\Theta_2}(\omega), \dots, \Xi_{\Theta_S}(\omega)), \quad (3.2)$$

який визначає відповідно векторний випадковий процес або векторне випадкове поле.

Випадкову функцію можна стисло позначити $\xi(\omega, \theta)$, $\omega \in \Omega$, $\theta \in \Theta$; у прикладних роботах часто не позначається аргумент ω ; лише на словах кажуть, що розглядаються випадкова функція.

Як і усяку функцію, випадкову функцію строго визначають як відображення, коли елементу однієї множини поставлено у відповідність елементи іншої. При цьому сім'ю $\{\xi(\omega, t), \omega \in \Omega, t \in T\}$ можна охарактеризувати принаймні так: або як упорядковану відносно параметра $t \in T$ множину випадкових величин; або як множину числових функцій, кожна з яких розглядають як одну із елементарних подій, що відповідає деякому $\omega \in \Omega$; або як вимірну функцію на добутку $\Omega \times T$. Подамо відповідні означення стосовно випадкового процесу.

Означення 1 Випадковим процесом називається функція, яка відображає множину T у простір S випадкових величин, визначених на імовірнісному просторі (Ω, F, P) : $T \xrightarrow{\xi} S$ (рис. 3.1).

Значення $\xi(\omega, t)$ у точках $t \in T$ є випадковими величинами.

Це означення будемо називати *S-означенням*.

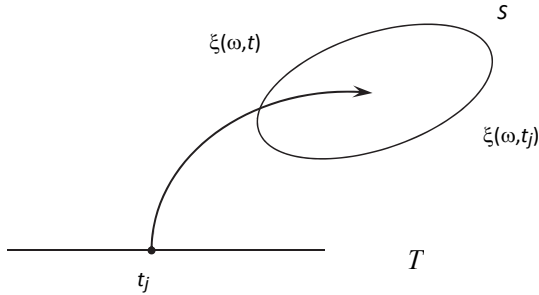


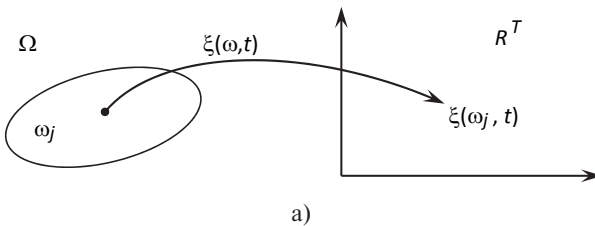
Рис.3.1. – Ілюстрація S -означення випадкового процесу як відображення множини T у простір S випадкових величин

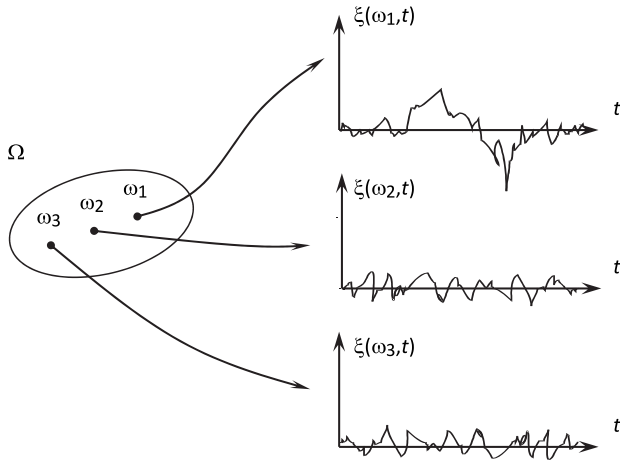
Означення 2. Випадковий процес є вимірною функцією, яка відображує ймовірнісний простір (Ω, \mathcal{F}, P) у вибірковий простір $(R^T, \mathcal{J}^T) : \Omega \xrightarrow{\xi} R^T$. Тут $R^T = \prod_{t \in T} R^t$ – прямий добуток незчисленних множин R^t – є простором усіх дійсних функцій; \mathcal{J}^T – σ -алгебра, побудована на множинах із R^T .

Значення процесу у точках ω називають *вибірковими функціями*, реалізаціями, траєкторіями випадкового процесу.

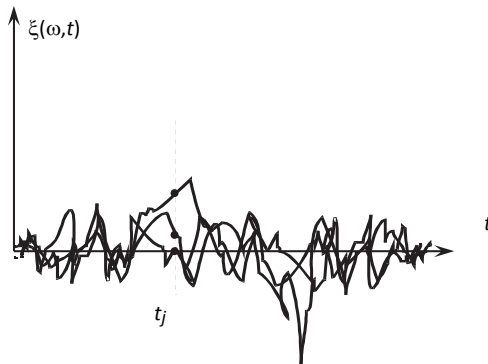
Означення 3. $F \times B$ – *вимірним випадковим процесом* називають вимірну функцію, яка відображує $(\Omega \times T, F \times B)$ у простір $(R^1, \mathcal{J}^1) : \Omega \times T \xrightarrow{\xi} R^1$.

Тут $\Omega \times T$ – прямий добуток множин Ω і T , що є сукупністю пар (ω, t) , $\omega \in \Omega$, $t \in T$; $F \times B$ – σ -алгебри підмножин із $\Omega \times T$, породжена F і B .





б)



в)

Рис.3.2. — Ілюстрація вибіркового означення випадкового процесу: а) відображення Ω у простір R^T усіх функцій; б) ілюстрація реалізацій процесу як елементарних подій; в) ілюстрація випадкової величини як перерізу процесу при $t=t_j$.

Кожний переріз у точці t ($F \times B$)-вимірною процесу є випадковою величиною, переріз у точці ω дає реалізацію процесу, яка є B -вимірною функцією. При цьому говорять про вимірний процес. Для застосувань найважливішим є таке означення.

Означення 4. (Справджується для сепарабельних випадкових процесів, які визначаються зчисленою множиною їх значень). Випадковим процесом $\xi(\omega, t)$, який заданий на множині $T (t \in T)$ і набуває дійсних значень, називають сім'ю розподілів

$$\{F_n(x_1, \dots, x_n; t_1, \dots, t_n)\}, \quad t_j \in T,$$

які задовольняють дві умови, сформульовані Колмогоровим:

а) узгодженості:

$$F_{n+p}\left(x_1, \dots, x_n, \underbrace{\infty, \dots, \infty}_p; t_1, \dots, t_n, t_{n+1}, \dots, t_{n+p}\right) = F_n(x_1, \dots, x_n; t_1, \dots, t_n); \quad p=1, 2, \dots;$$

б) симетрії:

$$F_n(x_{i_1}, \dots, x_{i_n}; t_{i_1}, \dots, t_{i_n}) = F_n(x_1, \dots, x_n; t_1, \dots, t_n),$$

де i_1, i_2, \dots, i_n – будь яке переставлення індексів.

Функцію

$$F_n(x_1, \dots, x_n; t_1, \dots, t_n) = P\{\xi(t_1) \leq x_1, \dots, \xi(t_n) \leq x_n\} \quad (3.3)$$

називають *скінченновимірним розподілом процесу* Рис.3.3.

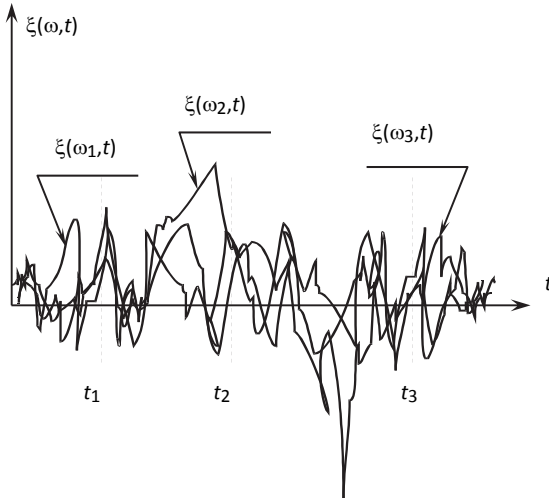


Рис. 3.3. – Ілюстрація впровадження n -вимірної функції розподілу n -вимірного вектора, який визначається n перерізами випадкового процесу у точках t_1, t_2, \dots, t_n .

3.1.2. Основні підходи до вивчення випадкових процесів

1. S -означення ВП

Коли дають S -означення випадкового процесу $T \xrightarrow{\xi} S$, у загальному випадку вважають, що множина T – це деяка множина евклідової прямої R^1 , а простір S – простір випадкових величин. Звичайно враховують момент розподілу лише n -го чи навіть другого порядку. В останньому разі використовують гільбертів простір випадкових величин $L_{2\xi} = L_2(\Omega, F, P)$. Випадковий процес називають *гільбертовим*, якщо при кожному $t \in T$ $\xi(\omega, t) \in L_{2\xi}$. Коли t пробігає всі значення з T , значення процесу описують криву в просторі $L_{2\xi}$ (рис. 3.4).

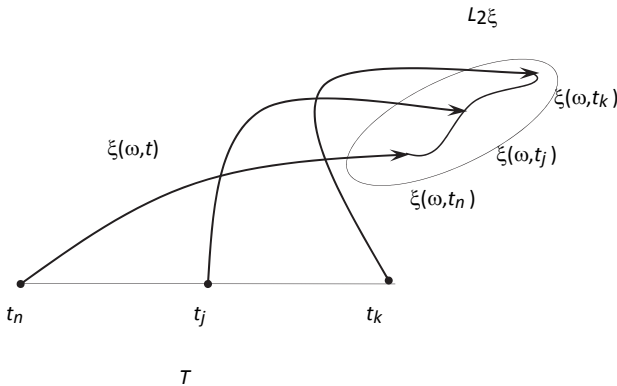


Рис. 3.4. – Зображення випадкового процесу як кривої у гільбертовому просторі випадкових величин; t_n – початкове значення параметра, t_k – кінцеве значення параметра

Випадкові процеси як криві в гільбертовому просторі вивчає кореляційна теорія, яка враховує моменти перших двох порядків – математичне сподівання $M\xi(\omega, t)$ і кореляційну функцію

$$\begin{aligned} R(t, s) &= M[\xi(\omega, t) - M\xi(\omega, t)][\overline{\xi(\omega, s) - M\xi(\omega, s)}] = \\ &= M[\xi(\omega, t)\overline{\xi(\omega, s)}] - M\xi(\omega, t)M\overline{\xi(\omega, s)}. \end{aligned} \quad (3.4)$$

Останню можна подати через скалярний добуток у $L_{2\xi}$:

$$R(t, s) = (\xi(\omega, t), \xi(\omega, s)) - (\xi(\omega, t), 1)(\overline{\xi(\omega, s)}, 1). \quad (3.4a)$$

У теоретичних дослідженнях звичайно покладають $M\xi(\omega, t) = 0$. Це ми і будемо робити далі.

Кореляційна теорія випадкових процесів – це частина теорії, яка повністю визначається властивостями кореляційної функції. Кажуть також: теорія ВП другого порядку; L_2 – теорія випадкових процесів. Це синоніми.

Характеристичною властивістю кореляційної функції $R(t,s)$ є її невід’ємна визначеність: якими б не були n , t_1, \dots, t_n і числа a_1, \dots, a_n виконується умова

$$\sum_{j,k=1}^n R(t_j, t_k) a_j \overline{a_k} \geq 0. \quad (3.5)$$

Якщо нерівність (3.5) строго виконується для всіх ненульових a_j, a_k , функція $R(t,s)$ додатньо визначена. При цьому припускається, що $R(t,s)$ неперервна і обмежена.

Щоб підкреслити важливість цієї властивості кореляційної функції, наведемо такий класичний результат:

Теорема 3.1. Визначена на $T \times T$ функція $R(t,s)$ є кореляційною функцією тоді і тільки тоді, коли вона невід’ємно визначеного типу.

З цієї теореми витікає, що для будь-якого процесу другого порядку кореляційна функція завжди невід’ємно визначена. Навпаки, якщо взяти будь-яку функцію з класу функцій невід’ємно визначеного типу, існує випадковий процес другого порядку, для якого вибрана функція є кореляційною.

Усі часткові властивості кореляційної функції витікають з характеристичної властивості (3.5).

Коли вивчається випадковий процес у рамках кореляційної теорії, його властивості визначаються властивостями кореляційної функції.

Демо визначення аналітичних властивостей ВП у середньоквадратичному – с.к. властивості.

Випадковий процес середньоквадратично неперервний у точці t_0 (с.к. неперервний), якщо

$$M|\xi(\omega, t) - \xi(\omega, t_0)|^2 \rightarrow 0 \text{ при } t \rightarrow t_0. \quad (3.6)$$

Умова (3.6) може бути більш розгорнуто записана так:

$$\lim_{t \rightarrow t_0} \int_{\Omega} |\xi(\omega, t) - \xi(\omega, t_0)|^2 dP = 0. \quad (3.6a)$$

Стисло прийнято записувати:

$$l.i.m._{t \rightarrow t_0} \xi(\omega, t) = \xi(\omega, t_0), \quad (3.6b)$$

де *l.i.m.* — *limit in medio*, тобто границя у середньому.

Якщо процес с.к. неперервний у кожній точці $t \in T$, його називають с.к. неперервним.

Використовуючи (3.6) неважко довести такий результат:

Теорема 3.2. Для с.к. неперервності випадкового процесу у точці t_0 необхідно і достатньо неперервність кореляційної функції $R(t,s)$ у точці (t_0, t_0) .

Аналогічно с.к. неперервності ВП вводять його с.к. диференційовність: випадковий процес $\xi(\omega, t)$ називається с.к. диференційовним у точці t_0 , якщо існує границя

$$M \left| \frac{\xi(\omega, t_0 + h) - \xi(\omega, t_0)}{h} - \xi'(\omega, t_0) \right|^2 \rightarrow 0 \quad (3.7)$$

при $h \rightarrow 0$. Випадкову величину $\xi'(\omega, t_0)$ називають с.к. похідною випадкового процесу в точці t_0 .

Границю (3.7) можна записати у іншому вигляді

$$\lim_{h \rightarrow \infty} \int_{\Omega} \left| \frac{\xi(\omega, t_0 + h) - \xi(\omega, t_0)}{h} - \xi'(\omega, t_0) \right|^2 dP = 0 \quad (3.7a)$$

або у символічному стислому запису :

$$l.i.m. \frac{\xi(\omega, t_0 + h) - \xi(\omega, t_0)}{h} = \xi'(\omega, t_0). \quad (3.7b)$$

Якщо процес $\xi(\omega, t)$ с.к. диференційовний у кожній точці $t \in T$, кажуть, що ВП — с.к. диференційовний.

Із співвідношень (3.7) можна одержати такий результат:

Теорема 3.3. Нехай $\xi(\omega, t)$, $t \in T$ гільбертів випадковий процес, для якого існує при кожному $t \in T$ узагальнена похідна кореляційної функції

$$\left. \frac{\partial^2 K(t,s)}{\partial t \partial s} \right|_{t=s}^0.$$

Тоді процес $\xi(\omega, t)$ с.к. диференційовний на T .

Аналогічно вводиться с.к. інтегровність випадкового процесу: випадковий процес с.к. інтегровний, якщо існує границя інтегральних сум

$$M \left| \sum_{k=1}^n \xi(\omega, t_{n_k}) \Delta t_{n_k} - J(\omega) \right|^2 \rightarrow 0 \quad (3.8)$$

при $n \rightarrow \infty$, $\Delta t_{n_k} \rightarrow 0$, $\Delta t_{n_k} = t_{n_k} - t_{n_{k-1}}$, $a = t_{n_0} < t_{n_1} < \dots < t_{n_n} = b$;

тут $J(\omega) = \int_a^b \xi(\omega, t) dt$.

Як і раніше, співвідношення (3.8) можна записати у розгорнутому вигляді:

$$\lim_{\substack{n \rightarrow \infty \\ \Delta t_{n_k} \rightarrow 0}} \int_{\Omega} \left| \sum_{k=1}^n \xi(\omega, t_{n_k}) \Delta t_{n_k} - J(\omega) \right|^2 dP = 0 \quad (3.8a)$$

або символічно –

$$l.i.m. \sum_{k=1}^n \xi(\omega, t_{n_k}) \Delta t_{n_k} = \int_a^b \xi(\omega, t) dt. \quad (3.8б)$$

Можна довести такий факт:

Теорема 3.4. Для с.к. інтегровності випадкового процесу $\xi(\omega, t)$ необхідно і достатньо інтегровності за Ріманом кореляційної функції $R(t, s)$, $a \leq t, s \leq b$.

Випадковий процес називають стаціонарним у широкому розумінні, якщо

$$M \xi(\omega, t) = const, \quad (3.9)$$

$$R(t, s) = R(t - s) = R(\tau), \quad \tau = t - s,$$

тобто його перші дві моментні функції не залежать від початку відліку.

Теорія стаціонарних у широкому розумінні випадкових процесів – це найпоширеніша частина кореляційної теорії.

Для стаціонарних у широкому розумінні процесів кореляційна функція має такі властивості (рис. 3.5)

$$R(0) \geq 0,$$

$$\overline{R(\tau)} = R(-\tau),$$

$$|R(\tau)| \leq R(0), \quad (3.10)$$

$$\sum_{j,k=1}^n R(t_j - t_k) \alpha_j \overline{\alpha_k} \geq 0.$$

Наведені вище умови тих чи інших аналітичних властивостей у середньоквадратичному при цьому також справджуються відповідним чином.

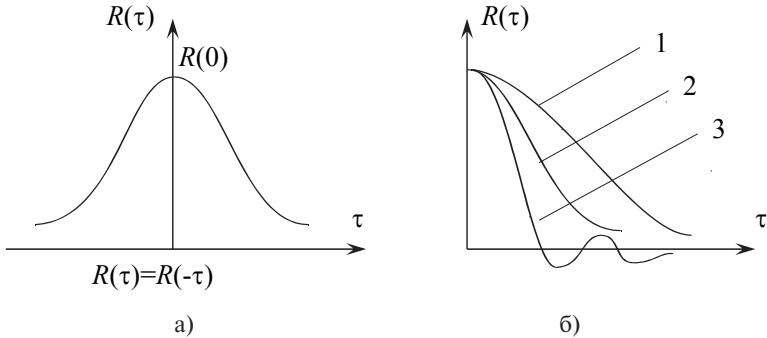


Рис.3.5. – Типовий вигляд кореляційних функцій дійсних стаціонарних (ергодичних) випадкових процесів: а) симетрія $R(\tau)$; б) загальний вигляд кореляційних функцій різних процесів

2. Вибіркове означення ВП

Вибіркове означення випадкового процесу $\Omega \xrightarrow{\xi} R^T$ надає інші можливості описати його. При цьому ВП трактується як вимірне відображення імовірнісного простору (Ω, \mathcal{F}, P) у простір (R^T, \mathcal{J}^T) .

Таке означення є прямим узагальненням означень випадкової величини та вектору: їх визначають як вимірне відображення імовірністності простору у R^1 та R^n . Для ВП відображення здійснюється у R^T – незчисленновимірний простір. З точки зору прикладника це найбільш природне означення, тому що вони безпосередньо приводить до реалізації ВП, з якими має справу прикладник. Однак у деякому розумінні воно найбільш складне, тому що необхідно «будувати» індуковану міру на просторі (R^T, \mathcal{J}^T) .

Згідно з вибірковим означенням вводять новий імовірнісний простір – вибірковий простір $(R^T, \mathcal{J}^T, P^T)$. Міра P^T індукується при вимірному відображенні Ω на R^T . Її задають так, що міра кожної вимірної множини $B \in \mathcal{J}^T$ визначається як імовірність її повного прообразу

$$P^T(B) = P(\{\omega: \xi(\omega, \cdot) \in B\}),$$

тобто це міра P (з імовірнісного простору (Ω, \mathcal{F}, P)) множини точок $\omega \in \Omega$, які відображаються функцією ξ у множину реалізацій B (рис.3.6). Складність цього підходу полягає у необхідності конструктивного завдання міри P^T на досить «широких» σ -алгебрах J^T . Міра P^T дає змогу визначити узгоджену сім'ю скінченновимірних розподілів процесу.

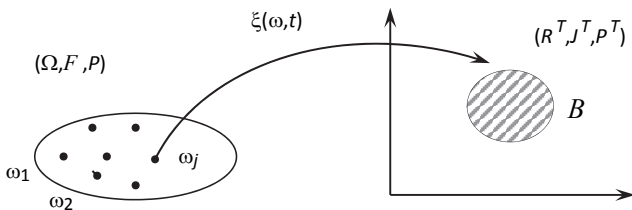


Рис.3.6. – Знаходження індукованої міри $P^T(B)$ як міри P повного прообразу множини B : $P^T\{\xi^{-1}(B)\} = P\{\omega: \xi(\omega) \in B\}$

Вибірковий простір R^T є простором усіх дійсних функцій. Він дуже «широкий». Звичайно розглядають «вужчі» простори, зокрема гільбертів простір L_{2X} реалізацій випадкового процесу.

Випадковий процес можна також повністю описати характеристичним функціоналом, який визначається за відомою мірою

$$f(u) = \int \exp[i(u, x)] dP^T. \quad (3.11)$$

За розглянутого підходу вводять означення *стаціонарності* (у вузькому розумінні) та *ергодичності* процесу.

Уведемо оператор зсуву T_h , який діє на вибірковому просторі (R^T, J^T, P^T) так, що для реалізації $x(t)$ процесу виконується умова

$$T_h x(t) = x(t+h).$$

Якщо для всіх множин $\Phi \in J^T$ і всіх дійсних h виконується рівність

$$P^T(\Phi) = P^T(T_h \Phi), \quad (3.12)$$

процес називають стаціонарним (у вузькому розумінні). При цьому всі скінченновимірні розподіли процесу не залежать від моменту відліку. Стаціонарність у широкому розумінні, введена раніше, впливає зі стаціонарності у вузькому розумінні, але не навпаки.

Множину $A \in J^T$ називають інваріантною відносно зсуву, якщо

$$T_h A = A \quad (3.13)$$

при всіх h . Процес називається ергодичним, якщо кожна інваріантна множина має ймовірність 0 або 1.

Ймовірнісні характеристики ергодичного процесу можна визначити за будь-якою реалізацією з інваріантної множини одиничної ймовірності. Ці реалізації називають типовими на відміну від нетипових, які належать до інваріантної множини нульової ймовірності.

Зокрема, математичне сподівання ергодичного процесу

$$m = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \xi(\omega_j, t) dt, \quad (3.14)$$

кореляційна функція

$$R(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \overset{\circ}{\xi}(\omega_j, t) \overset{\circ}{\xi}(\omega_j, t + \tau) dt, \quad (3.15)$$

де $\overset{\circ}{\xi}(\omega, t) = \xi(\omega, t) - m$.

Для того щоб виконувалися ці умови для стаціонарних у широкому розумінні процесів, необхідно й достатньо, щоб

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T R(\tau) \left(1 - \frac{|\tau|}{T}\right) d\tau = 0.$$

Достатня умова – рівність нулю середнього значення кореляційної функції:

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T R(\tau) d\tau = 0.$$

Ергодичність – властивість специфічніша, ніж стаціонарність. Для стаціонарності достатньо, щоб ймовірність будь-якої події з J^T не залежала від зсуву, хоча сама множина Φ при зсуві може змінюватися. Для ергодичності сама подія з J^T при зсуві не повинна змінюватися, причому її ймовірність має бути або 1 або 0.

3. Означення вимірного ВП

$(F \times B)$ – вимірний випадковий процес – це вимірне відображення простору $(\Omega \times T, F \times B)$ у простір стану (R', J') .

Тут не віддається переваги ні ω ні t . На відміну від попереднього тут (як і раніше) для кожного значення t дістаємо випадкову величину, але додатково для кожного ω_j одержуємо вимірну функцію часу $\xi(\omega_j, t)$.

Це зрештою дає змогу ввести інтегрування реалізацій. Вимірний випадковий процес $\xi(\omega, t)$ виявляється функцією двох змінних ω і t .

Розглянемо простір $(\Omega \times T, \mathcal{F} \times \mathcal{B}, P \times \mu)$, де $\Omega \times T$ – прямий добуток простору елементарних наслідків Ω та параметричної множини T ; $\mathcal{F} \times \mathcal{B}$ – добуток σ -алгебр \mathcal{F} і \mathcal{B} ; $P \times \mu$ – міра на $\mathcal{F} \times \mathcal{B}$; μ – міра Лебега на \mathcal{B} . Уведемо обмеження

$$\int_T M\xi(\omega, t) d\mu < \infty. \quad (3.16)$$

Тоді за теоремою Фубіні (яка широко застосовується у сучасній теорії ймовірностей) маємо

$$\int_{\Omega T} \xi(\omega, t) d\mu dP = \int_T \xi(\omega, t) dP d\mu. \quad (3.17)$$

Згідно з цією рівністю

$$\begin{aligned} \int_{\Omega} \xi(\omega, t) dP < \infty, \\ \int_T \xi(\omega, t) d\mu < \infty, \end{aligned} \quad (3.18)$$

тобто процес $\xi(\omega, t)$ для всіх значень $t \in T$ є вимірною функцією ω (інакше кажучи для будь-якого t переріз процесу є випадкова величина), а майже для всіх ω – вимірна функція t . Остання властивість, як вже підкреслювалось вище, характерна саме для вимірних процесів. При цьому майже для всіх реалізацій існує у звичайному розумінні інтеграл Лебега (вибірковий інтеграл)

$$\int_T \xi(\omega, t) d\mu. \quad (3.19)$$

Якщо енергія випадкового процесу скінченна

$$E_{\xi} = \int_T M|\xi(\omega, t)|^2 d\mu = \int_T R(t, t) dt < \infty, \quad (3.20)$$

то

$$\int_T M|\xi(\omega, t)|^2 d\mu = M \int_T |\xi(\omega, t)|^2 d\mu.$$

При цьому з імовірністю 1 існує інтеграл Лебега

$$\int_T |\xi(\omega, t)|^2 d\mu < \infty, \quad (3.21)$$

тобто майже всі реалізації процесу мають скінченну енергію й належать до гільбертового простору L_{2X} , який фізично тлумачать як простір реалізацій випадкового процесу.

3.1.3. Приклад: гаусові випадкові процеси

Гаусові процеси займають особливе місце серед інших випадкових процесів. Це обумовлено їх властивостями. Так, гаусові процеси повністю визначені, коли відомі їх кореляційні функції; для них стаціонарність у вузькому розумінні збігається зі стаціонарністю у широкому розумінні; їх лінійні перетворення не змінюють закону розподілу тощо.

Випадковий процес $\xi(\omega, t)$ називається гаусовим, якщо кожна його скінченновимірна функція розподілу

$$F_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P\{\xi(\omega, t_1) \leq x_1, \xi(\omega, t_2) \leq x_2, \dots, \xi(\omega, t_n) \leq x_n\}$$

є гаусова.

Наведемо основні властивості гаусова випадкового вектора

$$\bar{\xi} = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega)).$$

Випадковий вектор $\bar{\xi}$ має гаусів (нормальний) розподіл, якщо характеристична функція розподілу зображується у вигляді

$$f(u_1, u_2, \dots, u_n) = Me^{j(\bar{u}, \bar{x})} = e^{j(\bar{m}, \bar{x}) - \frac{1}{2}(R\bar{x}, \bar{x})}, \quad (3.22)$$

де $\bar{m} = (m_1, m_2, \dots, m_n)$, $\bar{x} = (x_1, x_2, \dots, x_n)$ – n -вимірні вектори, R – невід’ємно визначена дійсна симетрична матриця, $R = [r_{ik}]$, $i, k = 1, \dots, n$.

$$\begin{aligned} \text{Тут} \quad (\bar{m}, \bar{x}) &= \sum_{k=1}^n m_k x_k; \\ (R\bar{x}, \bar{x}) &= \sum_{j,k=1}^n r_{jk} x_j x_k. \end{aligned} \quad (3.23)$$

відповідні скалярні добутки.

Доводиться, що необхідною та достатньою умовою, щоб характеристична функція (3.22) була х.ф. деякого n -вимірного розподілу,

є властивість матриці R бути невід'ємно визначеною та симетричною. Крім того, неважко також довести, що вектор \vec{m} у співвідношенні (3.22) є вектором математичного сподівання, а R – кореляційною матрицею:

$$\vec{m} = M\vec{\xi};$$

$$R = [r_{jk}], \quad r_{jk} = M\left\{\left(\xi_j - m_j\right)\left(\xi_k - m_k\right)\right\}. \quad (3.24)$$

Звичайно у прикладній літературі не підкреслюється такий факт: якщо кореляційна матриця R гаусового випадкового вектора $\vec{\xi}$ невинроджена, то існує n -вимірна щільність розподілу $p(\vec{x})$:

$$p(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n |R|}} \exp\left\{-\frac{1}{2}\left(R^{-1}(\vec{x} - \vec{m}), (\vec{x} - \vec{m})\right)\right\}, \quad (3.25)$$

де R^{-1} – матриця, обернена кореляційній матриці R ;
 $|R|$ – визначник матриці R .

Співвідношення (3.25) зручно записувати у векторно-матрицевому вигляді

$$p(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n |R|}} \exp\left\{-\frac{1}{2}\left((\vec{x} - \vec{m})^T R^{-1}(\vec{x} - \vec{m})\right)\right\} \quad (3.25a)$$

на відміну від використання символу скалярного добутку у (3.25).

Скінченновимірна характеристична функція гаусового випадкового процесу при будь-якому n має вигляд (3.22), а його щільність імовірності за умову невинродженості кореляційної матриці – (3.25)–(3.25a).

Вищі моменти гаусового випадкового процесу можуть бути одержані із розкладу характеристичної функції. Обмежуючись випадком центральних моментів, тобто покладаючи $\vec{m} = \vec{0}$, маємо

$$f(u_1, \dots, u_n; x_1, \dots, x_n) = 1 - \frac{1}{2}(R\vec{u}, \vec{u}) + \frac{1}{2!2^2}(R\vec{u}, \vec{u})^2 - \dots + (-1)^n \frac{1}{2^n n!}(R\vec{u}, \vec{u})^n + \dots, \quad (3.26)$$

де $R = [R(t_j, t_k)]_{j,k=1}^n$.

Звідси для моментної функції непарного порядку одержуємо

$$\mu_{j_1 \dots j_s}(t_1, \dots, t_s) = 0, \quad \text{коли } \sum_{k=1}^s j_k = 2n + 1;$$

а для моментних функцій парного порядку –

$$\mu_{j_1 \dots j_s}(t_1, \dots, t_s) = \frac{\partial^{2n}}{\partial u_1^{j_1} \dots \partial u_s^{j_s}} \frac{1}{2^n n!} (R\vec{u}, \vec{u})^n, \quad \sum_{k=1}^n j_k = 2n,$$

тобто маємо співвідношення

$$\mu_{j_1 \dots j_s}(t_1, \dots, t_s) = \sum \prod R(t_p, t_q). \quad (3.27)$$

Наприклад:

$$\mu_4(t) = 3R^2(t, t);$$

$$\mu_{31}(t_1, t_2) = 3R(t_1, t_1)R(t_1, t_2),$$

$$\mu_{1111}(t_1, t_2, t_3, t_3) = R(t_1, t_2)R(t_3, t_3) + R(t_1, t_3)R(t_2, t_3) + R(t_1, t_4)R(t_2, t_3).$$

Звідси витікає, що завдання кореляційної функції гаусового процесу дає можливість обчислювати моментні функції довільного порядку; більш того, як вже відмічено вище, можна довести, що такий процес задається у цілому, коли відома лише його кореляційна функція.

3.2. Інтегральні зображення випадкових процесів зі скінченними енергетичними характеристиками

Найістотніші класичні результати у теорії випадкових процесів другого порядку дістали в результаті використання математичного апарату гільбертових просторів. Сучасна енергетична теорія випадкових процесів також є L_2 -теорією, хоча й використовує нові варіанти гільбертових просторів. Вона надає класичним положенням кореляційної теорії досить завершеного характеру.

Одним з найважливіших результатів кореляційної теорії випадкових процесів є інтегральні зображення, зокрема гармонічні зображення. Останні мають у теорії електрозв'язку фундаментальне значення – вони у відповідній формі поширюють ідеї перетворення Фур'є детермінованих сигналів на випадкові, тобто дають можливість описувати сигнали (як фізичні процеси) на шкалі частот.

3.2.1. Класичні інтегральні зображення гільбертових процесів

Говоритимемо, що випадковий процес $\xi(\omega, t)$ другого порядку, центрований математичним сподіванням, припускає інтегральне зображення, коли існує випадкова функція (міра) $Z(\omega, \lambda)$, $\lambda \in \Lambda$ другого порядку, така що

$$\xi(\omega, t) = \int_{\Lambda} \varphi(t, \lambda) dZ(\omega, \lambda), \quad (3.28)$$

де $\varphi(t, \lambda)$ – деяка числова функція.

Говоритимемо також, що кореляційна функція $R(t_1, t_2)$ процесу $\xi(\omega, t)$ припускає інтегральне зображення, якщо існує комплекснозначна невід'ємно визначена функція $\Phi(\lambda_1, \lambda_2)$, така що

$$R(t_1, t_2) = \iint_{\Lambda \times \Lambda} \overline{\varphi(t_1, \lambda_1)} \varphi(t_2, \lambda_2) d^2 \Phi(\lambda_1, \lambda_2) \quad (3.29)$$

де $\Phi(\lambda_1, \lambda_2) = MZ(\omega, \lambda_1) \overline{Z(\omega, \lambda_2)}$ – кореляційна функція для $Z(\omega, \lambda)$.

Класичним є такий результат (де у означенні вживатимемо більш точні терміни).

Теорема 3.5 (Карунена). Для того, щоб випадковий процес $\xi(\omega, t)$ другого порядку допускав інтегральне зображення (3.28), де $Z(\omega, \lambda)$ – випадкова ортогональна міра, а $\varphi(t, \lambda) \in L_2(\Lambda)$ (як функція λ), необхідно і достатньо, щоб його кореляційну функцію можна було подати у вигляді

$$R(t_1, t_2) = \int_{\Lambda} \overline{\varphi(t_1, \lambda)} \varphi(t_2, \lambda) d\Phi(\lambda). \quad (3.30)$$

Тут випадкова міра $Z(\omega, \lambda)$ ортогональна у тому розумінні, що функція множин $Z(\omega, \Delta)$ задовольняє умови:

а) $Z(\omega, \emptyset) = 0$ (з імовірністю 1);

б) $Z(\omega, \Delta_1 \cup \Delta_2) = Z(\omega, \Delta_1) + Z(\omega, \Delta_2)$, якщо $\Delta_1 \cap \Delta_2 = \emptyset$,

Δ_1, Δ_2 підмножини із Λ .

в) $MZ(\omega, \Delta_1) \overline{Z(\omega, \Delta_2)} = 0$, якщо $\Delta_1 \cap \Delta_2 = \emptyset$;

$M|Z(\omega, \Delta)|^2 = \Phi(\Delta)$ – звичайна міра.

Властивість «в» міри $Z(\omega, \lambda)$ в решті решт дає можливість співвідношення (3.29) для кореляційної функції $R(t_1, t_2)$ привести до вигляду (3.30). Тут параметр λ має розуміння частоти; зображення процесу (3.28) при цьому називають спектральним. Однак є можливість побудувати зображення, коли параметри процесів ξ і Z мають однакове розуміння та змінюються у той самій області:

$$\xi(\omega, t) = \int_T \varphi(t, \tau) d\eta(\omega, \tau), \quad t, \tau \in T, \quad (3.31)$$

$$R(t_1, t_2) = \int_T \overline{\varphi(t_1, \tau)} \varphi(t_2, \tau) d\Phi(\tau). \quad (3.32)$$

Процеси, для яких припустимі такі зображення, називаються лінійними.

У цьому разі ортогональна випадкова міра $Z(\omega, \lambda)$ задається приростами процесу $\eta(\omega, \tau)$ з некорельованими приростами, тобто при

$$M|\eta(\omega, \tau_1) - \eta(\omega, \tau_2)| < \infty, \tau_1, \tau_2 \in T$$

виконується умова

$$M\{\overline{[\eta(\omega, \tau_4) - \eta(\omega, \tau_3)]} \overline{[\eta(\omega, \tau_2) - \eta(\omega, \tau_1)]}\} = M[\eta(\omega, \tau_4) - \eta(\omega, \tau_3)]M[\overline{[\eta(\omega, \tau_2) - \eta(\omega, \tau_1)]}],$$

$$\tau_1 < \tau_2 \leq \tau_3 < \tau_4,$$

або ця міра задається приростами процесу $\eta(\omega, \tau)$ з незалежними приростами, коли для будь-яких значень параметра $\tau_1 < \tau_2 < \dots < \tau_n$ ($n \geq 3$) прирости процесу $\eta(\omega, \tau_2) - \eta(\omega, \tau_1)$, $\eta(\omega, \tau_3) - \eta(\omega, \tau_2)$, ..., $\eta(\omega, \tau_n) - \eta(\omega, \tau_{n-1})$, взаємно незалежні.

Деякі часткові випадки лінійних процесів в останній період дістали широке поширення у теорії електрозв'язку.

3.2.2. Гармонічні зображення випадкових процесів

Серед інтегральних зображень ВП особливо важливими є гармонічні зображення.

Випадковий процес другого порядку, центрований математичним сподіванням, називають *гармонізовним (за Лоевим)*, якщо існує така випадкова функція другого порядку $Z(\omega, \lambda)$, що

$$\xi(\omega, t) = \int_{\Lambda} \exp(it\lambda) dZ(\omega, \lambda). \quad (3.33)$$

Кореляційну функцію $R(t_1, t_2)$ процесу $\xi(\omega, t)$ називають гармонізовною (за Лоевим), якщо існує така кореляційна функція $\Phi(\lambda_1, \lambda_2)$ з обмеженою варіацією на $\Lambda \times \Lambda$, що виконується с.к. перетворення Фур'є-Стілтгеса

$$R(t_1, t_2) = \iint_{\Lambda \times \Lambda} \exp[i(t_1\lambda_1 - t_2\lambda_2)] d^2\Phi(\lambda_1, \lambda_2). \quad (3.34)$$

Для гармонізовних (за Лоевим) випадкових процесів справджується таке твердження:

Теорема 3.6. (Лоева). Випадковий процес гармонізовний (за Лоевим) тоді і тільки тоді, коли гармонізовна його кореляційна функція.

Зокрема, стаціонарні у широкому розумінні процеси характеризуються тим, що функція $\Phi(\lambda_1, \lambda_2)$ задана на діагоналі простору, причому

$$d^2\Phi(\lambda_1, \lambda_2) = \delta(\lambda_1 - \lambda_2) d\Phi(\lambda_1) d\lambda_2. \quad (\text{рис. 3.7.})$$

Тоді вирази (3.33) і (3.34) набувають вигляду

$$\xi(\omega, t) = \int_{-\infty}^{\infty} \exp(it\lambda) dZ(\omega, \lambda), \quad (3.33a)$$

$$R(t_1, t_2) = \int_{-\infty}^{\infty} \exp[i(t_1 - t_2)\lambda] d\Phi(\lambda), \quad (3.34a)$$

де $R(t_1, t_2) = R(t_1 - t_2) = R(\tau)$, $\tau = t_1 - t_2$, $M|Z(\omega, \lambda)|^2 = \Phi(\lambda)$, $\Phi(\lambda)$ – функція з обмеженою варіацією.

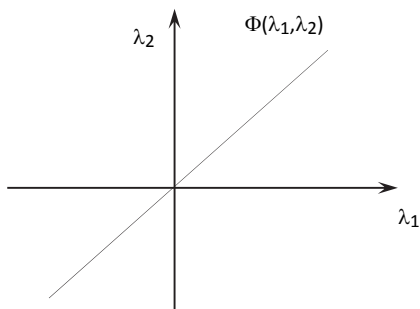


Рис.3.7. – Завдання двочастотної спектральної функції $\Phi(\lambda_1, \lambda_2)$ нестационарного процесу на площині $\Omega \times \Omega$ та одночастотної спектральної функції стаціонарного процесу на прямій при $\lambda_1 = \lambda_2$:

$$\lambda_1 = \lambda_2 \Rightarrow d^2\Phi(\lambda_1, \lambda_2) = \delta(\lambda_1 - \lambda_2) d\Phi(\lambda_1) d\lambda_2$$

Зображення (3.33a) називають зображенням Крамера, а зображення (3.34a) – зображенням Вінера-Хінчина (Бохнера-Вінера-Хінчина). Класичні результати можуть бути сформульовані у такій формі:

Теорема 3.7. (Крамера). Для того щоб стаціонарний випадковий процес другого порядку, центрований математичним сподіванням, мав гармонічний ортогональний розклад (3.33a), необхідно і достатньо, щоб він був с.к. неперервним.

Теорема 3.8. (Вінера-Хінчина). Для того щоб кореляційна функція стаціонарного випадкового процесу другого порядку, центрованого математичним сподіванням, допускала гармонічне зображення (3.34a), необхідно і достатньо, щоб процес був с.к. неперервним.

Функцію $\Phi(\lambda)$ називають *спектральною*. У загальному випадку функція $\Phi(\lambda)$ допускає канонічне зображення

$$\Phi(\lambda) = \Phi_1(\lambda) + \Phi_2(\lambda) + \Phi_3(\lambda),$$

де $\Phi_1(\lambda)$ – функція стрибків для $\Phi(\lambda)$;

$\Phi_2(\lambda)$ – абсолютна неперервна складова $\Phi(\lambda)$;

$\Phi_3(\lambda)$ – сингулярна функція.

Якщо $\Phi(\lambda) = \Phi_2(\lambda)$, вираз (3.34а) зводиться до вигляду

$$R(\tau) = \int_{-\infty}^{\infty} \exp(i\tau\lambda) G(\lambda) d\lambda, \quad (3.35)$$

де $G(\lambda) = \frac{d\Phi(\lambda)}{d\lambda}$ – спектральна густина (щільність), що, як впливає

$$\text{з рівності } R(0) = \int_{-\infty}^{\infty} G(\lambda) d\lambda = P_{\xi},$$

характеризує розподіл потужності P_{ξ} випадкового процесу на осі частот. За додатковим припущенням, що

$$\int_{-\infty}^{\infty} |R(\tau)| d\tau < \infty,$$

існує обернене перетворення Фур'є

$$G(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\tau\lambda) R(\tau) d\tau. \quad (\text{рис. 3.8}) \quad (3.36)$$

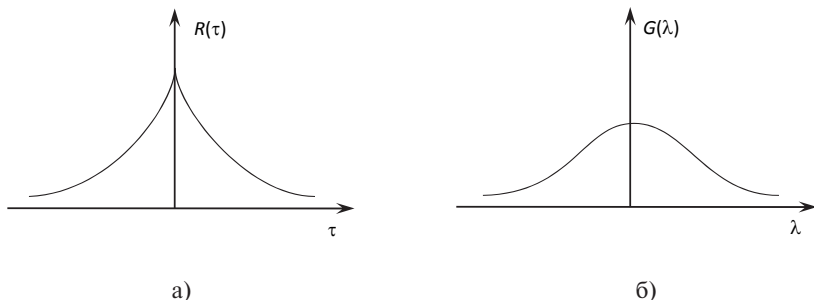


Рис. 3.8. – Приклад спектральної густини $G(\lambda)$ – «б», яка відповідає заданій кореляційній функції $R(\tau)$ – «а»

Розглянемо частинний випадок зображення Крамера випадкового процесу, коли він має фінітний спектр:

$$\xi(\omega, t) = \int_{-B}^B \exp(it\lambda) dZ(\omega, \lambda), \quad (3.33б)$$

$$R(\tau) = \int_{-B}^B \exp(i\tau\lambda) d\Phi(\lambda). \quad (3.34б)$$

У цьому разі функцію $\xi(\omega, t)$ можна подати у вигляді

$$\xi(\omega, t) = \sum_{k=-\infty}^{\infty} \xi\left(\omega, k \frac{\pi}{B}\right) \frac{\sin(Bt - k\pi)}{Bt - k\pi}, \quad (3.33в)$$

що є результатом теореми Котельникова-Шеннона: с.к. неперервний випадковий процес другого порядку, який центрований математичним сподіванням і має фінітний спектр, допускає зображення (3.33в) у вигляді ряду, який називається рядом Котельникова.

Згідно з цією теоремою випадковий процес $\xi(\omega, t)$, $t \in (-\infty, \infty)$ з неперервним часом еквівалентний послідовності випадкових величин $\{\xi(\omega, k\Delta t)\}_{k=-\infty}^{\infty}$, які є відліками процесу, взятими з періодом

$\Delta t = \frac{\pi}{B} = \frac{1}{2F}$, де $F = \frac{B}{2\pi}$. Інакше кажучи, у оговорених умовах процес з

неперервним часом еквівалентний відповідному процесу з дискретним часом. (За деяких умов теорему можна також застосовувати майже до всіх реалізацій процесу – рис. 3.9).

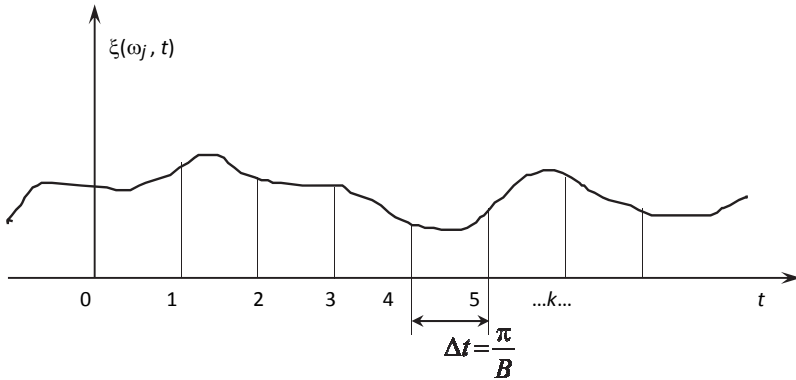


Рис. 3.9. – Дискретизація реалізації $\xi(\omega_j, t)$ випадкового процесу $\xi(\omega, t)$ згідно з теоремою Котельникова

Класичні результати гармонічного зображення стаціонарних та нестаціонарних випадкових процесів припускають скінченність їх потужності, тобто належності до гільбертового (сепарабельного) простору $L_{2\xi}$ випадкових величин. Сучасні результати щодо гармонізованості розглядаються в умовах скінченності їх середньої потужності:

$$\bar{P}_\xi = \lim_{T \rightarrow \infty} M \frac{1}{T} \int_{-T/2}^{T/2} |\xi(\omega, t)|^2 dt < \infty. \quad (3.37)$$

Справджується таке твердження: «випадковий процес гармонізований (за Драганом) тоді і тільки тоді, коли він має скінченну середню потужність. Відповідно, з точки зору функціонального аналізу, тут використовується апарат несепарабельних гільбертових просторів.

Класичні спектральні інтегральні зображення випадкових процесів можна конкретизувати для процесів з дискретним часом.

3.2.3. Гармонічні зображення випадкових послідовностей

Розглянемо випадкову послідовність $\xi(\omega, k)$, $\omega \in \Omega$, $k \in Z = \{0, \pm 1, \pm 2, \dots\}$ другого порядку, центровану математичним сподіванням, тобто

$$M |\xi(\omega, k)|^2 < \infty, \quad M \xi(\omega, k) = 0.$$

Кореляційна функція такої послідовності

$$R(n, m) = M \xi(\omega, k) \overline{\xi(\omega, m)}, \quad n, m = 0, \pm 1, \pm 2, \dots$$

Для стаціонарних у широкому розумінні послідовностей

$$R(n, n+s) = M \xi(\omega, n) \overline{\xi(\omega, n+s)} = R(s)$$

гармонічне зображення стаціонарного випадкового процесу та його кореляційної функції (3.33 а) і (3.34 а) для послідовностей має вигляд

$$\xi(\omega, k) = \int_{-\pi}^{\pi} \exp(ik\lambda) dz(\omega, \lambda), \quad (3.38)$$

$$R(k) = \int_{-\pi}^{\pi} \exp(ik\lambda) d\Phi(\lambda). \quad (3.39)$$

Частинний випадок теореми Вінера-Хінчина стосовно стаціонарної випадкової послідовності називають теоремою Герглотца.

Теорема 3.9. (Герглотца). Для всякої стаціонарної випадкової послідовності справджується інтегральне зображення (3.39) для кореля-

ційної функції $R(k)$, причому $\Phi(\lambda)$ – неспадна, обмежена на $[-\pi, \pi)$ функція.

Характерно, що спектр стаціонарної випадкової послідовності завжди зосереджений на скінченному інтервалі (рис.3.10).

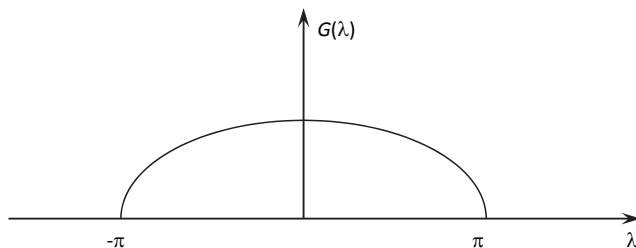


Рис.3.10. – Спектральна густина стаціонарної послідовності, яка зосереджена на скінченному інтервалі частот $-\pi \leq \lambda \leq \pi$

3.3. Зображення рядами випадкових процесів зі скінченними енергетичними характеристиками

Хоч гармонізованість стаціонарних процесів і особливо їх кореляційних функцій дуже широко застосовується у теорії електровз'язку, насправді зображення процесів (як стаціонарних, так і нестаціонарних) рядами використовується ще ширше. Це більш прості у теоретичному та практичному розумінні зображення ВП. Розглянемо декілька з них.

3.3.1. Зображення стаціонарних періодичних випадкових процесів

Багато з фізичних явищ у галузі електровз'язку має періодичний характер. Для детермінованих процесів ми маємо можливість використовувати добре та давно розвинений математичний апарат рядів Фур'є. Виникає бажання щось подібне мати і для дослідження випадкових процесів.

Розглянемо для початку стаціонарний періодичний випадковий процес другого порядку, центрований математичним сподіванням:

$$\xi(\omega, k), \omega \in \Omega ; M \xi(\omega, k) = 0 ; R(t_1, t_2) = R(t_1 - t_2) = R(\tau).$$

Випадковий стаціонарний у широкому розумінні процес називається періодичним з періодом $T > 0$, якщо його кореляційна функція $R(\tau)$ задовольняє умову

$$R(\tau + T) = R(\tau). \quad (3.40)$$

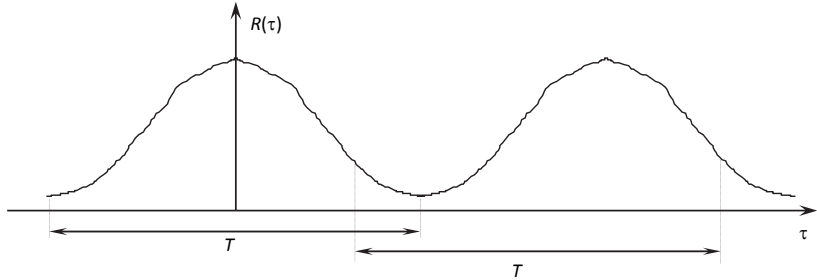


Рис. 3.11. – Кореляційна функція стаціонарного періодичного випадкового процесу з періодом T

Як навідні міркування, нагадаємо властивості інтегральних зображень ВП з підрозділу 3.2: зображення кореляційної функції у межах кореляційної теорії породжує відповідні зображення процесу. Для періодичних ВП, як це очевидно, функцію $R(\tau)$ можна розкласти у ряд Фур'є за гармоніками і цей розклад, як підказує інтуїція, повинен породжувати відповідний розклад процесу. Дійсно, має місце такий факт.

Теорема 3.10. Нехай $\xi(\omega, t)$ – стаціонарний періодичний випадковий процес з періодом (рис. 3.11), кореляційна функція $R(\tau)$ якого зображена рядом

$$R(\tau) = \sum_{k=-\infty}^{\infty} g_k e^{ik\tau}, \quad (3.41)$$

де
$$\lambda = \frac{2\pi}{T}, g_k = \frac{1}{T} \int_0^T R(\tau) e^{-ik\lambda\tau} d\tau.$$

Тоді випадковий процес $\xi(\omega, t)$ може бути зображений у вигляді

$$\xi(\omega, t) = \sum_{k=-\infty}^{\infty} d_k(\omega) e^{ik\lambda t}, \quad (3.42)$$

де $d_k(\omega) = \frac{1}{T} \int_0^T \xi(\omega, t) e^{-ik\lambda t} dt$; причому випадкові величини $d_k(\omega)$

парами некорельовані і при $k \neq 0$ $M d_k(\omega) = 0$, $D[d_k(\omega)] = g_k$.

Ми не будемо доводити це твердження. Дамо деякі фізичні пояснення на більш простому прикладі коливань з випадковими параметрами

$$\xi(\omega, t) = \sum_{k=1}^n \gamma_k(\omega) e^{i\lambda_k t}, \quad (3.43)$$

тобто процесу, який є сумою n гармонічних коливань з випадковими амплітудами $\gamma_k(\omega)$ та частотами λ_k . Як стверджується у теоремі 3.10 ми тут припускаємо, що комплексні амплітуди $\gamma_k(\omega)$ мають нульові математичні сподівання і некорельовані поміж себе:

$$M\gamma_k(\omega) = 0, \quad M\gamma_k(\omega)\overline{\gamma_r(\omega)} = 0, \quad k \neq r.$$

Тоді кореляційна функція процесу дістає таке зображення

$$R(t_1, t_2) = M\xi(\omega, t_1)\overline{\xi(\omega, t_2)} = M\left[\sum_{k,r=1}^n \gamma_k(\omega)\overline{\gamma_r(\omega)} e^{i(\lambda_k t_1 - \lambda_r t_2)} \right] = \sum_{k=1}^n M|\gamma_k(\omega)|^2 e^{i\lambda_k(t_1 - t_2)},$$

тобто

$$R(t_1, t_2) = R(t_1 - t_2) = R(\tau) = \sum_{k=1}^n g_k e^{i\lambda_k \tau}, \quad (3.44)$$

$$g_k = M|\gamma_k(\omega)|^2.$$

Ця формула при $\lambda_k = k\lambda$ і $n \rightarrow \infty$ переходить у розклад (3.41) кореляційної функції стаціонарного періодичного процесу. Формули (3.44) і (3.41) дають спектральне зображення кореляційної функції. Вони стверджують, що спектр випадкового процесу – це сукупність частот λ_k ($k\lambda$ – відповідно), $k = 1, 2, \dots$

Можна довести, що величини g_k мають фізичне розуміння як середнє значення потужності гармонічних складових з частотами λ_k . Вони одержуються, як це підкреслюється формулою (3.37) шляхом подвійного усереднення: усереднення у часі та у теоретико-ймовірністному розумінні, тобто знаходження математичного сподівання.

3.3.2. Ряди Фур'є випадкових процесів зі скінченною потужністю

Тут ми будемо розглядати нестационарні випадкові процеси другого порядку з нульовим середнім, які згідно з \mathcal{S} -означенням можна трактувати як криві у гільбертовому просторі випадкових величин $L_{2\xi}$:

$$\xi(\omega, t), \quad \omega \in \Omega, \quad t \in T;$$

$$M\xi(\omega, t) = 0; \quad M|\xi(\omega, t)|^2 = P_\xi < \infty;$$

$$\xi(\omega, t) \in L_{2\xi}, \quad t \in T.$$

У гільбертовому просторіві $L_{2\xi}$ існує ортонормований базис $\{\zeta_k(\omega)\}$, де

$$(\zeta_k, \zeta_n) = M\zeta_k \overline{\zeta_n} = \delta_{kn} = \begin{cases} 1, & k=n; \\ 0, & k \neq n. \end{cases} \quad (3.45)$$

Довільний елемент гільбертового простору (згідно з теорією рядів Фур'є у H) можна розкласти у ряд Фур'є за ортонормованим базисом (3.45). Тобто довільний елемент $\xi_t(\omega)$ гільбертової кривої $\xi(\omega, t)$ у просторі $L_{2\xi}$ можна зобразити узагальненим рядом Фур'є

$$\xi_t(\omega) = \sum_{k=1}^{\infty} C_{kt} \zeta_k(\omega), \quad (3.46)$$

де $C_{kt} = (\xi_t, \zeta_k) = M\xi_t(\omega) \overline{\zeta_k(\omega)}$ – узагальнений коефіцієнт Фур'є.

Ряд Фур'є (3.46) існує для будь-якого значення параметра, тобто стверджується рівність

$$\begin{aligned} \xi(\omega, t) &= \sum_{k=1}^{\infty} C_{kt} \zeta_k(\omega), \\ C_{kt} &= M\xi_t(\omega) \overline{\zeta_k(\omega)}. \end{aligned} \quad (3.47)$$

Співвідношення (3.47) – це ряд Фур'є випадкового процесу другого порядку $\xi(\omega, t)$. З точки зору теорії сигналів у електрозв'язку це – розклад у ряд Фур'є випадкового сигналу. Однак все тут незвично: базис, за яким виконується розклад, – це послідовність ортогональних (некорельованих) випадкових величин, а коефіцієнти Фур'є – це детерміновані функції часу. Такі розклади іноді зустрічаються у теорії електрозв'язку, але загалом ми привели його, щоб підкреслити дійсну структуру ряду Фур'є випадкового сигналу (процесу). Те, що у прикладній літературі вже давно трактується як «ряд Фур'є випадкового сигналу» насправді зовсім інший розклад – ортогональний розклад.

3.3.3. Ортогональні розклади випадкових процесів зі скінченною енергією

Класичні зображення гільбертових процесів у вигляді суми ряду дістали й дослідили багато авторів у межах кореляційної теорії в 40-50-ті роки 20-го століття. Теорію другого порядку таких розкладів було завершено, головним чином, у 80-ті роки. Розглянемо деякі її положення.

Простори випадкових процесів зі скінченною енергією. Ми вже раніше зустрічалися з різними реалізаціями абстрактного гільбертового простору у межах теорії випадкових процесів – простором $L_{2\xi}$ випадкових величин та простором L_{2x} реалізацій випадкових процесів.

Зараз ми зупинимося на просторі, елементами (точками, векторами) якого є самі випадкові процеси. Для цього необхідно ввести додаткову умову – скінченність енергії процесу.

Випадковість процесу $\xi(\omega, t)$, $\omega \in \Omega$, $t \in T$ має скінченну енергію, якщо

$$E_\xi = \int_T M |\xi(\omega t)|^2 dt < \infty. \quad (3.48)$$

Розглянемо клас випадкових процесів зі скінченною енергією як відображень Ω у простір $L_2(T)$. Уведемо гільбертів простір таких відображень $L_2(\Omega, L_2(T))$:

$$\begin{aligned} (\xi, \eta)_{L_2(\Omega, L_2(T))} &= M \int_T \xi(\omega, t) \overline{\eta(\omega, t)} dt, \\ \|\xi\|_{L_2(\Omega, L_2(T))} &= \sqrt{M \int_T |\xi(\omega, t)|^2 dt}, \\ \rho(\xi, \eta)_{L_2(\Omega, L_2(T))} &= \sqrt{M \int_T |\xi(\omega, t) - \eta(\omega, t)|^2 dt}. \end{aligned} \quad (3.49)$$

Це L_2 – простір над гільбертовим простором $L_2(T)$. Процес $\xi(\omega, t)$, є елементом простору $L_2(\Omega, L_2(T))$; майже всі реалізації процесу належать основному простору $L_2(T)$.

Використовуючи цю термінологію та відповідні позначення, наведемо основні результати стосовно ортогональних розкладів.

Ортогональні розклади випадкових процесів. На відміну від інтегральних зображень випадкових процесів, де у взаємозв'язку розглядалися зображення процесів і їх кореляційних функцій, тут розглядаються розклади процесу, його кореляційних функцій, а також реалізації процесу.

Практичний критерій можливості виконання ортогонального розкладу випадкового процесу формулюється у такому вигляді:

Теорема 3.11. Якщо випадковий процес $\xi(\omega, t)$ належить L_2 – простору над гільбертовим простором $L_2(\Omega, L_2(T))$, то він в цьому просторі зображується рядом

$$\begin{aligned} \xi(\omega, t) &= \sum_{j=1}^{\infty} d_j(\omega) \psi_j(t), \\ \text{де} \end{aligned} \tag{3.50}$$

$$d_j(\omega) = \int_T \xi(\omega, t) \overline{\psi_j(t)} dt,$$

який збігається за нормою цього простору. Зауважимо, що в прикладній літературі функції $\psi_j(t)$ з ортонормованого базису звичайно називають координатними (або базисними), а випадкові величини $d_j(\omega)$ – коефіцієнтами розкладу.

Наслідок 3.1. Якщо ряд (3.50) збігається за нормою простору $L_2(\Omega, L_2(T))$ до випадкового процесу $\xi(\omega, t)$, то майже для всіх реалізацій $\xi(\omega_s, t)$ процесу $\xi(\omega, t)$ справджуються узагальнені ряди Фур'є

$$\begin{aligned} \xi(\omega_s, t) &= \sum_{j=1}^{\infty} d_j(\omega_s) \psi_j(t), \\ \text{де} \end{aligned} \tag{3.51}$$

$$d_j(\omega_s) = \int_T \xi(\omega_s, t) \overline{\psi_j(t)} dt$$

за ортонормованим базисом $\{\psi_j(t)\}_{j=1}^{\infty}$.

Характерно, що ортогональний розклад процесу не є його рядом Фур'є: структура ряду Фур'є процесу наведена у пункті 3.3.2. Однак у теорії електров'язку саме розклад майже всіх реалізацій процесу у ряд Фур'є ототожнювали з розкладом процесу у той самий ряд. Це врешті рещт пов'язано з невірним розумінням випадкового процесу просто як множини реалізацій.

Термін «ортогональний розклад» пов'язаний з тим, що члени $\zeta_j(\omega, t) = d_j(\omega) \psi_j(t)$, $j=1, 2, \dots$ розкладу (3.50) ортогональні в метриці простору $L_2(\Omega, L_2(T))$:

$$\begin{aligned} (\zeta_j, \zeta_k)_{L_2(\Omega, L_2(T))} &= M \int_T d_j(\omega) \psi_j(t) \overline{d_k(\omega) \psi_k(t)} dt = \\ &= dM_j(\omega) \overline{d_k(\omega)} \int_T \psi_j(t) \overline{\psi_k(t)} dt = dM_j(\omega) \overline{d_k(\omega)} \delta_{jk}. \end{aligned}$$

Ортогональність в метриці гільбертова простору над гільбертовим простором досягається за рахунок ортогональності координатних

функцій $\{\psi_j(t)\}_{j=1}^{\infty}$ незалежно від корельованості або некорельованості коефіцієнтів розкладу із системи $\{d_j(\omega)\}_{j=1}^{\infty}$.

Енергетичні характеристики випадкових процесів. У рамках кореляційної теорії класичні результати в області розкладів випадкових процесів розклад процесу пов'язують з відповідним зображенням його кореляційної функції. Аналогічний результат існує і для процесів зі скінченною енергією.

Теорема 3.12. Випадковий процес зі скінченною енергією допускає розклад

$$\xi(\omega, t) = \sum_{j=1}^{\infty} d_j(\omega) \varphi_j(t), \quad (3.52)$$

який збігається за нормою простору $L_2(\Omega, L_2(T))$, тоді й тільки тоді, коли його кореляційна функція допускає розклад

$$R(t_1, t_2) = \sum_{k, j=1}^{\infty} g_{kj} \varphi_k(t_1) \overline{\varphi_j(t_2)}, \quad (3.53)$$

який збігається за нормою простору $L_2(T \times T)$. Тут:

$$d_j(\omega) = \int_T \xi(\omega, t) \overline{q_j(t)} dt,$$

$\{q_j(t)\}_{j=1}^{\infty}$ – базис, ортогонально спряжений базису

$$\{\varphi_j(t)\}_{j=1}^{\infty}, \quad (\varphi_j, q_k) = \delta_{jk};$$

g_{kj} – елементи матриці $[g_{kj}]_{k, j=1}^{\infty} = g$, $g_{kj} = \iint_{T \times T} R(t_1, t_2) q_k(t) \overline{q_j(t)} dt_1 dt_2$.

Розклади (3.50) і (3.52) не співпадають: ряд (3.50) охоплює два типи розкладів, а саме, розклади з ортогональними координатними функціями та корельованими або некорельованими коефіцієнтами розкладів; ряд (3.52) – чотири типи розкладів, серед яких крім означених є ще розклади з неортогональними (тобто лінійно незалежними) координатними функціями та корельованими або некорельованими коефіцієнтами розкладів.

Серед чотирьох типів розкладів у (3.52) три ортогональних; неортогональним є розклад з неортогональними координатними функціями та корельованими коефіцієнтами розкладу.

Ортогональні розклади характеризуються тим, що для них виконується принцип адитивного розподілу енергії між складовими розкладу. Тому енергетичні властивості випадкового процесу, породжені ортогональними розкладами, дають змогу ввести енергетичний спектр процесу в різних базисах. Введемо відповідні означення.

Енергетичним спектром випадкового процесу зі скінченною енергією у довільному ортонормованому базисі $\{\psi_j(t)\}_{j=1}^{\infty}$ називають розподіл його енергії E_{ξ} між складовими ортогонального розкладу (3.50) – рис. 3.12:

$$G_{\Psi}(j) = g_j = M|d_{\Psi_j}(\omega)|^2, \quad j=1,2,\dots$$

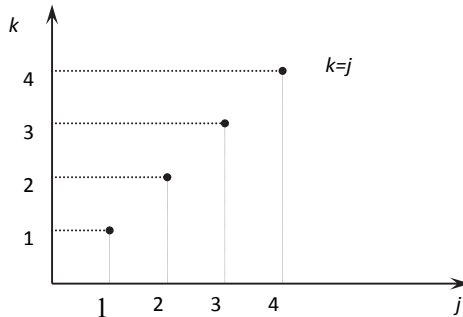


Рис. 3.12. – Носій енергетичного спектру $G_{\Psi}(j)$ нестационарного випадкового процесу у вибраному ортонормованому базисі, заданого у точках $k = j : G_{\Psi}(j) = M|d_{\Psi_j}(\omega)|^2$

Справджується така теорема.

Теорема 3.13. Енергія випадкового процесу визначена через енергетичний спектр у довільному ортонормованому базисі таким чином:

$$E_{\xi} = \sum_{j=1}^{\infty} G_{\Psi}(j). \quad (3.54)$$

Енергетичний спектр нестационарного випадкового процесу в довільному ортонормованому базисі характеризує розподіл енергії процесу між зчисленною множиною складових розкладу. Для обчислення цього енергетичного спектру необхідно виконувати відповідні усереднення у часі та у теоретико-ймовірнісному розумінні (тобто по ансамблю реалізацій).

На прикладному рівні можна трактувати, що енергетичний спектр стаціонарного випадкового процесу (у розумінні Вінера-Хінчина) є спектром відносно незчисленного базису гармонік, а розглянутий спектр процесу зі скінченною енергією є спектром відносно довільного зчисленого ортонормованого базису. Таких базисів – незчислена множина (множина потужності континууму). Енергетичний спектр у розумінні Вінера-Хінчина у частковому граничному випадку збігається з останнім. Тобто він являється частковим випадком цього спектра. Існує також спектральна область реалізацій. Можна показати, що аналогічно випадковому процесу зі скінченною енергією інші випадкові функції зі скінченною енергією (векторні процеси, поля, векторні поля, матричні функції) мають аналогічний енергетичний спектр і спектральну область реалізацій.

Приклади ортогональних розкладів

1. Розклад Карунена-Лоева.

Серед усіх ортогональних розкладів розклад Карунена-Лоева займає особливе місце. Для нього координатними функціями є власні функції кореляційного оператора:

$$\int_T R(t_1, t_2) \psi_{jont}(t_2) dt_2 = \lambda_j \psi_{jont}(t_1),$$

де λ_j – власні числа, відповідні власним функціям $\psi_{jont}(t)$.

Ортогональний розклад (3.52) процесу та його кореляційної функції (3.53) мають вигляд

$$\xi(\omega, t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \chi_j(\omega) \psi_{jont}(t), \quad (3.55)$$

де $\sqrt{\lambda_j} \chi_j(\omega) = \int_T \xi(\omega, t) \overline{\psi_{jont}(t)} dt$;

$$i \quad R(t_1, t_2) = \sum_{j=1}^{\infty} \lambda_j \psi_{jont}(t_1) \overline{\psi_{jont}(t_2)}. \quad (3.56)$$

Координатні функції у розкладі Карунена-Лоева ортогональні, а коефіцієнти розкладу некорельовані.

Розклад Карунена-Лоева займає особливе місце серед усіх інших розкладів завдяки його оптимізаційним властивостям:

а) Якщо задати число членів N часткової суми ряду, цей розклад порівняно з іншими ортогональними розкладами забезпечить най-

меншу середньоквадратичну помилку апроксимації у метриці простору $L_2(\Omega, L_2(T))$:

$$M \int_T \left| \xi(\omega, t) - \sum_{j=1}^N d_j(\omega) \psi_{jont}(t) \right|^2 dt = \varepsilon_{min}^2.$$

б) Якщо зафіксувати точність апроксимації ε_{min}^2 порівняно з іншими ортогональними розкладами розклад Карунена-Лоева забезпечить найменше число членів розкладу:

$$\varepsilon^2 = M \int_T \left| \xi(\omega, t) - \sum_{j=1}^{N_{min}} d_j(\omega) \psi_{jont}(t) \right|^2 dt.$$

в) Розподіл енергії процесу між складовими розкладу Карунена-Лоева приймає вигляд:

$$E_\xi = \sum_{j=1}^{\infty} \lambda_j,$$

де енергія найкращим чином концентрується у перших членах розкладу. Порівняно з іншими ортогональними розкладами швидкість спадання енергетичного спектру у «базисі Карунена-Лоева» найбільша.

2. Розклад Райса.

Це розклад нестационарного процесу $\xi(\omega, t)$ за координатними функціями $\left\{ \exp\left(i \frac{2\pi}{T} jt\right) \right\}_{j=1}^{\infty}$ на скінченному інтервалі:

$$\xi(\omega, t) = \sum_{j=1}^{\infty} d_j(\omega) \exp\left(i \frac{2\pi}{T} jt\right), \quad (3.57)$$

$$R(t_1, t_2) = \sum_{j,s=1}^{\infty} g_{js} \exp\left\{i \frac{2\pi}{T} (jt_1 - st_2)\right\}. \quad (3.58)$$

Тут

$$d_j(\omega) = \int_T \xi(\omega, t) \exp\left(-i \frac{2\pi}{T} jt\right) dt;$$

$$g_{js} = \iint_{T \times T} R(t_1, t_2) \exp\left\{i \frac{2\pi}{T} (jt_1 - st_2)\right\} dt_1 dt_2.$$

Цей розклад дає узагальнення розкладу стаціонарного періодичного процесу (3.41), (3.42) на нестаціонарний процес. На відміну від (3.42) коефіцієнти розкладу (3.57) корельовані.

3.3.4. Розклади періодично корельованих та споріднених з ними процесів зі скінченною середньою потужністю

Це підкласи «нестаціонарних» випадкових процесів, які об'єднують властивості періодичності та стохастичності. Розглядаються процеси, які задовольняють умову (3.37) скінченності середньої потужності.

Простори випадкових процесів зі скінченною середньою потужністю. Особливості обчислення характеристик розглядуваних процесів полягає у тому, що виконується усереднення за ансамблем реалізацій, у часі, а також, крім цього, виконується граничний перехід, коли інтервал «спостереження» прямує до нескінченності. Так, середня потужність, згідно з (3.37) дорівнює

$$\bar{P} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} M \int_{-\tau/2}^{\tau/2} |\xi(\omega, t)|^2 dt.$$

Введемо спочатку гільбертів простір B_2 , де скалярний добуток, норма та відстань задаються, відповідно, формулами:

$$(\xi, \eta)_{B_2} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{-\tau/2}^{\tau/2} \xi(\omega, t) \overline{\eta(\omega, t)} dt;$$

$$\|\xi\|_{B_2} = \sqrt{\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{-\tau/2}^{\tau/2} |\xi(\omega, t)|^2 dt};$$

$$\rho(\xi, \eta)_{B_2} = \sqrt{\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{-\tau/2}^{\tau/2} |\xi(\omega, t) - \eta(\omega, t)|^2 dt}.$$

Така метрика носить назву метрики Бора-Безиковича. Простір B_2 , на відміну від розглядуваного раніше простору L_2 є несепара-

бельним гільбертовим простором, тобто у ньому не існує зчисленного базису.

Крім простору B_2 детермінованих функцій розглядається також L_2 –простір над гільбертовим простором випадкових процесів $L_2(\Omega, B_2)$, де скалярний добуток, норма й відстань визначаються так:

$$(\xi, \eta)_{L_2(\Omega, B_2)} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} M \int_{-\tau/2}^{\tau/2} \xi(\omega, t) \overline{\eta(\omega, t)} dt;$$

$$\|\xi\|_{L_2(\Omega, B_2)} = \sqrt{\lim_{\tau \rightarrow \infty} \frac{1}{\tau} M \int_{-\tau/2}^{\tau/2} |\xi(\omega, t)|^2 dt};$$

$$\rho(\xi, \eta)_{L_2(\Omega, B_2)} = \sqrt{\lim_{\tau \rightarrow \infty} \frac{1}{\tau} M \int_{-\tau/2}^{\tau/2} |\xi(\omega, t) - \eta(\omega, t)|^2 dt}.$$

Випадкові процеси зі скінченною середньою потужністю належать цьому просторові, причому їх середня потужність дорівнює

$$\bar{P}_\xi = \|\xi\|_{L_2(\Omega, B_2)}^2 = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} M \int_{-\tau/2}^{\tau/2} |\xi(\omega, t)|^2 dt;$$

майже всі реалізації процесу належать простору B_2 . L_2 –простір над гільбертовим простором $L_2(\Omega, B_2)$ – несепарабельний.

Періодично корельовано випадкові процеси. Процеси зі скінченною середньою потужністю називаються періодично корельованими (ПКВП), якщо їх кореляційні функції задовольняють умову

$$R(t+T, s+T) = R(t, s),$$

при деякому фіксованому $T > 0$, яке називається періодом корельованості. Сумарна потужність таких процесів може бути обчислена за періодом як

$$\bar{P}_\xi^T = \frac{1}{T} \int_0^T R(t, t) dt.$$

Для періодично корельованих процесів розглядуваного класу справджується таке твердження: ПКВП зображується у вигляді

$$\xi(\omega, t) = \sum_{k=-\infty}^{\infty} \exp(i\Lambda kt) \xi_k(\omega, t)$$

тоді і тільки тоді, коли їх кореляційна функція зображується у вигляді

$$R(t,s) = \sum_{j,k=-\infty}^{\infty} D_{kj}(t-s) \exp\{i\Lambda(kt-js)\}.$$

Тут $\Lambda = \frac{2\pi}{T}$; $[\xi_k(\omega,t)]_{k=0,\pm 1,\pm 2,\dots}^T$ – векторний зчисленновимірний стаціонарний випадковий процес з кореляційною матрицею $[D_{kj}(t-s)]_{k,j=0,\pm 1,\pm 2,\dots}^T$.

Таке зображення ПКВП називають зображенням через стаціонарні компоненти, тобто розклад установлює відповідну еквівалентність часткового випадку «нестационарного» процесу $\xi(\omega,k)$ та векторного стаціонарного процесу $\bar{\xi}(\omega,t) = [\xi_k(\omega,t)]_{k=0,\pm 1,\dots}$.

Як і у довільному «нестационарному» процесі спектр ПКВП двочастотний, але спектральна міра зосереджена на множині прямих $\lambda - \mu = k\Lambda$, $k=0,\pm 1,\pm 2,\dots$, паралельних бісектрисі I–III квадрантів. У процесів ПКВП корельовані лише гармоніки, віддалені на відстані, кратні $\Lambda = \frac{2\pi}{T}$.

У частковому випадку ПКВП з дискретним часом зображення процесу спрощується. Наведемо відповідні результати.

Випадкова послідовність називається періодично корельованою, якщо її кореляційна функція задовольняє умову

$$R(k+N, l+N) = R(k, l),$$

де – період корельованості.

Справджується таке твердження:

Періодично корельована випадкова послідовність допускає зображення

$$\xi(\omega, k) = \sum_{p=0}^{N-1} \exp\left(i \frac{2\pi}{N} pk\right) \xi_p(\omega, k)$$

тоді і тільки тоді, коли її кореляційна функція допускає зображення у вигляді

$$R(k, j) = \sum_{p,q=0}^{N-1} \exp\left[i \frac{2\pi}{T} (pk - qj)\right] D_{pq}(k - j).$$

Тут ПКВП зображується через стаціонарні компоненти $\xi_p(\omega, k)$, $p=0,1,\dots,N-1$, які утворюють векторну N -вимірну стаціонарну послідовність з кореляційною матрицею

$$D_{pq}(k-j) = \xi_p(\omega, k) \overline{\xi_q(\omega, j)}.$$

Спектральна двочастотна міра періодично корельованої випадкової послідовності періоду зосереджена на відрізках прямих

$$\lambda - \mu = \frac{2\pi}{N}k, \quad k = -N+1, \dots, N-1,$$

які містяться в середині квадрата $[0, 2\pi) \times (0, 2\pi]$.

Важливою для практичних застосувань є властивість відліків значень періодично корельованої випадкової послідовності. Множина таких значень, узятих через період корельованості, за всякої початкової фази $j=0, 1, \dots, N-1$, тобто $\eta_j(\omega, k) = \xi(\omega, j+kN)$, $k=0, \pm 1, \pm 2, \dots$, утворює стаціонарну випадкову послідовність. Тому сукупність всіх таких стаціонарних послідовностей утворює N -вимірну векторну стаціонарну послідовність.

Бі-полі-періодично корельовано випадкові процеси. ПКВП мають ряд узагальнень про які кажуть «споріднені з ПКВП». Широке узагальнення визначається розкладом кореляційної функції у ряд

$$R = (t+v, s+v) = \sum_{k=-\infty}^{\infty} C_k(t, s) \exp(i\lambda kv),$$

що визначає клас майже періодично корельованих процесів. Більш простим випадком є поліперіодично корельовані випадкові процеси, для яких $\lambda_k = \sum_{j=1}^M U_{kj} \Lambda_j$, де U_{kj} – цілі числа, Λ_j – задовольняють умову

$$\sum_{j=1}^N \alpha_j \Lambda_j = 0 \quad \text{лише тоді, коли } \alpha_j = 0, \quad j = \overline{1, N}.$$

Якщо $M=2$, дістають бі-періодично корельовані випадкові процеси. Найбільш просту форму мають розклади бі-ПКВП з дискретним часом.

Такі бі-ПКВП мають розклад

$$\xi(\omega, k) = \sum_{k_0, k_1} \exp[i(k_0 \Lambda_0 + k_1 \Lambda_1)] \xi_{k_1, k_0}(\omega, k),$$

що еквівалентно розкладу їх кореляційних функцій:

$$R(k, j) = \sum_{k_0, k_1, j_0, j_1} \exp\{i(k_0 \Lambda_0 + k_1 \Lambda_1)k + (j_0 \Lambda_0 + j_1 \Lambda_1)j\} D_{k_0, k_1, j_0, j_1}(k-j).$$

Тут $[\xi_{k_1, k_0}(\omega, k)]$ – матрична стаціонарна випадкова послідовність. Її можна розглядати як модульовану двома множинами гармонік, кратних періодам $\frac{2\pi}{\Lambda_0}$ і $\frac{2\pi}{\Lambda_i}$, оскільки

$$\exp[i(k_0\Lambda_0 + k_1\Lambda_1)k] = \exp(ik_0\Lambda_0k)\exp(ik_1\Lambda_1k).$$

Якщо $M=1$, дістають періодично корельовані випадкові послідовності, які були розглянуті вище.

3.4. Приклади класів випадкових процесів

Класична та сучасна кореляційна (енергетична) теорія випадкових процесів дає можливість вивчати сигнали, як фізичні процеси, тільки на рівні їх других моментів розподілу. Для ряду задач практики цього недостатньо і виникає необхідність описувати їх повністю з точки зору можливостей випадкових процесів. Особливо важливі у застосуванні такі класи випадкових процесів, як білі шуми, марківські процеси та лінійні випадкові процеси.

3.4.1. Білі шуми

Білий шум – це випадковий процес з незалежними чи некорельованими значеннями. Білі шуми розглядають у вузькому (власному), а також у широкому (у межах кореляційної теорії) розумінні. Вони бувають з неперервним та дискретним часом, стаціонарними та не-стаціонарними, гаусовими та негаусовими.

Білий шум з неперервним часом. Білий шум з неперервним часом – це узагальнений випадковий процес. Для нього не має сенсу вираз «значення $\xi(\omega, t)$ процесу ξ в точці t ». Він є лінійним функціоналом

$$\xi(\omega, \varphi) = \int_T \xi(\omega, t)\varphi(t)dt. \quad (3.59)$$

Узагальненим випадковим процесом, заданим на імовірнісному просторі (Ω, τ, P) із параметричною множиною T , називають випадковий лінійний функціонал $\xi(\omega, t)$ з простором основних функцій Φ .

Випадковий лінійний функціонал (3.59) визначає сім'ю звичайних випадкових величин

$$\Xi_{\Phi}(\omega) = \{\xi(\omega, \varphi), \omega \in \Omega, \varphi \in \Phi\},$$

яка лінійно залежить від φ . Вона стандартним чином описується узгодженою сім'єю скінченновимірних розподілів $\{F_n(\vec{x}, \vec{\varphi})\}_{n=1}^{\infty}$, де:

$$\vec{x} = (x_1, x_2, \dots, x_n),$$

$$\vec{\varphi} = (\varphi_1, \varphi_2, \dots, \varphi_n),$$

$$F_n(x_1\varphi_1; x_2\varphi_2; \dots, x_n\varphi_n) = P\{\xi(\omega, \varphi_1) \leq x_1; \xi(\omega, \varphi_2) \leq x_2, \dots, \xi(\omega, \varphi_n) \leq x_n\}.$$

Додатково вимагається, щоб існувала неперервна залежність від φ як у просторі скінченновимірних розподілів, так і в просторі основних функцій.

Білий шум можна дістати як узагальнену похідну випадкового процесу з незалежними приростами. Якщо процес однорідний, білий шум буде стаціонарним, якщо неоднорідний – нестаціонарним. Однорідність процесу – це стаціонарність відносно його приростів. Так, випадковий процес $\{\eta(\omega, t), P[\eta(\omega, 0) = 0] = 1, t \in T\}$ з незалежними приростами називають однорідним (або зі стаціонарними приростами), якщо його прирости $\{\eta(\omega, t + \tau) - \eta(\omega, t), t, t + \tau \in T\}$ мають розподіл, який не залежить від t . Інакше кажучи, при всіх η таких, що $t + \eta \in T$, характеристична функція процесу $\eta(\omega, t)$ задовольняє співвідношення

$$f(u; s, \tau) = f(u, s + \eta, \tau + \eta) = f(u, \tau - s).$$

Наведемо приклад гаусового процесу з незалежним приростами, який застосовують, щоб дістати білий шум.

Вінеровим $\hat{W}(\omega, t)$ називають однорідний випадковий процес

$$\{\hat{W}(\omega, t), P[\hat{W}(\omega, 0) = 0] = 1, t \in [0, \infty)\}$$

з незалежними приростами. Характеристична функція його приростів

$$f_{\hat{W}}(u; t_1, t_2) = M \exp\{ju[\hat{W}(\omega, t_2) - \hat{W}(\omega, t_1)]\} = \exp\left\{ju(t_2 - t_1)a - \frac{\sigma^2 t}{2}(t_2 - t_1)\right\},$$

тобто прирости вінерова процесу є гаусовими з параметрами

$$\{a(t_2 - t_1), \sigma\sqrt{t_2 - t_1}\}, \quad t_2 > t_1,$$

де a – швидкість змінювання математичного сподівання,

σ^2 – швидкість змінювання дисперсії.

Вінерів процес нестаціонарний. Він неперервний за ймовірністю, а також з ймовірністю 1. Останнє означає, що майже всі реалізації процесу є неперервними. Проте вони майже скрізь недиференційов-

ні. Крім того, вони мають необмежену варіацію (довжину) Рис.3.13 покаже типову реалізацію вінерового процесу.

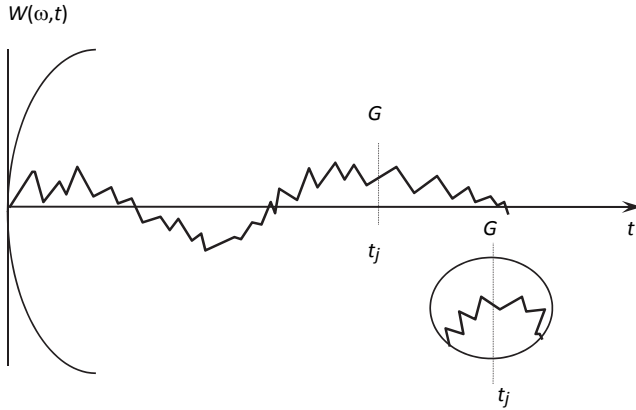


Рис. 3.13. – Типова реалізація вінерового процесу

Якщо вінерів процес центрований математичним сподіванням, його узагальнена похідна $W'(omega, t) = v_W(omega, t)$ визначає стаціонарний гаусів (вінерів) білий шум з кореляційною функцією

$$R_W(t_1, t_2) = \sigma^2 \delta(t_1 - t_2). \quad (3.60)$$

Його характеристичний функціонал має вигляд

$$f[u(t)] = \exp\left[-\frac{N_0}{2} \int_{-\infty}^{\infty} u^2(x) dx\right], \quad (3.61)$$

де $N_0 = \sigma^2$ – спектральна густина білого шуму.

Наведемо ще один приклад однорідного процесу з незалежними приростами.

Однорідний випадковий процес $\pi(omega, t)$, $t \in [0, \infty)$ називають пуассоновим (простим), якщо від має такі властивості:

$$P[\pi(omega, 0) = 0] = 1;$$

приріст $\pi(omega, t_2) - \pi(omega, t_1)$, $t_1, t_2 \in [0, \infty)$, $t_1 < t_2$ має розподіл Пуассона з параметрами $\lambda(t_2 - t_1)$, тобто

$$P[\pi(omega, t_2) - \pi(omega, t_1) = n] = \frac{[\lambda(t_2 - t_1)]^n}{n!} e^{-\lambda(t_2 - t_1)}, \quad n = 0, 1, 2, \dots;$$

для будь-яких $0 = t_0 < t_1 < t_2 < \dots < t_n$ прирости

$\pi(\omega, t_1) - \pi(\omega, t_0), \pi(\omega, t_2) - \pi(\omega, t_1), \dots, \pi(\omega, t_n) - \pi(\omega, t_{n-1})$ незалежні.

Характеристична функція значень пуассонового процесу має вигляд

$$f_{\pi}(n) = \exp\left[\lambda \sum_{i=1}^n (e^{iu} - 1)\right]$$

Реалізації процесу являють собою функції, які складаються лише з ділянок сталості і стрибків, що дорівнюють одиниці (рис. 3.14). Проте сам процес неперервний за ймовірністю. (Підкреслимо, що це є яскрава ілюстрація, що випадковий процес не можна ототожнювати з його реалізаціями – пуассонів процес та його реалізації мають діаметрально протилежні аналітичні властивості: якщо процес неперервний у відповідному розумінні, майже всі його реалізації розривні у своєму розумінні).

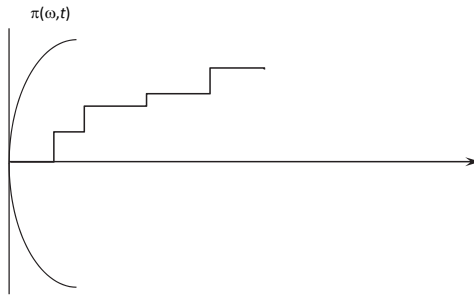


Рис. 3.14. – Типова реалізація пуассонового процесу

Уводять також пуассонові процеси $\hat{\pi}(\omega, t)$, які мають випадкові величини стрибків.

Узагальнена похідна пуассонового процесу $\pi'(\omega, t) = v_{\pi}(\omega, t)$ визначає пуассонів білий шум з кореляційною функцією

$$R_{\pi}(t_1, t_2) = \lambda \delta(t_1 - t_2) \quad (3.62)$$

і характеристичним функціоналом

$$f[n(t)] = \exp\left[\lambda \int_0^{\infty} [e^{iu(t)} - 1] dt\right]. \quad (3.63)$$

Вибравши процес з незалежними приростами, який має інший закон розподілу, дістанемо білий шум з новим законом розподілу. Характерна особливість білого шуму – його δ -корельованість. Кореляційна функція як δ -функція є детермінованою узагальненою функцією.

Важлива властивість звичайних (не узагальнених) стаціонарних процесів – існування спектральної густини як перетворення Фур'є кореляційної функції. Для узагальнених кореляційних функцій (3.60) і (3.62) можна строго виконати перетворення Фур'є у термінах теорії узагальнених функцій. Однак виконуючи формально перетворення Фур'є функцій (3.60) і (3.62), можна також одержати вірний результат – спектральна густина білого шуму буде рівномірною на всіх частотах (рис. 3.15). Саме так у технічній літературі частіше вводився білий шум, тобто тільки у рамках кореляційної теорії. Якщо додавався термін «гаусів білий» шум, він не підкреслювався ніяким математичним співвідношенням. Сучасна теорія білих шумів дає можливість їх розділяти на рівні кореляційної теорії, де для всіх білих шумів кореляційна функція є δ -функцією з деяким коефіцієнтом пропорційності – (3.60), (3.62), а спектральна густина – рівномірна за своєю величиною $N_0 = \sigma^2$ або $N_0 = \lambda$, однак ця теорія дає також можливість повністю описувати білий шум його характеристичним функціоналом – (3.61), (3.63). Характерно, білі шуми і в термінах кореляційної теорії і при повному їх описуванні визначаються характеристиками і параметрами процесів з незалежними приростами, з яких вони одержані.

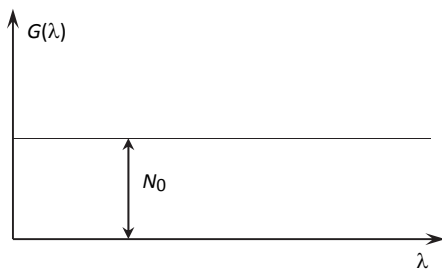


Рис. 3.15. – Спектральна густина будь-якого білого шуму з неперервним часом

Білий шум з дискретним часом. Білий шум з дискретним часом – це послідовність незалежних або некорельованих випадкових величин

$$\Xi_Z(\omega) = \{\xi(\omega, k), \omega \in \Omega, k \in Z\}, \quad Z = \{0, \pm 1, \pm 2, \dots\}. \quad (3.64)$$

Цей білий шум звичайний (не узагальнений) випадковий процес, який звичайним чином можна повністю описати сім'єю скінченновимірних розподілів або в межах кореляційної теорії. Вважають, що $M\xi(\omega, k) = 0$.

Як усякий сепарабельний випадковий процес, білий шум з дискретним часом повністю описується узгодженою сім'єю скінченновимірних розподілів $\{F_n(\vec{x}, \vec{n})\}_{n=1}^{\infty}$, де

$$\vec{x} = (x_1, x_2, \dots, x_n),$$

$$\vec{n} = (1, 2, \dots, n),$$

$$F_n(x_1, 1, x_2, 2, \dots, x_n, n) = P\{\xi(\omega, 1) \leq x_1, \xi(\omega, 2) \leq x_2, \dots, \xi(\omega, n) \leq x_n\}.$$

Згідно з незалежністю випадкових величин маємо

$$F_n(\vec{x}, \vec{n}) = \prod_{i=1}^n F(x_i, i). \quad (3.65)$$

Для абсолютно неперервних функцій розподілу дістаємо

$$P_n(\vec{x}, \vec{n}) = \prod_{i=1}^n P(x_i, i). \quad (3.66)$$

де $P_n(\vec{x}, \vec{n}) = \frac{\delta^n F_n(\vec{x}, \vec{n})}{\delta x_1 \dots \delta x_n}$; $P(x_i, i) = \frac{dF(x_i, i)}{dx_i}$ – щільність імовірності відпо-

відно n та одновимірна.

Із означенням білого шуму з дискретним часом впливає, що багатовимірна функція розподілу повністю визначається одновимірною. Це характеристична особливість білого шуму у власному розумінні.

Стационарний у вузькому розумінні випадковий процес з дискретним часом характеризується інваріантністю функції розподілу відносно часу, тобто

$$\begin{aligned} P(x_i, i) &= P(x_i), \\ P_n(\vec{x}, \vec{n}) &= [P(x_i)]^n. \end{aligned} \quad (3.67)$$

Із стаціонарності процесу у вузькому розумінні впливає стаціонарність у широкому розумінні. Тому кореляційна функція такого білого шуму має вигляд

$$R(k_1, k_2) = R(0)\delta_{k_1, k_2} = \sigma^2 \delta_{k_1, k_2}. \quad (3.68)$$

Це опис білого шуму в межах кореляційної теорії.

Зображення кореляційної функції стаціонарної послідовності, центрованої математичним сподіванням, згідно з теоремою Герглота визначається виразом

$$R(s) = \int_{-\pi}^{\pi} \exp(js\lambda) d\Phi(\lambda).$$

Звідси, враховуючи δ -корельованість (у розумінні δ -функції Кронекера) випадкової послідовності, дістаємо рівномірну в смузі $[-\pi, \pi]$ спектральну густину $G(\lambda) = \frac{d\Phi(\lambda)}{d\lambda}$, де $\Phi(\lambda)$ – абсолютно неперервна спектральна функція. Рівномірність спектральної густини білого шуму з дискретним часом у скінченній смузі частот є його характеристичною властивістю в межах кореляційної теорії (рис. 3.16)

Випадковий процес з дискретним часом називають гаусовим, якщо будь-яка – вимірна функція розподілу є гаусовою.

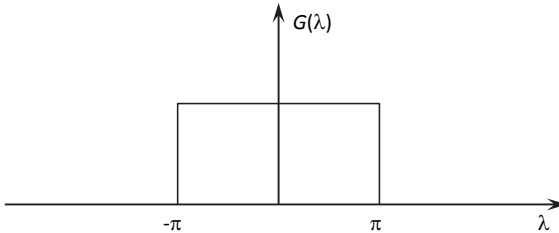


Рис. 3.16. – Спектральна густина будь-якого білого шуму з дискретним часом

Запишемо n -вимірну щільність ймовірності гаусового білого шуму з дискретним часом:

$$P_n(\vec{x}, \vec{n}) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{x_j^2}{2\sigma_j^2}\right\}. \quad (3.69)$$

У стаціонарному випадку, тому

$$P_n(\vec{x}, \vec{n}) = \left[\frac{1}{\sqrt{2\pi}\sigma}\right]^n \prod_{j=1}^n \exp\left\{-\frac{x_j^2}{2\sigma^2}\right\} = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{j=1}^n x_j^2\right\}. \quad (3.70)$$

Стаціонарний білий шум з дискретним часом еквівалентний квазібілому шуму з неперервним часом, який має рівномірну спектральну густину у смузі $\sigma \leq \lambda \leq \Lambda$, тобто

$$G_{\text{кв.б.}}(\lambda) = \begin{cases} N_0, & \text{якщо } \sigma \leq \lambda \leq \Lambda; \\ 0, & \text{якщо } \Lambda < \lambda < \infty, \end{cases} \quad (3.71)$$

і відповідно кореляційну функцію

$$R(\tau) = N_0 \frac{\sin \Lambda \tau}{\Lambda \tau}. \quad (3.72)$$

Такий випадковий процес згідно з теоремою Котельникова—Шеннона можна зобразити у вигляді ряду

$$\xi(\omega, t) = \sum_{k=-\infty}^{\infty} \xi(\omega, k\Delta t) \frac{\sin \Lambda(t - k\Delta t)}{\Lambda(t - k\Delta t)}, \text{ де } \Delta t = \frac{\pi}{\Lambda}.$$

Відліки процесу $\{\xi(\omega, k\Delta t)\}_{k=-\infty}^{\infty}$ є некорельованими величинами, тобто визначають білий шум з дискретним часом $\xi(\omega, k\Delta t)$, $k \in Z$, заданий у межах кореляційної теорії, через зміну масштабу часу $\Delta t = \frac{1}{2F}$ (де $F = \frac{\Lambda}{2\pi}$), масштаб частот змінюється в

$\frac{1}{\Delta t} = 2F$ раз, і на відміну від спектра розглянутого раніше білого шуму з дискретним часом, заданого в смузі $-\pi \leq \lambda < \pi$, тут спектр задано в смузі $-2\pi F \leq \lambda \leq 2\pi F$.

Якщо квазібілий шум гаусів, його відліки, взяті за теоремою Котельникова—Шеннона, визначають еквівалентний гаусів білий шум з дискретним часом.

3.4.2 Марковські процеси

Марковські процеси є складнішим класом випадкових процесів порівняно з класом білих шумів. Білий шум (з дискретним часом) повністю визначений, якщо відомі всі його одновимірні функції, марковські випадкові процеси можна задати, коли відомі всі їх двовимірні функції розподілу (або одновимірні та умовні).

Марківські процеси з неперервним часом, які приймають аналогові значення. Сепарабельний випадковий процес повністю описується узгодженою сім'єю скінченномірних розподілів

$$\{F_n(x_1, t_1; x_2, t_2; \dots; x_n, t_n)\}_{n=1}^{\infty},$$

$$\text{де } F_n(x_1, t_1; x_2, t_2; \dots; x_n, t_n) = P\{\xi(\omega, t_1) \leq x_1; \xi(\omega, t_2) \leq x_2; \dots; \xi(\omega, t_n) \leq x_n\}.$$

Оскільки

$$\begin{aligned}
& P\{\xi(\omega, t_1) \leq x_1; \xi(\omega, t_2) \leq x_2; \dots; \xi(\omega, t_n) \leq x_n\} = \\
& = P\{\xi(\omega, t_1) \leq x_1\} P\{\xi(\omega, t_2) \leq x_2, \dots; \xi(\omega, t_n) \leq x_n | \xi(\omega, t_1) \leq x_1\} = \\
& = P\{\xi(\omega, t_1) \leq x_1\} \times P\{\xi(\omega, t_2) \leq x_2 | \xi(\omega, t_1) \leq x_1\} \times \\
& \quad \times P\{\xi(\omega, t_3) \leq x_3, \dots, \xi(\omega, t_n) \leq x_n | \xi(\omega, t_1) \leq x_1, \xi(\omega, t_2) \leq x_2\} = \dots \\
& = P\{\xi(\omega, t_1) \leq x_1\} \prod_{j=2}^n P\{\xi(\omega, t_j) \leq x_n | \xi(\omega, t_1) \leq x_1; \dots; \xi(\omega, t_j) \leq x_{j-1}\},
\end{aligned}$$

його n -вимірну функцію розподілу можна записати у вигляді

$$F_n(x_1, t_1; \dots; x_n, t_n) = F(x_1, t_1) \prod_{j=2}^n F(x_j, t_j | x_1, t_1; x_2, t_2; \dots; x_{j-1}, t_{j-1}). \quad (3.73)$$

Для абсолютно неперервних розподілів використовується відповідна рівність щодо щільності ймовірності:

$$p_n(x_1, t_1; x_2, t_2; \dots; x_n, t_n) = p(x_1, t_1) \prod_{j=2}^n p(x_j, t_j | x_1, t_1; \dots; x_{j-1}, t_{j-1}). \quad (3.74)$$

Це відповідно n -вимірна функція розподілу й відповідна n -вимірна щільність ймовірності довільного випадкового процесу, записані через умовні ймовірності, де умова охоплює всі попередні значення параметру t , починаючи від першого $- t_1$.

Випадковий процес називають марківським, якщо його багатовимірні функції визначаються одновимірною функцією розподілу і функціями розподілу переходів, тобто

$$F_n(x_1, t_1; \dots; x_n, t_n) = F(x_1, t_1) \prod_{j=2}^n F(x_j, t_j | x_{j-1}, t_{j-1}). \quad (3.75)$$

або в термінах багатовимірних щільностей імовірності

$$p_n(x_1, t_1; \dots; x_n, t_n) = p(x_1, t_1) \prod_{j=2}^n p(x_j, t_j | x_{j-1}, t_{j-1}) \quad (3.76)$$

Вирази (3.75)–(3.76) називають марківською властивістю. Такий процес ще називають однозв'язним на відміну від багатозв'язного процесу, у якому ймовірність переходу залежить від скінченного числа попередніх значень процесу.

Щільність ймовірностей переходу марківського процесу, крім загальних вимог до щільності, має задовольняти інтегральне рівняння Колмогорова-Чепмена

$$p_n(x_3, t_3 | x_1, t_1) = p(x_3, t_3 | x_2, t_2) p(x_2, t_2 | x_1, t_1) dx_2, \quad t_1 < t_2 < t_3. \quad (3.77)$$

Розв'язування інтегрального рівняння є складною задачею. Тому важливими є процеси, для яких рівняння можна звести до диференціального. Це можна зробити для дифузійних процесів, які мають з імовірністю 1 неперервні реалізації та щільності ймовірності переходу яких задовольняють такі умови: для будь-якого δ існують границі

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{|y-x|>\delta} p(y,t|x,t-\Delta t) dy = 0, \quad (3.78)$$

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{|y-x|\leq\delta} (y-x)^k p(y,t|x,t-\Delta t) dy = A_k(x,t), \quad k=1,2; \quad A_k(x,t) \equiv 0, \quad k \geq 3. \quad (3.79)$$

Згідно з (3.78)–(3.79) моменти зміщень першого і другого порядку обмежені, а моменти вищих порядків дорівнюють нулю; момент зміщення першого порядку $A_1(x,t)$ називають коефіцієнтом зносу, а момент зміщення другого порядку $A_2(x,t) \geq 0$ – коефіцієнтом дифузії.

Для дифузійних процесів рівняння (3.77) переходять у диференціальне рівняння в частинних похідних другого порядку

$$\frac{\partial_p p(y,t|x,s)}{\partial t} = -\frac{\partial}{\partial y} [A_1(y,t) p(y,t|x,s)] + \frac{1}{2} \frac{\partial^2}{\partial y^2} [A_2(y,t) p(y,t|x,s)]. \quad (3.80)$$

Рівняння (3.80) називають прямим рівнянням Колмогорова, або рівнянням Фоккера-Планка.

Марківський процес називають стаціонарним, якщо розподіл імовірностей інваріантний відносно часового зсуву.

Щільність імовірності переходу стаціонарного дифузійного процесу залежить від різниці $\tau = t - s$, а коефіцієнти $A_1(y,t)$, $A_2(y,t)$ не залежать від часового параметру.

Рівняння Колмогорова-Чепмена для гаусового марківського процесу набирає вигляду

$$R(s, \tau) = R(s, t) R(t, \varepsilon), \quad (3.81)$$

а в стаціонарному випадку –

$$R(t + \tau) = R(t) R(\tau), \quad t > 0, \tau > 0. \quad (3.81a)$$

Рівняння (3.81a) має єдиний розв'язок

$$R(\tau) = e^{-\lambda \tau}, \quad \lambda > 0, \tau \geq 0.$$

Цей результат характеризує теорему Дуба: стаціонарний гаусів процес є марківським тоді й тільки тоді, коли його кореляційна функція експоненціальна.

У загальному випадку гаусів процес з незалежними приростами є марківським і, взагалі, клас стаціонарних марківських процесів збігається з класом однорідних процесів із незалежними приростами. Зокрема, вінерів і пуассонів процеси є марківськими.

Марківські процеси набули дуже широке поширення у теорії електровз'язку. Ці моделі розглядають і як класичні, і як сучасні – нові задачі практики ефективно розв'язуються з використанням цих моделей.

Ланцюги Маркова. Розглядають марківські процеси у загальному випадку, як це описано вище, марківські процеси з неперервним часом, які набувають дискретні значення, а також процеси з дискретним часом, які приймають неперервні або дискретні значення. У останньому випадку, коли марківські процеси з дискретним часом приймають дискретні значення, використовується спеціальний термін – ланцюги Маркова. Розглянемо такі процеси.

Таким чином, ланцюг Маркова – це марківський процес з дискретним часом, який набуває скінченної чи зчисленної множини значень. Такі процеси у застосуваннях називають цифровими. Відповідно у теорії електровз'язку говорять і про цифрові марківські сигнали.

Нехай ланцюг Маркова набуває значень $x_1, x_2, x_1, \dots, x_n$ із скінченної чи зчисленної множини. У цьому разі марківську властивість записують так. Враховуючи рівність

$$P\{\xi(\omega, t_{k_n})=x_{k_n} | \xi(\omega, t_{k_1})=x_{k_1}, \dots, \xi(\omega, t_{k_{n-1}})=x_{k_{n-1}}\}, \quad (3.82)$$

одержуємо

$$P\{\xi(\omega, t_{k_n})=x_{k_n} | \xi(\omega, t_{k_{n-1}})=x_{k_{n-1}}\}, \quad (3.82a)$$

де $t_1 < t_2 < \dots < t_n$ і $k_1 < k_2 < \dots < k_n$ – будь які цілі додатні числа.

Рівняння Колмогорова-Чепмена (3.77) для ланцюга Маркова записують так:

$$P(t_3, j | t_1, i) = \sum_k P(t_3, j | t_2, k) P(t_2, k | t_1, i). \quad (3.83)$$

Ланцюг Маркова називають однорідним, якщо ймовірність переходу залежить лише від різниці $t_j - t_i$.

Наведемо приклади застосування ланцюга Маркова у теорії електровз'язку.

Один із них – модель джерела телеграфного тексту. Так, якщо алфавіт містить 32 літери, для побудови марківської моделі досить задати ймовірність появи літер $\{P_j\}_{j=1}^{32}$ і матрицю $[P(k|j)]_{k,j=1}^{32}$ ймовірностей переходу, де $P(k|j)$ – ймовірність появи k -ї літери, якщо попередньою була j -та літера.

Марківська модель джерела дискретних повідомлень набула дуже широкого поширення.

Другий приклад – модель двійкового каналу зв'язку з пам'яттю.

Нагадаємо, що модель каналу зв'язку – це трійка: множина вхідних сигналів та розподіл ймовірностей вихідного сигналу, коли вхідний сигнал заданий. У дискретному каналі вхідні та вихідні сигнали – послідовність кодових символів. У двійковому каналі символи можуть приймати тільки два значення – 0 чи 1. Кожен символ може бути переданий правильно або ні.

Марківська модель двійкового каналу з пам'яттю визначається матрицею перехідних ймовірностей

$$P = \begin{bmatrix} 1-P_1 & P_1 \\ P_2 & 1-P_2 \end{bmatrix},$$

де P_1 – умовна ймовірність прийняти $(i+1)$ -й символ помилково, якщо попередній прийнято правильно; $1-P_1$ – умовна ймовірність прийняти $(i+1)$ -й символ правильно, якщо попередній прийнято правильно; P_2 – умовна ймовірність прийняти $(i+1)$ -й символ помилково, якщо попередній прийнято помилково; $1-P_2$ – умовна ймовірність прийняти $(i+1)$ -й символ правильно, якщо попередній прийнято помилково.

Ця найпростіша модель дискретного каналу з пам'яттю. Тут ймовірність помилки створює простий ланцюг Маркова, тобто залежить від того, правильно чи помилково прийнятий попередній символ, але не залежить від того, який символ передається.

3.4.3. Лінійні випадкові процеси

У теорії та практиці електрозв'язку часто зустрічаються випадкові процеси, котрі являють собою суми досить великого числа елементарних складових. Раніше у цих випадках припускалося, що виконуються умови нормалізації й розглядалися гаусові процеси. У другій половині 20-го десятиріччя істотно розвинено математичний апарат, який дає змогу дослідити більш широке коло процесів, які мають

безмежно подільний розподіл. Відповідні процеси дістали назву лінійних випадкових процесів. Клас цих процесів замкнений відносно лінійних операцій.

Лінійні випадкові процеси з неперервним часом. Дійсний сепарабельний випадковий процес $\xi(\omega, t)$, $\omega \in \Omega$, $t \in T$ на ймовірністному просторі (Ω, \mathcal{F}, P) називають лінійним, якщо його можна подати у вигляді

$$\xi(\omega, t) = \lim_{n \rightarrow \infty} \sum_{k=1}^n G_{n_k}(\omega, t), \quad (3.84)$$

де $\{G_{n_k}(\omega, t), k=1, n, n=1, 2, \dots\}$ – нескінченна послідовність незалежних для кожного фіксованого n процесів, доданки послідовностей сум яких задовольняють умову рівномірної нескінченної малюзи або умову граничної нехтуваності:

$$\lim_{n \rightarrow \infty} \max_k P\{G_{n_k}(\omega, t) \geq \varepsilon\} = 0 \quad (3.85)$$

для кожного $\varepsilon > 0$. Це означає, що при $n \rightarrow \infty$ роль кожного доданка в дограничній сумі для $\xi(\omega, t)$ стає як завгодно малою.

Послідовності сум незалежних випадкових величин

$$G_n(\omega, t) = \sum_{k=1}^n G_{n_k}(\omega), \quad (3.86)$$

як відзначалося у розділі 2, мають граничні закони розподілу, клас яких збігається з класом усіх безмежно подільних розподілів. Тобто згідно з означенням лінійного процесу (3.84), його одновимірні закони розподілу є безмежно подільними. Цьому класові законів розподілу, окрім гаусового, належать пуассонів, гамма-розподіл, логорифмічно-нормальний, Лапласа та ін.

Нехай випадковий процес $\xi(\omega, t)$, $\omega \in \Omega$, $t \in T$ подано у вигляді

$$\xi(\omega, t) = \int_{-\infty}^{\infty} \varphi(\tau, t) d\eta(\omega, \tau), \quad t \in T, \quad (3.87)$$

де $\varphi(\tau, t)$ – не випадкова функція, котра набуває при кожному $t \in T$ рівномірно за τ скінченних значень; $\eta(\omega, \tau)$, $\tau \in (-\infty, \infty)$ – стохастично неперервний сепарабельний випадковий процес з незалежними приростами.

Можна довести, що за цих умов процес (3.87) – лінійний процес. Більш того, справджується таке твердження:

Теорема 3.14.

Нехай

$$\xi(\omega, t) = \int_{-\infty}^{\infty} \varphi(\tau, t) d\eta(\omega, \tau), \quad t \in (-\infty, \infty), \quad (3.88)$$

де $\varphi(\tau, t)$, $t \in (-\infty, \infty)$ – дійсна не випадкова числова вимірною інтегрована в квадраті за τ при всіх t функція; $\{\eta(\omega, t), P[\eta(0)=0]=1, \tau \in (-\infty, \infty)\}$ – дійсний однорідний стохастично неперервний гільбертів процес з незалежними приростами. Тоді $\xi(\omega, t)$ є безмежно подільним лінійним випадковим процесом.

Такий лінійний процес називають лінійним у вузькому розумінні, або безмежно подільним лінійним процесом. Всі його скінченновимірні характеристичні функції безмежно подільні.

Зауважимо, що коли в (3.88) функція $\eta(\omega, t)$ є процесом з некорельованими приростами, процес $\xi(\omega, t)$ є лінійним у широкому розумінні. Він, як і процес, поданий в інтегральному вигляді (3.31) згідно з теоремою Карунена, вивчається в межах кореляційної теорії.

Особливості достатньо повного завдання лінійного безмежно подільного процесу полягають у тому, що він природно описується послідовністю характеристичних функцій різного порядку; можливо також задавати мішані кумулянти для значень процесу $\xi(\omega, t)$ n -го порядку в моменти часу $t_1; \dots; t_n$:

$$\chi_n[\xi(\omega, t_1), \dots, \xi(\omega, t_n)] = \chi_n[\eta(\omega, 1)] \int \prod_{k=1}^n \varphi(\tau, t_k) d\tau, \quad n=1, 2, \dots,$$

де $\chi_n[\eta(\omega, 1)]$ – n -й кумулянт випадкової величини $\eta(\omega, 1)$.

Властивості лінійних випадкових процесів визначаються властивостями функції, яку називають ядром інтегрального зображення процесу $\xi(\omega, t)$ і функцією $\eta(\omega, \tau)$ – так званим породжуючим процесом. Зокрема лінійний процес буде стаціонарним у вузькому розумінні, якщо

$$\xi(\omega, t) = \int_{-\infty}^{\infty} \varphi(t - \tau) d\eta(\omega, \tau), \quad (3.89)$$

де $d\eta(\omega, \tau)$ – однорідний процес, $\varphi(\tau, t) = \varphi(t - \tau)$.

Записавши (3.88) у вигляді

$$\xi(\omega, t) = \int_{-\infty}^{\infty} \varphi(\tau, t) v_{\eta}(\omega, \tau) d\tau, \quad (3.88a)$$

де $v_{\eta}(\omega, \tau)$ – білий шум, дістають інше тлумачення лінійного випадкового процесу, а саме як результат перетворення білого шуму лінійним фільтром (рис.3.17). Справді, згідно з (3.88а) можна вважати, що задано лінійний фільтр з часовою характеристикою $\varphi(\tau, t)$, який називається формуючим фільтром. Вхідним впливом є білий шум $v_{\eta}(\omega, \tau)$, який називається породжуючим процесом. У результаті $\xi(\omega, t)$ виявляється відгуком фільтра на такий вхідний вплив (рис. 3.17). Якщо породжуючий процес – гаусів білий шум, випадковий процес $\xi(\omega, t)$ теж гаусів; у разі, коли породжуючий процес не є гаусовим (наприклад – це пуассонів білий шум), вихідний процес $\xi(\omega, t)$ має деякий закон розподілу, який належить класу безмежно подільних законів.

Згідно з (3.88а) лінійний процес можна розглядати як результат перетворення білого шуму лінійним фільтром.

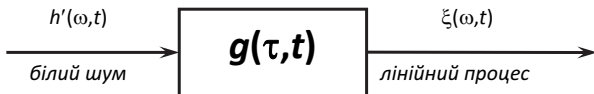


Рис. 3.17. – Формування лінійного процесу лінійним фільтром з білого шуму на його вході

Лінійні випадкові процеси з дискретним часом. Випадковий процес $\xi(\omega, k)$, $\omega \in \Omega$, $k \in Z$ з дискретним часом, як і процес з неперервним часом, називають лінійним, якщо його можна подати у вигляді

$$\xi(\omega, k) = \lim_{n \rightarrow \infty} \sum_{j=1}^n G_{n_j}(\omega, k), \quad (3.90)$$

де $\{G_{n_j}(\omega, k), j = \overline{1, n}, n = 1, 2, \dots\}$ – нескінченна послідовність незалежних для кожного фіксованого n процесів з дискретним часом. Як і у попередньому випадку вимагається, щоб доданки послідовностей сум задовольняли умову граничної нехтуваності. Клас усіх граничних законів розподілу у (3.90), як і у випадку неперервного часу, збігається з класом безмежно подільних розподілів.

Лінійний випадковий процес з дискретним часом (3.90) аналогічно (3.88а) можна подати у вигляді відгуку лінійного фільтра в разі впливу на його вхід білого шуму:

$$\xi(\omega, k) = \sum_{j=-\infty}^{\infty} \varphi(k-j)v(\omega, j) = \sum_{j=0}^{\infty} \varphi(j)v(\omega, k-j). \quad (3.91)$$

Тут $\varphi(j)$, $j=1,2,\dots$ ($j \in Z$) – вагова послідовність, яка характеризує лінійний фільтр; $\sum_{j=0}^{\infty} [\varphi(j)]^2 < \infty$, $v(\omega, j)$ – білий шум з дискретним часом; $Mv(\omega, j) = 0$.

Згідно з (3.91) лінійний випадковий процес з дискретним часом можна повністю визначити, якщо відома послідовність $j \in Z$ вагових коефіцієнтів і одновимірні закони розподілу білого шуму. Такий лінійний процес подається як зважена сума теперішнього та минулого значень білого шуму з дискретним часом.

Лінійний процес (3.91) можна також подати у вигляді зваженої суми його минулих значень і білого шуму:

$$\xi(\omega, k) = \sum_{j=1}^{\infty} \varphi(j) \xi(\omega, k-j) + v(\omega, k). \quad (3.92)$$

Тут $\xi(\omega, k)$ розглядають як авторегресію на минулі значення $\xi(\omega, k-1)$, $\xi(\omega, k-2)$, ...

Звичайно лінійні випадкові процеси розглядають у межах кореляційної теорії. Кореляційна функція такого стаціонарного процесу з дискретним часом має вигляд:

$$R(l) = M[\xi(\omega, k) \xi(\omega, k+l)] = \sigma_v^2 \sum_{j=1}^{\infty} \varphi(j) \varphi(j+l), \quad (3.93)$$

де σ_v^2 – дисперсія значень білого шуму $v(\omega, k)$, $k \in Z$.

Дисперсія процесу (3.92) визначається виразом

$$\sigma_{\xi}^2 = R(0) = \sigma_v^2 \sum_{j=0}^{\infty} [\varphi(j)]^2;$$

а його спектральна густина – виразом

$$G(f) = 2\sigma_v^2 \left/ \sum_{j=1}^{\infty} \varphi(j) e^{-i2\pi f j} \right|^2, \quad f \in [0, 1/2).$$

Практично особливо важливим є скінченновимірні варіанти лінійних випадкових процесів з дискретним часом – процеси авторегресії, процеси ковзного середнього, а також мішані процеси авторегресії та ковзного середнього.

Процес авторегресії p -го порядку описується рівнянням

$$\xi(\omega, k) = \sum_{j=1}^p \varphi(j) \xi(\omega, k-j) + v(\omega, k). \quad (3.94)$$

Такий процес стаціонарний, якщо корені характеристичного рівняння

$$z^p - \varphi(1)z^{p-1} - \dots - \varphi(p) = 1$$

лежать усередині одиничного круга $|z|=1$.

Для кореляційної функції такого процесу справджується рівняння

$$R(k) = \varphi(1)R(k-1) + \varphi(2)R(k-2) + \dots + \varphi(p)R(k-p),$$

і його спектральна густина допускає параметричне зображення

$$G(f) = 2\sigma_v^2 / \left| 1 - \varphi(1)e^{i2\pi f} - \dots - \varphi(p)e^{i2\pi fp} \right|^2, \quad 0 \leq f < \frac{1}{2}.$$

Процес авторегресії (3.94) можна дістати на виході лінійного фільтра з частотною характеристикою

$$\varphi(e^{-i2\pi f}) = \frac{1}{1 - \varphi(1)e^{-i2\pi f} - \dots - \varphi(p)e^{-i2\pi fp}},$$

на вході якого діє білий шум з дискретним часом. Такий фільтр у загальному випадку має імпульсну характеристику нескінченної довжини. Марківська послідовність – частинний випадок процесу (3.94), якщо $p=1$.

Процес ковзного середнього дістають з (3.91) при ненульових значеннях перших вагових коефіцієнтів $\varphi(j)$, $j = \overline{1, q}$:

$$\xi(\omega, k) = v(\omega, k) + \varphi(1)v(\omega, k-1) + \dots + \varphi(q)v(\omega, k-q). \quad (3.95)$$

Його також можна тлумачити як відгук лінійного фільтра в разі впливу на вхід білого шуму з дискретним часом.

Процес ковзного середнього завжди стаціонарний. Аналогічно випадку процесу авторегресії знаходиться його кореляційна функція та спектральна густина.

Між процесами авторегресії та ковзного середнього існує тісний зв'язок, породжений загальною вихідною моделлю лінійного процесу. Так процеси авторегресії можна описати як процеси ковзного середнього нескінченного порядку, тобто лінійні процеси загального вигляду.

Звичайно класичні лінійні процеси із дискретним часом розглядають у межах кореляційної теорії. Коли необхідно використовувати їх розподіли, звичайно розглядають гаусові процеси. Тобто останнім часом виникла потреба використовувати також негаусові процеси. Сучасна теорія лінійних процесів з дискретним часом дає таку можливість. Зауважимо, що на відміну від процесів з неперервним часом лінійні процеси х дискретним часом у загальному випадку не обов'язково мають бути безмежно подільними. Тільки при виконанні умов, обумовлених означеннями (3.90), процес стає лінійним безмежно подільним процесом з дискретним часом. Означення (3.90) еквівалентне вимозі, щоб процес $\xi(\omega, k)$, $k \in Z$ при кожному k був послідовністю безмежно подільних випадкових величин.

Характеристична функція лінійного безмежно подільного процесу з дискретним часом має вигляд

$$f_{\xi}(n) = \prod_{j=-\infty}^{\infty} f_{\nu}[u\varphi(j)]. \quad (3.96)$$

Канонічна форма для логарифма х.ф. (3.96) у формі Колмогорова записується так:

$$\ln f_{\xi}(u) = imu \sum_{j=-\infty}^{\infty} \varphi(j) + \int_{-\infty}^{\infty} \left[\frac{e^{iux\varphi(j)} - 1 - iux\varphi(j)}{x^2} \right] dK(x), \quad (3.97)$$

де

$$m = -i \frac{d}{du} \ln f_{\nu}(u) \Big|_{u=0};$$

$$K(x) = \frac{1}{2\pi} \lim_{y \rightarrow \infty} \int_{-\infty}^{\infty} \frac{e^{-iuy} - e^{-iux}}{iu} \frac{d^2 \ln f_{\nu}(u)}{du^2} du.$$

У частковому випадку процесу ковзного середнього (3.95) звідси одержуємо логарифм характеристичної функції безмежно подільного процесу ковзного середнього

$$\ln f_{\xi}(u) = imu \sum_{j=1}^q \varphi(j) + \sum_{j=1}^q \int_{j=-\infty}^{\infty} \left[\frac{e^{iux\varphi(j)} - 1 - iux\varphi(j)}{x^2} \right] \frac{1}{x^2} dK(x). \quad (3.97a)$$

Аналогічно можна визначити безмежно поділені процеси авторегресії та мішані процеси авторегресії та ковзного середнього.

Висновки

1. Випадковий процес як математичний об'єкт означається як параметрична сім'я випадкових величин, які задовольняють деякі випадкові умови. Існує принаймні чотири означення випадкового про-

цесу, кожне з яких визначає свій напрямок визначення процесів і має свої межі застосування.

Серед основних напрямків вивчення випадкових процесів найважливішими з практичної точки зору є кореляційна теорія та теорія ергодичних процесів; з теоретичної точки зору свої особливості мають вимірні процеси; найпростіше описування процесів сім'єю скінченновимірних розподілів мають сепарабельні випадкові процеси.

2. Інтегральні зображення випадкових процесів зі скінченними енергетичними характеристиками у відповідній формі узагальнюють інтегральні зображення детермінованих функцій. Найважливішими серед них є гармонічне зображення, яке у відповідній формі поширює ідеї перетворення Фур'є на випадкові процеси. Характерною особливістю гармонічних зображень випадкових процесів є їх зв'язок з відповідними гармонічними зображеннями кореляційних функцій. Найважливіший класичний результат у цій галузі є теорема Вінера-Хінчина щодо гармонічного зображення кореляційної функції стаціонарного у широкому розумінні випадкового процесу.

3. Зображення випадкових процесів зі скінченними енергетичними характеристиками – це більш прості зображення порівняно з інтегральними як у теоретичному, так і у практичному розумінні.

До класичних відносять зображення стаціонарних періодичних процесів та зображення нестаціонарних процесів у вигляді розкладів Райса та Карунена-Лоева. Більш сучасними є ортогональні розклади процесів зі скінченною енергією та зображення періодично корельованих (та споріднених з ними) процесів зі скінченного середньою потужністю.

Ортогональні розклади нестаціонарного випадкового процесу пов'язані з відповідними розкладами його кореляційної функції, з одного боку, а також з зображенням рядами Фур'є майже всіх реалізацій цього процесу – з другого. Вони характеризуються такими енергетичними властивостями: енергія процесу адитивно розподіляється поміж складовими розкладу, що визначає енергетичний спектр процесу у обраному ортонормованому базисі.

Розклад періодично корельованого (нестаціонарного) процесу встановлює еквівалентність цього нестаціонарного процесу і деякого векторного стаціонарного процесу – це розклад на стаціонарні компоненти.

4. Найважливішими класами випадкових процесів, які є одночас як класичними, так і сучасними, є білі шуми, марківські процеси, лінійні процеси.

Сучасна теорія білих шумів розглядає білі шуми з неперервним і дискретним часом, гаусові і негаусові.

З практичної точки зору білі шуми з неперервним часом розглядають як узагальнену похідну процесів з незалежними або некорельованими приростами. Відповідно, вибираючи необхідний процес з незалежними приростами, одержують гаусів (вінерів) білий шум, пуассонів білий шум тощо. Вони можуть бути повністю описані характеристичними функціоналами, куди входять параметри законів розподілу відповідних процесів з незалежними приростами. У межах кореляційної теорії всі ці білі шуми описуються однаково — їх кореляційна функція визначається δ -функцією, а спектральна густина стала на всіх частотах.

Білий шум з дискретним часом — це послідовність незалежних або некорельованих випадкових величин. Характеристичною особливістю таких білих шумів є можливість їх повного завдання, коли відомі лише одновимірні функції розподілу. У межах кореляційної теорії усі ці білі шуми мають однакові характеристики — їх кореляційна функція визначається δ -функцією Кронекера, а спектральна густина стала і зосереджена на скінченному інтервалі — $-\pi \leq \lambda < \pi$.

Марківські процеси повністю визначаються, якщо відомі одновимірні та умовні функції розподілу — ймовірність переходу. Щільність імовірностей переходу марківських процесів задовольняє інтегральне рівняння Колмогорова-Чепмена, а для дифузійних процесів — диференціальне рівняння, яке називають прямим рівнянням Колмогорова, або рівнянням Фоккера-Планка. Рівняння особливо спрощується для гаусового марковського процесу.

Використовують чотири типи марківських процесів — процесів з неперервним часом, які приймають континуальні або дискретні значення та процеси з дискретним часом, які також набувають континуальні або дискретні значення. Найпростіший — останній випадок, що має назву ланцюгів Маркова.

Класичні лінійні випадкові процеси з дискретним часом — процеси авторегресії, ковзного середнього, мішані процеси авторегресії та ковзного середнього мають широке узагальнення — лінійні безмежно подільні процеси з дискретним, а також неперервним часом. Всі вони розглядаються як результат перетворення білого шуму (гаусова або негаусова) відповідним лінійним фільтром. Особливість лінійних безмежно подільних процесів полягає у тому, що їх природно описувати послідовністю характеристичних функцій; зручно також використовувати кумулянти різних порядків.

Запитання, задачі й вправи

1. Дайте основні визначення випадкового процесу.
2. Що таке стаціонарні у вузькому та широкому розумінні випадкові процеси?
3. Назвіть основні властивості кореляційної функції.
4. Що таке ергодичні випадкові процеси? Дайте умови ергодичності процесу.
5. Що таке гаусів випадковий процес? Наведіть його основні властивості.
6. Сформулюйте теорему Карунена. Чим відрізняється спектральне зображення процесу від часового згідно з теоремою Карунена?
7. Дайте означення гармонізованості (за Лоевим) випадкового процесу та його кореляційної функції. Сформулюйте теорему Лоева щодо гармонізованості процесу.
8. Сформулюйте теореми Крамера та Вінера-Хінчина. Чим відрізняються гармонічні зображення стаціонарних та нестаціонарних випадкових процесів?
9. Сформулюйте теорему Герглотца. У чому різниця гармонічних зображень кореляційних функцій процесів з неперервним та дискретним часом (тобто – послідовностей).
10. Чим характеризується розклад стаціонарного періодичного процесу за гармоніками?
11. Охарактеризуйте ряд Фур'є процесу зі скінченною потужністю у гільбертовому просторі випадкових величин.
12. Що таке ортогональний розклад випадкового процесу зі скінченною енергією?
13. Як пов'язаний ортогональний розклад випадкового процесу та розклади його кореляційної функції та реалізацій?
14. Що таке енергетичний спектр випадкового процесу зі скінченною енергією у довільному ортонормованому базисі?
15. Охарактеризуйте класичні ортогональні розклади випадкового процесу – розклади Карунена-Лоева та Райса.
16. Наведіть розклад періодично корельованого випадкового процесу зі скінченною середньою потужністю. Як пов'язаний цей нестаціонарний випадковий процес з відповідними стаціонарними?
17. Охарактеризуйте білий шум з неперервним часом. Як одержати гаусів (вінерів) та пуассонів білий шум?
18. Дайте повне описування гаусова (вінерова) та пуассонова білого шуму та їх описування в межах кореляційної теорії.

19. Що таке марківська властивість, характерна для марківських процесів? Запишіть відповідні співвідношення у загальному випадку марківського процесу та ланцюгу Маркова.

20. Запишіть рівняння Фоккера-Планка для дифузійних марковських процесів та ланцюгу Маркова.

21. Що таке лінійні безмежно подільні процеси та послідовності?

22. Наведіть інтегральне зображення лінійного процесу та зображення лінійної послідовності у вигляді відповідної суми.

23. Наведіть зображення лінійного процесу та послідовності у вигляді результату перетворення лінійним фільтром відповідного білого шуму.

24. Охарактеризуйте процеси авторегресії, ковзного середнього та суміш процесів авторегресії та ковзного середнього.

25. Компоненти випадкового вектора $[\xi(\omega), \eta(\omega)]$ набувають можливих значень (x_i, y_j) з імовірностями

$$P_{ij} = P\{\omega: \xi(\omega) = x_i, \eta(\omega) = y_j\}, \quad i = \overline{1, m}, \quad j = \overline{1, n}.$$

Яку умову мають задовольняти імовірності P_{ij} ?

26. За умов попередньої задачі знайти закони розподілу для кожного з компонентів $\xi(\omega)$ і $\eta(\omega)$ випадкового вектора $[\xi(\omega), \eta(\omega)]$.

27. Дано випадковий вектор $\vec{\xi}(\omega) = [\xi_1(\omega), \xi_2(\omega)]$ із гаусовим розподілом. Знайти аналітичний вираз для ліній з однаковим значенням щільності ймовірності.

28. Характеристична функція випадкової величини має вигляд:

$$f(u) = \int_{-\infty}^{\infty} \exp(iux) p(x) dx = (1 - iu/\beta)^{-\alpha}, 0$$

де $p(x)$ – щільність її розподілу. Чи є вона безмежно подільною? Чому?

29. Зобразити графічно реалізації випадкових процесів

а) $\xi(\omega, t) = a(\omega) \cos(\lambda t)$;

б) $\xi(\omega, t) = a(\omega) \cos(\lambda t - \varphi(\omega))$;

в) $\xi(\omega, t) = a(\omega) \exp(\beta t)$;

г) $\xi(\omega, t) = \exp(\beta t) \cos(\lambda t - \varphi(\omega))$,

де $a(\omega)$, $\varphi(\omega)$ – випадкові величини; β, λ – невідповідні величини. Проілюструвати на прикладах а) і б) чотири означення випадкових

процесів, вважаючи, що випадкова величина $a(\omega)$ має рівномірний або експоненціальний закон розподілу.

30. Знайти одновимірний закон розподілу випадкового процесу виду $\xi(\omega, t) = a(\omega)t + b$, де $a(\omega)$ — гаусова випадкова величина з математичним сподіванням m і дисперсією σ^2 ; b — не випадкова величина.

Проілюструвати різні означення випадкового процесу.

31. Які з поданих далі функцій параметра τ не можуть бути кореляційними функціями стаціонарних випадкових процесів? Пояснити чому.

а) $\exp(-\tau^2)$;

б) $|\tau| \exp(-|\tau|)$;

в) $10 \exp[-(\tau+2)]$;

г) $\sin^2 \pi\tau / (\pi\tau)^2$;

д) $(\tau^2 + 4) / (\tau^2 + 8)$.

32. Кореляційна функція стаціонарного випадкового процесу $\xi(\omega, t)$ має вигляд $R(\tau) = \exp(-|\tau|) \cos^2 \tau$. Виписати кореляційну матрицю для чотирьох відліків випадкового процесу в моменти часу, відокремлені інтервалом $\Delta t = 0,5$.

33. Кореляційна матриця відліків стаціонарного випадкового процесу $\xi(\omega, t)$ має вигляд:

$$R = \begin{bmatrix} 1,0 & 0,6 & 0,4 & - \\ - & 1,0 & 0,6 & - \\ 0,4 & 0,6 & - & 0,6 \\ 0,2 & - & - & 1,0 \end{bmatrix}.$$

Заповнити пропуски в матриці.

34. Нехай задано випадковий процес

$$\xi(\omega, t) = a \cos[\lambda(\omega)t + \varphi(\omega)],$$

де a — випадкова величина; $\lambda(\omega)$, $\varphi(\omega)$ — незалежні випадкові величини, причому $\varphi(\omega)$ рівномірно розподілена в інтервалі $[0, 2\pi]$. Довести, що випадковий процес є стаціонарним у широкому розумінні.

35. Чи є диференційовними стаціонарні випадкові процеси, кореляційні функції яких визначаються виразами

а) $R_1(\tau) = \exp\{-\alpha|\tau|\}$;

$$\text{б) } R_2(\tau) = \exp\{-\alpha\tau^2\};$$

$$\text{в) } R_3(\tau) = \cos(\alpha\tau),$$

де параметр $\alpha > 0$.

36. Показати, що узагальнений телеграфний сигнал $\xi(\omega, t)$ з $M\xi(\omega, t) = 0$ і $R(t_1, t_2) = \exp\{-2\lambda|t_1 - t_2|\}$, де $\lambda > 0$, недиференційовний.

37. Кореляційна функція стаціонарного випадкового процесу описується виразом $R(\tau) = C \exp\{-\alpha|\tau|\}$, де $C > 0$, $\alpha > 0$. Записати вираз для спектральної щільності процесу.

38. Спектральна щільність стаціонарного випадкового процесу описується виразом

$$G(\lambda) = \begin{cases} G_0, & \text{якщо } |\lambda| \leq \Lambda; \\ 0, & \text{якщо } |\lambda| > \Lambda, \end{cases}$$

де $\Lambda > 0$, $G_0 > 0$. Записати вираз для кореляційної функції.

39. Випадкову функцію $\xi(\omega, t)$ задано у вигляді найпростішого елемента простору

$$L_2(\Omega, L_2(T)): \xi(\omega, t) = \zeta(\omega)y(t), \quad M[\zeta(\omega)] = 0; \quad M[\zeta^2(\omega)] = \sigma^2, \quad y(t) \in L_2(T).$$

Визначити математичне сподівання $M[\xi(\omega, t)]$, кореляційну функцію $R_\xi(t, t')$ процесу $\xi(\omega, t)$.

40. Дійсне випадкове поле $\xi(\omega, \vec{s})$ задано розкладом за найпростішими елементами простору $L_2(\Omega, L_2(D))$ з некорельованими коефіцієнтами

Записати вираз для кореляційної функції $R_\xi(\vec{\rho}_1, \vec{\rho}_2)$.

$$\xi(\omega, \vec{\rho}) = \sum_{j=1}^{\infty} d_j(\omega) \varphi_j(\vec{\rho}), \quad M[d_j(\omega)d_k(\omega)] = \sigma_k^2 \delta_{j,k}, \quad \delta_{j,k} = \begin{cases} 1, & \text{якщо } j=k; \\ 0, & \text{якщо } j \neq k; \end{cases}$$

$$M \xi(\omega, \vec{\rho}) = 0.$$

41. Визначити кореляційну функцію для випадкового процесу $\eta(\omega, t)$, здобутого в результаті інтегрування $\eta(\omega, t) = \int_0^t \xi(\omega, \tau) d\tau$ випадкового процесу $\xi(\omega, t)$, заданого зображенням

$$\xi(\omega, t) = \sum_{j=1}^n d_j(\omega) e^{-\alpha_j t} \quad (t > 0),$$

де $d_j(\omega)$ – центровані некорельовані випадкові величини
 $M[d_j^2(\omega)] = \sigma_j^2$, $d_j > 0$, $j = \overline{1, n}$.

42. Нехай стаціонарний гаусів шум $\xi(\omega, t)$ має рівномірну спектральну щільність у смузі шириною Λ

$$G(\lambda) = \begin{cases} N_0/2\pi, & \text{якщо } -\Lambda \leq \lambda < \Lambda; \\ 0, & \text{якщо } \lambda < -\Lambda \text{ або } \lambda > \Lambda. \end{cases}$$

Довести, що значення шуму $\xi(\omega, t)$ у моменти часу $\Delta t_k = k\pi/\Lambda$, $k = 1, 2, \dots$, статистично незалежні.

4. ОСНОВИ ТЕОРІЇ ІНФОРМАЦІЇ

Одна з сучасних гілок теорії ймовірності є теорія інформації. Вона виникла у сорокові роки 20-го сторіччя з потреб техніки зв'язку і перетворилася у математичну теорію у наступний період. З нею пов'язана теорія кодування. У цих теоріях особливо важлива фізична мотивація – інакше теорію кодування можна розглядати як теорію перетворення символів, а теорію інформації – як виявлення потенційних можливостей таких перетворень. Тому у цьому розділі поряд з математичними поняттями і термінами особливо часто ми будемо використовувати фізичні, такі як джерело повідомлень, канал зв'язку, одержувач повідомлень, передавання інформації від джерела до одержувача тощо.

Ми далі кілька разів будемо розглядати з фізичної точки зору однакове коло питань (для різних каналів зв'язку) – впровадимо поняття ентропії, кількості інформації, дамо основні теореми кодування для відповідних каналів. При цьому ми почнемо з найпростішого з математичної точки зору випадку. Хоч далі буде значно ускладнюватись математичний апарат, з фізичної точки зору результати лише будуть відображати особливості відповідних каналів.

4.1. Передавання інформації по дискретному каналу без завад

З методичних міркувань ми розглядаємо канал зв'язку без завад, хоча результати, перш за все, стосуються тільки джерела повідомлень. У цьому підрозділі вивчаються можливості кодування повідомлень джерела з метою узгодження джерела повідомлень з каналом зв'язку.

4.1.1. Математична модель джерела повідомлень

З фізичної точки зору під джерелом повідомлень розуміють джерело реальних повідомлень – телеграфного тексту, мовного сигналу, цифрових даних тощо. У теорії інформації використовується математична модель такого джерела. У даному разі розглядаємо джерело дискретних повідомлень.

Задамо спочатку простий ансамбль дискретних повідомлень (X, P) . Тут $X = \{x_1, x_2, \dots, x_M\}$ – множина елементарних подій, які можна уявити собі як букви алфавіту, з яких складається повідомлення, а $P = (P_1, P_2, \dots, P_M)$ – їх ймовірності. Тобто ансамбль дискретних повідомлень (стисло-дискретний ансамбль) – це дискретна множина X разом із заданою ймовірнісною мірою. У найпростіших випадках мож-

на обмежитись розгляданням простого ансамблю, однак у загальному випадку слід разом розглядати два, три і більше простих ансамблів (наприклад, ансамбль переданих та прийраних повідомлень).

Введемо дві множини $X = \{x_1, \dots, x_M\}$, $Y = \{y_1, \dots, y_N\}$ та їх прямиий добуток XY . Задамо ансамбль дискретних повідомлень (XY, P_{XY}) , де P_{XY} – імовірнісна міра, яка надає кожному елементові $(x_i, y_j) \in XY$ імовірність $P(x_i, y_j)$. При цьому визначаються

$$P(x_i) = \sum_{y_j \in Y} P(x_i, y_j),$$

та
$$P(y_j) = \sum_{x_i \in X} P(x_i, y_j),$$

тобто імовірнісні міри P_X та P_Y на множинах X та Y .

Кажуть, що у даному разі сумісно задані ансамблі (X, P_X) та (Y, P_Y) .

У загальному випадку можна ввести прямиий добуток n множин $X_1 X_2 \dots X_n$, задати складний ансамбль $(X_1 X_2 \dots X_n, P_{X_1 X_2 \dots X_n})$, який дає можливість визначити n сумісно заданих ансамблів, (X_i, P_{X_i}) , $i=1, n$, де

$$P_{X_i} = \sum_{X_1} \sum_{X_2} \dots \sum_{X_{i-1}} \sum_{X_{i+1}} \dots \sum_{X_n} P(x_1, \dots, x_n).$$

З фізичної точки зору складний ансамбль $(X_1 X_2 \dots X_n, P_{X_1 X_2 \dots X_n})$ описує поведінку довільного повідомлення на виході джерела тривалістю n літер-елементів алфавіту. У загальному випадку задати математичну модель джерела повідомлень – це значить охарактеризувати випадковий процес на його виході.

Нехай U_X – дискретне джерело, яке вибирає літери з алфавіту X . Кажуть, що джерело повідомлень задано, якщо для довільних $n=1, 2, \dots$ та будь-яких $i=0, \pm 1, \pm 2, \dots$ задана сім'я розподілів імовірностей

$$F(x_1^{(i+1)}, x_2^{(i+2)}, \dots, x_n^{(i+n)}), x_j^{(i+j)} \in X,$$

яка задовольняє умови узгодженості А.М. Колмогорова.

Таке визначення джерела дає можливість природно ввести не-стаціонарні, стаціонарні, ергодичні, марківські та інші джерела повідомлень.

4.1.2. Ентропія

Введемо для дискретного ансамблю (X, P_X) деяку формальну характеристику – ентропію, яка визначається формулою

$$H(X) = -\sum_{i=1}^M P_i \log P_i. \quad (4.1)$$

Розглянемо властивості ентропії.

1. Ентропія дискретного ансамблю невід’ємна $H(X) \geq 0$. Рівність нулю досягається для вірогідного стану, коли $P_j = 1$; при цьому для ймовірностей решти станів $P_i = 0$, $i \neq j$.

Дійсно, при $0 < P_i < 1$ $P_i \log P_i < 0$ і тому

$$H(X) = -\sum_{i=1}^n P_i \log P_i.$$

При наявності вірогідного стану x_j маємо $P_j \log P_j = 0$; $P_i \log P_i = 0$ при $i \neq j$, тому що згідно з правилом Лопітала

$$\lim_{P \rightarrow 0} P \log P = \lim_{P \rightarrow 0} \frac{1/P}{-1/P^2} \log e = 0.$$

2. Ентропія дискретного ансамблю сягає максимуму, коли всі стани рівноймовірні.

Розглянемо спочатку двійковий ансамбль, коли $M = 2$. При цьому

$$H(X) = -P \log P - Q \log Q, \quad (4.2)$$

$P + Q = 1$. Визначаючи Q у вигляді $Q = 1 - P$ та знаходячи максимум (4.2) з умови

$$\frac{dH(X)}{dP} = \frac{d}{dP} [-P \log P - (1-P) \log(1-P)] = 0, \text{ одержуємо } P = 1/2.$$

На рис. 4.1 показана залежність ентропії двійкового ансамблю від ймовірності. Із першої властивості виходить, що $H(X) = 0$ досягається при $P = 0$ та $P = 1$. $H_{\max}(X)$ досягається при $P = 1/2$.

Тепер доведемо, що ентропія ансамблю зі скінченним числом станів сягає максимуму, коли всі символи алфавіту рівноймовірні.

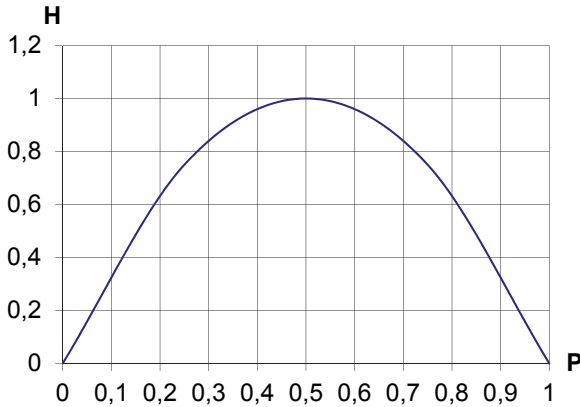


Рис. 4.1. – Ентропія двійкового ансамблю

Знайдемо максимум

$$H(X) = -\sum_{i=1}^M P_i \log P_i \quad (4.3)$$

за умовою $\sum_{i=1}^M P_i = 1$.

Користуючись методом невизначених множників Лагранжа, знайдемо екстремум функції

$$f(P_1, P_2, \dots, P_M) = -\sum_{i=1}^M P_i \log P_i - \lambda \sum_{i=1}^M P_i \quad (4.4)$$

Диференціювання (4.4) дає можливість одержати систему рівнянь

$$\log P_i + \log e + \lambda = 0, \quad (i = \overline{1, M}),$$

тобто

$$\log P_i = -\lambda - \log e = 0, \quad (i = \overline{1, M}).$$

Звідси випливає, що максимум досягається при $P_1 = P_2 = \dots = P_M$. Тому, враховуючи умову в (4.3), остаточно одержуємо $P_i = 1/M$. Тому максимальне значення ентропії

$$H_{\max}(X) = \log M \quad (4.5)$$

дорівнює логарифму числа станів ансамблю (числу різних символів алфавіту) і досягається, коли вони рівноймовірні. Величина $H_{\max}(X)$ збільшується зі зростанням числа станів.

Таким чином, ентропія ансамблю повідомлень невід’ємна, дорівнює нулеві, коли його стан наперед визначений, та збільшується зі зростанням числа станів. Ці властивості дають можливість розглядати ентропію як міру невизначеності ансамблю.

Введена характеристика ансамблю відноситься до одного перетину випадкового процесу, який формується на виході джерела повідомлень. Якщо цей процес є стаціонарний цифровий білий шум з дискретним часом, він є моделлю стаціонарного джерела дискретних повідомлень без пам’яті (див. рис. 4.2).

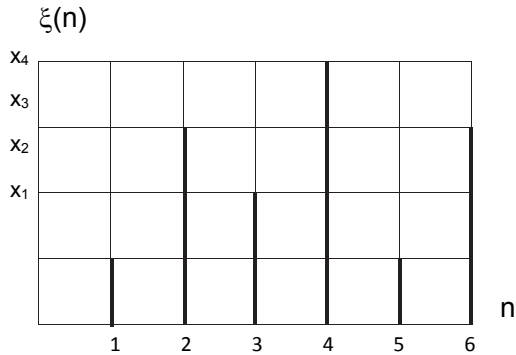


Рис. 4.2. а)

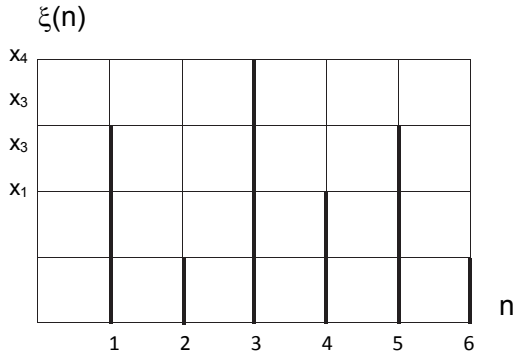


Рис. 4.2. б)

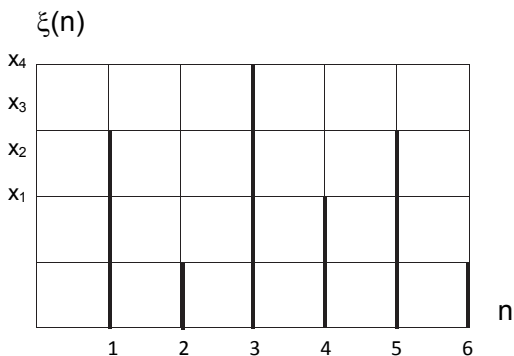


Рис. 4.2. в)

Рис. 4.2. – Ілюстрація реалізацій а)-в) цифрового процесу (цифрового білого шуму з дискретним часом)

Тоді ентропія ансамблю є також ентропією джерела. У загальному випадку це не так. Необхідно окремо розглядати ентропію джерела або, що те саме, – ентропію випадкового процесу на його виході.

Введемо спочатку ентропію складних ансамблів. Нехай $(X_1 X_2, P_{X_1 X_2})$ – ансамбль, що визначає два сумісно заданих ансамблі (X_1, P_{X_1}) та (X_2, P_{X_2}) , які можна ототожнювати з двома перетинами випадкового процесу на виході джерела. Умовний розподіл імовірностей для i -ї події:

$$P(x_i | x_j) = \frac{P(x_i, x_j)}{P(x_j)}.$$

Тому аналогічно (4.1) можна визначити умовну ентропію ансамблю X_i відносно ансамблю X_j :

$$H(X_i | X_j) = - \sum_{x_i} \sum_{x_j} P(x_i, x_j) \log P(x_i | x_j). \quad (4.6)$$

Можна показати, що $H(X_i) \geq H(X_i | X_j)$, причому знак рівності досягається тоді і тільки тоді, коли ансамблі X_i та X_j статистично незалежні. Інакше кажучи, рівність характерна для стаціонарного джерела без пам'яті (тобто для процесу у вигляді стаціонарного білого шуму).

У загальному випадку, коли розглядається ансамбль $(X_1 \dots X_{i-s}, P_{X_1 \dots X_{i-s}})$, який утворений s перетинами довільного випадкового процесу на виході джерела, маємо

$$H(X_i | X_{i-1} \dots X_{i-s}) = - \sum_{X_i} \dots \sum_{X_{i-s}} P(x_i, x_{i-1}, \dots, x_{i-s}) \log P(x_i | x_{i-1}, \dots, x_{i-s}). \quad (4.7)$$

Можна довести, що

$$H(X_i) \geq H(X_i | X_{i-1}) \geq \dots \geq H(X_i | X_i^{s-1}) \geq H(X_i | X_i^s),$$

де $X_i^s = X_{i-1} X_{i-2} \dots X_{i-s}$ – прямий добуток s множин X .

Вираз (4.7) можна розглядати як ентропію багатозв'язного марківського процесу, в якого майбутні значення залежать від s минулих значень. У частинному випадку $s=1$ звідси можна одержати ентропію звичайного (однозв'язного) марківського процесу, в якому майбутні значення залежать від минулих тільки через сучасне:

$$H(X_i | X_i^s) = - \sum_{X_i} \dots \sum_{X_{i-s}} P(x_i, x_{i-1}, \dots, x_{i-s}) \log P(x_i | x_{i-1}) = H(X_i | X_{i-1}).$$

Для довільного стаціонарного джерела, враховуючи рівність

$$P(x_i, x_{i-1}, \dots, x_{i-s}) = P(x_j, x_{j-1}, \dots, x_{j-s}), \quad s < i,$$

маємо

$$H(X_i | X_i^s) = H(X_j | X_j^s).$$

При цьому, як можна довести, умовні ентропії, коли відомі усі попередні події, утворюють монотонно незростаючу послідовність

$$H(X_i) \geq H(X_2 | X_1) \geq \dots \geq H(X_s | X_s^{s-1}).$$

Більш того, існує границя

$$\lim_{s \rightarrow \infty} H(X_s | X_s^{s-1}) = H(X | X^\infty),$$

де $H(X | X^\infty)$ – деяке число, яке характеризує стаціонарне джерело і задовольняє нерівність $0 \leq H(X | X^\infty) \leq \log M$.

4.1.3. Кількість інформації

З фізичної точки зору процес передавання інформації розглядається як процес її одержання споживачем від джерела. Спочатку обговоренням раніше стаціонарним джерелом без пам'яті, яке повністю визначається ансамблем повідомлень. До приймання повідомлення невизначеність ансамблю дорівнює $H(X)$. Після

приймання вона зменшиться. Кількість інформації визначають через величину зменшення цієї невизначеності.

Кількість інформації – це міра знятої невизначеності, спричиненої прийманням повідомлення.

Для каналу без завад стан ансамблю (X, P_X) після приймання повідомлення визначиться повністю, тобто

$$I(X) = H(X) - 0 = -\sum_{i=1}^M P_i \log P_i. \quad (4.8)$$

Кількість інформації, яку одержує споживач при вичерпному з'ясуванні стану дискретного ансамблю повідомлень, дорівнює ентропії ансамблю.

Розглянемо (4.8) з інших позицій. Відповідно з цим виразом на ймовірнісному просторі (X, P_X) задана випадкова величина $J(X)$, яка набуває значення

$$J(x_1) = -\log P_1, \dots, J(x_M) = -\log P_M$$

з імовірностями P_1, \dots, P_M . Математичне сподівання цієї випадкової величини дорівнює

$$I(X) = -\sum_{i=1}^M P_i \log P_i = -M \log J(X). \quad (4.8a)$$

Вираз (4.8a) – кількість інформації – має тлумачення середнього значення функції $J(X)$; тому впроваджують також власну інформацію у конкретному повідомленні $x_i \in X$:

$$J(x_i) = \log \frac{1}{P_i}, \quad i = \overline{1, M}. \quad (4.9)$$

Ця трактовка дає можливість надати інформаційний зміст ентропії ансамблю – ентропія являє собою середню кількість інформації у повідомленні дискретного ансамблю.

Дотепер при означенні ентропії та інформації не домовлялися відносно основи логарифму у відповідних формулах. Вона впливає на одиницю вимірювання кількості інформації. Залежно від вибору основи можна впровадити різні одиниці вимірювання.

Нехай ансамбль двійковий, а його стани рівномірні, тобто $M=2$, $P_1=P_2=1/2$. Тоді згідно з (4.5)

$$I(X) = \log_2 2 = 1, \quad (4.10)$$

якщо використовувати двійкові логарифми. У цьому разі одиниця вимірювання має назву «дв.од.» (двійкова одиниця) або «біт» – від англ. «binary digit.» Біт – це кількість інформації, яка одержується при спостереганні двійкового ансамблю з рівноймовірними станами.

Підкреслимо ще раз, що джерело з математичної точки зору тожне з випадковим процесом, який формується на його виході. Стационарне джерело без пам'яті еквівалентне стационарному цифровому білому шуму. Саме в цьому випадку розглянуті характеристики дискретного ансамблю збігаються з характеристиками перетину випадкового процесу у довільну мить. Тобто вище по суті знайдені інформаційні характеристики такого випадкового процесу. У загальному випадку довільного джерела повідомлень аналогічно впроваджуються відповідні характеристики, або, інакше кажучи, – інформаційні характеристики довільного випадкового процесу.

4.1.4. Кодування дискретних джерел

Наведені результати дають можливість довести теорему кодування. Залежно від прийнятого шляху викладання результатів, її називають теоремою Шеннона для дискретного каналу без завад або теоремою Шеннона про ефективне кодування. Підкреслюючи, що цей результат є першим серед найважливіших результатів Клода Шеннона, її також називають першою теоремою Шеннона.

Розглянемо проблему зображення символів алфавіту джерела за допомогою іншої системи символів з точки зору граничних можливостей кодування – забезпечення максимально можливої швидкості, інакше кажучи, це проблема мінімізації середнього числа символів на повідомлення. Спочатку торкнемось питання, саме за рахунок чого це фізично можна досягти.

Надмірність джерела. Візьмемо n суміжних перетинів випадкового процесу на виході стационарного джерела без пам'яті. Введемо складний ансамбль (X^n, P_{X^n}) . Завдяки статистичній незалежності перетинів процесу цей ансамбль визначає n спільно заданих незалежних ансамблів (X_i, P_{X_i}) , $i=\overline{1, n}$. Вони однакові. Тому середня кількість інформації, яка міститься у ньому, дорівнює

$$I(X^n) = H(X^n) = nH(X). \quad (4.11)$$

Впровадимо тепер інший ансамбль, який визначає n_{\min} спільно заданих незалежних ансамблів з рівноймовірними елементами

$(X^{n_{\min}}, P_{X^{n_{\min}}})$ за умовою, що середня кількість інформації, яка міститься у ньому, залишається тою самою:

$$I(X^{n_{\min}}) = I(X^n).$$

Для цього ансамблю маємо

$$I(X^{n_{\min}}) = H(X^{n_{\min}}) = n_{\min} H_{\max}(X).$$

Тому правильна рівність

$$nH(X) = n_{\min} H_{\max}(X). \quad (4.12)$$

Тут число елементів позначається n_{\min} , тому що згідно з умовою рівномірності елементів ентропія досягає максимального значення $H_{\max}(X)$ (див.4.5).

Величина, яка визначається за (4.12)

$$r = \frac{n - n_{\min}}{n} = 1 - \frac{H(X)}{H_{\max}(X)} \quad (4.13)$$

називається надмірністю джерела. Надмірність – це відносна кількість елементів повідомлення, від яких можна звільнитися без зменшення кількості середньої інформації у ньому. Згідно з (4.5) формула (4.13) може бути також зображена у вигляді:

$$r = 1 - \frac{H(X)}{\log M}. \quad (4.13a)$$

У загальному випадку довільного джерела, символи якого на виході статистично залежні, надмірність визначається більш складним співвідношенням.

Теорема Шеннона про ефективне кодування

Теорема 4.1. (Перша теорема Шеннона). При заданому ансамблі X з M повідомлень і ентропією $H(X)$ та алфавіті, який складається з m символів, існує можливість так закодувати повідомлення ансамблю з допомогою послідовності символів, які належать заданому алфавіту, що середнє число символів на повідомлення (середня тривалість кодівих слів) задовільняє нерівність

$$H_m(X) \leq L < H_m(X) + 1. \quad (4.14)$$

Число L не може бути меншим від нижньої границі у виразі (4.14).

Тут $H_m(X)$ — ентропія ансамблю X , в якій логарифм обчислюється на основі m :

$$H_m(X) = \frac{H(X)}{\log_2 m}.$$

Перш, ніж доводити теорему, уточнимо термінологію, пов'язану з кодуванням джерела. Позначимо через $A = \{a_1, \dots, a_m\}$ деяку множину елементів, які називають кодовими символами. Послідовність кодових символів називають кодовим словом. Кодуванням повідомлень із ансамблю X за допомогою коду називають відображення множини повідомлень у множину кодових слів.

Розрізняють рівномірні і нерівномірні коди. Код називають рівномірним, якщо всі кодові слова однакової довжини. Якщо кодові слова мають різну довжину, кажуть про нерівномірні коди.

Результати дослідження рівномірного кодування показують, що швидкість створення інформації дискретним джерелом дорівнює його ентропії. Це відноситься як до стаціонарних джерел без пам'яті, де $H = H(X)$, так і до більш широкого класу ергодичних джерел, де

$$H = H(X|X^\infty).$$

При нерівномірному кодуванні враховуються ймовірності P_i виникнення символів x_i джерела. Основною характеристикою такого кодування є кількість символів, які приходяться на кожне повідомлення джерела. Міра якості кодування — середня кількість символів на повідомлення. Оптимальним називають код, середня довжина кодових слів якого мінімально можлива. Принципове питання тут — знаходження мінімального значення середньої тривалості кодового слова. Зупинимось на цьому більш детально.

При обговоренні питання знадобляться наступні факти, які приймемо без доведення.

1. Нерівність Крафта:

$$K = \sum_{i=1}^M \frac{1}{m^{l_i}} \leq 1,$$

яка встановлює необхідну і достатню умову існування префіксного коду у алфавіті об'єму m з довжинами кодових слів l_1, \dots, l_M . (Замітимо: префіксний код характеризується тим, що ніяке кодове слово не є початком більш тривалого кодового слова).

2. Так звана фундаментальна нерівність

$$\sum_{i=1}^M x_i \log\left(\frac{y_i}{x_i}\right) \leq 0,$$

де

$$\sum_{i=1}^M x_i = 1, \quad \sum_{i=1}^M y_i = 1.$$

Тепер можна перейти до безпосереднього доведення теореми 4.1, яку розглянемо у деякому частинному випадку.

Для знаходження мінімального значення середньої тривалості кодових слів введемо число

$$Q_i = m^{-l_i} / K.$$

Тоді $\sum_{i=1}^M Q_i = 1$ і можна застосувати фундаментальну нерівність:

$$\sum_{i=1}^M P_i \log\left(\frac{Q_i}{P_i}\right) \leq 0.$$

Звідси одержуємо

$$-\sum_{i=1}^M P_i \log P_i \leq \sum_{i=1}^M P_i \log\left(\frac{1}{Q_i}\right)$$

або

$$H(X) \leq \sum_{i=1}^M P_i \log\left(\frac{1}{Q_i}\right).$$

Тоді, враховуючи як введено число Q_i , приводимо остаточною нерівність до вигляду

$$H(X) \leq \sum_{i=1}^M P_i (\log K - \log m^{-l_i}) \leq \log K + \sum_{i=1}^M P_i l_i \log m.$$

Згідно з нерівністю Крафта $\log K \leq 0$, і тому відкидання цього члена тільки підсилить нерівність

$$H(X) \leq \sum_{i=1}^M (P_i l_i) \log m = L \log m,$$

де $L = \sum_{i=1}^M P_i l_i$ – середня тривалість кодових слів.

Нагадаємо, що під \log умовилися розуміти \log_2 , тому

$$\frac{H(X)}{\log_2 m} = H_m(X),$$

де $H_m(X)$ – ентропія ансамблю X , у якому логарифм взято на основі m .

Враховуючи це означення, остаточно одержуємо шуканий результат:

$$H_m(X) \leq L,$$

тобто ентропія є нижньою границею середньої тривалості L для однозначно декодуємої системи. Цей результат є частковим випадком сформульованої теореми 4.1 для кодування джерела повідомлень (або для каналу без завад).

Наведемо приклад оптимального коду.

Приклад 4.1. Метод Шеннона-Фано. Розглянемо ансамбль (X, P_X) . Пропонується така процедура кодування.

1-й крок. З ансамблю (X, P_X) створюємо новий ансамбль з двома рівноймовірними групами повідомлень. Для цього розбиваємо множину X на дві рівноймовірні групи. Приймаємо, що перший символ кодових слів є нулем для всіх повідомлень першої групи та одиницею для всіх повідомлень другої групи.

2-й крок. Побудований на першому кроці кодування ансамбль перетворюємо на новий ансамбль за таким правилом. Кожну з груп, яка одержана на першому кроці, розбиваємо на рівноймовірні підгрупи. При цьому другим символом кодових слів вибираємо нуль для всіх повідомлень «верхніх» підгруп другого кроку та одиницю для всіх повідомлень відповідних «нижніх» підгруп.

Процес продовжуємо, поки у підгрупах, що формуються, не залишиться по одному повідомленню. Якщо на попередніх кроках у деякій підгрупі залишається одне повідомлення, для нього процес кодування завершується.

Числовий приклад побудови кода Шеннона-Фано наведено у табл.4.1.

Таблиця 4.1

Повідомлення	Імовірність	Процедура кодування	Кодове слово
x_1	1/2		0
-----		1-й крок	
x_2	1/4		10
.....		2-й крок	
x_3	1/8		110
~~~~~		3-й крок	
$x_4$	1/16		1110
~~~~~		4-й крок	
x_5	1/32		11110
.....		5-й крок	
x_6	1/32		11111

Якщо до цього коду підходити з позиції кодування джерела, можна обчислити середню тривалість кодового слова:

$$L = \sum_{i=1}^6 P_i l_i = 1 + \frac{20}{32}.$$

Ентропія джерела:

$$H(X) = -\sum_{i=1}^6 P_i \log P_i = 1 + \frac{30}{32},$$

тобто середня тривалість L збігається з ентропією ансамблю; згідно з теоремою Шеннона вона мінімально можлива – виконується рівність у нерівності (4.14).

Розглянемо цей приклад з позицій передавання інформації у дискретному каналі без завад.

Обчислимо індивідуальну інформацію, яка міститься у деякому повідомленні x_i :

$$J(x_i) = -\log P_i.$$

Імовірність P_i можна знайти так:

$$P_i = [1/2]^{n_i},$$

де n_i – номер кроку кодування, або, інакше кажучи, кількість символів у кодовому слові, яке кодує повідомлення x :

Тому для $J(x_i)$ одержуємо рівність:

$$J(x_i) = -\log[1/2]^{n_i} = n_i.$$

Таким чином, індивідуальна інформація $J(x_i)$ у повідомленні x_i чисельно збігається з кількістю двійкових знаків у кодовому слові. Тому кожен двійковий символ кодового слова містить рівно 1 біт інформації, тобто швидкість передавання інформації стає максимально можливою і дорівнює пропускній здатності каналу (див. далі означення – формулу 4.34).

Розглянемо код Шеннона-Фано у загальному випадку, коли при кодуванні використовується m -знаковий код.

Для кожного значення ймовірності P_i повідомлення x_i з ансамблю (X, P_X) знайдеться ціле число l_i таке, що

$$\log_m(1/P_i) \leq l_i < \log_m(1/P_i) + 1.$$

Тому
$$\frac{1}{P_i} \leq m^{l_i} \leq \frac{m}{P_i},$$

або
$$P_i \geq \frac{1}{m^{l_i}} > \frac{P_i}{m}.$$

Підсумовуючи по i від 1 до M , одержуємо

$$\sum_{i=1}^M P_i = 1 \geq \sum_{i=1}^M \left(\frac{1}{m^{l_i}}\right) > \frac{1}{m}.$$

Це нерівність Крафта, яка забезпечує існування коду, який має кодові слова вказаної тривалості.

Помножуючи на P_i першу із вказаних нерівностей та підсумовуючи по i , одержуємо:

$$H_m(X) = \sum_{i=1}^M P_i \log_m P_i \leq \sum_{i=1}^M P_i l_i < H_m(X) + 1.$$

Ураховуючи позначення середньої тривалості L кодового слова, остаточно одержуємо

$$H_m(X) \leq L < H_m(X) + 1,$$

тобто ентропія коду Шеннона-Фано дає нижню границю цієї середньої тривалості. Якщо кожна з імовірностей P_i являє собою від'ємний цілий степінь основи m , нерівність переходить у рівність. При цьому $L=H_m(X)$.

У розглянутому прикладі (табл.4.1) для двійкового коду, коли $m=2$, ця умова виконується і $L=H_m(X)$.

Зробимо основні висновки щодо коду Шеннона-Фано, підкреслюючи його фізичні характеристики.

1. Код оптимальний. Він забезпечує найменшу середню тривалість кодового слова або найбільшу швидкість передавання інформації.

2. Підвищення швидкості передавання інформації з фізичної точки зору забезпечується тим, що повідомленням з найбільшою ймовірністю (тобто повідомленням, які зустрічаються найчастіше) приписуються кодові слова найменшої тривалості, і навпаки, повідомленням з найменшою ймовірністю (тобто повідомленням, які зустрічаються найбільш рідко) кодуються словами максимальної тривалості. Тому у середньому швидкість передавання кодових слів підвищується. Така ідея на якісному рівні використовувалась і у перших відомих кодах, але у коді Шеннона-Фано з неї було одержано все, що принципово можливо.

3. При передаванні кодових слів не потрібні додаткові розділові символи поміж словами, тому що, як видно з табл.4.1, ніяке ходове слово меншої тривалості не є початком довільного кодового слова більшої тривалості (тобто це префіксний код).

Наведемо ще один приклад побудови оптимального коду.

Приклад 4.2 Метод Хаффмена. Код Шеннона-Фано дозволяє визначити тривалість кодового слова, відповідаючого кожному повідомленню, безпосередньо по його імовірності. Є можливість зробити інакше – визначити тривалість кодового слова по всій сукупності ймовірностей повідомлень. Так поводяться згідно із запропонованою Д.А. Хаффменом процедурою кодування. Розглянемо приклад. Нехай заданий ансамбль (X, P_X) , наведений у табл.4.2. Розмістимо повідомлення у послідовності зменшення їх імовірностей (крок 1).

2-й крок. Згрупуємо разом m_0 повідомлень з найменшими ймовірностями. Знайдемо їх сумарну ймовірність. Тут $2 \leq m_0 \leq m$, $\frac{M - m_0}{m - 1} = a$, a -ціле додатне число. При $m=2$ маємо $m_0=2$. Для простоти припус-

каємо, що розглядається двійковий код. При цьому групується два повідомлення з найменшою ймовірністю.

3-й крок. Сформуємо перший допоміжний ансамбль, розглядаючи два об'єднаних повідомлення як нове повідомлення. Знову упорядкуємо повідомлення зі зменшенням їх ймовірностей.

4-й крок. Об'єднаємо два найменш ймовірних повідомлень нового ансамблю та обчислимо їх сумісну ймовірність.

5-й крок. Сформуємо другий допоміжний ансамбль, розглядаючи об'єднання двох повідомлень, введених на 4-му кроці, як нове повідомлення. Розмістимо елементи цього ансамблю у послідовності зменшення ймовірностей.

Продовжимо цей процес, доки у ансамблі не залишиться єдине повідомлення з ймовірністю одиниця. Процес об'єднання ймовірностей приводить до кодового дерева, яке однозначно визначає код. Для знаходження кодового слова, наприклад, руху вгору по дереву слід зіставляти символ 1, руху вниз – символ 0 (рух здійснюється справа наліво від вершини до повідомлення і символи кодового слова знаходяться, починаючи з молодших розрядів).

Таблиця 4.2

Повідомлення	Ймовірність	1	2	3	4	5	6	Кодове слово
x_1	0,3							11
x_2	0,2							01
x_3	0,15							101
x_4	0,15							100
x_5	0,1							001
x_6	0,05							0001
x_7	0,05							0000

Аналогічно попередньому, можна довести, що цей процес кодування приводить до оптимального коду.

Код Хаффмена з практичної точки зору більш прийнятний у порівнянні із кодом Шеннона-Фано.

4.2. Передавання інформації по дискретному каналу із завадами

У цьому підрозділі розглядається те саме коло питань, що і у попередньому – ентропія, кількість інформації, теорема кодування. Однак раніше цей шлях був обраний лише із методичних міркувань, щоб у найпростішому у математичному розумінні випадку розглянути поняття ентропії та кількості інформації. Насправді фізично результати стосувалися лише джерела повідомлень. У даному підрозділі фізично йде мова про передавання інформації від джерела до одержувача через так званий канал зв'язку.

4.2.1. Математичні моделі дискретного каналу

З фізичної точки зору канал зв'язку – це сукупність технічних засобів, які забезпечують передавання та приймання сигналів від джерела до одержувача повідомлень.

У теорії інформації використовується математична модель каналу – трійка $[X, Y, P_{Y|X}]$, де $X = \{x_1, x_2, \dots, x_m\}$ – множина, з якої вибираються символи, що поступають на вхід каналу, $Y = \{y_1, y_2, \dots, y_n\}$ – множина вихідних символів; $P_{Y|X}$ – відповідна ймовірнісна міра, яка задає ймовірність будь-якого символу y_j на виході каналу за умови, що відомий символ x_i на його вході.

Відмінність прийнятих символів від переданих фізично трактується як наявність завади у каналі.

Канал називають дискретним, якщо множина X вхідних, множина Y вихідних символів і час дискретні. Взагалі кажучи, до дискретних відносять канали, у яких множини X та Y або одна з них зліченні. У випадку скінченних множин кількість елементів у X та Y може бути різною.

Нехай заданий вхідний ансамбль символів (X, P_X) та множина вихідних символів. Дискретний канал називають каналом без пам'яті, якщо для довільних q та j , а також послідовностей

$$\begin{aligned}\bar{x}(j) &= (x^{(j)}, x^{(j+1)}, \dots, x^{(j+q-1)}), \\ \bar{y}(j) &= (y^{(j)}, y^{(j+1)}, \dots, y^{(j+q-1)}),\end{aligned}$$

має місце рівність:

$$P(\vec{y}^{(j)}|\vec{x}^{(j)}) = \prod_{i=j}^{j+q-1} P_i(y^{(i)}|x^{(i)}).$$

У загальному випадку, коли задані ансамбль вхідних символів (X, P_X) та множина Y вихідних символів, а також заданий канал, можна визначити спільний розподіл

$$P(\vec{x}^{(j)}, \vec{y}^{(j)}) = P(\vec{x}^{(j)})P(\vec{y}^{(j)}|\vec{x}^{(j)})$$

та розподіл ймовірностей вихідної послідовності

$$P(\vec{y}^{(j)}) = \sum_{X^q} P(\vec{x}^{(j)})P(\vec{y}^{(j)}|\vec{x}^{(j)}) = \sum_{X^q} P(\vec{x}^{(j)}, \vec{y}^{(j)}).$$

Кажуть, що дискретний канал без пам'яті задовольняє умову стаціонарності, якщо розподіл імовірності $P(\vec{y}^{(j)}|\vec{x}^{(j)})$ не залежить від початку відліку « j ».

Марківським називають канал з пам'яттю, в якому ймовірність похибки характеризує простий ланцюг Маркова (тобто ймовірність не залежить від того, який символ передається, а залежить від того, правильно чи помилково прийнятий попередній символ).

Аналогічно задаються математичні моделі інших конкретних каналів зв'язку.

4.2.2. Ентропія складного ансамблю

У п.4.1.1 ми вже торкалися питання складного ансамблю повідомлень. Однак у першому підрозділі зусилля були зосереджені на впровадженні через ентропію власної (та середньої) інформації ансамблю повідомлень. У цьому підрозділі складні ансамблі ми будемо аналізувати з позицій передавання інформації.

Розглянемо ентропію складного ансамблю (XY, P_{XY}) , який, як підкреслювалось, визначає два сумісно заданих ансамблів (X, P_X) та (Y, P_Y) , де

$$P_X = \sum_Y P_{XY}, \quad P_Y = \sum_X P_{XY}.$$

З фізичної точки зору ансамблі (X, P_X) та (Y, P_Y) зручно уявляти собі як ансамблі повідомлень (символів, сигналів) на вході та відповідно виході каналу.

Визначимо ентропію складного ансамблю (XY, P_{XY}) або, що те саме, ентропію сумісно заданих ансамблів (X, P_X) та (Y, P_Y) :

$$H(X, Y) = - \sum_{i=1}^m \sum_{j=1}^n P(x_i, y_j) \log P(x_i, y_j). \quad (4.15)$$

Як і раніше, можна ввести випадкову величину, яка задана на ансамблі (XY, P_{XY}) і приймає значення

$$J(x_i, y_j) = -\log P(x_i, y_j), \quad i = \overline{1, m}, \quad j = \overline{1, n}. \quad (4.16)$$

Вираз (4.15) визначає математичне сподівання цієї випадкової величини:

$$H(X, Y) = -M \log P(X, Y). \quad (4.15a)$$

Розглянемо властивості ентропії складного ансамблю.

1 Ентропії незалежних ансамблів складаються:

$$H(X, Y) = H(X) + H(Y). \quad (4.17)$$

Незалежні ансамблі вводяться так: формується прямий добуток множин X та Y як сукупність упорядкованих пар (x_i, y_j) , імовірність на добутку множин задається як добуток імовірностей:

$$P(x_i, y_j) = P(x_i)P(y_j),$$

тобто

$$P_{XY} = P_X P_Y. \quad (4.18)$$

Вираз (4.18), що визначає імовірнісну міру $P_{XY} = P(x, y)$ на множині XY , вводиться як добуток імовірнісних мір $P_X = P(x)$ та $P_Y = P(y)$ на множинах X та Y і по суті є означенням статистичної незалежності.

Підставляючи вираз (4.18) у (4.15a), одержуємо властивість ансамблю (4.17):

$$\begin{aligned} H(X, Y) &= -M \log P(X, Y) = -M \log P(X)P(Y) = -M \log P(X) - M \log P(Y) = \\ &= H(X) + H(Y). \end{aligned}$$

2. Ентропія складного ансамблю дорівнює ентропії однієї зі складових плюс умовна ентропія другої частини відносно першої:

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y).$$

У загальному випадку складний ансамбль будується так: вводять прямий добуток множин $X = \{x_1, x_2, \dots, x_m\}$ та $Y = \{y_1, y_2, \dots, y_n\}$ як мно-

жину упорядкованих пар (x_i, y_j) , а ймовірність на них задається таким чином

$$P(x_i, y_j) = P(x_i)P(y_j|x_i) = P(y_j)P(x_i|y_j). \quad (4.20)$$

Ймовірність (4.20) визначає статистично залежні події x_i та y_j . У разі їх незалежності і як частинний випадок одержується попередня рівність $P(x_i, y_j) = P(x_i)P(y_j)$.

Вираз (4.20) дає можливість у загальному випадку записати ймовірнісну міру складного ансамблю

$$P_{XY} = P_X P_{Y|X} = P_Y P_{X|Y}. \quad (4.21)$$

Підставляючи (4.21) у (4.15а), легко одержати (4.19):

$$\begin{aligned} H(X, Y) &= -M \log P(X, Y) = -M \log P(X)P(Y|X) = -M \log P(X) - M \log P(Y|X) = \\ &= H(X) + H(Y|X); \end{aligned}$$

$$\begin{aligned} H(X, Y) &= -M \log P(X, Y) = -M \log P(Y)P(X|Y) = -M \log P(Y) - M \log P(X|Y) = \\ &= H(Y) + H(X|Y). \end{aligned}$$

У разі незалежних ансамблів маємо

$$H(Y|X) = H(Y), \quad H(X|Y) = H(X).$$

Тому одержані вирази приводять до (4.17):

$$H(X, Y) = H(X) + H(Y).$$

3. Ентропія N ансамблів не перевищує суми їх ентропій:

$$H(X_1, \dots, X_N) \leq \sum_{i=1}^N H(X_i). \quad (4.22)$$

Розглянемо складний ансамбль $(X_1, X_2, \dots, X_N, P_{X_1, X_2, \dots, X_N})$, який визначає N сумісно заданих ансамблів (X_i, P_{X_i}) , $i = \overline{1, N}$.

Ураховуючи, що ймовірність $P(x_1, x_2, \dots, x_N)$ можна зобразити у вигляді

$$P(x_1, x_2, \dots, x_N) = P(x_1) \prod_{j=1}^N P(x_j | x_{j-1}, \dots, x_1),$$

легко одержати вираз для ентропії $H(X_1, \dots, X_N)$:

$$H(X_1, \dots, X_N) = H(X_1) + H(X_2 | X_1) + H(X_3 | X_2, X_1) + \dots + H(X_N | X_{N-1}, \dots, X_1).$$

Ураховуючи, що

$$H(X_i|X_i^s) \leq H(X_i), \quad (4.23)$$

де $X_i^s = X_{i-1}X_{i-2}\dots X_{i-s}$, одержуємо шукану нерівність:

$$H(X_1, \dots, X_N) \leq \sum_{i=1}^N H(X_i).$$

Знак рівності у (4.23) досягається при незалежних ансамблях, тому і у (4.22) він досягається за тією ж умовою.

Аналогічно обчислюється ентропія сумісного ансамблю вхідних та вихідних послідовностей символів для стаціонарного дискретного каналу без пам'яті, марківського та інших каналів.

4.2.3. Взаємна інформація та її властивості

Нехай (X, P_X) – ансамбль повідомлень на вході, а (Y, P_Y) – ансамбль на виході каналу. Знайдемо для них повну та індивідуальну взаємну інформацію.

Повна взаємна інформація.

Кількість інформації щодо ансамблю X , яка одержується в результаті спостереження ансамблю Y – це міра знятої невизначеності ансамблю X у результаті спостереження ансамблю Y :

$$I_{X \rightarrow Y} = H(X) - H(X|Y), \quad (4.24)$$

де $H(X)$, $H(X|Y)$ – відповідно початкова та кінцева невизначеність ансамблю X .

Функцію $I_{X \rightarrow Y}$ називають повною (середньою) інформацією щодо ансамблю X , яка міститься у ансамблі Y . Знайдемо інші важливі зображення цієї функції. Згідно (4.19) маємо

$$H(X) - H(X|Y) = H(Y) - H(Y|X).$$

Тому можна ввести функцію

$$I_{X \rightarrow Y} = H(Y) - H(Y|X), \quad (4.25)$$

яка визначає повну інформацію відносно ансамблю Y , яка міститься у ансамблі X . Характерно, що $I_{X \rightarrow Y} = I_{Y \rightarrow X}$. Тому вводять позначення

$$I(X; Y) = I_{X \rightarrow Y} = I_{Y \rightarrow X};$$

функцію $I(X; Y)$ називають повною (середньою) взаємною інформацією.

Ураховуючи (4.19), співвідношення для $I(X;Y)$ можна також записати у вигляді

$$I(X;Y) = H(X) + H(Y) - H(X,Y). \quad (4.26)$$

Це співвідношення дає можливість дати також тлумачення кількості інформації як міри відповідності двох випадкових об'єктів, числове значення якої визначається виразом (4.26).

Таким чином маємо основні визначення середньої взаємної інформації:

$$I(X;Y) = \begin{cases} H(X) - H(X|Y), \\ H(Y) - H(Y|X), \\ H(X) + H(Y) - H(X,Y). \end{cases} \quad (4.27)$$

Вираз (4.27) дає можливість дати фізичну трактовку процесу передавання інформації по каналу із завадами, що ілюструє рис. 4.3.

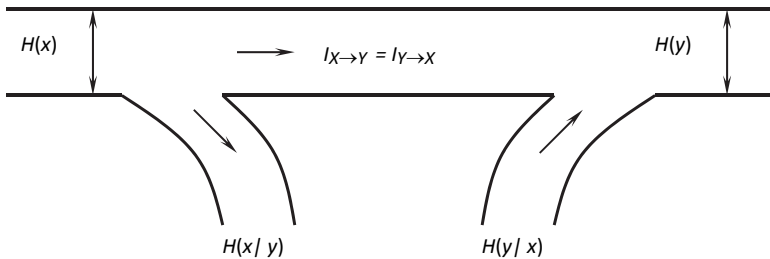


Рис. 4.3. – Ілюстрація процесу передавання інформації по каналу із завадами

Джерело повідомлень характеризується ентропією $H(X)$, котра визначає середню кількість інформації, яка припадає на повідомлення джерела у каналі без завад. Коли завада є, стан джерела визначається в результаті його спостереження не повністю – існує «залишкова» невизначеність $H(X|Y)$. Одержувачу повідомлень доступна різниця цих ентропій

$$I_{X \to Y} = H(X) - H(X|Y).$$

З іншого боку, одержувач повідомлень характеризується загальною невизначеністю $H(Y)$ – ентропією ансамблю прийнятих повідомлень. Останнє складається з інформації $I_{Y \to X}$, яку може одержати ансамбль Y щодо ансамблю X , а також ентропії завади $H(Y|X)$:

$$H(Y) = I_{X \to Y} + H(Y|X).$$

Ентропія завади $H(Y|X)$ – це «дезинформація», котра вноситься завадою і заважає передаванню інформації від джерела.

Співвідношення (4.27) характеризують закономірності передавання та приймання інформації:

$$I_{X \rightarrow Y} = H(X) - H(X|Y) = H(Y) - H(Y|X) = I_{Y|X} = I(X; Y).$$

Розглянемо властивості повної взаємної інформації для різних ансамблів:

1. Якщо ансамблі X та Y незалежні,

$$I(X; Y) = 0.$$

Дійсно, для незалежних ансамблів

$$H(X; Y) = H(X) + H(Y).$$

Тому згідно з (4.26), $I(X; Y) = 0$.

2. Якщо стан ансамблю X повністю визначає стан ансамблю Y

$$I(X; Y) = H(X) = H(Y). \quad (4.28)$$

Дійсно, у цьому випадку $H(X) = H(Y)$, тоді $H(X|Y) = H(Y|X) = 0$. Згідно з (4.24) – (4.25)

$$I(X; Y) = I(X) = I(Y) = H(X) = H(Y).$$

Це канал без завад, розглянутий у підрозділі 4.1. Кількість інформації, впроваджена там, згідно з (4.28) – це повна власна інформація відносно самого себе.

Індивідуальна взаємна інформація. Введемо індивідуальну взаємну інформацію, починаючи з її середнього значення.

Задамо складний ансамбль (XY, P_{XY}) , що визначає сумісно задані ансамблі (X, P_X) та (Y, P_Y) , де $X = \{x_1, \dots, x_m\}$, $Y = \{y_1, \dots, y_n\}$. Конкретизуємо для них загальні співвідношення (4.27):

$$I(X; Y) = \sum_{i=1}^m \sum_{j=1}^n P(x_i, y_j) \log \frac{P(x_i|y_j)}{P(x_i)}, \quad (4.29)$$

$$I(X; Y) = \sum_{i=1}^m \sum_{j=1}^n P(x_i, y_j) \log \frac{P(y_j|x_i)}{P(y_j)}, \quad (4.30)$$

$$I(X; Y) = \sum_{i=1}^m \sum_{j=1}^n P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)}, \quad (4.31)$$

Співвідношення (4.29) – (4.31) еквівалентні; формулу (4.31) можна записати, помножуючи чисельник і знаменник виразу під логарифмом у (4.29) на $P(y_j)$, а у (4.30) – на $P(x_i)$. Обміркуємо докладніше, наприклад, співвідношення (4.29).

Впровадимо на ймовірнісному просторі (XY, P_{XY}) випадкову величину

$$J(X, Y) = \log \frac{P(X|Y)}{P(X)}, \quad (4.32)$$

що приймає значення $\log \frac{P(x_i|y_j)}{P(x_i)}$, $i = \overline{1, m}$, $j = \overline{1, n}$.

Її називають взаємною інформацією. Це індивідуальна взаємна інформація поміж повідомленнями x_i та y_j , яка існує у повідомленні y_j щодо повідомлення x_i

Її середнє значення

$$I(X; Y) = M \log \frac{P(X|Y)}{P(X)}, \quad (4.29a)$$

Аналогічно, розглядаючи (4.30), одержуємо

$$I(X; Y) = M \log \frac{P(Y|X)}{P(Y)}, \quad (4.30a)$$

а розглядаючи (4.31), дістаємо

$$I(X; Y) = M \log \frac{P(X, Y)}{P(X)P(Y)}, \quad (4.31a)$$

Співвідношення (4.29a)-(4.31a) характеризують повну (середню) інформацію як математичне сподівання випадкових величин

$$J(X; Y) = \log \frac{P(X|Y)}{P(X)}, \quad J(X; Y) = \log \frac{P(Y|X)}{P(Y)}, \quad \text{та} \quad J(X; Y) = \log \frac{P(X, Y)}{P(X)P(Y)}.$$

Наведемо основні властивості середньої взаємної інформації

1. $I(X; Y) \geq 0$.
2. $I(X; Y) = 0$. тоді і тільки тоді, коли ансамблі X та Y незалежні.
3. $I(X; Y) = I(Y; X)$.

$$4. I(X; Y) = \begin{cases} H(X) - H(X|Y), \\ H(Y) - H(Y|X), \\ H(X) + H(Y) - H(X, Y). \end{cases}$$

Властивості 1-4 обмірковані раніше. Наведемо без доведення ще деякі властивості функції $I(X;Y)$.

5. $I(X;Y)$ – опукла функція.

Нагадаємо, що функція $f(x)$ називається опуклою вгору у деякій області G , якщо для будь-яких $n=1,2,\dots$, невід’ємних чисел $\lambda_1,\dots,\lambda_n$, що задовольняють рівність $\sum_{i=1}^n \lambda_i = 1$, та будь-яких чисел $x_1, x_2, \dots, x_n \in G$ виконується нерівність

$$\sum_{i=1}^n \lambda_i f(x_i) \geq f\left(\sum_{i=1}^n \lambda_i x_i\right).$$

Якщо виконується протилежна нерівність, функція називається опуклою донизу.

6. Середнє значення взаємної інформації, яке міститься у вихідній послідовності \vec{y} відносно вхідної \vec{x} , задовольняє нерівність

$$I(X^q; Y^q) \leq \sum_{i=1}^q I(X_i; Y_i). \quad (4.33)$$

Тут $I(X_i; Y_i)$ – середня взаємна інформація поміж i -м вхідним та i -м вихідним символами.

4.2.4. Кодування у дискретному каналі із завадами

Результат, який обмірковується у цьому пункті є центральним у теорії інформації. Відповідна теорема кодування називається теоремою Шеннона для дискретного каналу із завадами або другою теоремою Шеннона; кажуть також про оптимальне за Шенноном завадостійке кодування.

Для розглядання суті теореми Шеннона введемо інформаційну пропускну здатність каналу.

Пропускна здатність каналу. Пропускна здатність каналу визначають як максимально можливу швидкість передавання інформації:

$$C = \max R \left[\frac{\partial v.од.}{c} \right], \quad (4.34)$$

де $R = \frac{I_T}{T}$ – швидкість передавання інформації.

Конкретизуємо (4.34) стосовно дискретного каналу із завадами.

Нехай заданий вхідний ансамбль (X, P_X) , множина Y та дискретний канал. Середнє значення взаємної інформації вихідної послідовності символів $\vec{y} = (y^{(1)}, \dots, y^{(q)})$ відносно вхідної послідовності $\vec{x} = (x^{(1)}, \dots, x^{(q)})$ дорівнює $I(X^q, Y^q)$. Використовуючи означення R як кількості інформації в одиницю часу та впроваджуючи швидкість передавання символів в одиницю часу $V \left[\frac{\text{симв.}}{c} \right]$, одержуємо співвідношення

$$C = V \max_{P_{X,q}} \left[\frac{I(X^q, Y^q)}{q} \right], \quad (4.35)$$

де максимум (або супремум) береться по всіляких розподілах ймовірностей P_X та всіх цілих додатних q .

Величину:

$$C_{\text{симв}} = \max_{P_{X,q}} \left[\frac{I(X^q, Y^q)}{q} \right] \quad (4.36)$$

називають пропускну здатністю каналу на символ.

У стаціонарному дискретному каналі без пам'яті

$$P(\vec{y} | \vec{x}) = \prod_{i=1}^q P(y^{(i)} | x^{(i)}),$$

тому:

$$I(X^q; Y^q) = \sum_{i=1}^q I(X_i; Y_i) = qI(X; Y),$$

і формула (4.36) набуває вигляду

$$C_{\text{симв}} = \max_{P_X} I(X; Y). \quad (4.36a)$$

У частковому випадку симетричного стаціонарного каналу без пам'яті, для якого (при $m=n$)

$$P(y^{(i)} | x^{(i)}) = \begin{cases} P/(m-1), & i \neq j, \\ 1-P, & i = j, \end{cases}$$

із (4.36a) легко одержати формулу

$$C_{\text{симв}} = \log m + P \log \frac{P}{m-1} + (1-P) \log(1-P). \quad (4.37)$$

Для двійкового симетричного стаціонарного каналу без пам'яті одержуємо $C_{симв}$ з (4.37) при $m=2$:

$$C_{симв} = 1 + P \log P + (1 - P) \log(1 - P). \quad (4.37a)$$

Пропускна здатність у двійкових одиницях на одиницю часу для останнього випадку

$$C = VC_{симв} = V[1 + P \log P + (1 - P) \log(1 - P)]. \quad (4.38)$$

Графік залежності C/V від імовірності похибки приймання символів наведено на рис. 4.4

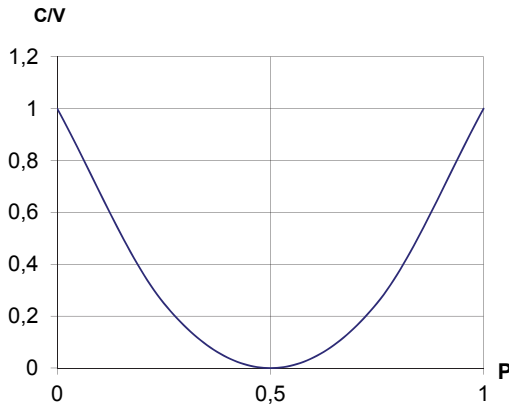


Рис. 4.4. — Залежність пропускної здатності стаціонарного двійкового симетричного каналу без пам'яті від імовірності похибки приймання символів

Теорема Шеннона для каналу із завадами. Основний результат теорії інформації полягає у тому, що наявність завади у каналі не обмежує досягаємої вірності приймання, а саме: по такому каналу можна передавати інформацію з ненульовою швидкістю та як завгодно малою ймовірністю помилки.

Сформулюємо цей результат. Додатково впровадимо продуктивність джерела R' як кількість інформації, яка видається джерелом в одиницю часу.

Теорема 4.2 (друга теорема Шеннона). Нехай ϵ джерело дискретних повідомлень з продуктивністю R' та дискретний канал із завадами з пропускною здатністю C . Тоді, якщо $R' \leq C - \epsilon$, де ϵ — довільна як завгодно мала величина, то завжди існує спосіб кодування та декодування, який дає можливість передати по каналу всі повідомлення

джерела (без зростаючої затримки у часі), забезпечуючи яку завгодно малу ймовірність помилкового декодування. Якщо $R' > C$, тоді таких способів не існує.

Теорема про завадостійке кодування включає як прямий результат, стверджуючий існування потрібного коду, а також обернений результат, що при недодержанні потрібних умов відповідний код не існує.

Розглянемо на фізичному рівні елементи доведення теореми 4.2.

Нехай повідомлення джерела X передаються по каналу із завадами, який має алфавіти

$A = \{a_i\}$ та $B = \{b_j\}$ на вході і виході та пропускну здатність на символ $C_{\text{симв}}$.

Процедура кодування та декодування полягає у наступному.

Зіставляються різні послідовності символів джерела та кодові послідовності тривалістю q . Припускається, що вибрано та перенумеровано M кодових комбінацій \vec{a}_i тривалістю q . Їх називають дозволеними комбінаціями блочного q -значного коду. Завдяки дії завади замість їх може бути прийнята довільна з m^q кодових послідовностей \vec{b}_i тривалістю q .

Декодування полягає у тому, що вся множина $B^{[q]}$, яка складається із m^q послідовностей \vec{b}_i на виході каналу, розбивається на M підмножин $B_1^{[q]}, B_2^{[q]}, \dots, B_M^{[q]}$. Якщо прийнята кодова комбінація попадає у область $B_i^{[q]}$, приймається рішення, що передана комбінація \vec{a}_i . Доводиться, що ймовірність помилкового рішення може бути зроблена як завгодно малою, якщо величина q достатньо велика, а на число M накладені відповідні обмеження. Справджується таке твердження.

Лема 4.1. Якщо канал має пропускну здатність на символ $C_{\text{симв}}$ і задані довільні $\epsilon > 0$ і $H(X) < C$, тоді завжди знайдеться таке q_0 , що при будь-якому $q > q_0$ існує блочний код тривалістю q , який складається із $M = 2^{qH}$ комбінацій, та метод декодування, які забезпечують виконання нерівності

$$P(B_i^{[q]} | \vec{a}_i) \geq 1 - \epsilon \quad (i=1, \overline{M}). \quad (4.39)$$

Якщо $H(X) > C_{\text{симв}}$, тоді ця нерівність при задовільному ϵ не виконується, як би велике не було q .

У цій лемі вхід каналу ще не пов'язаний з джерелом повідомлень. Це можна зробити, використовуючи твердження теореми 4.1, що

можна узгодити джерело із каналом (як при кодуванні у каналі без завад).

При цьому число різних символів алфавіту каналу приймається рівним M .

Згідно з теоремою 4.1 існує спосіб кодування повідомлень джерела такий, що на кожне елементарне повідомлення джерела у середньому приходиться

$$L \geq \frac{H(X)}{\log M}$$

кодових комбінацій тривалістю q .

Число кодових комбінацій згідно з лемою 4.1 дорівнює $M = 2^{qH}$, де $H(X) < C_{\text{симв}}$, тобто $L > \frac{H(X)}{qC_{\text{симв}}}$.

Тому умова узгодженості у часі джерела з даним кодом має вигляд

$$H(X)V_{\text{дж}} \leq C_{\text{симв}}V_{\text{к}},$$

де $V_{\text{дж}}$ – швидкість знаків джерела,

$V_{\text{к}}$ – швидкість знаків каналу.

Остаточно одержуємо $R' < C$,

де $R' = H(X)V_{\text{дж}}$, $C = C_{\text{симв}}V_{\text{к}}$.

Ця нерівність і встановлює другу теорему Шеннона.

У лемі 4.1 по суті використовується поняття типових і нетипових послідовностей джерела, які є наслідком властивості асимптотичної рівномірності, яка витікає із теореми про великі числа. Наведемо відповідні дані.

Усі послідовності джерела тривалістю q можна розбити на дві групи – типові і нетипові. Зі зростанням q імовірність того, що джерело видасть типову послідовність, прагне до одиниці, а ймовірності різних нетипових послідовностей – до нуля, тобто вони стають практично неможливими. Справджується таке твердження.

Лема 4.2. Для довільних як завгодно малих ϵ і δ можна знайти таке число $q_0 = q_0(\epsilon, \delta)$, що з ймовірністю, більш ніж $1 - \delta$, послідовність символів джерела \bar{x} тривалістю $q \geq q_0$ має ймовірність $P(\bar{x})$, що задовольняє нерівність

$$\left| \frac{1}{q} \log \frac{1}{P(\bar{x})} - H(X) \right| < \varepsilon.$$

З цієї леми витікає, що число типових послідовностей тривалістю q асимптотично прагне до $N_{\text{тип}}$:

$$N_{\text{тип}}(X) = 2^{qH(X)}.$$

4.3. Передавання інформації по неперервному каналу

У попередніх підрозділах перед доведенням теореми кодування було обрано такий шлях викладання результатів теорії інформації: спочатку вводили ентропію, далі – середню, а потім – індивідуальну інформацію. У розглянутих найпростіших випадках не має значення, з чого починати чи спочатку впроваджувати ентропію, а під кінець – індивідуальну інформацію, чи починати з індивідуальної інформації. Для неперервних каналів це вже істотно і коректніше починати із введення індивідуальної взаємної інформації.

Цей підрозділ присвячено особливостям неперервних каналів при використанні неперервних джерел з дискретним та неперервним часом.

4.3.1. Математична модель джерела повідомлень

Математичним еквівалентом поняття ансамблю повідомлень є ймовірнісний простір (Ω, \mathcal{F}, P) . При завданні дискретного ансамблю робили так: задавали ансамбль як скінченну множину X разом з ймовірнісною мірою P_X , або тлумачили його як випадкову величину, яка приймає значення з тієї ж множини X з ймовірностями P_i , $i=1, M$, які визначаються мірою P_X . У загальному випадку, коли множина X не є дискретною, розглядають ймовірнісний простір $(X, \mathcal{B}, F(x))$, де X – дійсна вісь R , події з класу \mathcal{B} – борелівські множини, $F(x)$ – функція розподілу, яка задає ймовірнісну міру на R . У теорії інформації звичайно під неперервним ансамблем розуміють ймовірнісний простір $(X, \mathcal{B}, F(x))$, записуючи його стисло $(X, F(x))$.

Неперервним ансамблем будемо називати пару $(X, F(x))$, де X – множина дійсних чисел; $F(x)$ – функція розподілу. Часто під неперервним ансамблем розуміють пару $(X, p(x))$, де $p(x)$ – щільність ймовірності.

Нагадаємо, що $p(x) = \frac{dF(x)}{dx}$ вводить для абсолютно неперервних функцій розподілу.

Основні результати теорії інформації для неперервних ансамблів звичайно викладаються при запису ансамблю у вигляді $(X, p(x))$, але це лише традиція; їх можна одержати, якщо ансамбль задавати у вигляді $(X, F(x))$ або (X, P_X) .

Заданий неперервний ансамбль з точки зору теорії випадкових процесів можна розглядати як переріз процесу у фіксовану мить часу.

Задамо складні ансамблі.

У найпростішому випадку розглядання двох перерізів процесу маємо $(XY, F(x, y))$ або $(XY, p(x, y))$, де

$$p(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} -$$

двовимірна щільність імовірності. Складний ансамбль визначає сумісно задані ансамблі $(X, p(x))$ та $(Y, p(y))$,

$$\begin{aligned} \text{де} \quad p(x) &= \int_{-\infty}^{\infty} p(x, y) dy; \\ p(y) &= \int_{-\infty}^{\infty} p(x, y) dx. \end{aligned}$$

У загальному випадку вводять складний ансамбль

$$(X_1 X_2 \dots X_n, F(x_1, x_2, \dots, x_n)), \quad (4.40)$$

який визначає n сумісно заданих ансамблів

$$(X_i, p(x_i)), \quad i = \overline{1, n}, \quad (4.41)$$

$$\text{де} \quad p(x_i) = \int_{X_1 \dots X_{i-1} X_{i+1} \dots X_n} p(x_1, \dots, x_n) d_1 \dots d_{i-1} d_{i+1} \dots d_n -$$

одновимірна щільність імовірності.

Такий ансамбль можна трактувати як n перерізів випадкового процесу.

Нарешті можна дати означення неперервного джерела (джерела аналогових повідомлень).

$\xi_1(n)$

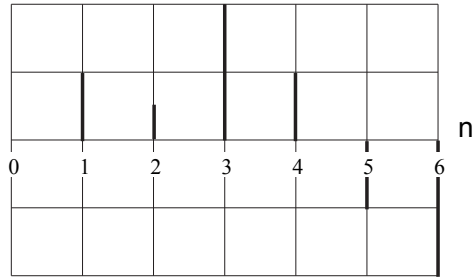


Рис. 4.5. а)

$\xi_2(n)$

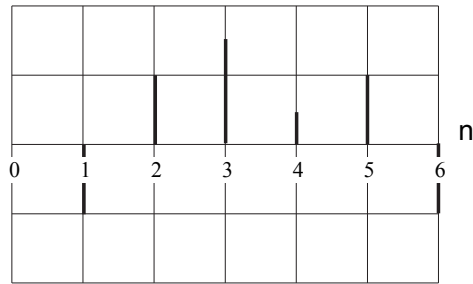


Рис. 4.5. б)

$\xi_3(n)$

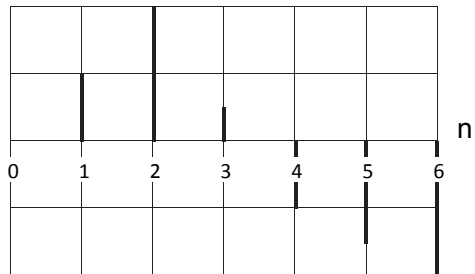


Рис. 4.5. в)

Рис. 4.5. — Ілюстрація реалізацій випадкового процесу з дискретним часом на виході джерела аналогових повідомлень з дискретним часом

Нехай U_X — джерело аналогових повідомлень (з дискретним — рис.4.5 —або неперервним часом). Кажуть, що джерело аналогових повідомлень задано, якщо задана сім'я скінченновимірних розподілів випадкового процесу на його виході, що задовольняє умови узгодженості Колмогорова.

4.3.2. Кількість інформації та ентропія

Для впровадження індивідуальної взаємної інформації використаємо результати, одержані для дискретних ансамблів. При цьому спочатку апроксимуємо неперервний ансамбль дискретним, а далі повернемося до неперервного шляхом відповідного граничного переходу.

Розглянемо складний ансамбль $(XY, F(x, y))$, який визначає два сумісно заданих ансамблів $(X, F(x))$ і $(Y, F(y))$. Припускається, що $F(x, y)$ — абсолютно неперервна функція розподілу, тобто існує двовимірна щільність імовірності $p(x, y)$, а отже, $p(x)$ та $p(y)$.

Використаємо індивідуальну взаємну інформацію для дискретних ансамблів

$$J(X, Y) = \log \frac{P(X, Y)}{P(X)P(Y)}, \quad (4.42)$$

яка приймає значення

$$J(x_i, y_j) = \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)}, \quad i = \overline{1, m}, \quad j = \overline{1, n}. \quad (4.42a)$$

Дискретизуємо ансамбль $(XY, F(x, y))$.

На рис. 4.6 дана ілюстрація дискретизації неперервного одновимірного ансамблю. Для деякого Δ згідно з (4.42a) можна знайти індивідуальну взаємну інформацію для дискретизованих ансамблів:

$$\begin{aligned} J_{\Delta}(x_i, y_j) &= \log \frac{P\{x_i - \Delta x < x \leq x_i, y_j - \Delta y < y \leq y_j\}}{P\{x_i - \Delta x < x \leq x_i\}P\{y_j - \Delta y < y \leq y_j\}} = \\ &= \log \frac{F(x_i, y_j) - F(x_i - \Delta x, y_j - \Delta y)}{[F(x_i) - F(x_i - \Delta x)][F(y_j) - F(y_j - \Delta y)]}. \end{aligned} \quad (4.43)$$

Після ділення чисельника та знаменника (4.43) на $\Delta x \Delta y$ та граничного переходу одержуємо

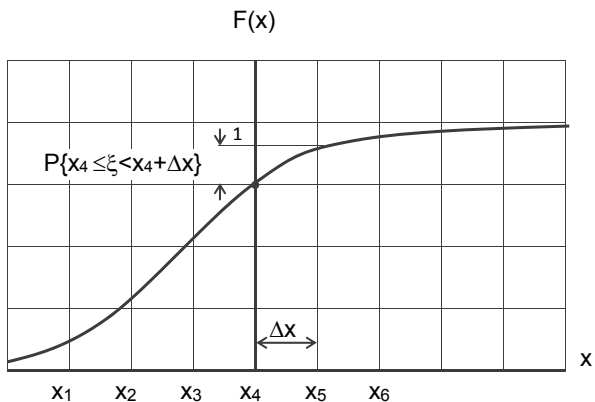


Рис. 4.6. а)

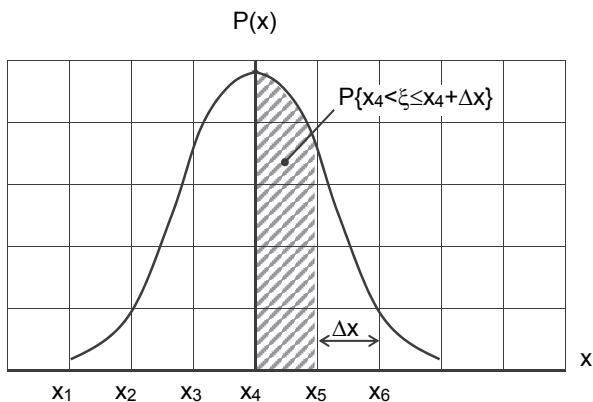


Рис. 4.6. б)

Рис. 4.6. – Ілюстрація дискретизації неперервних ансамблів $(X, F(x))$ та $(X, p(x))$:

а) дискретизація функції розподілу; б) дискретизація щільності імовірності

$$\begin{aligned}
J(x_i, y_j) &= \lim_{\Delta x \rightarrow 0} \lim_{\Delta y \rightarrow 0} I_{\Delta}(x_i, y_j) = \lim_{\Delta x \rightarrow 0} \lim_{\Delta y \rightarrow 0} \log \left[\frac{\frac{F(x_i, y_j) - F(x_i - \Delta x, y_j - \Delta y)}{\Delta x \Delta y}}{\left[\frac{F(x_i) - F(x_i - \Delta x)}{\Delta x} \right] \left[\frac{F(y_j) - F(y_j - \Delta y)}{\Delta y} \right]} \right] = \\
&= \log \left[\frac{\lim_{\Delta x \rightarrow 0} \lim_{\Delta y \rightarrow 0} \frac{F(x_i, y_j) - F(x_i - \Delta x, y_j - \Delta y)}{\Delta x \Delta y}}{\lim_{\Delta x \rightarrow 0} \frac{F(x_i) - F(x_i - \Delta x)}{\Delta x} \lim_{\Delta y \rightarrow 0} \frac{F(y_j) - F(y_j - \Delta y)}{\Delta y}} \right] = \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}, \quad i = \overline{1, m}, \quad j = \overline{1, n}
\end{aligned}$$

Таким чином, індивідуальна взаємна інформація поміж двома неперервними ансамблями задається формулою

$$J(X, Y) = \log \frac{p(X, Y)}{p(X)p(Y)}. \quad (4.44)$$

Ця функція – випадкова величина. Її середнє значення

$$I(X; Y) = MJ(X, Y) = \iint_{-\infty}^{\infty} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (4.45)$$

характеризує повну (середню) взаємну інформацію поміж ансамблями $(X, p(x))$ та $(Y, p(y))$.

Для неперервних ансамблів немає особливої необхідності вводити відповідні ентропії та виражати через них повну взаємну інформацію, як це дається для дискретних ансамблів формулами (4.27). Однак для зручності використання однакових за виглядом формул у різних випадках це іноді роблять.

Формально можна ввести ентропії

$$\begin{aligned}
H^*(X) &= \int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} dx; \\
H^*(Y) &= \int_{-\infty}^{\infty} p(y) \log \frac{1}{p(y)} dy; \\
H^*(X|Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{1}{p(x|y)} dx dy; \\
H^*(Y|X) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{1}{p(y|x)} dx dy, \quad (4.46)
\end{aligned}$$

які називають диференційними. Тоді згідно з (4.45) можна одержати формули, аналогічні (4.27):

$$I(X;Y)=\begin{cases} H^*(X)-H^*(X|Y), \\ H^*(Y)-H^*(Y|X), \\ H^*(X)+H^*(Y)-H^*(X,Y). \end{cases} \quad (4.47)$$

Розглянемо довільні складні ансамблі $(X^q, p(x_1, \dots, x_q))$ та $(Y^q, p(y_1, \dots, y_q))$, де $X^q = X_1 X_2 \dots X_q$; $Y^q = Y_1 Y_2 \dots Y_q$ – прямий добуток q копій відповідно множин X_i та Y_i ; $p(x_1, \dots, x_q)$ та $p(y_1, \dots, y_q)$ – q -вимірні щільності ймовірностей.

Для таких ансамблів аналогічно (4.44) можна ввести індивідуальну взаємну інформацію як випадкову величину

$$I(\bar{X}, \bar{Y}) = \log \frac{P(\bar{X}, \bar{Y})}{P(\bar{X})P(\bar{Y})}. \quad (4.48)$$

Повна взаємна інформація поміж ансамблями X^q та Y^q дорівнює математичному сподіванню (4.48)

$$I(X^q; Y^q) = M \log J(\bar{X}, \bar{Y}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\bar{x}, \bar{y}) \log \frac{p(\bar{x}, \bar{y})}{p(\bar{x})p(\bar{y})} d\bar{x} d\bar{y}. \quad (4.49)$$

Можна також, аналогічно (4.47), для обчислення $I(X^q; Y^q)$ користуватися відповідними диференційними ентропіями

$$I(X^q; Y^q) = \begin{cases} H^*(X^q) - H^*(X^q|Y^q), \\ H^*(Y^q) - H^*(Y^q|X^q), \\ H^*(X^q) + H^*(Y^q) - H^*(X^q, Y^q). \end{cases} \quad (4.50)$$

У загальному випадку, коли неперервний ансамбль характеризується імовірнісною мірою чи функцією розподілу, яка не є абсолютно неперервною, кількість інформації задають більш загальним способом. Зробимо деякі зауваження щодо цього випадку.

Нехай (XY, P_{XY}) – довільний ансамбль з імовірнісною мірою P_{XY} , яка може бути задана також і функцією розподілу $F(x, y)$. Розглядають розбиття X на $D_X = (D_{X_1}, \dots, D_{X_n})$ та $D_Y = (D_{Y_1}, \dots, D_{Y_n})$. Ці розбиття дозволяють одержати дискретний ансамбль $(X_{D_X} Y_{D_Y}, P_{D_X D_Y})$, який задає ансамблі (X_{D_X}, P_{D_X}) та (Y_{D_Y}, P_{D_Y}) .

Взаємну інформацію поміж ансамблями (X_{D_X}, P_{D_X}) та (Y_{D_Y}, P_{D_Y}) визначають таким чином:

$$I(X;Y) = \sup I(X_{D_X}; Y_{D_Y}), \quad (4.51)$$

де

$$I(X_{D_X}; Y_{D_Y}) = \sum_{i,j} P(D_{X_i}, D_{Y_j}) \log \frac{P(D_{X_i}, D_{Y_j})}{P(D_{X_i})P(D_{Y_j})}, \quad (4.52)$$

причому верхня грань береться по всіх розбиттях ансамблю X та всіх розбиттях ансамблю Y .

Функцію $I(X;Y)$ із (4.51) можна розуміти як границю відповідним чином вибраних все більш тонких розбиттів. При цьому для дискретних ансамблів та ансамблів, які мають щільності ймовірностей, з (4.51) одержуються відповідні співвідношення (4.31) та (4.45).

4.3.3. Кодування у неперервних каналах з дискретним часом

Математичні моделі неперервного каналу. Математична модель каналу – це трійка $[X, Y, P_{Y|X}]$, тобто множини вхідних, вихідних сигналів та умовна ймовірність $P(y|x)$. Теорему кодування можна розглядати для конкретних X, Y та типу ймовірнісної міри $P_{Y|X}$. Нехай множини X та Y сигналів на вході та виході каналу – це множини дійсних чисел.

Кажуть, що канал заданий, якщо для довільних $q=1,2,\dots$, $i=0,\pm 1,\pm 2,\dots$ та довільних послідовностей вхідних і вихідних сигналів

$$\begin{aligned} \vec{x}^{(i)} &= (x_1^{(i+1)}, x_2^{(i+2)}, \dots, x_q^{(i+q)}), \\ \vec{y}^{(i)} &= (y_1^{(i+1)}, y_2^{(i+2)}, \dots, y_q^{(i+q)}) \end{aligned}$$

задані умовні функції розподілу

$$F_q(\vec{y}^{(i)} | \vec{x}^{(i)}) = F(y_1^{(i+1)}, \dots, y_q^{(i+q)} | x_1^{(i+1)}, \dots, x_q^{(i+q)}),$$

де i – початок відліку.

Неперервний канал називають каналом без пам'яті, якщо для всіх $q=1,2,\dots$, $i=0,\pm 1,\pm 2,\dots$, $\vec{x} \in X^q$, $\vec{y} \in Y^q$ багатовимірна умовна функція розподілу дорівнює добутку одновимірних функцій розподілу. Такий канал називають стаціонарним, якщо, крім того, одновимірні функції розподілу не залежать від часу, тобто не залежать від початку відліку:

$$F_q(\vec{y}^{(l)}|\vec{x}^{(l)}) = \prod_{j=1}^q F(y_j|x_j). \quad (4.53)$$

Кажуть, що у каналі діє адитивний шум, якщо $y = x + z$, тобто при незалежності випадкових величин X та Z умовна функція розподілу $F(y|x)$ залежить тільки від різниці y та x :

$$F(y|x) = \Phi(y-x).$$

Пропускною здатністю на символ $C_{\text{симв}}$ стаціонарного неперервного каналу з дискретним часом при обмеженні P_c на середню потужність вхідних сигналів та адитивному гаусовому шумі називається максимум взаємної інформації поміж вхідними та вихідними елементарними повідомленнями

$$C_{\text{симв}} = \max_{\Phi(P_c)} I(X;Y). \quad (4.54)$$

Тут максимум береться по всіх розподілах з класу $\Phi(P_c)$ сигналів з обмеженням на середню потужність P_c .

Теорема кодування для неперервного каналу без пам'яті з адитивним гаусовим шумом. Розглядаємо неперервний стаціонарний канал без пам'яті з дискретним часом, в якому діє адитивний гаусів білий шум, статистично незалежний з вхідним сигналом.

Тут вихідний сигнал має вигляд

$$\vec{Y} = \vec{X} + \vec{Z}, \quad (4.55)$$

$$\vec{Y} = (Y_1, \dots, Y_q);$$

де

$$\vec{X} = (X_1, \dots, X_q);$$

$$\vec{Z} = (Z_1, \dots, Z_q)$$

q -вимірні випадкові вектори відповідно вихідної, вхідної послідовності та шуму.

Для стаціонарного каналу без пам'яті при статистичній незалежності сигналу і шуму згідно з умовою (4.53) одержуємо

$$p_Z(\vec{y}|\vec{x}) = p_Z(\vec{y}-\vec{x}) = \prod_{j=1}^q p(y_j - x_j), \quad (4.56)$$

$$p_Z(\vec{z}) = \prod_{j=1}^q p_Z(z_j),$$

де

$$p_Z(z_j) = \frac{1}{\sqrt{2\pi\sigma_{uu}^2}} \exp\left\{-\frac{1}{2} \frac{z_j^2}{\sigma_{uu}^2}\right\}$$

– гаусова щільність імовірності шуму, $\sigma_{uu}^2 = P_{uu}$ – потужність шуму.

Для пропускної здатності каналу вірний такий результат:

Теорема 4.3. Пропускна здатність на символ стаціонарного неперервного каналу без пам'яті з дискретним часом з адитивним гаусовим шумом та обмеженням P_c на середню потужність вхідних сигналів визначається співвідношенням

$$C_{\text{симв}} = \frac{1}{2} \log\left(1 + \frac{P_c}{P_{uu}}\right). \quad (4.57)$$

Дійсно, згідно з (4.54) з урахуванням (4.47) та (4.55) маємо

$$C_{\text{симв}} = \max_{\Phi(P_c)} [H^*(Y) - H^*(Y|X)] = \max_{\Phi(P_c)} [H^*(Y) - H^*(Z)]. \quad (4.58)$$

Обчислимо диференційну ентропію гаусова шуму

$$\begin{aligned} H^*(Z) &= \int_{-\infty}^{\infty} p(z) \log \frac{1}{p(z)} dz = - \int_{-\infty}^{\infty} p(z) \log \left[\frac{1}{\sqrt{2\pi\sigma_{uu}^2}} \exp\left\{-\frac{1}{2} \frac{z^2}{\sigma_{uu}^2}\right\} \right] dz = \\ &= \log \sqrt{2\pi\sigma_{uu}^2} \int_{-\infty}^{\infty} p(z) dz + \int_{-\infty}^{\infty} p(z) \frac{z^2}{2\sigma_{uu}^2} \log e dz = \log \sqrt{2\pi e \sigma_{uu}^2}. \end{aligned} \quad (4.59)$$

При фіксованій величині $H^*(Z)$ для максимізації різниці у (4.58) необхідно знайти максимум диференційної ентропії $H^*(Y)$.

Відомо, що максимум ентропії $H^*(Y)$ по всіх розподілах з класу $\Phi(P_c)$ розподілів з обмеженням по потужності досягається на щільності гаусового розподілу. При цьому, як відомо, якщо сума випадкових величин $Y = X + Z$ гаусова, то кожна складова має гаусів розподіл, тобто у цьому разі і вхідний сигнал має гаусів розподіл.

Диференційна ентропія прийнятого сигналу $H^*(Y)$ у даному випадку

$$H^*(Y) = \int_{-\infty}^{\infty} p(y) \log \frac{1}{p(y)} dy = \log \sqrt{2\pi e (P_c + P_{uu})}, \quad (4.60)$$

$$\text{де } p(y) = \frac{1}{\sqrt{2\pi(P_c + P_{ш})}} \exp\left\{-\frac{1}{2} \frac{y^2}{P_c + P_{ш}}\right\}$$

– щільність імовірності гаусового розподілу прийнятого сигналу.

Остаточно, підставляючи (4.59) та (4.60) у (4.58), одержуємо формулу (4.57), що і завершує доведення теореми:

$$C_{\text{симв}} = \log \sqrt{2\pi e(P_c + P_{ш})} - \log \sqrt{2\pi e P_{ш}} = \frac{1}{2} \log \left(1 + \frac{P_c}{P_{ш}}\right).$$

Обчислення пропускної здатності каналу з негаусовим шумом навіть для стаціонарного каналу без пам'яті – досить складна задача. Однак гаусів шум є найгіршим з точки зору пропускної здатності – величина $C_{\text{симв}}$ для каналу з довільним шумом не менш, ніж (4.57):

$$C_{\text{симв}} \geq \frac{1}{2} \log \left(1 + \frac{P_c}{P_{ш}}\right).$$

Тому співвідношення (4.57) можна розглядати як оцінку знизу для пропускної здатності на символ каналу з негаусовим шумом.

Для розглянутого неперервного каналу, для якого обчислена пропускна здатність на символ (4.57), вірна теорема кодування, аналогічна теоремі Шеннона щодо завадостійкого кодування для дискретного каналу із завадами.

Теорема 4.4. Нехай є неперервний стаціонарний канал без пам'яті з дискретним часом, обмеженням на середню потужність вхідних сигналів, з адитивним гаусовим шумом, який має пропускну здатність на символ $C_{\text{симв}}$. При довільній швидкості коду $\tilde{R} < C_{\text{симв}}$ та довільній як завгодно малій додатній величині δ існує код, що задовольняє обмеження на середню потужність кодових слів, дозволяє кодувати та декодувати всі повідомлення з максимальною ймовірністю помилки, яка не перевищує δ . Якщо $\tilde{R} > C_{\text{симв}}$, то такого коду не існує. Пропускна здатність $C_{\text{симв}}$ каналу обчислюється за формулою (4.57).

Зауважимо, що під швидкістю коду \tilde{R} розуміється величина $\tilde{R} = \frac{\log M}{q}$, яка визначає кількість двійкових символів на повідомлення джерела, тут q – тривалість коду.

4.3.4. Кодування у неперервних каналах з неперервним часом

Для обмірковування питань передавання інформації по неперервному каналу треба, як і раніше, дати спочатку визначення джерела повідомлень та неперервного каналу зв'язку.

Математична модель джерела, як неодноразово підкреслювалось, вводиться так, щоб його означення збігалось з визначенням випадкового процесу на його виході. Неперервний канал заданий, якщо для кожної реалізації $x(t)$ вхідного сигналу $X(t) \equiv x(w, t)$, $w \in W$, $t \in T$ на виході заданий вихідний випадковий процес $Y(t) \equiv y_x(w, t)$ своєю послідовністю скінченновимірних функцій розподілу.

Нехай $X_T(t)$, $t \in T$ – випадковий процес на вході каналу, заданий на континуальній множині T ; $Y_T(t)$ – процес на виході каналу. Тоді можна ввести середню взаємну інформацію $I(X_T(t); Y_T(t))$ і далі пропустити здатність на символ $C_{\text{симв}}$.

$$C_{\text{симв}} = \sup_{T; P[X_T(t)]} \frac{1}{T} I(X_T(t); Y_T(t)),$$

де верхня грань береться по всіх T , а також по всіх розподілах $P[X_T(t)]$ процесів $X_T(t)$, які задовольняють обмеження на середню потужність P_c вхідних сигналів.

Пропускна здатність каналу з адитивним білим гаусовим шумом і скінченною смугою пропускання. Пропускна здатність в одиницю часу пов'язана з пропускною здатністю на символ рівністю $C = VC_{\text{симв}}$ (див. 4.35), де V – швидкість передавання символів. Тут для обчислення пропускної здатності C необхідно обчислити V та знайти $C_{\text{симв}}$.

Обмежуючись рівнем фізичної строгості, дискретизуємо випадковий процес зі скінченною смугою частот згідно з теоремою Шеннона-Котельнікова. У розглядуваних умовах такий сигнал на інтервалі T визначається $N = 2FT + 1$ відліками. Тоді при $2FT \gg 1$ одержуємо швидкість передавання повідомлень.

$$V = \frac{2FT}{T} = 2F. \quad (4.61)$$

Використовуючи співвідношення (4.57) для пропускної здатності на символ неперервного каналу з дискретним часом і враховуючи (4.61), відразу одержуємо

$$C = F \log \left(1 + \frac{P_c}{P_{\text{ш}}} \right), \quad (4.62)$$

де $P_{\text{ш}} = FN_0$, N_0 – спектральна густина потужності гаусова білого шуму.

Формула (4.62) незалежно одержана Н. Вінером та К. Шенноном. Її звичайно називають формулою Шеннона. Вона дає на рівні фізичної строгості значення пропускної здатності неперервного каналу з адитивним білим гаусовим шумом при обмеженні P_c на середню потужність та обмеженні F на смугу пропускання.

Проаналізуємо формулу Шеннона та зробимо відповідні висновки.

1. Згідно з формулою (4.62) відношення сигнал/шум (P_c/P_u) може бути довільним, у тому разі і меншим за одиницю.

2. Величина пропускної здатності C зростає зі збільшенням (P_c/P_u) практично згідно з логарифмічним законом.

3. При зростанні F пропускна здатність C збільшується до деякої границі (рис.4.7), яку можна обчислити за правилом Лопітала:

$$C_\infty = \lim_{F \rightarrow \infty} F \log \left(1 + \frac{P_c}{FN_0} \right) = \lim_{F \rightarrow \infty} \frac{\log \left(1 + \frac{P_c}{FN_0} \right)}{1/F} = \frac{P_c}{N_0} \log e. \quad (4.63)$$

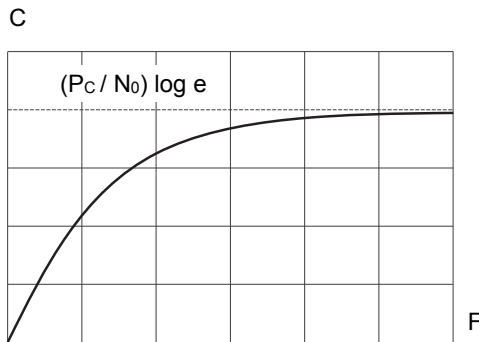


Рис. 4.7. – Пропускна здатність каналу з гаусовим білим шумом при збільшуванні смуги частот

4. Для передавання кожної двійкової одиниці (біта) інформації необхідна деяка енергія сигналу, тобто без енергетичних витрат передавання інформації неможливе.

Дійсно, зі співвідношення (4.63) випливає, що

$$PI(X_T; Y_T) < TC_\infty = T \left(\frac{P_c}{N_0} \log e \right).$$

Тому для передавання 1 біта інформації необхідна енергія сигналу

$$E_c = P_c T > N_0 / \log e = N_0 \ln 2 = 0,693 N_0.$$

Особливості кодування дискретних повідомлень у неперервних каналах з неперервним часом. Неперервний канал можна вважати частиною дискретного каналу і вивчати закономірності передавання повідомлень від дискретних джерел. У цьому випадку вірний результат, аналогічний другій теоремі Шеннона.

Теорема 4.5. Повідомлення кожного дискретного джерела з продуктивністю R' можна закодувати неперервними сигналами та передати по неперервному каналу з пропускною здатністю C з імовірністю помилки менше довільного наперед заданого додатного числа δ , якщо $R' < C$. Це неможливо, якщо $R' > C$. Пропускна здатність обчислюється за формулою (4.62).

Зауважимо, що пропускна здатність дискретного каналу не перевищує пропускну здатність його частини – неперервного каналу ($C_{\text{дискр}} < C_{\text{непер}}$).

У випадку, коли неперервний канал розглядається з джерелом аналогових повідомлень, також правильна аналогічна теорема кодування. Для неї потрібні нові поняття, які ми розглянемо у наступному пункті.

4.3.5. Кодування джерела аналогових повідомлень із заданим критерієм якості

Повідомлення джерел аналогових повідомлень у каналі із завданнями неможливо відтворити абсолютно точно. З другого боку, одержувачу повідомлень достатньо передати їх з деякою похибкою, яка не перевищує допустимої величини. Тому виникає задача кодування джерела за додатковою умовою, що повідомлення приймаються з допустимою похибкою. На відміну від кодування джерела, яке розглянуто у підрозділі 4.1, задача у даному випадку полягає у тому, щоб задати мінімально можливе число символів на повідомлення, при якому похибка не перевищує допустимого рівня. Таку задачу, взагалі кажучи, можна ставити не тільки для неперервних, але й дискретних джерел повідомлення.

Загальні положення щодо кодування джерел аналогових повідомлень із заданим критерієм якості. Кожне джерело за скінченний інтервал часу практично видає скінченну кількість інформації. Тому для одер-

жувача повідомлень деякі повідомлення по суті не відрізняються одне від одного та сприймаються як еквівалентні. Це навідне міркування підказує, що з математичної точки зору нам необхідно тим чи іншим способом ввести поняття еквівалентності повідомлень. Ясна річ, для різних джерел, різних каналів, різних одержувачів повідомлень еквівалентність повідомлень розуміється по-різному. Введемо спочатку відповідну міру еквівалентності, як це роблять у теоретичних дослідженнях, обмежившись неперервним джерелом з дискретним часом. Далі дамо деякі узагальнення.

Нехай джерело U_X у кожному мить часу вибирає повідомлення з множини $X=R$. Множина послідовностей повідомлень $\vec{x}=(x^{(1)}, x^{(2)}, \dots, x^{(q)})$, породжених джерелом U_X , належить $X^q=R^q$. Ці послідовності замінюються скінченним числом «апроксимуючих» послідовностей $\vec{y}_1, \vec{y}_2, \dots, \vec{y}_M$, $\vec{y}_i=(y_i^{(1)}, \dots, y_i^{(q)})$.

Введемо величину похибки

$$d_q(\vec{x}, \vec{y}) = \frac{1}{q} \sum_{j=1}^q d(x^{(j)}, y^{(j)}), \quad (4.64)$$

де $d(x^{(j)}, y^{(j)})$ – критерій якості, який задається відповідно до особливостей конкретної задачі передавання повідомлень. Це співвідношення по-різному використовують для порівняння послідовностей на виході джерела.

Так, введемо ансамбль

$$(X^q Y^q, p(\vec{x}, \vec{y})),$$

де

$$p(\vec{x}, \vec{y}) = p(\vec{x})p(\vec{y}|\vec{x}).$$

Знайдемо середню похибку як математичне сподівання випадкової величини $D_q(\vec{x}, \vec{y})$ на $X^q Y^q$:

$$\bar{d}_q = MD_q(\vec{x}, \vec{y}) = \int \int_{X^q Y^q} p(\vec{x}, \vec{y}) d_q(\vec{x}, \vec{y}) d\vec{x} d\vec{y}. \quad (4.65)$$

Формалізація поняття критерію якості, наприклад, у вигляді (4.65), дає можливість обмірковувати особливості введення поняття ентропії, інформації та кодування повідомлень розглядуваного джерела. Розглянемо спочатку найпростіший випадок стаціонарного джерела без пам'яті.

Епсілон-ентропією джерела називається мінімальна кількість інформації, яка утримується у випадковій величині Y відносно X , при якій вони ще еквівалентні, тобто

$$H(\varepsilon) = \min_{\Phi(\varepsilon)} I(X; Y), \quad (4.66)$$

де $\Phi(\varepsilon)$ — множина одновимірних функцій розподілу, для яких середня похибка не перевищує заданої величини ε .

Співвідношення (4.66) визначає мінімальну кількість інформації, яку необхідно передати по каналу зв'язку для відтворення одного відліку сигналу з точністю, не гіршою за допустиму. Тобто, поняття епсілон-ентропії у відповідній формі об'єднує поняття ентропії та інформації, які були введені раніше.

Запишемо (4.66) у вигляді, необхідному для конкретних обчислень. Взаємна інформація $I(X, Y)$ з урахуванням розуміння випадкових величин X та Y — звичайна взаємна інформація, яка вивчалась у пункті 4.32 для неперервного каналу з дискретним часом. Використовуючи (4.47), запишемо її у вигляді:

$$H(\varepsilon) = \min_{\Phi(\varepsilon)} [H^*(X) - H^*(X|Y)], \quad (4.67)$$

де $H^*(X)$, $H^*(X|Y)$ — відповідні диференційні ентропії.

Це співвідношення дає можливість робити безпосередні обчислювання.

Введемо тепер ε — ентропію у більш загальному випадку. Для цього знадобляться більш спеціальні означення.

Дамо їх.

Кодом для кодування із заданим критерієм якості послідовностей тривалістю q називається довільна множина $T_q = \{\vec{y}_1, \vec{y}_2, \dots, \vec{y}_M\} \in Y^q$ апроксимуючих послідовностей. Кодуванням для коду T_q називається довільне відображення $U(\vec{x})$ множини X^q на T_q . Число

$$\tilde{R} = \frac{\log M}{q}$$

називається швидкістю коду. Кодування називається оптимальним для коду T_q та заданого критерію якості $d(x^{(j)}, y^{(j)})$, якщо мінімізується середня похибка декодування.

$$\bar{d}_q = MD_q(\vec{x}, u(\vec{x})).$$

Відповідно до цього означення швидкість кодування \tilde{R} — це кількість двійкових символів на повідомлення джерела, при якому його повідомлення апроксимуються зі середньою похибкою \bar{d}_q . Код позначається (\tilde{R}, \bar{d}_q) .

Введемо клас $\Phi_q(\varepsilon)$ функцій розподілу $p(\bar{y}|\bar{x})$ такий, що для довільної $p(\bar{y}|\bar{x})$ величина $\varepsilon = \Phi_q(\varepsilon)$ середньої похибки не перевищує ε . Позначимо

$$H_q(\varepsilon) = \min_{\Phi_q(\varepsilon)} \frac{1}{q} I(X^q; Y^q). \quad (4.68)$$

Тоді можна ввести таке означення: ε -ентропією неперервного стаціонарного джерела U_X відносно критерію якості $d(x^{(j)}, y^{(j)})$ називається нижня грань $H_q(\varepsilon)$:

$$H(\varepsilon) = \inf_q H_q(\varepsilon), \quad q=1,2,\dots \quad (4.69)$$

Легко бачити, що звідси безпосередньо отримується для стаціонарного джерела без пам'яті означення ε -ентропії (4.66).

Поняття ε -ентропії найважливіше при вивченні кодування джерела із заданим показником якості. У термінах ε -ентропії доводиться теорема кодування для джерела із заданим показником якості, аналогічна теоремі кодування для дискретного каналу із завадами. Її суть така.

Теорема 4.6. Для кожного $\tilde{R} > H(\varepsilon)$ знайдеться код зі швидкістю \tilde{R} , що кодує послідовність повідомлень тривалістю q , для якого середня похибка відносно критерію якості $d(x^{(j)}, y^{(j)})$ не перевищує ε . При $\tilde{R} < H(\varepsilon)$ для всіх q та для всіх кодів зі швидкістю \tilde{R} середня похибка відносно критерію якості $d(x^{(j)}, y^{(j)})$ перевищує ε .

Епсілон-ентропія стаціонарного гаусова джерела без пам'яті. Розглянемо тут і далі окремі часткові випадки, які найчастіше зустрічаються на практиці.

Стаціонарне гаусове джерело з дискретним часом формує на виході послідовності гаусових незалежних випадкових величин. Без обмеження загальності можна вважати, що вони мають нульові середні значення.

Таке джерело можна задати, впроваджуючи ансамбль повідомлень $(X, p(x))$, де

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{x^2}{2\sigma_X^2}\right\},$$

де $\sigma_X^2 = P_c$ – потужність випадкового процесу на виході джерела.

Введемо квадратичний критерій якості: $d(x,y) = (x-y)^2$. Тоді ε -ентропія стаціонарного джерела визначається формулою (4.66), де $H^*(X)$ – диференційна ентропія гаусового розподілу:

$$H^*(X) = \log \sqrt{2\pi e \sigma_X^2}.$$

Мінімум різниці у (4.66) при цьому досягається, якщо максимізувати $H^*(X|Y)$ у класі $\Phi(\varepsilon)$:

$$H(\varepsilon) = \log \sqrt{2\pi e \sigma_X^2} - \max_{\Phi(\varepsilon)} H^*(X|Y), \quad (4.70)$$

де $\Phi(\varepsilon)$ – клас одновимірних розподілів $p(y|x)$, що задовольняють умову

$$\bar{d} = M(X-Y)^2 = \iint_{XY} (x-y)^2 p(y|x) dx dy \leq \varepsilon.$$

Додатково припустимо, що шум відновлення виявляється, як деяка адитивна випадкова величина Z , яка додається до вихідного сигналу Y , тобто $X = Y + Z$. Внаслідок умови, що X – гаусова величина, величини Y та Z також гаусові. Тоді

$$\max H^*(X|Y) = H^*(Z) = \log \sqrt{2\pi e P_\varepsilon}. \quad (4.71)$$

Тут використаний гаусів розподіл шуму відновлення з дисперсією $\sigma_{ш}^2 = P_\varepsilon$:

$$P(z) = \sqrt{2\pi e P_\varepsilon} \exp\left\{-\frac{z^2}{2P_\varepsilon}\right\}.$$

Згідно з (4.71) співвідношення (4.70) приводиться до виду

$$H(\varepsilon) = \log \sqrt{2\pi P_c} - \log \sqrt{2\pi e P_\varepsilon} = \frac{1}{2} \log \frac{P_c}{P_\varepsilon}. \quad (4.72)$$

Величина $P_c | P_\varepsilon$ у (4.72) характеризує мінімальне відношення сигнал/шум, при якому випадкові величини X та Y залишаються ще еквівалентними.

Рівність (4.72) вірна при $P_\varepsilon \leq P_c$. Можна довести, що у загальному випадку вірна формула

$$H(\varepsilon) = \frac{1}{2} \log \max \left\{ 1, \frac{P_c}{P_\varepsilon} \right\}. \quad (4.73)$$

Епсілон-ентропія гаусова вектора. У більш загальному випадку, коли джерело являє собою джерело довільного гаусова процесу з дискретним часом, повідомлення скінченної тривалості — це реалізація скінченновимірної гаусова вектора.

Розглянемо гаусів вектор $\vec{X} = (X^{(1)}, X^{(2)}, \dots, X^{(q)})^{tr}$ як вектор-стовпець з нульовим середнім. Це, як і раніше, не обмежує загальності.

Щільність імовірності \vec{X} така:

$$p(\vec{x}) = \frac{1}{(2\pi |R_X|)^{q/2}} \exp \left\{ -\frac{1}{2} \vec{x}^{tr} R_X^{-1} \vec{x} \right\}. \quad (4.74)$$

де R_X — кореляційна матриця вектора

R_X^{-1} — матриця, обернена до матриці R_X ;

$|R_X|$ — визначник матриці R_X .

Введемо вектор $\vec{Y} = (Y^{(1)}, Y^{(2)}, \dots, Y^{(q)})^{tr}$, апроксимуючий первинний вектор \vec{X} із заданою середньоквадратичною похибкою $\bar{d} \leq \varepsilon$, де $\bar{d}_q = MD_q(\vec{x}, \vec{y})$, $D_q(\vec{x}, \vec{y})$ — випадкова величина, яка приймає значення $d_q(\vec{x}, \vec{y})$ із (4.64).

Тепер можна ставити задачу знаходження ε -ентропії гаусова вектора. Згідно з (4.68) ε -ентропію можна обчислити за формулою

$$H_q(\varepsilon) = \min_{\Phi_q(\varepsilon)} \frac{1}{q} I(X^q; Y^q) = \min_{\Phi_q(\varepsilon)} \frac{1}{q} I(X_1, \dots, X_q; Y_1, \dots, Y_q),$$

де $\Phi_q(\varepsilon)$ — клас функцій (4.74), для яких середня квадратична похибка не перевищує ε :

$$\bar{d}_q = MD_q(\vec{x}, \vec{y}) = M \frac{1}{q} \sum_{j=1}^q [x^{(j)} - y^{(j)}]^2 \leq \varepsilon.$$

Тобто з математичної точки зору стоїть задача знаходження мінімуму:

$$H_q(\varepsilon) = \min_{\Phi_q(\varepsilon)} \frac{1}{q} I(X_1, \dots, X_q; Y_1, \dots, Y_q)$$

при обмеженні

$$\bar{d}_q = M \frac{1}{q} \sum_{j=1}^q [x^{(j)} - y^{(j)}]^2 = M \frac{1}{q} (\bar{X} - \bar{Y})^T \times (\bar{X} - \bar{Y}) \leq \varepsilon.$$

Для розв'язання цієї задачі можна виконати таке невироджене ортогональне перетворення векторів, щоб координати перетвореного вектора були лінійно незалежними. Це перетворення у власних координатах. Тоді, використовуючи відому властивість інформації

$$I(X;Y) = I(AX;Y) = I(X;BY),$$

де A, B – довільні лінійні невироджені перетворення, а також співвідношення (4.73) для ентропії стаціонарного гаусова джерела без пам'яті, можна одержати остаточне співвідношення:

$$H_q(\varepsilon) = \frac{1}{2q} \sum_{j=1}^q \log \max \left\{ 1, \frac{\lambda_j}{\theta} \right\}. \quad (4.75)$$

Тут θ – корінь рівняння

$$\frac{1}{q} \sum_{j=1}^q \min(\theta, \lambda_j) = \varepsilon,$$

$\lambda_j, j=1, \bar{q}$ – власні числа кореляційної матриці R_X .

Епсілон-ентропія довільного стаціонарного гаусова джерела. Розглянемо більш загальний випадок джерела стаціонарного гаусова процесу спочатку з дискретним, а далі – з неперервним часом. Намітимо шлях одержання відповідних результатів. Для стаціонарного джерела з дискретним часом згідно з означенням (4.69)

$$H(\varepsilon) = \inf_q H_q(\varepsilon), \quad q=1,2,\dots.$$

Тут $H_q(\varepsilon)$ – епсілон-ентропія скінченновимірною вектора, яка у гаусовому випадку згідно з (4.75) дорівнює

$$H_q(\varepsilon) = \frac{1}{2q} \sum_{j=1}^q \log \max \left(1, \frac{\lambda_j^{(q)}}{\theta} \right). \quad (4.76)$$

де $\lambda_j^{(q)}, j=1, \bar{q}$ – власні числа кореляційної матриці відрізка процесу. Підставляючи формулу (4.76) у співвідношенні для $H(\varepsilon)$, та виконуючи граничний перехід, можна одержати такий результат.

Теорема 4.7. Для джерела стаціонарного гаусова процесу з дискретним часом, який має інтегровану спектральну густину потужності, ε -ентропія при квадратичному показнику якості дорівнює:

$$H(\varepsilon) = \frac{1}{2} \int_{-1/2}^{1/2} \log \max \left[1, \frac{G(f)}{\theta} \right] df, \quad (4.77)$$

$$\int_{-1/2}^{1/2} \min[\theta, G(f)] df = \varepsilon.$$

Тут $G(f_j) = \lim_{q \rightarrow \infty} \lambda_j^{(q)}$, $G(f)$ – спектральна густина потужності стаціонарної послідовності.

У загальному випадку стаціонарного гаусова джерела з неперервним часом для обчислення ε -ентропії спочатку вводять ε -швидкість створення інформації відносно квадратичного критерію якості $d_T(x, y)$:

$$H_T(\varepsilon) = \inf_{\Phi_T(\varepsilon)} \frac{1}{2} I(X_T(t); Y_T(t)). \quad (4.78)$$

Нижня грань у (4.78) береться по гаусовим розподілам $\Phi_T(\varepsilon)$ за умовою, що $\overline{d_T}(x, y) \leq \varepsilon$, де $\overline{d_T}(x, y) = MD_T(x, y)$;

$D_T(x, y)$ – випадкова величина з реалізаціями $d_T(x, y)$:

$$d_T(x, y) = \frac{1}{2} \int_T [x(t) - y(t)]^2 dt.$$

Епсілон-ентропія такого джерела вводиться через ε – швидкість створення інформації:

$$H(\varepsilon) = \lim_{T \rightarrow \infty} H_T(\varepsilon). \quad (4.79)$$

Використовуючи розклад процесу $X_T(t)$ на інтервалі T у ряд Карунена-Лоева, можна знайти його коефіцієнти розкладу, які є гауссовими статистично незалежними випадковими величинами. При цьому нову задачу зводять до розглянутої вище. У термінах коефіцієнтів ε -швидкість створення інформації визначають співвідношенням:

$$H_T(\varepsilon) = \inf \frac{1}{T} I(\bar{C}_X; \bar{C}_Y),$$

а ε -ентропія – співвідношенням

$$H(\varepsilon) = \lim_{T \rightarrow \infty} \inf \frac{1}{T} I(\bar{C}_X; \bar{C}_Y).$$

Тут $\vec{C}=(C^{(1)}, C^{(2)}, \dots)$ – послідовність коефіцієнтів розкладу процесу.

Нижня грань у цих співвідношеннях береться по класу гаусових розподілів, для яких C_{X_j} – незалежні гаусові випадкові величини з дисперсіями λ_j .

Після відповідних обчислень можна одержати узагальнення співвідношень (4.77):

$$H(\varepsilon)=\frac{1}{2} \int_{\{f:G_X(f)>0\}} \log \frac{G_X(f)}{\theta} df, \quad (4.80)$$

$$\varepsilon = \int_{\{f:G_X(f)>0\}} df + \int_{\{f:G_X(f)\leq 0\}} G_X(f) df.$$

Тут $G_X(f)$ – спектральна густина потужності стаціонарного процесу $X(t)$.

Висновки

1. Дискретний канал зв'язку без завад розглядається, головним чином, із методичних міркувань. По суті розв'язується задача ефективного кодування джерела дискретних повідомлень. Для розв'язання теоретичних задач вводяться математичні моделі джерела і каналу.

У найпростішому випадку стаціонарного джерела без пам'яті джерело повністю описується ансамблем повідомлень. Ентропія ансамблю визначає міру його невизначеності, з інформаційної точки зору ентропія збігається з середньою інформацією на символ повідомлення. Кількість інформації визначається мірою знятої невизначеності, спричиненої прийманням повідомлення.

Перша теорема Шеннона щодо ефективного кодування стверджує існування за певних умов, коду, який чисельно зменшує середню тривалість кодового слова до ентропії. Таке зменшення середньої тривалості кодового слова (або підвищення швидкості передавання інформації) фізично забезпечується за рахунок зменшення надмірності джерела.

2. У дискретному каналі із завадами вводяться ентропія складних сигналів, середня взаємна інформація, індивідуальна взаємна інформація, доводиться друга теорема Шеннона щодо завадостійкого кодування.

Середня взаємна інформація визначається принаймні трьома засобами – через ансамбль переданих повідомлень, ансамбль прийнятих повідомлень, а також при сумісному розгляданні обох ансамблів.

Друга теорема Шеннона, стверджує існування, за певних умов, оптимального завадостійкого коду, який забезпечує передавання усіх повідомлень джерела по каналу із завадами та їх приймання з якою завгодно малою помилкою декодування. Зменшення помилки декодування фізично забезпечується за рахунок відповідного введення надмірності у кодові слова.

3. Неперервний канал зв'язку використовується як частина дискретного каналу при передаванні повідомлень дискретного джерела, а також при передаванні аналогових повідомлень неперервного джерела.

У першому випадку вводять аналогічно дискретному каналу поняття ентропії, середньої взаємної інформації, індивідуальної взаємної інформації. Коректніше при цьому починати з введення індивідуальної взаємної інформації, через неї одержувати середню взаємну інформацію; при необхідності можна ввести також і ентропії – як відповідні диференційні ентропії. Важливим результатом для цих каналів є формула Шеннона для пропускної здатності гаусового каналу зв'язку.

Для неперервного каналу вірна теорема, аналогічна другій теоремі Шеннона, коли повідомлення дискретного джерела передаються неперервними сигналами.

Якщо передаються аналогові повідомлення, вводяться характеристики, які враховують заданий показник якості (епсілон – швидкість створення інформації, епсілон-ентропія) та доводиться відповідний аналог другої теореми Шеннона.

Запитання, задачі й вправи

1. Що таке математичні моделі джерела повідомлень та каналу зв'язку?

Наведіть приклади математичних моделей джерела.

2. У якому випадку ансамбль повідомлень повністю визначає джерело ?

3. Що таке ентропія дискретного ансамблю і який інформаційний зміст вона має ?

4. Наведіть означення кількості інформації. Що таке один біт інформації?

5. Сформулюйте першу теорему Шеннона (теорему про ефективне кодування джерела).

6. Як визначається кількість інформації у дискретному каналі із завадами? Наведіть основні формули її визначення та обчислення.

7. Що таке середня взаємна інформація, індивідуальна взаємна інформація і як вони пов'язані?

8. Сформулюйте другу теорему Шеннона (теорема про завадостійке кодування).

9. Дайте означення індивідуальної та середньої взаємної інформації для неперервного каналу. Як визначається середня взаємна інформація через відповідні диференційні ентропії ?

10. Запишіть формулу Шеннона для пропускної здатності гаусова каналу і дайте її аналіз.

11. Охарактеризуйте аналоги другої теореми Шеннона для неперервних каналів.

12. У чому особливість кодування джерел аналогових повідомлень (неперервних джерел) ?

13. Що таке епсілон-ентропія? Наведіть формули для обчислення ϵ -ентропії декількох конкретних джерел аналогових повідомлень.

14. Сформулюйте аналог другої теореми Шеннона для випадку кодування неперервних джерел із заданим показником якості.

15. Джерело повідомлень без пам'яті видає символи із ансамблю $X = \{x_i\}$, $i=1,2,3,4$ з імовірностями

$$p(x_1)=0,2 \text{ , } p(x_2)=0,3 \text{ , } p(x_3)=0,4 \text{ , } p(x_4)=0,1.$$

Знайдіть кількість інформації, яка міститься у кожному елементарному повідомленні джерела, обчисліть ентропію та надмірність джерела.

16. Доведіть, що для джерела без пам'яті з об'ємом алфавіту M ентропія досягає максимального значення $H_{\max}(X)=\log M$ при рівноімовірних символах.

17. Закодуйте двійковим кодом Шеннона-Фано ансамбль повідомлень $\{x_i \text{ , } i=1\dots 8\}$, якщо ймовірності символів мають такі значення: $p(x_1)=p(x_2)=1/4$, $p(x_3)=p(x_4)=1/8$, $p(x_5)=p(x_6)=p(x_7)=p(x_8)=1/16$.

Знайдіть середню тривалість кодової комбінації. Покажіть, що такий код близький до оптимального.

18. Закодуйте кодом Хаффмена символи алфавіту ансамблю $\{x_i \text{ , } i=1\dots 9\}$ з імовірностями:

$$p(x_1)=0,2, p(x_2)=0,15, p(x_3)=0,15, p(x_4)=0,12, p(x_5)=0,1,$$

$$p(x_6)=0,1, p(x_7)=0,08, p(x_8)=0,06, p(x_9)=0,04.$$

Обчисліть середню тривалість кодового слова. Покажіть, що код близький до оптимального.

19. Покажіть, що для m -кового симетричного каналу без пам'яті і стирання з ймовірностями переходу

$$P(b_j^* | b_j) = \begin{cases} 1-p, & j=i ; \\ \frac{p}{m-1}, & j \neq i \end{cases}$$

пропускна здатність визначається співвідношенням:

$$C = V \left[\log m + (1-p) \log(1-p) + p \log \frac{p}{m-1} \right].$$

Спростіть цю формулу для двійкового каналу. Побудуйте графік залежності пропускної здатності двійкового каналу від ймовірності помилки.

20. Знайдіть диференційну ентропію гаусова випадкового процесу з математичним сподіванням m та дисперсією σ^2 .

21. Знайдіть максимально можливе значення епсилон-ентропії при заданій середній потужності P_c сигналу та середній потужності шуму спотворень $P_{ш}$.

5. МАТЕМАТИЧНА СТАТИСТИКА. ТЕОРІЯ СТАТИСТИЧНИХ РІШЕНЬ

Теорія ймовірностей вивчає сукупність наслідків експерименту, властивості ймовірнісних мір, заданих на цих наслідках, та різні перетворення цих множин наслідків, а також породжені цими перетвореннями трансформації мір. Існує така відповідність поміж об'єктами реальності та теорією ймовірностей: з одного боку стоять об'єкти реального світу, а з другого – описуючі їх об'єкти теорії ймовірностей. Для їх узгодження необхідна ще одна гілка математики, якою є математична статистика; вона дає теорію вимірювання кількісних характеристик випадкових об'єктів.



Рис. 5.1. – Співвідношення поміж об'єктами реального світу та теорії ймовірностей

Ця теорія встановлює правила і закони, дозволяючи приписати ймовірнісні міри тим чи іншим наслідкам експерименту.

Наведемо деякі приклади класичних задач математичної статистики.

1. Оцінка по наслідках експерименту ймовірності випадкової події.
2. Знаходження по значеннях незалежних випробувань над випадковою величиною її функції розподілу.
3. Знаходження невідомих параметрів заданої функції розподілу.

Тут припускається, що функція розподілу випадкової величини відома, але невідомі декілька її параметрів. Виконуючи послідовні спостереження випадкової величини, треба по результатах експерименту знайти значення цих параметрів.

4. Перевірка статистичних гіпотез.

У цій задачі припускається, що на основі деякого міркування можна вважати, що функція розподілу випадкової величини $\xi \in F(x)$. Питається, чи сумісні спостереження значень випадкової величини ξ з гіпотезою, що ξ дійсно має розподіл $F(x)$?

5. Фільтрація випадкових процесів, коли один із них спостерігається на фоні іншого і його необхідно відокремити.

У теорії електров'язку є ці та інші задачі, де необхідно використовувати методи математичної статистики. Однак для галузі оптимального приймання сигналів (статистичного синтезу приймачів) перш за все необхідні перевірка статистичних гіпотез, оцінювання параметрів розподілу та фільтрація випадкових процесів.

5.1. Постановка задач статистичного синтезу

Канал зв'язку (як математичний об'єкт) вважається заданим, якщо відома трійка $[X, Y, P_{Y|X}]$ – множини X та Y вхідних і вихідних сигналів каналу та ймовірнісна міра $P_{Y|X}$. Задача оптимального приймання сигналів полягає у тому, щоб при заданому каналі найкращім (у деякому розумінні) чином прийняти переданий сигнал.

Математична статистика одержує певні висновки з експериментальних даних. Тому для використання методів математичної статистики завдання каналу недостатньо – для розв'язання задач оптимального приймання сигналів необхідна відповідна додаткова інформація. Перш за все, припускається, що відома реалізація прийнятого сигналу, яка використовується безпосередньо, або у вигляді деяких її відліків. Згідно з умовою завдання каналу припускається, що відома вся множина реалізацій з відповідною ймовірнісною мірою, або випадкові величини – перерізи випадкового процесу. Крім цього, ще необхідна деяка інша інформація, яка обговорена нижче.

5.1.1. Особливості задач статистичного синтезу

Простір реалізацій $y(t)$ процесу $Y(t) = y(\omega, t)$, $\omega \in \Omega$, $t \in T$, що спостерігається, або множину відповідних відліків називають *простором спостережень*. Тобто у загальному випадку цей простір не збігається з множиною реалізацій процесу, спостережуваного на вході приймача (рис.5.2).

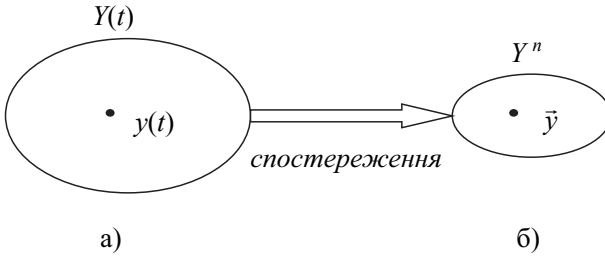


Рис. 5.2. – Ілюстрація відмінності (у загальному випадку) прийнятого сигналу та вектору спостережень: а) простір реалізацій прийнятого сигналу; б) простір спостережень, $\bar{y} = (y_1, \dots, y_n)$, $y_i = y(t_i)$

Як і раніше ω – елементарна подія з простору Ω елементарних подій; T – час спостереження сигналу; множина T – континуальна, коли реалізації процесу $Y(t)$ спостерігаються повністю, і скінченна або зліченна, коли процес спостерігається у дискретних точках. За скінченної множини T вектор $\bar{y} = (y_1, \dots, y_n)$ – *скінченновимірний вектор спостережень*,

$$y_i = y(t_i), y_i \in Y, t_i \in T, i = \overline{1, n},$$

де Y – підмножина скінчено вимірного евклідова простору. Вектор \bar{y} називається *вибіркою розміром n* ; Y^n – *вибірковим простором*.

Вибірка \bar{y} в загальному випадку відрізняється від прийнятого сигналу. Її часто розглядають як незалежну, коли вибіркові значення $y(t_i)$, $i = \overline{1, n}$ вибираються статистично незалежними. Кажуть, що *вибірка однорідна*, коли її елементи y_i мають однаковий розподіл $F(y)$.

При заданому каналі зв'язку відома міра $P_{Y|X}$, тобто розподіл прийнятого сигналу для кожного переданого сигналу, тому припускається відомою імовірнісна міра на вибірковому просторі. Це функція аргументу \bar{y} при відомому \bar{x} . Однак особливість методів математичної статистики полягає в тому, що висновки робляться за деяким фіксованим спостереженням \bar{y} , і тому цю міру розглядають як функцію \bar{x} при відомому \bar{y} . Відповідно функцію в математичній статистиці називають *функцією правдоподібності*. Вводячи замість загального запису переданого сигналу \bar{x} векторний параметр \bar{v} , щіль-

ність імовірності вибірки, яка розглядається як функція аргументу v , мають функцію правдоподібності $W(\bar{y}|\bar{v}), y \in Y^n, \bar{v} \in \theta$. При цьому розглядають два класи функцій правдоподібності: *параметричний*, коли для кожного вектора параметрів \bar{v} мають відому функцію розподілу аргументу \bar{y} , та *непараметричний*, коли така функція для кожного $\bar{v} \in \theta$ не відома.

Серед задач статистичного синтезу найважливішими для теорії електрозв'язку є такі три: *перевірка статистичних гіпотез* (коли відносно характеристик розподілу й імовірностей висувуються несумісні гіпотези H_0, H_1, \dots, H_m і за вектором спостережень вибирається одна з них); *оцінювання параметрів розподілу*; *фільтрація повідомлення з прийнятою реалізацією сигналу*.

Крім вибіркового простору Y^n та функції правдоподібності $W(\bar{y}|\bar{v})$ у математичній статистиці вводять також *простір Γ рішень*, з якого в результаті спостереження вибираються деякі рішення. У задачі перевірки гіпотез простір Γ містить $M+1$ елемент:

$$\gamma_i \in \Gamma, i = \overline{0, M};$$

у задачах оцінювання параметрів розподілів простір рішень збігається з простором параметрів:

$$\bar{\gamma} \in \bar{v}^* \in \theta = \Gamma,$$

де \bar{v}^* — оцінка параметра \bar{v} .

Задачі фільтрації сигналів характеризуються тим, що простір рішень — це простір функцій часу, спостережуваних на скінченному чи нескінченному інтервалі T ; він збігається з простором переданих повідомлень:

$$\gamma(t) = s^*(t) \in S^* = \Gamma,$$

де $s^*(t)$ — оцінка переданого повідомлення.

З геометричної точки зору вибір рішення означає відображення простору спостережень у простір рішень за допомогою вирішальної функції $D(\bar{y})$ (рис. 5.3, а).

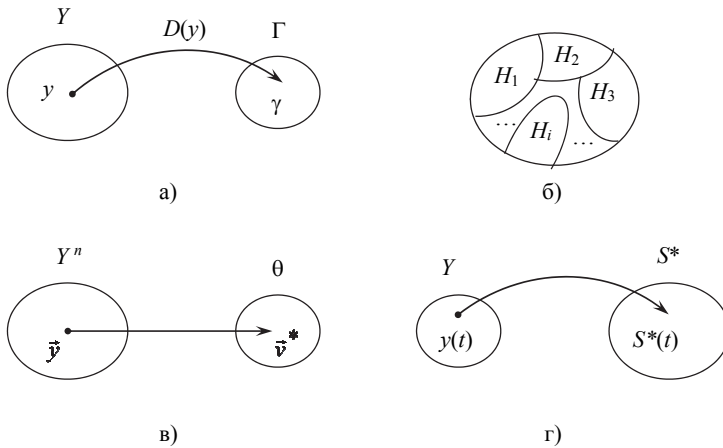


Рис. 5.3. – Прийняття рішення: а) геометрична трактовка прийняття рішення; б) рішення у задачах перевірки гіпотез; в) рішення у задачах оцінювання параметрів; г) рішення у задачах фільтрації

У задачах перевірки гіпотез $H_i, i=\overline{0, M}$, прийняття рішення геометрично означає розбиття простору спостережень на $M+1$ область, що не перетинаються (рис. 5.3, б):

$$Y_i \in Y^n, i=\overline{0, M}, \bigcup_{i=0}^M Y_i = Y^n, Y_i \cap Y_j = \emptyset, i \neq j.$$

У цій задачі i -те рішення приймається, коли вектор спостережень потрапляє у область Y_i . При оцінюванні параметра розподілу за спостереженням \bar{y} з простору Y^n знаходиться оцінка параметра \bar{v}^* , що належить простору параметрів θ (рис.5.3, в). У задачах фільтрації (рис. 5.3, г) за прийнятою реалізацією $y(t)$ знаходиться оцінка $s^*(t)$ переданого повідомлення з простору s^* . У математичній статистиці крім простору спостережень Y^n та функції правдоподібності $W(\bar{y}|\bar{v})$ до апріорної інформації відносять також так звану *функцію втрат*, яка характеризує «платню» за правильне чи помилкове рішення. Вона задається для кожної пари . Прийняте рішення – істинне твердження. Для задач перевірки гіпотез $H_i, i=\overline{0, M}$ – це матриця втрат $B = [B_{ij}]_{i,j=0}^M$, для задач оцінювання параметрів – функція втрат

$B=(\bar{v}^*, \bar{v})$. Елемент B_{ij} матриці B означає «платню» за вибирання гіпотези H_i , коли істинна гіпотеза H_j . Невід’ємна функція $B=(\bar{v}^*, \bar{v})$ означає «платню» за вибирання оцінки \bar{v}^* , коли істинне значення параметра дорівнює \bar{v} . Для того, щоб порівняти рішення, у математичній статистиці вибирають ті чи інші, як кажуть, *критерії (чи показники) якості* – критерії якості правила вибору рішень.

5.1.2. Постановка задачі перевірки вибирання гіпотез

Для постановки задачі перевірки обирання гіпотез необхідно задати відповідну апріорну інформацію, об’єм якої може бути різним для різних показників (критеріїв) якості. Розглянемо декілька з них, які найчастіше використовуються при статистичному синтезі.

Показники (критерії) якості рішень

1. Показник середнього ризику.

У задачах перевірки гіпотез $H_i, i=0, M$ має бути задана матриця втрат $B=[B_{ij}]_{i,j=0}^M$. Тут ще припускається, що крім заданого каналу

$[X, Y, P_{X|Y}]$, відома також імовірнісна міра P_X . При цьому виявляються відомими ймовірності значень дискретного параметра v , або що те саме, ймовірності гіпотез – $P(H_i), i=0, M$, коли значення параметра ототожнюється з гіпотезою.

Середній ризик вводиться як математичне сподівання матриці втрат: $R=M[B_{ij}] = \sum_{i=0}^M \sum_{j=0}^M B_{ij} P\{H_i \cap H_j\} = \sum_{i=0}^M \sum_{j=0}^M B_{ij} P(H_j) P\{\bar{y} \in Y_i | H_j\}$.

Враховуючи, що імовірності $P\{\bar{y} \in Y_i | H_j\}$ можна обчислити через функцію правдоподібності

$$P\{\bar{y} \in Y_i | H_j\} = \int_{Y_i} W(\bar{y} | H_j) d\bar{y},$$

остаточно маємо

$$R = \sum_{i=0}^M \sum_{j=0}^M B_{ij} P(H_j) \int_{Y_i} W(\bar{y} | H_j) d\bar{y}. \quad (5.1)$$

2. Показник середньої ймовірності помилки.

Середній ризик враховує як помилки, коли номер рішення γ_j не збігається з номером істинної гіпотези H_j , так і правильні рішення, коли $i=j$. В окремому випадку, коли матриця втрат проста –

$B_{ij}=1-\delta_{ij}$, де δ_{ij} – символ Кронекера, з (5.1) одержуємо імовірність середньої помилки

$$P_{ном} = \sum_{i=0}^M \sum_{\substack{j=0 \\ j \neq i}}^M P(H_j) \int_{Y_i} W(\bar{y}|H_j) d\bar{y}. \quad (5.2)$$

Замість $P_{ном}$ можна використовувати еквівалентний показник якості – ймовірність правильного рішення:

$$P_{np} = 1 - P_{ном} = 1 - \sum_{i=0}^M \sum_{j=0}^M P(H_j) \int_{Y_i} W(\bar{y}|H_j) d\bar{y} = \sum_{j=0}^M P(H_j) \int_{Y_j} W(\bar{y}|H_j) d\bar{y}. \quad (5.3)$$

3. Показник апостеріорної ймовірності гіпотези.

Матриця втрат B – це додаткова апіорна інформація, що може бути не задана. У цьому разі раціонально вибирати критерій, в якій вона не фігурує. Це може бути апостеріорна імовірність гіпотези $P(H_i|\bar{y})$, що обчислюється за формулою Байєса:

$$P(H_i|\bar{y}) = \frac{W(\bar{y}|H_i)P(H_i)}{\sum_{i=0}^M W(\bar{y}|H_i)P(H_i)}. \quad (5.4)$$

Використовують і інші показники. Досить часто за показник якості приймають саму функцію правдоподібності, коли невідома апіорна інформація, щодо імовірностей гіпотез.

Розглянуті та інші показники якості використовують для формулювання критеріїв оптимальності рішень.

Критерії оптимальності рішень

Наведемо критерії оптимальності рішень, які базуються на розглянутих і інших показниках їх якості.

1. Байєсівський критерій оптимальності

Критерієм оптимальності є вимога мінімізації середнього ризику (5.1) або у загальному випадку – забезпечення нижньої точної грані:

$$R_B = \min_D R = \min_D \sum_{i=0}^M \sum_{j=0}^M B_j P(H_j) \int_{Y_i} W(\bar{y}|H_j) d\bar{y}. \quad (5.5)$$

Рішення – це гіпотеза H_j , що забезпечує мінімум ризику. Останній шукається у множині $D(\bar{y})$ відображень простору спостережень Y^n у простір рішень Γ . Нагадаємо, що аргумент функції правдопо-

дібності — це значення параметра v (або номер гіпотези). Тому (5.5) записують також у вигляді

$$D(\bar{y})_B = \arg \min_D R. \quad (5.5a)$$

2. Критерій мінімуму середньої імовірності помилки (критерій Зігерта-Котельникова, або критерій ідеального спостерігача).

У цьому разі використовується показник якості рішень (5.2). Критерій оптимальності вимагає мінімізації величини середньої імовірності помилки:

$$R_{зк} = \min_D P_{ном} = \min_D \sum_{i=0}^M \sum_{j=0, j \neq i}^M P(H_j) \int_{Y_i} W(\bar{y}|H_j) d\bar{y}, \quad (5.6)$$

або

$$D(\bar{y})_{зк} = \arg \min_D P_{ном}. \quad (5.6a)$$

Назва критерію оптимальності пояснюється тим, що його використовував В.А. Котельников у «Теорії потенційної завадостійкості» для знаходження оптимальних приймачів систем зв'язку; критерій називають також критерієм «ідеального спостерігача», тому що можна уявити собі, що деякий спостерігач задає вагову матрицю $[B_{ij}]$ так, що її члени завжди нульові $B_{ij} = 0$, коли приймається правильне рішення, а коли виникає помилка він не цікавиться тим, як саме вона виникла, і завжди задає однаковий ваговий коефіцієнт $B_{ij} = 1, i \neq j$.

Критерій мінімуму імовірності помилки є окремим випадком

байєсівського критерію мінімуму середнього ризику за простої функції (матриці) втрат

$$B_{ij} = 1 - \delta_{ij}.$$

Іноді зручніше використовувати замість $\min_D P_{ном}$ еквівалентний критерій максимуму імовірності правильного рішення:

$$\max_D P_{np} = \max_D \sum_{j=0}^M P(H_j) \int_{Y_i} W(\bar{y}|H_j) d\bar{y}. \quad (5.7)$$

3. Критерій максимуму апостеріорної ймовірності.

Згідно з показником якості (5.4), критерій оптимальності рішень задається так: серед гіпотез $H_j, j = \overline{0, M}$ вибирається такий номер « i », що забезпечує максимум у (5.4):

$$i = \arg \max_{0 \leq j \leq M} P(H_j | \bar{y}) = \arg \max_j \frac{W(\bar{y} | H_j) P(H_j)}{\sum_{j=0}^M W(\bar{y} | H_j) P(H_j)}. \quad (5.8)$$

4. Мінімаксний критерій оптимальності.

Введені вище критерії оптимальності рішень по суті вимагали задання розподілу P_X переданого сигналу, що дає змогу ввести ймовірності гіпотез $P(H_j)$. Коли розподіл P_X невідомий, можна врахувати найгірший випадок – мінімізувати середній ризик в умовах найгіршого (з точки зору величини ризику) розподілу:

$$\max_{0 \leq j \leq M} \min_D R = \max_{0 \leq j \leq M} \min_D \sum_{i=0}^M \sum_{j=0}^M B_{ij} P(H_j) \int_{Y_i} W(\bar{y} | H_j) d\bar{y}. \quad (5.9)$$

Можна довести, що рішення буде таке саме якщо використовувати умовні ризики $r_i = \sum_{i=0}^M B_{ij} \int_{Y_i} W(\bar{y} | H_j) d\bar{y}$ та вимагати щоб рішення шукалось з умови:

$$\min_D \max_{0 \leq j \leq M} \sum_{i=0}^M B_{ij} \int_{Y_i} W(\bar{y} | H_j) d\bar{y}. \quad (5.9a)$$

Мінімаксний критерій оптимальності приводить до байєсівського рішення в умовах найгіршого розподілу параметра (переданого сигналу).

5. Критерій оптимальності Неймана-Пірсона.

Стисло розглянемо двоальтернативну задачу. Яку називають задачею перевірки гіпотез. Гіпотезу H_0 називають основною, а H_1 – альтернативною. Ставиться задача перевірки гіпотези H_0 проти альтернативи H_1 . Часто гіпотези несиметричні і зручно основну увагу приділяти одній із них. Саме таку гіпотезу називають основною і позначають H_0 .

У задачі перевірки гіпотези H_0 проти альтернативи H_1 мають місце дві помилки

$$\alpha = P(H_1 | H_0) = \int_{Y_1} W(\bar{y} | H_0) d\bar{y}$$

та

$$\beta = P(H_0 | H_1) = \int_{Y_0} W(\bar{y} | H_1) d\bar{y}.$$

У прикладній літературі застосовують фізичну термінологію (взяту із галузі радіолокації). Ситуація, коли приймається гіпотеза H_1 за істинної гіпотези H_0 , означає, що дійсно сигналу нема (існує тільки шум) але приймається рішення про існування сигналу. Тому α називають умовною ймовірністю хибної тривоги. У математичній статистиці її називають умовною ймовірністю помилки першого роду. У разі коли приймається гіпотеза H_0 при істинній гіпотезі H_1 , фізично сигнал існує (тобто радіолокаційна ціль присутня), але приймається хибне рішення, що сигналу нема. Тому β називають умовною ймовірністю пропуску цілі; у математичній статистиці її називають умовною ймовірністю помилки другого роду.

Крім ймовірностей помилок α та β у задачі перевірки гіпотези H_0 проти альтернативи H_1 розглядають також ймовірності правильних рішень

$$P(H_1|H_1) = 1 - \beta,$$

$$P(H_0|H_0) = 1 - \alpha.$$

Критерій оптимальності рішень Неймана-Пірсона використовує два показника якості рішень – умовні ймовірності хибної тривоги та пропуску цілі.

Критерій Неймана-Пірсона вимагає знаходження рішення, що забезпечує мінімальне значення умовної ймовірності пропуску цілі

$$\beta_{\text{НП}} = \min_D \beta = \min_D \int_{Y_0} W(\bar{y}|H_1) d\bar{y}, \quad (5.10)$$

за умову $\alpha \leq \alpha_0$, тобто при обмеженні умовної ймовірності хибної тривоги величиною α_0 .

Замість (5.10) часто використовують умову максимізації ймовірності правильного рішення про наявність цілі

$$\max_D (1 - \beta) \quad (5.10a)$$

при обмеженні $\alpha \leq \alpha_0$.

5.1.3. Постановка задачі оцінювання параметрів розподілів

Задачі оцінювання параметрів розподілів ймовірностей має деякі відмінності порівняно із задачею перевірки гіпотез. Різниця у тому, що параметр функції правдоподібності $W(\bar{v}|\bar{y})$ у задачах вибору гіпотез має дискретний характер (і значення параметра ототожнюються з гіпотезами), а в задачах оцінювання параметрів він звичайно набирає

значення з континуальної множини. Це відбивається як на вигляді показників якості рішення, так і на вигляді критеріїв оптимальності. Спинимося на них.

Показники (критерії) якості рішень

1. Показник середнього ризику.

Середній ризик – це середнє значення функції втрат:

$$R = M\{B(\bar{v}^*; \bar{v})\} = \int_{Y^n \theta^m} B[\bar{v}^*(\bar{y}), \bar{v}] W(\bar{v}, \bar{y}) d\bar{v} d\bar{y} = \int_{Y^n \theta^m} B[\bar{v}^*(\bar{y}), \bar{v}] W(\bar{y}|\bar{v}) W(\bar{v}) d\bar{v} d\bar{y}. \quad (5.11)$$

Тут припускається, що вимірність m вектора параметрів \bar{v} у загальному випадку не збігається з вимірністю n вектора спостережень \bar{y} .

2. Показник середньоквадратичної похибки.

В окремому випадку середньоквадратичної функції втрат

$$B[\bar{v}^*(\bar{y}), \bar{v}] = [\bar{v}(\bar{y}) - \bar{v}]^2 = \varepsilon^2$$

середній ризик приводить до середньоквадратичної похибки оцінювання скалярного параметру

$$R = M\varepsilon^2 = M[\bar{v}^*(\bar{y}) - \bar{v}]^2 = \int_{Y^n \theta^n} [\bar{v}^*(\bar{y}) - \bar{v}]^2 W(\bar{y}|\bar{v}) W(\bar{v}) d\bar{v} d\bar{y}. \quad (5.12)$$

Величина цієї похибки і використовується як показник якості рішення.

3. Показник апостеріорної щільності ймовірностей.

Для завдання цього показника якості використовують відповідну формулу Байеса. Аналогічно (5.4),

$$W(\bar{v}|\bar{y}) = \frac{W(\bar{y}|\bar{v})W(\bar{v})}{\int_{\theta^m} W(\bar{y}|\bar{v})W(\bar{v})d\bar{v}}. \quad (5.12)$$

Існують й інші показники якості. Наведені показники якості рішень дають змогу ввести відповідні критерії оптимальності.

Критерії оптимальності рішень

1. Байєсівський критерій оптимальності.

Аналогічно (5.5), байєсівський критерій оптимальності характеризується умовою мінімізації середнього ризику (5.11):

$$R_B = \min_D R = \min_D \int_{\gamma^n} \int_{\theta^m} B[\bar{v}^*(\bar{y}), \bar{v}] W(\bar{y}|\bar{v}) W(\bar{v}) d\bar{v} d\bar{y}. \quad (5.13)$$

Враховуючи, що

$$W(\bar{y}|\bar{v}) W(\bar{v}) = W(\bar{v}|\bar{y}) W(\bar{y}) = W(\bar{y}, \bar{v}),$$

співвідношення (5.13) можна записати так:

$$R_B = \min_D \int_{\gamma^n} \int_{\theta^m} B[\bar{v}^*(\bar{y}), \bar{v}] W(\bar{v}|\bar{y}) W(\bar{y}) d\bar{y} d\bar{v} = \min_D \int_{\gamma^n} W(\bar{y}) d\bar{y} \int_{\theta^m} B[\bar{v}(\bar{y}), \bar{v}] W(\bar{v}|\bar{y}) d\bar{v}.$$

Можна довести, що оцінка, яка мінімізує функціонал

$$I[\bar{v}^*(\bar{y}), \bar{v}] = \int_{\theta^m} B[\bar{v}^*(\bar{y}), \bar{v}] W(\bar{v}|\bar{y}) d\bar{v},$$

мінімізує також і середній ризик R , що має назву апостеріорного ризику.

2. Критерій мінімізації середньоквадратичної похибки.

Тут вимагається мінімізація величини похибки $M\varepsilon^2$:

$$\min_D M\varepsilon^2 = \min_D \int_{\gamma^n} \int_{\theta^m} [\bar{v}^*(\bar{y}) - \bar{v}]^2 W(\bar{y}|\bar{v}) W(\bar{v}) d\bar{v} d\bar{y}.$$

3. Критерій максимуму апостеріорної щільності ймовірності.

У задачі оцінювання параметрів цей критерій набирає такого вигляду:

$$\bar{v}_{м.а.щ}^* = \arg \max_{\bar{v} \in \theta^m} W(\bar{v}|\bar{y}) = \arg \max_{\bar{v} \in \theta^m} \frac{W(\bar{y}|\bar{v}) W(\bar{v})}{\int_{\theta^m} W(\bar{y}|\bar{v}) W(\bar{v}) d\bar{v}}. \quad (5.15)$$

Оцінка $\bar{v}_{м.а.щ}^*$ має назву оцінки максимальної апостеріорної щільності ймовірності оцінювання параметра.

Аналогічно задачі вибору гіпотез можна розглядати мінімаксний критерій, критерій максимальної правдоподібності тощо.

У теорії оцінювання параметрів розподілів важливі якісні характеристики одержуваних оцінок, основними з яких є: незсуненість, ефективність, обґрунтованість.

Оцінка, математичне сподівання якої за будь якого значення параметра збігається з істинним значенням параметра

$$M\bar{v}(\omega) = \bar{v}, \quad (5.16)$$

називається *незсуненою*.

Нагадаємо, що оцінка – це функція спостереження $\bar{v}^* = \bar{v}^*(\bar{y})$. На множені спостережень Y^n задана імовірнісна міра і тому можна розглядати одержувану оцінку \bar{v}^* як реалізацію випадкової величини $\bar{v}^*(\omega)$. Відповідно можна знаходити її математичне сподівання (5.16).

Для порівняння різних оцінок вводять ту чи іншу міру розкиду. Так, для скалярного параметра v використовують другий момент $M(v^*(y) - v)^2 = m_2$. Якщо оцінка незсунена, ця величина збігається з дисперсією D_{*v^*} .

Оцінка v_1^* більш ефективна, порівняно з оцінкою v_2^* , якщо

$$M(v_1^* - v)^2 < M(v_2^* - v)^2.$$

Введемо нижню грань $\inf M(v^* - v)^2$ для всіляких оцінок v^* . Оцінка v^* називається *ефективною*, якщо для неї досягається нижня грань:

$$M(v^* - v)^2 = m_{2\min}. \quad (5.17)$$

Нарешті, оцінка v^* параметра \bar{v} називається *обґрунтованою*, якщо за умови $n \rightarrow \infty$ вона збігається за імовірністю з \bar{v} , тобто якщо

$$\bar{v}^* \xrightarrow[n \rightarrow \infty]{P} \bar{v}. \quad (5.18)$$

5.1.4. Постановка задачі фільтрації повідомлень

У теорії електрозв'язку розглядаються особливості передавання різних повідомлень (дискретних, аналогових) різними методами. При передаванні дискретних повідомлень чи при передаванні аналогових повідомлень з використанням цифрових та аналогових методів модуляції задачі приймання сигналів можна розв'язувати методами теорії перевірки (вибору) гіпотез та оцінювання параметрів. Проте, у загальному випадку, передавання аналогових повідомлень з використанням аналогових методів модуляції цих методів недостатньо. Необхідно використовувати більш спеціальні методи – методи фільтрації повідомлень.

Прийнятий сигнал $u(t)$ у загальному випадку може бути описаний операторним рівнянням

$$y = V[Us, \xi], \quad (5.19)$$

де $s(t)$ – повідомлення, U_S – переданий сигнал (U – оператор, формуючий сигнал), $\xi(t)$ – завада. Залежно від методів модуляції та завадових умов залежність прийнятого сигналу від переданого повідомлення може бути різною.

У лінійних методах модуляції в умовах адитивних завад можна розглядати найпростішу задачу адитивної взаємодії $s(t)$ та $\xi(t)$:

$$y(t) = s(t) + \xi(t). \quad (5.19a)$$

У нелінійних методах модуляції ця взаємодія більш складна.

Задача фільтрації повідомлень полягає у тому, що за прийнятою реалізацією $y(t)$ випадкового сигналу $Y(t) = y(t, \omega)$ необхідно знайти оцінку $S^*(t)$ реалізації переданого повідомлення $S(t) = s(t, \omega)$, $\omega \in \Omega$; оцінка має бути оптимальною у тому чи іншому розумінні.

Як і в задачах перевірки гіпотез та оцінювання параметрів, насамперед необхідно ввести показники (критерії) якості рішень та критерії їх оптимальності.

Якщо розв'язувати задачу фільтрації при кожному фіксованому значенні часу t , можна використовувати всі показники, введені в задачах оцінювання параметрів: середнього ризику, середньоквадратичної похибки, показник правдоподібності тощо. Проте, оцінювання повідомлення виконується на деякому фіксованому інтервалі часу і тому іноді придатними будуть інші показники, що враховують якість відновлення реалізації у цілому.

Найчастіше використовують показник середньоквадратичної похибки, що вводиться для кожного t на інтервалі T :

$$\varepsilon^2(t) = M[S(t) - S^*(t)]^2, \quad t \in T. \quad (5.20)$$

Використання цього показника якості рішення вимагає завдання каналу $[X, Y, P_{X|Y}]$ та міри P_X (чи P_S):

$$\varepsilon^2(t) = \iint_{YS} [S(t) - S^*(t)]^2 W(y, s) ds dy, \quad (5.21)$$

де

$$W(y, s) = W(y|s)W(s) = W(s|y)W(y); \quad S^*(t) = S^*[y(t)].$$

Співвідношення (5.21) можна також подати так:

$$\varepsilon^2(t) = \iint_{YS} [S(t) - S^*(t)]^2 W(s|y)W(y) ds dy. \quad (5.21a)$$

Відповідно, критерій оптимальності рішення задається у вигляді мінімізації похибки (5.21а):

$$\min_{S^*(t)} \varepsilon^2(t) = \min_{S^*(t)} \int \int_{\mathcal{Y}_S} \{ [S(t) - S^*(t)]^2 W(s|y) ds \} W(y) dy. \quad (5.22)$$

Мінімум $\varepsilon^2(t)$ забезпечується, якщо мінімізувати функціонал

$$I = \int_S [S(t) - S^*(t)]^2 W(s|y) ds, \quad (5.23)$$

тобто якщо забезпечити мінімізацію середньоквадратичної похибки при кожному спостереженні $y(t)$. Результат рішення – оцінка $S^*(t)$ переданого повідомлення.

Найпростіша задача (5.19а) приводить до лінійної фільтрації, загальний випадок (5.19) – до нелінійної фільтрації. Перші фундаментальні результати з теорії лінійної фільтрації (для процесів з дискретним часом) дістав А.Н. Колмогоров у 1939 році, а для процесів з неперервним часом – Н. Вінер у 1942 році. Досить завершена теорія лінійної фільтрації гаусових процесів одержана у роботах американських вчених Р.Є. Калмана та Р.С. Бьюсі у 60-ті роки. Фундаментальні результати із теорії нелінійної фільтрації марківських процесів дістав російський вчений Р.Л. Стратанович (починаючи з 1959 року).

Отже, ми розглянули показники якості та критерії оптимальності для, так званих, однокрокових методів статистичних рішень. Вони характеризуються тим, що час спостереження фіксований і за одержаним спостереженням остаточно приймається те чи інше рішення. Проте також існують багатокрокові методи статистичних рішень. Для них характерно, що час спостережень не є фіксованим. Наприклад, у задачі перевірки гіпотези H_0 проти альтернативи H_1 на кожному кроці спостереження або вибирається одна з гіпотез H_0 та H_1 , якщо спостереження достатньо для остаточного вибору гіпотези, або приймається рішення, щодо продовження спостереження. Серед цих, так званих послідовних методів обробки сигналів, що мінімізують час спостереження при обмеженнях на ймовірності похибок першого та другого роду, класичним є алгоритм Вальда. Далі ми розглядатимемо насамперед однокрокові методи статистичних рішень.

5.2. Перевірка (вибирання) гіпотез

Фізичні задачі виявлення та розрізнення сигналів в електров'язку математично формуються як задачі перевірки гіпотез, щодо параметрів розподілу.

Найпростішим з математичної точки зору є випадок приймання сигналів, коли сигнал задано, тобто відома трійка $[X, Y, P_{Y|X}]$, а також задана міра P_X . При цьому, як обговорено у підрозділі 5.1, можна вважати, що відомі функції правдоподібності та імовірності гіпотез: $W(\vec{y}|H_i)$, $P(H_i)$, $i = 1, M-1$.

5.2.1. Перевірка гіпотези проти однієї альтернативи

Припускається, що відома апріорна інформація, необхідна для використання критерію середнього ризику: функція правдоподібності $W(\vec{y}|H_i)$, імовірності гіпотез $P(H_i) = P_i$, $i = 0, 1$, та матриця втраг $B = [B_{ij}]$, $i, j = 0, 1$. Спостерігається реалізація процесу $y(t)$, що зображується у вигляді вибірки $\vec{y} = (y_1, \dots, y_n)$ об'єму n . На основі цього спостереження висувається гіпотеза H_0 проти альтернативи H_1 . Якщо гіпотеза H_i , $i = 0, 1$, повністю визначає функцію правдоподібності $W(\vec{y}|H_i)$, вона називається *простою*. Коли це не так, вона називається *складною* (наприклад, вибірка може характеризувати сім'ю функцій правдоподібності).

Розглянемо спочатку задачу перевірки простої гіпотези H_0 проти простої альтернативи H_1 , оптимізуючи спочатку рішення за байєсівським критерієм оптимальності.

Байєсівський критерій оптимальності. За правилом вибору рішень простір Y^n розбивається на дві області – Y_0 та Y_1 : $Y_0 \cup Y_1 = Y^n$. Основна гіпотеза H_0 відкидається кожного разу, коли точка \vec{y} попадає в область Y_1 , і приймається, якщо $\vec{y} \in Y_0$. Відповідно до байєсівського критерію оптимальності (5.5) для знаходження оптимального правила вибору рішень необхідно мінімізувати середній ризик:

$$R = \sum_{i=0}^1 \sum_{j=0}^1 B_{ij} P_i \int_{Y_i} W(\vec{y}|H_j) d\vec{y}, \quad (5.24)$$

де
$$B = [B_{ij}] = \begin{bmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{bmatrix};$$

$B_{01} > B_{00} \geq 0; B_{10} > B_{11} \geq 0; B_{00}, B_{11}$ – «платня» за правильні рішення; B_{01}, B_{10} – «платня» за помилки.

Ураховуючи, що об'єднання областей Y_0 та Y_1 дає весь вибірковий простір і, відповідно одну з них можна зобразити у вигляді різниці всього простору Y^n та другої області, співвідношення для середнього ризику (5.24) не важко привести до такого вигляду:

$$R = P_0 B_{00} + P_1 B_{01} - \int_{Y_1} [P_1(B_{01} - B_{11})W(\bar{y}|H_1) - P_0(B_{10} - B_{00})W(\bar{y}|H_0)]d\bar{y}. \quad (5.24a)$$

Досить просто доводиться, що мінімум R досягається тоді й тільки тоді, коли підінтегральна функція в (5.24a), тобто співвідношення у прямих дужках, невід'ємна:

$$P_1(B_{01} - B_{11})W(\bar{y}|H_1) - P_0(B_{10} - B_{00})W(\bar{y}|H_0) \geq 0. \quad (5.25)$$

Середній ризик мінімізується, коли область Y_1 прийняття гіпотези H_1 містить у собі всі точки \bar{y} , для яких (5.25) приймає невід'ємні значення, а область Y_0 прийняття гіпотези H_0 – всі точки \bar{y} , для яких (5.25) приймає невід'ємні значення.

З (5.25) одержуємо такий алгоритм прийняття рішення:

$$P_1(B_{01} - B_{11})W(\bar{y}|H_1) - P_0(B_{10} - B_{00})W(\bar{y}|H_0) \underset{\gamma_0}{\overset{\gamma_1}{\geq}} 0. \quad (5.26)$$

Впроваджуючи відношення правдоподібності:

$$I(\bar{y}) = \frac{W(\bar{y}|H_1)}{W(\bar{y}|H_0)}, \quad (5.27)$$

байєсівське правило (5.26) можна записати також у вигляді

$$I(\bar{y}) \underset{\gamma_0}{\overset{\gamma_1}{\geq}} I_{пор}, \text{ де } I_{пор} = \frac{P_0(B_{10} - B_{00})}{P_1(B_{01} - B_{11})} - \text{поріг відношення правдоподібнос-}$$

ті. У багатьох випадках зручно замість функції $I(\bar{y})$ користуватися її логарифмом (який, як і кожна строго монотонна функція, не змінює знака нерівності):

$$\ln I(\bar{y}) \underset{\gamma_0}{\overset{\gamma_1}{\geq}} \ln I_{пор}.$$

Таким чином, байєсівський алгоритм перевірки простої гіпотези H_0 проти простої альтернативи H_1 зводиться до обчислення за спостереженням \bar{y} значення функцій правдоподібності $l(\bar{y})$ та порівняння його з наперед установленою величиною порогу $I_{пор}$. Якщо $l(\bar{y}) < I_{пор}$, приймається гіпотеза H_0 , якщо $l(\bar{y}) > I_{пор}$ – гіпотеза H_1 .

Критерій оптимальності Зігерта–Котельникова. Згідно з критерієм Зігерта-Котельникова необхідно мінімізувати середню ймовірність помилки (5.6). Враховуючи, що $P_{пом}$ збігається із середнім ризиком при простій матриці втрат, коли $B_{00} = B_{11} = 0, B_{10} = B_{01} = 1$, правило вибирання рішень, оптимальне за цим критерієм, одержуємо з (5.26):

$$P_1 W(\bar{y} | H_1) - P_0 W(\bar{y} | H_0) \underset{\gamma_0}{\overset{\gamma_1}{\geq}} 0. \quad (5.28)$$

Впроваджуючи відношення правдоподібності, маємо:

$$l(\bar{y}) \underset{\gamma_0}{\overset{\gamma_1}{\geq}} \frac{P_0}{P_1}, \quad (5.28a)$$

або

$$\ln l(\bar{y}) \underset{\gamma_0}{\overset{\gamma_1}{\geq}} \ln \left[\frac{P_0}{P_1} \right]. \quad (5.28b)$$

Бачимо, що значення порогу тут $I_{пор} = P_0/P_1$.

Критерій оптимальності Неймана-Пірсона. Розглянемо алгоритм, оптимальний за критерієм Неймана-Пірсона. Тут окремо розглядаються помилки першого (помилки «хибної тривоги») та другого («пропуску цілі») роду.

Для основної гіпотези H_0 область Y_0 називається критичною областю. З фізичної точки зору, критерій вибору гіпотез дуже рідко має відкидати основну гіпотезу, коли вона істинна. Звичайна процедура перевірки гіпотези полягає у фіксуванні довільного малого числа α , яке називається *рівнем значущості*, та перевіріці вимоги, що ймовірність помилки першого роду не перевищувала α . Якщо критична область задовольняє цю умову, кажуть: вона відповідає рівню значущості α . Знаходження такої процедури – перший етап задачі. Другий етап виникає коли задача розв’язана на першому етапі. Тоді необхідно вибрати з усіх можливих критичних областей, що відповідають за-

даному рівню значущості, таку, яка найбільш задовільно контролює помилку другого роду.

На фізичній мові згідно з (5.10) задача ставиться як задача мінімізації умовної імовірності хибної тривоги

$$\min_D \beta, \alpha \leq \alpha_0 \cdot 0$$

Синтез алгоритму, оптимального за критерієм Неймана-Пірсона, виконується при мінімізації допоміжного функціоналу

$$\Phi = \beta + c\alpha, \quad (5.29)$$

де c - невизначений множник Лагранжа.

Ураховуючи, що

$$\alpha = \int_{Y_1} W(\bar{y}|H_0) d\bar{y},$$

$$\beta = \int_{Y_0} W(\bar{y}|H_1) d\bar{y},$$

та порівнюючи (5.29) з (5.24), бачимо, що функціонал Φ збігається з R , коли $P_0 = P_1 = 1/2$, $B_{00} = B_{11} = 0$; $B_{01} = 2$; $B_{10} = 2c$. Тобто із загального вигляду байєсівського правила з урахуванням співвідношень (5.29) та (5.24) одержуємо правило, оптимальне за критерієм Неймана-Пірсона:

$$I(\bar{y}) \begin{matrix} \gamma_1 \\ > \\ < \\ \gamma_0 \end{matrix} c. \quad (5.30)$$

Поріг c визначається з умови $\alpha \leq \alpha_0$, тобто:

$$P\{I(\bar{y}) \geq c | H_0\} = \alpha. \quad (5.31)$$

Звідси бачимо, що й оптимальний алгоритм за критерієм Неймана-Пірсона такий самий, як і в загальному випадку (5.26а); тут лише у спеціальний спосіб вибирається значення порогу.

Якщо знаходити алгоритми обробки сигналів, оптимальні за кожним з наведених вище критеріїв оптимальності, ми дійдемо аналогічних результатів. Усюди алгоритм має вигляд (5.26а), відміна лише у значенні порогу $I_{пор}$. Можна стверджувати, що одержуються байєсівські алгоритми, які відрізняються видом матриці витрат і значенням ймовірностей гіпотез. Так, критерій Зігерта-Котельникова дає байєсівський алгоритм при простій функції (матриці) витрат; такий самий результат дає критерій максимальної апостеріорної ймовірнос-

ті. Алгоритм максимальної правдоподібності є окремий випадок останнього при рівноймовірних гіпотезах; мінімаксний критерій приводить до байєсівського алгоритму при найгіршому розподілі ймовірностей гіпотез; критерій Неймана–Пірсона також дає байєсівський алгоритм при рівноймовірних гіпотезах і спеціальному вигляді матриці втрат. Тобто клас байєсівських алгоритмів дійсно є найважливішим у теорії рішень.

У таблиці (5.1) зведені значення матриць втрат та ймовірностей гіпотез і для розглянутих критеріїв оптимальності, за яких рішення, оптимальні за цими критеріями є байєсівськими.

Таблиця 5.1

Критерій оптимальності	Матриця втрат, імовірності гіпотез
Байєсівський	$B=[B_{ij}], B_{00}, B_{01}, B_{10}, B_{11}; P_0, P_1$
Зігерта-Котельнікова, максимуму апостеріорної імовірності гіпотези	$B_{01} = B_{10}; B_{00} = B_{11} = 0; P_0, P_1$
Макимуму правдоподібності	$B_{01} = B_{10} = 1; B_{00} = B_{11} = 0; P_0 = P_1 = 1/2$
Неймана-Пірсона	$B_{00} = B_{11} = 0; B_{01} = 2, B_{10} = 2c; P_0 = P_1 = 1/2$
Мінімаксний	$P_0 = P_{0HG}; P_{1HG} = 1 - P_{0HG}; B=[B_{ij}]$

У таблиці (5.1) значення втрат $B_{01} = 2c$, характерний для критерію критерій Неймана–Пірсона, визначається умовою (5.31); найгірша ймовірність P_{0HG} , характерна для мінімаксного критерію визначається з аналізу залежності байєсівського ризику від ймовірності P_0 , а саме – $R(P_{0HG})$ досягає максимального значення.

5.2.2. Багатоальтернативна задача перевірки гіпотез

Задача перевірки гіпотез проти однієї альтернативи – це формалізація багатьох задач прийняття сигналів електрозв’язку. До неї зводяться задачі прийняття сигналів за різних видів модуляції, цифрових видів модуляції, використання двійкових кодів, тобто у випадках, коли необхідно розрізнати (розпізнавати) два переданих сигнали, спотвореннях завадами. Ряд практичних задач не зводиться до цієї моделі, коли для передавання інформації використовуються кілька різних сигналів. Це насамперед m -кові коди. Такі задачі приводять до багато альтернативних задач перевірки гіпотез. Розглянемо відповідні узагальнення.

У найпростішому випадку припускається, що заданий канал, тобто трійка $[X, Y, P_{Y|X}]$, імовірнісна міра P_X , а також матриця втрат

$B = [B_{ij}]$. Щодо постановки відповідної задачі математичної статистики, то допускається висунення M гіпотез H_0, H_1, \dots, H_{M-1} , для яких відомі імовірності гіпотез $P(H_i) = P_i, i = \overline{0, M-1}$, і кожна з них однозначно характеризує функцію правдоподібності $W(\bar{y} | H_i), i = \overline{0, M-1}$. За прийнятим вектором спостережень необхідно вибрати одну з гіпотез.

Ця задача ставиться, як задача перевірки простої гіпотези H_0 проти $M-1$ простої альтернативи $H_i, i = \overline{1, M-1}$. У цьому випадку описані вище результати можна узагальнити.

Розглянемо спочатку правило вибирання рішення, оптимальне за мінімумом середнього ризику (див.(5.5)):

$$\min_D \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} B_{ij} P_j \int_{Y_i} W(\bar{y} | H_j) d\bar{y}.$$

Аналогічно попередньому випадку можна довести, що мінімізація середнього ризику забезпечується, коли вибірковий простір Y^n розбивається на M областей Y_i , які не перетинаються, $Y_i \cap Y_k = \emptyset, i \neq k$, причому $\bigcup_{j=0}^{M-1} Y_j = Y^n$, а всі спостереження \bar{y} , задовольняючи систему нерівностей

$$\sum_{j=0}^{M-1} (B_{kj} - B_{ij}) \frac{P_j W(\bar{y} | H_j)}{P_0 W(\bar{y} | H_0)} \geq B_{j0} - B_{k0}, \quad (5.32)$$

$$k = \overline{0, M-1}, k \neq j, i = \overline{0, M-1},$$

відносяться до області Y_i . Тобто гіпотеза $H_i, i = \overline{0, M-1}$ приймається, якщо виконується ця система нерівностей.

Впроваджуючи відношення правдоподібностей:

$$I_j(\bar{y}) = \frac{W(\bar{y} | H_j)}{W(\bar{y} | H_0)}, j = \overline{0, M-1},$$

систему нерівностей (5.32) можна записати також у вигляді

$$\sum_{j=0}^{M-1} (B_{kj} - B_{ij}) \frac{P_j}{P_0} I_j(\bar{y}) \geq B_{j0} - B_{k0},$$

$$k = \overline{0, M-1}, k \neq j, i = \overline{0, M-1}$$

Розглянутий вище випадок перевірки гіпотези H_0 проти альтернативи H_1 можна дістати, поклавши $j = 1$:

$$(B_{01} - B_{11}) \frac{P_1}{P_0} I_1(\bar{y}) \geq B_{10} - B_{00},$$

або інакше –

$$I_1(\bar{y}) \geq \frac{B_{10} - B_{00}}{B_{01} - B_{11}} \cdot \frac{P_0}{P_1}.$$

Тобто, дійсно байєсівське правило вибирання рішень (5.32а) для багато альтернативної задачі безпосередньо узагальнює правило (5.26а) вибирання рішень для одно альтернативної задачі.

Як і раніше, з байєсівського правила (5.32) у разі простої матриці втрат, коли $B_i = 0$, $i = \overline{0, M-1}$, $B_{ij} = 1$, $i \neq j$, можна вивести правило, оптимальне за критерієм Зігерта–Котельникова. Дійсно, у разі простої матриці втрат середній ризик (5,1) збігається з середньою ймовірністю помилки $P_{ном}$:

$$R = P_{ном} = \sum_{i=0}^{M-1} \sum_{\substack{j=0 \\ j \neq i}}^{M-1} P_j \int_{Y_i} W(\bar{y}|H_j) d\bar{y}.$$

Поступаючи аналогічно попередньому або безпосередньо підставляючи у вирішуючі правило (5.32) просту матрицю втрат, дістаємо:

$$P_i W(\bar{y}|H_i) - P_j W(\bar{y}|H_j) \stackrel{Y_i}{\geq} 0 \quad (5.33)$$

$$j = \overline{0, M-1}, i \neq j;$$

його можна записати також у вигляді

$$i = \arg \max_{0 \leq j \leq M-1} P_j W(\bar{y}|H_j). \quad (5.33a)$$

Відповідно до цих алгоритмів за одержаним спостереженням \bar{y} треба обчислювати значення відповідних статистик. Так, у випадку алгоритму (5.33) треба обчислити значення M статистик $P_j W(\bar{y}|H_j)$ та обрати серед них максимальну; її номер « i » визначає гіпотезу H_i , що вибирається.

Аналогічно узагальнюються правила вибору рішень, оптимальні за іншими критеріями – мінімаксним, максимальної апостеріорної ймовірності, критерієм Неймана-Пірсона тощо. Особливість задачі перевірки простої гіпотези H_0 проти простих альтернатив

$H_j, j = \overline{1, M-1}$, полягає у тому, що по суті одержуються відповідні узагальнення правил вибирання рішення, характерних для одно альтернативних задач.

5.2.3. Наявність заважаючи параметрів у задачах перевірки гіпотез

Ряд задач прийняття сигналів у теорії електрозв'язку характеризується тим, що необхідно розрізнити кілька сигналів, які мають випадкові параметри (наприклад, амплітуда квазігармонічного зображення сигналу, його початкова фаза тощо). З фізичної точки зору останні часто можна вважати заважаючими. Описану вище постановку задачі перевірки гіпотез не можна безпосередньо застосувати у цьому випадку, проте можливе відповідне узагальнення.

Вважається що згідно з фізичною постановкою задачі можна висунути M гіпотез $H_j, j = \overline{0, M-1}$, щодо параметрів функції розподілів, причому із заданими функціями правдоподібності $W(\bar{y} | H_j)$, ймовірностями гіпотез $P(H_j) = P_j, j = \overline{0, M-1}$, а також матрицею втрат $B = [B_{ij}], i, j = \overline{0, M-1}$. На відміну від попереднього випадку для кожної гіпотези замість однієї функції правдоподібності задається параметрична множина $W(\bar{y} | H_j, \vec{v}_3), \vec{v}_3 \in \Theta_3$, тобто гіпотеза складна. Додатково припускається, що задано розподіл заважаючого векторного параметра $W(\vec{v}_3)$.

Якщо не враховувати розподіл $W(\vec{v}_3)$, це задача перевірки складної гіпотези H_0 проти $M-1$ складної альтернативи $H_j, j = \overline{1, M-1}$. Така задача значно складніша за попередньою.

У конкретному випадку вона може бути зведена до задачі перевірки простої гіпотези H_0 проти $M-1$ простої альтернативи $H_j, j = \overline{1, M-1}$, усередненням функції правдоподібності за розподілом заважаючих параметрів:

$$W(\bar{y} | H_j) = \int_{\Theta_3} W(\bar{y} | H_j, \vec{v}_3) W(\vec{v}_3) d\vec{v}_3. \quad (5.34)$$

Байєсівське рішення типу (5.32) має вигляд

$$\sum_{j=0}^{M-1} (B_{kj} - B_{ij}) \frac{P_j \int_{\Theta_3} W(\bar{y} | H_j, \vec{v}_3) W(\vec{v}_3) d\vec{v}_3}{P_0 \int_{\Theta_3} W(\bar{y} | H_0, \vec{v}_3) W(\vec{v}_3) d\vec{v}_3} \geq B_{i0} - B_{k0}, \quad (5.35)$$

$$k = \overline{0, M-1}, k \neq j, i = \overline{0, M-1},$$

тобто воно збігається з правилом (5.32), де замість функції правдоподібності $W(\bar{y} | H_j)$ використовуються відповідні усередненні функції правдоподібності (5.34).

Правило вибирання рішень (5.32) може бути зображено аналогічно (5.32a) через усередненні відношення правдоподібності $I_j(\bar{y})$:

$$\sum_{j=0}^{M-1} (B_{kj} - B_{j0}) \frac{P_j}{P_0} \overline{I_j(\bar{y})} \geq B_{j0} - B_{k0}, \quad (5.35a)$$

$$k = \overline{0, M-1}, \quad k \neq j, \quad i = \overline{0, M-1},$$

де

$$\overline{I_j(\bar{y})} = \frac{\int_{\theta_3} W(\bar{y} | H_j, \vec{v}_3) W(\vec{v}_3) d\vec{v}_3}{\int_{\theta_3} W(\bar{y} | H_0, \vec{v}_3) W(\vec{v}_3) d\vec{v}_3}.$$

Застосування інших критеріїв оптимальності приводить при цьому до таких саме результатів, як і у звичайній задачі перевірки простої альтернативи H_j , $j = \overline{1, M-1}$; усяди лише змінюється кожна функція правдоподібності відповідного усередненого функцією правдоподібності.

5.2.4. Перевірка складних гіпотез

Вище ми вже дали такі означення: гіпотеза H_j називається *простою*, якщо вона однозначно визначає функцію правдоподібності $W(\bar{y} | H_j)$, у противному разі гіпотеза називається *складною*. При цьому розрізняють два класи функції правдоподібності: *параметричний* $W(\bar{y} | \vec{v})$, $\vec{y} \in Y^n$, $\vec{v} \in \theta$, коли при кожному фіксованім значенні параметра $\vec{v} \in \theta$ маємо фіксовану функцію правдоподібності, а також *непараметричний*, коли клас функцій правдоподібності не заданий.

Розглянемо деякі особливості задач перевірки гіпотез у цих випадках.

Параметричні класи функцій правдоподібності. Спинимось для конкретності на критерії Неймана-Пірсона. Необхідно перевірити гіпотезу H_0 , що визначає параметричну множину функцій правдоподібності $W(\bar{y} | H_0, \vec{v}_0)$, $\vec{v}_0 \in \theta_0$ проти альтернативи H_1 , що визначає параметричну множину $W(\bar{y} | H_1, \vec{v}_1)$, $\vec{v}_1 \in \theta_1$. У найпростішому випадку, коли θ_0 є єдиною точкою \vec{v}_0 , маємо задачу перевірки простої

гіпотези H_0 проти складної гіпотези H_1 . Спочатку розглянемо саме таку задачу.

За правилом вибирання рішень простір спостережень Y^n розбивається на дві області, Y_0 та Y_1 : $Y_0 \cup Y_1 = Y^n$. Як ми знаємо, гіпотеза H_0 приймається коли $\bar{y} \in H_0$, та відкидається, якщо $\bar{y} \in H_1$.

У математичній статистиці ймовірність правильного рішення про наявність цілі $P\{\bar{y} \in Y_1 | H_1\}$ має назву потужністю критерію; в прикладній літературі її називають потужністю рішення. Вона дорівнює

$$\begin{aligned} P\{\bar{y} \in Y_1 | H_1\} &= \int_{Y_1} W(\bar{y} | H_1, \bar{v}) d\bar{y} = \int_{Y^n | Y_0} W(\bar{y} | H_1, \bar{v}_1) d\bar{y} = \\ &= \int_{Y^n} W(\bar{y} | H_1, \bar{v}_1) d\bar{y} - \int_{Y_0} W(\bar{y} | H_1, \bar{v}_1) d\bar{y} = 1 - \beta(\bar{v}_1). \end{aligned}$$

Звідси видно, що потужність рішення залежить від конкретного значення параметра \bar{v}_1 з множини θ_1 .

Якщо розв'язувати задачу типу (5.10a), характерну для задачі перевірки простої гіпотези H_0 проти простої альтернативи, то слід знайти

$$\max(1 - \beta(\bar{v}_1)) \text{ при обмеженні } \alpha \leq \alpha_0. \quad (5.36)$$

Бачимо, що рішення залежатиме від конкретного значення параметра \bar{v}_1 з множини θ_1 . Якщо правило вибирання рішень таке, що ймовірність правильного рішення про наявність цілі (при обмеженні $\alpha \leq \alpha_0$ на ймовірність хибної тривоги) максимізується для кожного фіксованого значення параметра $\bar{v}_1 \in \theta_1$, такі рішення називаються рівномірно найбільш потужними (РНП).

РНП вирішуюче правило при виборі простої гіпотези проти складної альтернативи H_1 трапляється дуже рідко. Тому звичайно шукають вирішуючі правила з більш вузьких класів, які задовольняють деякі додаткові вимоги. Широко використовують, так звані, незсунені правила. Вони мають задовольняти умови, щоб ймовірність відкинути хибну гіпотезу була не менше ймовірності відкинути істину, тобто $\alpha \leq 1 - \beta(\bar{v}_1)$ для всіх значень \bar{v}_1 . Хоч РНП рішення у загальному випадку може не існувати, у класі незсунених рішень воно все ж таки існує.

У загальному випадку перевірки складної гіпотези H_0 проти складної альтернативи H_1 у задачі типу (5.36) не тільки ймовірність пропуску цілі β залежить від параметра

$\beta = \beta(\vec{v}_1)$, $\vec{v}_1 \in \theta_1$, а й імовірність хибної тривоги α також залежить від параметра $\alpha = \alpha(\vec{v}_0)$, $\vec{v}_0 \in \theta_0$. Щоб знайти відповідне вирішуюче правило, накладають додаткові (порівняно з попереднім) обмеження на відповідний клас правил.

Вирішуючі правила в цих випадках вже не мають структури байєсівських. Те саме стосується правил, оптимальних за іншими критеріями.

Непараметричні класи функцій правдоподібності. Кажуть, що гіпотези H_0 і H_1 непараметричні, якщо вони характеризують непараметричні класи W_0 , W_1 функцій правдоподібності.

На відміну від випадків перевірки простої гіпотези проти простої альтернативи, що характеризують параметричні класи, дотепер ще не існує загальних методів синтезу оптимальних методів перевірки непараметричних гіпотез. Алгоритми перевірки гіпотез у цьому випадку будуються на евристичній основі.

Непараметричний алгоритм має забезпечити постійне значення ймовірностей помилок по відношенню до кожної з непараметричних гіпотез H_0 та H_1 . До непараметричних відносяться також алгоритми, які забезпечують сталу величину ймовірності помилки відносно принаймні однієї з параметричних гіпотез.

Звичайно у непараметричних алгоритмах, коли це можливо, використовують незалежні однорідні вибірки. При цьому знаходять деякі статистики (тобто перетворення спостережень), для яких можливо обчислювання функцій правдоподібності. Популярними є, так звані, знакові, порядкові, рангові та інші статистики. Використовуючи їх на евристичній основі, знаходять непараметричні алгоритми перевірки гіпотез, що забезпечують сталу величину ймовірності хибної тривоги (при змінюванні обсягу вибірки та довільному розподілу з непараметричного класу); ймовірність помилки пропуску цілі, взагалі кажучи, для них більша від мінімально можливої.

5.2.5. Послідовні алгоритми прийняття рішення

Вище ми розглядали однокрокові алгоритми вибирання рішень, для яких час спостереження постійний; вони характеризуються тим, що за вибіркою фіксованого обсягу відразу вибирається одна з гіпотез. Але, як зазначалось, існують ще і багатокрокові алгоритми, для яких час спостереження не фіксований; для них на кожному кроці або вибирається одна з гіпотез, або приймається рішення, щодо про-

довження спостереження. Спинимося на особливостях послідовних алгоритмів для найпростішого випадку перевірки простої гіпотези проти постійної альтернативи.

У разі використання послідовних алгоритмів спочатку береться вибірка обсягу $n=1$, і за нього або вибирається одна з гіпотез (H_0 чи H_1), або приймається рішення продовжити спостереження. Якщо гіпотеза вибрана, спостереження припиняються, якщо ні – береться вибірка обсягу $n=2$ і все повторюється доти, поки не буде все ж таки вибрана одна з гіпотез. Мить припинення спостереження носить випадковий характер, але імовірність того, що число кроків у послідовній процедурі не обмежено, дорівнює нулю.

Особливість послідовних алгоритмів полягає у тому, що простір спостереження у задачі перевірки простої гіпотези H_0 проти простої альтернативи H_1 розбивається не на дві, а на три області (остання область проміжна): Y_0 , Y_1 , Y_{np} – відповідно прийняття гіпотези H_0 , прийняття гіпотези H_1 та прийняття рішення щодо продовження спостереження (рис. 5.4).

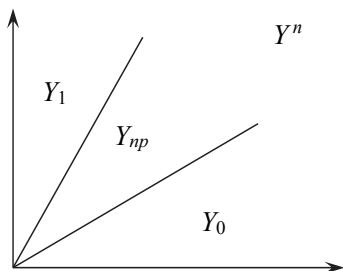


Рис. 5.4. – Ілюстрація розділення простору спостережень Y^n (за деякого обсягу вибірки « n ») у послідовних алгоритмах прийняття рішення:

Y_0 – приймається гіпотези H_0 , Y_1 – приймається гіпотези H_1 ;

Y_{np} – приймається рішення про продовження спостереження

А. Вальд запропонував критерій мінімуму середнього розміру вибірки (тобто часу спостереження) за умови, що імовірність помилки хибної тривоги не перевищує α_0 , а імовірність правильного виявлення сигналу (тобто потужність правила) не менше від $1 - \beta_0$.

Алгоритм Вальда для незалежних спостережень мінімізує величини розмірів вибірки $M[v | H_0]$ та $M[v | H_1]$ за умови, що імовірності помилок не перевищують α_0 та β_0 . Він характеризується тим, що відношення правдоподібності зрівнюється з двома порогами $I_{пор1}$ та $I_{пор2}$: приймається гіпотеза H_0 , якщо

$$I_{\text{пор}0} < I(Y_1, \dots, Y_k) < I_{\text{пор}1}, I(Y_1, \dots, Y_v) \leq I_{\text{пор}0}, k = \overline{1, v-1};$$

приймається гіпотеза H_1 , якщо

$$I_{\text{пор}0} < I(Y_1, \dots, Y_k) < I_{\text{пор}1}, I(Y_1, \dots, Y_v) \geq I_{\text{пор}0}, k = \overline{1, v-1};$$

приймається рішення продовжити спостереження, якщо

$$I_{\text{пор}0} < I(Y_1, \dots, Y_k) < I_{\text{пор}1}, k = \overline{1, v}.$$

Тут $M[v | H_0]$ та $M[v | H_1]$ – середні значення розміру v вибірки при гіпотезах H_0 і H_1 .

5.3. Оцінювання параметрів розподілів

Одна з найважливіших задач, що постають при обробці сигналів у електрозв'язку, – оцінювання параметрів сигналів. У математичній статистиці вона формулюється як задача оцінювання неперервних параметрів розподілів.

Найчастіше необхідно здійснити оцінювання параметрів розподілів з параметричної множини, що вище названа параметричним класом функцій правдоподібності.

5.3.1. Оцінювання скалярного параметра

З інформаційної точки зору повідомлення – це реалізація випадкового процесу, тому що детермінований сигнал не придатний для передавання інформації. У задачі перевірки гіпотез параметр (який ми ототожнювали з номером гіпотези) трактувався як дискретна випадкова величина. У задачі оцінювання параметра принципова інформаційна точка зору залишається такою самою, але тут параметр вважається випадковою величиною, яка набирає значення з деякої континуальної множини.

Вважаємо, що у загальному випадку у відповідному вигляді задані канал зв'язку трійкою $[X, Y, P_{XY}]$, міра P_X , а також функція втрат. На мові математичної статистики це означає, що задано параметричний клас функцій правдоподібності $W(\bar{y}|v)$, де $v \in \Theta$, для Θ – множина дійсних чисел; щільність імовірностей параметра $W(v)$, а також функція втрат $B(v^*; v)$.

Задача полягає у тому, щоб вибрати показник якості рішень, критерій їх оптимальності та синтезувати алгоритм оцінювання па-

раметра, який давав би «досить добру» оцінку (тобто щоб оцінка задовольняла деякі із сформульованих у п.5.1. вимоги незсуненості, ефективності, обґрунтованості або хоча б ці вимоги виконувались при збільшенні обсягу вибірки до нескінченності – асимптотично). Можна задати різні показники якості рішень і критерії їх оптимальності – баєсівський критерій оптимальності, критерій максимуму апостеріорної щільності, максимальної правдоподібності тощо. Розглянемо для конкретності останній – він вимагає знання лише функції правдоподібності.

Метод максимуму правдоподібності. Звичайно припускається, що вибірка однорідна, тобто з прийнятого сигналу беруться статистично незалежні відліки:

$$\vec{y}=(y_1,\dots,y_n), \quad W(\vec{y}|v)=\prod_{j=1}^n W(y_j|v), \quad v \in \Theta. \quad (5.38)$$

У методі максимуму правдоподібності оцінка параметра v – це те значення параметра функції правдоподібності $W(\vec{y}|v)$, за якого досягається її максимум. Така умова приводить до умови рівняння правдоподібності

$$\frac{\partial W(\vec{y}|v)}{\partial v} = 0. \quad (5.39)$$

Це може бути алгебраїчне, нелінійне або трансцендентне рівняння. Залежно від того, скільки максимумів і мінімумів має функція $W(\vec{y}|v)$ (як функція аргументу v), рівняння має відповідне число розв'язків. Тоді шукають розв'язок, що відповідає найбільшому значенню максимумів (рис. 5.5) – тут шукана оцінка v_5^* відповідає максимуму максимуму $W_{\max \max}$ функції $W(\vec{y}|v)$.

Враховуючи, що точки максимуму функції $W(\vec{y}|v)$ та $\ln W(\vec{y}|v)$ збігаються, рівняння (5.39) з використанням (5.38) звичайно приводять до вигляду

$$\sum_{j=0}^n \frac{\partial}{\partial v} \ln W(y_j|v) = 0. \quad (5.39a)$$

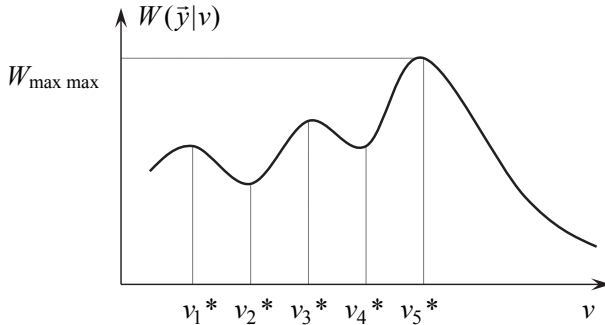


Рис. 5.5. — Ілюстрація неоднозначності розв'язування рівняння правдоподібності: розв'язки відповідають усім максимумам і мінімумам функції правдоподібності, v_5^* відповідає максимуму максимуму $W_{\max \max}$ функції $W(\bar{y}|v)$

1. Нижня границя дисперсії незсунених оцінок. Оцінки параметрів розподілу можна знаходити різними способами і вони можуть мати різні якості. Проти, якщо оцінка навіть не є ефективною, тобто її дисперсія не сягає можливого мінімуму, треба знайти значення цього мінімуму. Сформулюємо відповідні положення фундаментального характеру.

Лема 5.1

Нехай:

а) існують похідні $W'_v(\bar{y}|v)$ та $W''_v(\bar{y}|v)$, причому

$$M \left| \frac{\partial}{\partial v} \ln W(\bar{y}|v) \right| < \infty$$

$$M \left| \frac{\partial^2}{\partial v^2} \ln W(\bar{y}|v) \right| < \infty,$$

б)
$$M \left| \frac{\partial}{\partial v} \ln W(\bar{y}|v) \right|^2 < \infty.$$

Тоді
$$M \left| \frac{\partial}{\partial v} \ln W(\bar{y}|v) \right| = 0, v \in \Theta,$$

$$M \left[\frac{\partial^2}{\partial v^2} \ln W(\bar{y}|v) \right]^2 = -M \frac{\partial^2}{\partial v^2} \ln W(\bar{y}|v).$$

Тут: M – символ математичного сподівання, коли функція $W(\bar{y}|v)$ розглядається як функція \bar{y} :

$$M \left[\frac{\partial^2}{\partial v^2} \ln W(\bar{y}|v) \right]^2 = I(v)$$

– кількість інформації за Фішером (у деякому розумінні цю функцію можна розглядати як кількість інформації, що міститься у векторі \bar{y} щодо параметра v)

Теорема 5.1 (нерівність Крамера-Рао)

Нехай виконані припущення а) та б) леми 5.1, а $\eta(v)$ – диференційовна функція на Θ , для якої існує незсунена оцінка $h^*(\bar{y})$ зі скінчевою дисперсією, що задовольняє умову

$$\int_{R^n} \left| h^*(\bar{y}) \frac{\partial}{\partial v} W(\bar{y}|v) \right| d\bar{y} < \infty$$

для кожного $v \in \Theta$. Тоді для кожного $v \in \Theta$

$$M |h^*(\bar{y}) - \eta(v)|^2 \geq \frac{|h'(v)|^2}{I(v)}. \quad (5.40)$$

Знак рівності досягається тоді і тільки тоді, коли

$$\frac{\partial}{\partial v} \ln W(\bar{y}|v) = c(v) |h^*(\bar{y}) - \eta(v)|.$$

Зауважимо, що умови сформульовані в теоремі 5.1 часто називають *умовами регулярності*.

Якщо твердження цієї теореми застосувати до самого параметра v , тобто $\eta(v) = v$, нерівність (5.40) набере вигляду

$$M |h^*(\bar{y}) - v|^2 \geq \frac{1}{I(v)}. \quad (5.40a)$$

У разі, коли в нерівності (5.40a) досягається рівність, оцінка є незсуненою оптимальною оцінкою, тобто ефективною оцінкою.

2. Властивості оцінок максимальної правдоподібності. Оцінки максимальної правдоподібності мають ряд властивостей, що привели до значного поширення методу максимуму правдоподібності.

Наведемо їх:

- а) розв'язок рівняння (5.39) є обґрунтованою оцінкою;
- б) розв'язок рівняння є асимптотичне ефективною оцінкою параметра v ;
- в) розв'язок має асимптотичне нормальний розподіл (тобто розподіл оцінки як випадкової величини при $n \rightarrow \infty$ виявляється нормальним);
- г) якщо для параметра v існує деяка ефективна оцінка, то у цьому разі рівняння правдоподібності має єдиний розв'язок, що збігається з цією оцінкою.

5.3.2. Оцінювання векторного параметра

Узагальнемо наведені вище результати на випадок векторного параметра.

Розглядається параметричний клас функцій правдоподібності:

$$W(\bar{y}|\bar{v}), \quad \bar{y}=(y_1, \dots, y_n), \quad \bar{v}=(v_1, \dots, v_m), \quad \bar{v} \in \Theta^m.$$

У загальному випадку покладається, що завдані також щільність імовірності векторного параметра $\bar{v} - W(\bar{v})$ і функція втрат $B(\bar{v}^*; \bar{v})$. Залежно від того, яка задана апіорна інформація, використовується той чи інший показник якості рішень і критерій їх оптимальності. Застосування критерію максимуму правдоподібності, як сказано, не потребує знання щільності імовірності $W(\bar{v})$ та функції втрат $B(\bar{v}^*; \bar{v})$ — необхідно знати лише функцію правдоподібності.

Застосування критерію максимуму правдоподібності для знаходження m параметрів (v_1, \dots, v_m) приводить до системи рівнянь правдоподібності

$$\frac{\partial W(\bar{y}|v_1, \dots, v_m)}{\partial v_j} = 0, \quad j = \overline{1, m}. \quad (5.41)$$

Звідси можна дістати рівняння правдоподібності (5.39), коли $m=1$.

Розглянуті нижче положення щодо нижньої границі оцінки максимальної правдоподібності скалярного параметра мають відповідні узагальнення для векторного параметра. Спинимось на цьому.

Як і у лемі 5.1 припустимо, що: а) існують похідні $\frac{\partial}{\partial v} W(\bar{y}|\bar{v})$ функції правдоподібності $W(\bar{y}, v_1, \dots, v_m)$ по кожному параметру $v_j, j = \overline{1, m}$, причому

$$M \frac{\partial}{\partial v_j} |W(\bar{y}|\bar{v})| < \infty, \quad j = \overline{1, m}, \quad \bar{v} \in \Theta^m;$$

б) у матриці $I(\bar{v})$ з елементами

$$I_{ij}(\bar{v}) = M \frac{\partial}{\partial v_i} \ln W(\bar{y}|\bar{v}) \frac{\partial}{\partial v_j} \ln W(\bar{y}|\bar{v}), \quad i, j = \overline{1, m} \quad (5.42)$$

додано визначення для кожного $\bar{v} \in \Theta^m$.

(Зауважимо, що матриця $I(\bar{v})$ має назву *інформаційної матриці Фішера*; при $m=1$ вона збігається з кількістю інформації за Фішером $I(v)$ для скалярного параметра).

Припускається також, що є деяка векторна функція $\bar{\eta}^*(\bar{y}) = (\eta_1^*(\bar{y}), \dots, \eta_d^*(\bar{y}))$, яка являє собою незсунену оцінку d – вимірного параметра $\bar{\eta}(\bar{y}) = (\eta_1(\bar{y}), \dots, \eta_d(\bar{y}))$, причому існують частинні похідні $\frac{d}{dv_j} h_k(\bar{y})$.

Тоді справджується таке твердження.

Теорема 5.2

Нехай множина щільностей імовірностей $W(\bar{y}|\bar{v})$, $\bar{v} \in \Theta^m$ задовольняє умови а), б); $\bar{h}^*(\bar{y})$ – незсунена оцінка векторного параметра $\bar{h}(\bar{y})$, для якої

$$\int_{R^n} (\eta_k^*(\bar{y}) \frac{\partial}{\partial v_j} W(\bar{y}|\bar{v})) d\bar{y} < \infty, \quad j = \overline{1, m}, \quad k = \overline{1, d}, \quad \bar{v} \in \Theta^m.$$

Тоді матриця $D(\bar{v}) = H^{tr}(\bar{v}) I^{-1}(\bar{v}) H(\bar{v})$ невід’ємно визначена для всіх \bar{v} .

Тут:

$$\begin{aligned} D(\bar{v}) &= M(\bar{h}^*(\bar{y}) - \bar{h}(\bar{v}))(\bar{h}^*(\bar{y}) - \bar{h}(\bar{v}))^{tr}, \\ H(\bar{v}) &= M \nabla \ln W(\bar{y}|\bar{v}) \bar{h}^{*tr}(\bar{y}); \end{aligned} \quad (5.43)$$

$\nabla g(\bar{v})$ – позначення градієнта функції $g(\bar{v})$ аргументу \bar{v} :

$$\nabla g(\bar{v}) = \left(\frac{\partial g}{\partial v_1}, \dots, \frac{\partial g}{\partial v_m} \right);$$

tr – символ транспонування матриці.

Ця теорема узагальнює нерівність Крамера-Рао на векторний випадок.

Розглянемо два окремі випадки.

1. Функція $\eta(\vec{v})$ одновимірна. При цьому $D(\vec{v})$ – скалярна величина, а матриця $H(\vec{v})$ стає вектором – стовпцем з елементами

$$D(\vec{v}) = M \frac{\partial}{\partial v_j} W(\vec{y}|\vec{v}) \eta^*(\vec{y}), \quad j = \overline{1, m}.$$

З теореми 5.2 випливає нерівність

$$D(\vec{v}) \geq H^{tr}(\vec{v}) I^{-1}(\vec{v}) H(\vec{v}),$$

в якій знак рівності досягається тоді і тільки тоді, коли існує вектор $\vec{g}^*(\vec{v})$, такий що

$$\vec{g}^*(\vec{v}) \nabla \ln W(\vec{y}|\vec{v}) = \eta^*(\vec{y}) - \eta(\vec{v}).$$

У цьому випадку функція $\eta(\vec{v})$ має незсунену оцінку $\eta^*(\vec{y})$ з мінімальною дисперсією $H^{tr}(\vec{v}) I^{-1}(\vec{v}) H(\vec{v})$, де $H = I(\vec{v}) \vec{g}(\vec{v})$.

2. Функція $\eta^*(\vec{y}) = \eta(\vec{v})$, тобто - це сам параметр.

У цьому випадку $H(\vec{v})$ – одинична матриця. Тоді в умовах теореми 5.2 виконується рівність

$$\nabla \ln W(\vec{y}|\vec{v}) = I(\vec{v})(\vec{\eta}^*(\vec{y}) - \vec{v}). \quad (5.44)$$

Оцінки компонент вектора \vec{v} , для яких виконується рівність (5.44) називаються сумісно ефективними; для них $D(\vec{v}) = I^{-1}(\vec{v})$. Нажаль умова (5.44) виконується досить рідко. Тому часто розглядають асимптотичне сумісні ефективні оцінки для яких відповідна умова виконується асимптотичне при збільшенні обсягу вибірки до нескінченності – $n \rightarrow \infty$.

Оцінки максимальної правдоподібності (ОМП) обґрунтовані й асимптотично сумісно ефективні. Коли $n \rightarrow \infty$, оцінка векторного параметра $\vec{v}^* = (v_1^*, \dots, v_m^*)$ має m – вимірний закон розподілу з вектором середніх \vec{v} , кореляційна матриця якого визначається через інформаційну матрицю Фішера. Відомі умови, коли ОМП збігається з оцінками за максимумом апостеріорної імовірності або баєсівськими оцінками.

5.3.3. Достатні статистики

Функція спостережень називається *статистикою*. Винятково важливі ті статистики, що містять у собі (кажучи якісно) всю інформацію, необхідну для розв'язання відповідної задачі, але вимірність їх менше від вимірності первинного спостереження. Таке скорочення (стиснення) інформації дає змогу зменшити складність розв'язуваної задачі. Статистики, що зберігають усю необхідну інформацію, називаються *достатніми*. З ними вже зустрічалися раніше, хоча й не використовували відповідний термін.

Так, у задачі перевірки гіпотез достатньою статистикою було відношення правдоподібності: у відношенні правдоподібності $l(y_1, \dots, y_n)$ містилася вся інформація про гіпотези, яка є у спостереженні $\vec{y} = (y_1, \dots, y_n)$. При переході до відношення правдоподібності використовується також скорочення даних: замість n -вимірного вектора спостережень формується 1-вимірна функція (інакше кажучи, n -вимірний простір спостережень відображається на дійсну вісь – рис. 5.6).

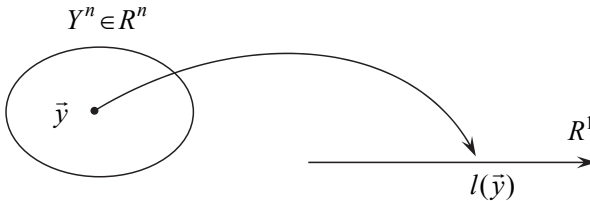


Рис. 5.6. – Стиснення даних при використанні достатньої статистики у вигляді відношення правдоподібності

Дамо строгі означення достатньої статистики.

Розглянемо параметричну множину щільностей імовірності $W(\vec{y}|\vec{v})$, $\vec{y} \in R^n$, $\vec{v} \in \theta^m$. Підкреслимо, що це щільність імовірності випадкового вектора \vec{Y} ; параметр ми розглядаємо як випадковий вектор $\vec{\theta} \in \theta^m$, що набирає у реалізаціях \vec{y} вектора \vec{Y} фіксовані значення \vec{v} .

Припустимо, що щільність імовірності $W(\vec{y}|\vec{v})$ може бути зображена у вигляді:

$$W(\vec{y}|\vec{v}) = \varphi(T(\vec{y}|\vec{v})h(\vec{y})), \quad (5.45)$$

де функції $T(\vec{y})$ і $h(\vec{y})$ не залежать від параметра, причому $T(\vec{y})$ – деяка вимірна функція.

Функція $T(\vec{y})$ називається *достатньою статистикою* для множини щільностей ймовірностей $W(\vec{y}|\vec{v})$, $\vec{v} \in \theta^m$ (або для параметра $\vec{\theta}$), якщо ця сім'я розподілів допускає зображення у вигляді (5.45).

Формула (5.45) називається *факторизацією щільності ймовірностей*.

Нажаль, не існує більш-менш загальних методів одержання достатніх статистик. Проте, коли останні знайдені, розв'язка задач оцінювання параметрів визначається через них. (Так, у теоремі 5.2 ми брали функцію $\bar{h}(\vec{v})$ параметра \vec{v} і називали $\bar{h}^*(\vec{y})$ оцінкою параметра; тут по суті використана достатня статистика для параметра \vec{v}).

При використанні для оцінювання параметрів методу максимальної правдоподібності, якщо існує достатня статистика, то кожний розв'язок рівняння правдоподібності (5.39) є функцією цієї достатньої статистики.

5.3.4. Послідовні алгоритми оцінювання параметрів

Аналогічно задачі перевірки гіпотез при розв'язанні задачі оцінювання параметрів, крім оцінювання параметра за спостереженням на фіксованому інтервалі часу, існують послідовні методи оцінювання. При цьому використовується математичний апарат послідовного статистичного аналізу для цілей оцінювання параметрів розподілів. Характерна риса такого підходу в тому, що достатньо обгрунтовано вибирається час спостереження реалізації прийнятого сигналу, за якого можна дістати оцінку параметрів необхідної якості.

Закінчуючи обговорення питань оцінювання параметрів, зауважимо, що крім розглянутого підходу, який у математичній статистиці називається *точковим*, використовується також *інтервальне оцінювання*. При інтервальному оцінюванні вказують інтервал, в якому з ймовірністю, не менше від заданої, міститься справжнє значення параметра. Ця задана ймовірність називається *коефіцієнтом довіри*, а вказаний інтервал значень параметра — *надійним інтервалом*. Підкреслимо, що при інтервальному оцінюванні випадковий інтервал покриває (у разі правильного рішення) чи не покриває (коли рішення хибне) істинне значення параметра. Надійний інтервал вибирається для того, щоб можливість хибних рішень можна було практично нехтувати.

5.4. Фільтрація повідомлень

У галузі обробки сигналів звичайно розглядають два відносно самостійних напрямки: розрізнення сигналів, спотворених завадами (яке містить у собі як частинний випадок задачу виявлення сигналів), а також фільтрацію повідомлень, у загальному випадку закодованих у сигналах і спотворених завадами (частинним випадком якої є оцінювання параметрів сигналів, що змінюються у часі). Задачу розрізнення сигналів можна розглядати як окремий випадок задачі оцінювання параметрів сигналів, тобто з математичної точки зору задача фільтрації повідомлень є найбільш загальною, а також, у деякому розумінні, і більш складною.

Розглянемо два окремих випадки: фільтрація повідомлень (стаціонарних і нестационарних), які передаються безпосередньо, та фільтрація повідомлень, що «закодовані» у сигналах із застосуванням нелінійних методів модуляції.

5.4.1. Оптимальна лінійна фільтрація аналогових повідомлень фільтром Колмогорова-Вінера

Теорія лінійної фільтрації була розроблена А.М. Колмогоровим для процесів з дискретним часом (1939 р.) та Н. Вінером для процесів з неперервним часом (1942 р.). У найпростішому випадку припускається, що реалізація спостережуваного сигналу є сумою реалізацій переданого повідомлення $s(t)$ та шуму $\xi(t)$:

$$y(t) = s(t) + \xi(t). \quad (5.46)$$

Випадкові процеси $S(t) = s(t, \omega)$ та $\Xi(t) = \xi(t, \omega)$ – стаціонарні у широкому розумінні процеси з нульовими середніми: $MS(t) = M\Xi(t) = 0$, $\omega \in \Omega$. Відомі їх моменти у межах теорії другого порядку, тобто процеси задані у межах кореляційної теорії.

Ставиться задача знаходження характеристик лінійного фільтра (рис. 5.7), який мінімізує середньоквадратичну похибку

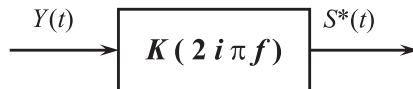


Рис. 5.7. – Оптимальний лінійний фільтр у задачі фільтрації повідомлень; « i » – уявна одиниця

$$\varepsilon^2 = M[S(t) - S^*(t)]^2 \quad (5.47)$$

відновлення переданого повідомлення.

Якщо не обмежуватися умовою реалізованості лінійного фільтра, випадковий сигнал на виході фільтра має вигляд:

$$S^*(t) = \int_{-\infty}^{\infty} Y(t-\tau) \cdot g(\tau) d\tau, \quad (5.48)$$

де $g(\tau)$ – часова характеристика фільтра.

Покладаючи, що повідомлення та шум некорельовані, легко довести, що кореляційні функції повідомлення $R_S(\tau)$, прийнятого сигналу $R_Y(\tau)$ та шуму $R_\xi(\tau)$ зв'язані рівнянням

$$R_S(\tau) = \int_{-\infty}^{\infty} R_Y(t-\tau) \cdot g(\tau) d\tau, \quad (5.49)$$

що називається *рівнянням Вінера-Хонфа*. Тут $R_Y(\tau) = R_S(\tau) + R_\xi(\tau)$.

Застосовуючи перетворення Фур'є до обох частин рівності (5.49), можна дістати відповідну рівність у частотній області:

$$G_x(f) = [G_S(f) + G_\xi(f)] K(f). \quad (5.50)$$

Якщо $G_S(f) + G_\xi(f) \neq 0$, $f \in (-\infty, +\infty)$, одержуємо коефіцієнт передачі оптимального лінійного фільтра

$$K(f) = \frac{G_S(f)}{G_S(f) + G_\xi(f)}. \quad (5.51)$$

У загальному випадку, коли враховується час затримки t_3 у фільтрі, маємо

$$K(f) = \frac{G_S(f)}{G_S(f) + G_\xi(f)} \cdot \exp[-i2\pi f t_3].$$

Фільтр з частотною характеристикою (5.51) називають *фільтром Колмогорова-Вінера*. Застосування такого фільтра забезпечує мінімум середньоквадратичної похибки, яку можна обчислити за формулою

$$\varepsilon_{\min}^2 = \int_{-\infty}^{\infty} \frac{G_S(f) \cdot G_\xi(f)}{G_S(f) + G_\xi(f)} df, \quad (5.52)$$

де $G_S(f)$, $G_\xi(f)$ – спектральна густина процесів відповідно $S(t)$ та $\Xi(t)$.

Цікаво, що похибка може бути і нульовою, коли спектри сигналу та шуму не перетинаються. Якщо співвідношення (5.52) подається у вигляді

$$\varepsilon_{\min}^2 = \int_{-\infty}^{\infty} \frac{G_S(f)}{1 + \frac{G_\xi(f)}{G_S(f)}} df, \quad (5.52a)$$

бачимо, що фільтр пропускає «спектральні складові» сигналу з тим меншою вагою, чим більше співвідношення $G_S(f) / G_\xi(f)$.

Зробимо деякі зауваження. Рівняння Вінера-Хопфа (5.48) природно узагальнюється на випадок реалізованих фільтрів. При цьому лише змінюється область інтегрування у (5.49): замість інтервалу $(-\infty, \infty)$ береться інтервал $(0, \infty)$.

Але розв'язання відповідного рівняння ускладнюється. Існує також просте узагальнення на випадок, коли сигнал і завада корельовані. Зауважимо також, що Вінер, розглядаючи задачу фільтрації, розробив метод розв'язання інтегрального рівняння Вінера-Хопфа шляхом переходу у частотну область. Він запропонував виконувати факторизацію відповідної спектральної густини – її зображення у формі добутку двох дзеркально симетричних співмножників. При цьому оптимальний фільтр визначається своєю частотною характеристикою i , крім того, розклад спектральної густини відповідає конструції характеристики породного фільтра, який формує відповідний випадковий сигнал при входному збуренні у вигляді білого шуму. У теорії фільтрації породні фільтри взагалі мають велике значення.

Розглянемо задачу фільтрації у більш формалізованому та узагальненому вигляді. З математичної точки зору природно ставити подібну задачу як задачу оцінювання випадкової величини. Тобто треба оцінити значення випадкової величини $\xi(\omega)$ через значення величин $\xi_t(\omega)$, заданих у точках $t \in T$. Це означає, що необхідно знайти функцію $f(\xi_t(\omega) | t \in T)$, яка залежить від множини величин $\xi_t(\omega)$, $t \in T$ з найменшою похибкою

$$\varepsilon^2 = M | \xi(\omega) - f[\xi_t(\omega) | t \in T] |^2. \quad (5.53)$$

Результат розв'язання такої задачі – це функція f , що характеризує відповідний фільтр.

В окремих випадках задача оцінювання випадкової величини містить у собі такі фізичні задачі:

- прогнозування (екстраполяцію) випадкового сигналу, коли треба оцінити значення сигналу у момент t за його значеннями на деякій множині попередніх моментів часу;

- фільтрація, коли на деякій множині t спостерігається суміш сигналу та шуму і треба відокремити сигнал від шуму.

У загальному випадку задача знаходження функції f_z (5.53) (тобто знаходження оптимального фільтра) має розв'язок у класі вимірних функцій. Фізично це означає, що оптимальний фільтр нелінійний. Однак часто ставлять більш вузьку задачу знаходження рішення у класі лінійних функцій (лінійних фільтрів). Це означає, що ми задовольняємося лінійним наближенням до нього.

Задача знаходження оптимального лінійного фільтра може розв'язуватись із застосуванням математичного апарату гільбертова простору. Це у даному разі гільбертів простір $L_{2\xi}$ випадкових величин зі скалярним добутком

$$(\xi, \zeta)_{L_{2\xi}} = M \xi \bar{\zeta}.$$

Саме такий простір й використовується для знаходження розв'язку задачі (5.53). У гільбертовому просторі випадкових величин вводять підпростір, що є лінійною оболонкою випадкових величин $\xi_t(\omega)$, $t \in T$ і константи та результатів їх граничних переходів. Найкращим лінійним наближенням $\xi^*(\omega)$ до випадкової величини $\xi(\omega)$ є той елемент побудованого підпростору, що перебуває від $\xi(\omega)$ на найменшій відстані. Це відома задача теорії апроксимації у гільбертовому просторі і вона завжди має єдиний розв'язок.

У позначеннях задачі фільтрації (5.46) – (5.47) розв'язок має таке геометричне розуміння (рис. 5.8).

Невідоме повідомлення \vec{S} – це деякий вектор гільбертова простору. У ньому побудований підпростір спостережуваних сигналів, де \vec{Y} – прийнятий сигнал. Тоді розв'язок задачі \vec{S}^* – це ортогональна проекція \vec{S} на підпростір \vec{Y} , тобто вектор похибки \vec{E} завжди має бути ортогональним вектору \vec{Y} . Це, так званий, принцип ортогональності у задачі лінійної фільтрації повідомлень.

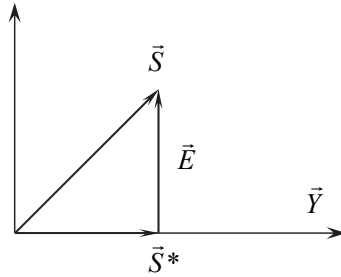


Рис. 5.8. — Геометрична ілюстрація принципу ортогональності у теорії лінійної фільтрації; випадкові вектори похибки \vec{E} та прийнятого сигналу \vec{Y} ортогональні

Хоч у загальному випадку задача оцінювання випадкової величини (і відповідно фільтрації повідомлень) має розв'язок у класі нелінійних функцій, лінійні фільтри знайшли дуже широке застосування завдяки своїй простоті. Ще дуже важлива така обставина: для гаусових сигналів похибка, що забезпечується оптимальним фільтром, збігається з похибкою, що забезпечується оптимальним лінійним фільтром. Інакше кажучи, нелінійна фільтрація гаусових сигналів не є необхідною.

Фільтр Колмогорова-Вінера придатний лише для стаціонарних випадкових процесів. Однак існує загальний метод фільтрації, який придатний як для лінійної фільтрації нестационарних процесів, так і для нелінійної фільтрації процесів, що не є гаусовими. Розглянемо його.

5.4.2. Оптимальна лінійна фільтрація фільтром Калмана-Бьюсі

Описаний вище метод привів до характеристик оптимального фільтра Колмогорова-Вінера. Можна поступити інакше — знайти диференціальні рівняння, які визначають оптимальний фільтр. Це робиться на основі метода змінних стану, який дає можливість розв'язати задачу вилучання з найменшою похибкою повідомлення із спостережуваного сигналу.

Нехай спостережуваний сигнал заданий рівнянням

$$Y(t) = S(t)f(t) + \Xi(t), \quad 0 \leq t \leq T, \quad (5.54)$$

де $\Xi(t)$ — білий гаусів шум з нульовим середнім та одnobічною спектральною густиною, $S(t)$ — повідомлення, яке породжується лінійним стохастичним диференціальним рівнянням

$$\frac{dS(t)}{dt} = -\alpha \cdot S(t) + \alpha \cdot \Xi_1(t). \quad (5.55)$$

Тут $\Xi_1(t)$ – білий шум з нульовим середнім та однобічною спектральною густиною N_1 ; $f(t)$ – переносник, який описується детермінованою відомою функцією.

Рівняння (5.55) означає, що повідомлення $S(t)$ розглядається як результат проходження білого шуму $\Xi_1(t)$ через лінійне коло – у даному разі інтегруючу схему RC . Для останній стала величина $\alpha = 1/RC$. Тоді повідомлення $S(t)$ є гаусовим марківським процесом з кореляційною функцією

$$R_S(\tau) = P_S \cdot e^{-\alpha|\tau|}, \quad P_S = \frac{N_1 \cdot \alpha}{t}$$

та спектральною густиною

$$G_S(f) = \frac{2 \cdot \alpha \cdot P_S}{\alpha^2 + 4 \cdot \pi^2 \cdot f^2}.$$

Рівняння (5.54) називають *рівнянням спостереження*, а (5.55) – *рівнянням стану*.

Оптимальна лінійна фільтрація забезпечується, якщо у гільбертовому просторі випадкових величин вектори похибки \vec{E} та прийнятого сигналу \vec{Y} ортогональні. Можна довести, що необхідною та достатньою умовою оптимальної лінійної текучої фільтрації є умова

$$M[E(t)Y(t-\tau)] = 0 \text{ для всіх } \tau. \quad (5.56)$$

Ця умова означає, що похибка $E(t)$ некорельована з сигналом $Y(t)$ у всі миті часу.

На основі умови (5.56) можна одержати рішення у вигляді рекурентного співвідношення

$$\frac{d\hat{S}}{dt} = -\alpha \cdot \hat{S}(t) + \frac{2 \cdot \delta}{N_0} \cdot f(t) \cdot [Y(t) - f(t) \cdot \hat{S}(t)], \quad (5.57)$$

$$\frac{d\delta}{dt} = \frac{N_1}{2} - 2 \cdot \alpha \cdot \delta - \frac{2}{N_0} \cdot f^2 \delta^2. \quad (5.58)$$

Рівняння (5.57) та (5.58) називають *рівняннями фільтра Калмана-Бьюсі*. При цьому рівняння (5.57) визначає алгоритм знаходження

оцінки (інакше кажучи – структурну схему фільтра), а (5.58) – похибку фільтрації

$$\delta = ME^2(t) = M[S(t) - S^*(t)]^2.$$

Тепер наведемо більш загальні результати. Ще раз підкреслимо, що сигнал, що спостерігається на вході приймача системи зв'язку, ми розглядаємо як вихідний сигнал деякої умовної системи. У цій системі є своє вхідне збурення у вигляді білого шуму, а також своє спотворення сигналу, що трактується як похибка вимірювання. Це також білий шум. Спостережувана система описується відповідним диференціальним рівнянням: сигнал, що формується на її виході, є сигналом на вході приймача реального каналу зв'язку.

Основний результат теорії лінійної фільтрації нестационарних процесів можна сформулювати у вигляді такого твердження.

Теорема 5.3 (Фільтр Калмана-Бьюсі)

Нехай стан $\vec{X}(t)$ стохастичної системи з неперервним часом і процес спостереження $\vec{Y}(t)$, $t_0 \leq \tau \leq t$, описуються співвідношеннями:

$$\vec{X}(t) = A(t) \cdot \vec{X}(t) + \vec{V}(t), \quad M\{\vec{X}(t_0)\} = 0;$$

$$\vec{Y}(t) = C(t) \cdot \vec{X}(t) + \vec{W}(t).$$

Тоді лінійна незсунена оцінка стану $\vec{X}^*(t)$ з мінімальною середньою квадратичною похибкою описується рівнянням

$$\frac{d}{dt} \vec{X}^*(t) = A(t) \cdot \vec{X}^*(t) + K(t) \{ \vec{Y}(t) - C(t) \cdot \vec{X}^*(t) \},$$

$$\vec{X}^*(t_0) = 0.$$

Матриця підсилення (коефіцієнт передачі) описується формулою

$$K(t) = \tilde{P}(t) \cdot C^T(t) + N^{-1}(t).$$

Кореляційна матриця похибок оцінювання описується матричним диференціальним рівнянням типу Ріккати:

$$\frac{d}{dt} \tilde{P}(t) = A(t) \cdot \tilde{P}(t) + \tilde{P}(t) \cdot A^T(t) - \tilde{P}^T(t) \cdot C^T(t) \cdot N^{-1}(t) \cdot C(t) \cdot \tilde{P}(t) + Q(t),$$

$$\tilde{P}(t_0) = P(t_0).$$

Тут: $\vec{X}(t)$ – оцінюваний вектор стану вимірності n , що з фізичної точки зору є відшукуваним повідомленням; його початкове значення

$\vec{X}(t_0)$ – випадковий вектор з нульовим математичним сподіванням і заданою кореляційною матрицею

$$R_x(t_0) = M\{\vec{X}(t_0) \cdot \vec{X}^{tr}(t_0)\} = P(t_0);$$

\vec{Y} – спостережуваний вектор вимірності m ; $\vec{V}(t)$, $\vec{W}(t)$ – відповідно вхідне збурення спостережуваної системи та похибки вимірювання. m – вимірні білі шуми з нульовими середніми та кореляційними матрицями

$$R_V(t, \tau) = M\{\vec{V}(t) \cdot \vec{V}^{tr}(\tau)\} = Q(t) \cdot \delta(t - \tau),$$

$$R_W(t, \tau) = M\{\vec{W}(t) \cdot \vec{W}^{tr}(\tau)\} = N(t) \cdot \delta(t - \tau).$$

Початкове значення $\vec{X}(t_0)$, вхідне збурення $\vec{V}(t)$ та похибка вимірювання $\vec{W}(t)$ взаємно некорельовані:

$$M\{\vec{X}(t_0) \cdot \vec{V}^{tr}(\tau)\} = 0,$$

$$M\{\vec{X}(t_0) \cdot \vec{W}^{tr}(\tau)\} = 0,$$

$$M\{\vec{V}(t_0) \cdot \vec{W}^{tr}(\tau)\} = 0.$$

Похибка оцінювання $\tilde{\vec{X}}(t) = \vec{X}(t) - \vec{X}^*(t)$; кореляційна матриця похибок для незсуненої оцінки

$$\tilde{P}(t) = M\{\tilde{\vec{X}}(t) \cdot \tilde{\vec{X}}^{tr}(t)\}.$$

Оцінка оптимальна у тому розумінні, що компоненти похибки оцінювання повинні мати мінімальну дисперсію, тобто

$$\text{spur} \tilde{P}(t) \rightarrow \min,$$

де *spur* – слід матриці.

Якщо матриці A , C , Q , N стали у часі, а $f_0 \rightarrow \infty$, спостережуваний процес буде стаціонарним, а сформульована постановка задачі фільтрації нестационарних процесів збігається з задачею Колмогорова-Вінера. Матриця коефіцієнта передачі фільтра Колмогорова-Вінера при цьому визначається стаціонарним розв'язком диференційного рівняння Ріккати:

Приклад 5.1

Розглянемо фільтр Колмогорова-Вінера першого порядку, що збігається з фільтром Калмана-Бьюсі.

Припустимо, що спостережуваний сигнал $Y(t)$ – це сума переданого повідомлення $S(t)$ та білого шуму $W(t)$:

$$Y(t) = S(t) + W(t).$$

Задані спектральні густини процесів $S(t)$ та $W(t)$:

$$G_s(f) = \frac{2}{1 + (2\pi)^2 \cdot f^2}; \quad G_w(f) = 1.$$

Необхідно знайти рівняння фільтра Калмана-Бьюсі та числові значення стаціонарного коефіцієнта передачі, що характеризує фільтр Колмогорова-Вінера.

Факторизуємо спектральну потужність сигналу $S(t)$:

$$G_s(f) = \frac{\sqrt{2}}{1 + i \cdot 2 \cdot \pi \cdot f} \cdot \frac{\sqrt{2}}{1 - i \cdot 2 \cdot \pi \cdot f}.$$

Перший співмножник – це частотна характеристика породного фільтра, що описується диференціальним рівнянням

$$\dot{S} + S = V,$$

де
$$G_v(f) = (\sqrt{2})^2 = 2.;$$

« i » – тут уявна одиниця.

Згідно з цим рівнянням на вході фільтра діє білий шум $V(t)$, а на його виході формується процес $S(t)$.

Отже, шукана модель спостережуваного процесу описується співвідношеннями:

$$\begin{aligned} \dot{S}(t) &= -S(t) + V(t); \\ Y(t) &= S(t) + W(t), \end{aligned}$$

де

$$\begin{aligned} M\{V(t) \cdot V(\tau)\} &= 2\delta \cdot (t - \tau); \\ M\{W(t) \cdot W(\tau)\} &= \delta \cdot (t - \tau). \end{aligned}$$

Тут $A = -1$; $C = 1$; $Q = 2$; $N = 1$.

Рівняння фільтра Калмана-Бьюсі випливає з теореми 5.3. Маємо:

$$\begin{aligned} \dot{S}^*(t) &= -S^*(t) + K(t)\{Y(t) - S^*(t)\}, \\ K(t) &= \tilde{p}(t), \end{aligned}$$

де \tilde{p} – розв’язок скалярного диференціального рівняння Ріккати:

$$\tilde{p}(t) = -2 \cdot \tilde{p}(t) - \tilde{p}^2(t) + 2.$$

Стационарний розв’язок визначається умовою $\dot{\tilde{p}}=0$, тобто

$$\tilde{p}^2 + 2 \cdot p - 2 = 0,$$

звідки
$$\tilde{p}_{1,2} = -1 \pm \sqrt{3}.$$

Значення \tilde{p} має бути (як дисперсія) невід’ємним, тому

$$\tilde{p}(\infty) = -1 + \sqrt{3} = 0,73.$$

Коефіцієнт передачі фільтра Колмогорова-Вінера $K(-\infty) = 0,73$. Фільтр можна замість диференціального рівняння описати також у вигляді частотної характеристики. Маємо

$$\dot{S}^*(t) = -S^*(t) + 0,73 \cdot \{Y(t) - S^*(t)\} = -1,73 \cdot S^*(t) + 0,73 \cdot Y(t).$$

Звідси одержуємо

$$\frac{F_S^*(i2\pi f)}{F_Y(i2\pi f)} = \frac{0,73}{1,73 + i2\pi f}.$$

Структурні схеми фільтра наведені на рис. 5.9.

Тут ліва частина схеми зображує математичну модель спостережуваного сигналу. Так, згідно з рис. 5.9, а) передавачем системи зв’язку передається повідомлення $S(t)$, що його ми формально трактуємо у вигляді вихідного сигналу породного фільтра з частотною характеристикою $\frac{1}{1+i2\pi f}$, на вході якого діє білий шум $V(t)$ зі спектральною густиною $G_V(2\pi f) = 2$. Це фізично передане повідомлення $S(t)$ спотворюється завадою у вигляді адитивного білого шуму $W(t)$ зі спектральною густиною $G_W(2\pi f) = 1$. Результат – прийнятий сигнал $Y(t)$ – подається на фільтр Колмогорова-Вінера, на виході якого формується оптимальна оцінка переданого повідомлення $S^*(t)$.

Аналогічно рис. 5.9, а) у схемі рис. 5.9, б) ліва частина характеризує математичну модель спостережуваного процесу $Y(t)$ (який фізично діє на вході приймача системи зв’язку), а права – дає фільтр Калмана-Бьюсі, який у даному випадку збігається з фільтром Колмогорова-Вінера описаного у вигляді диференціальних рівнянь.

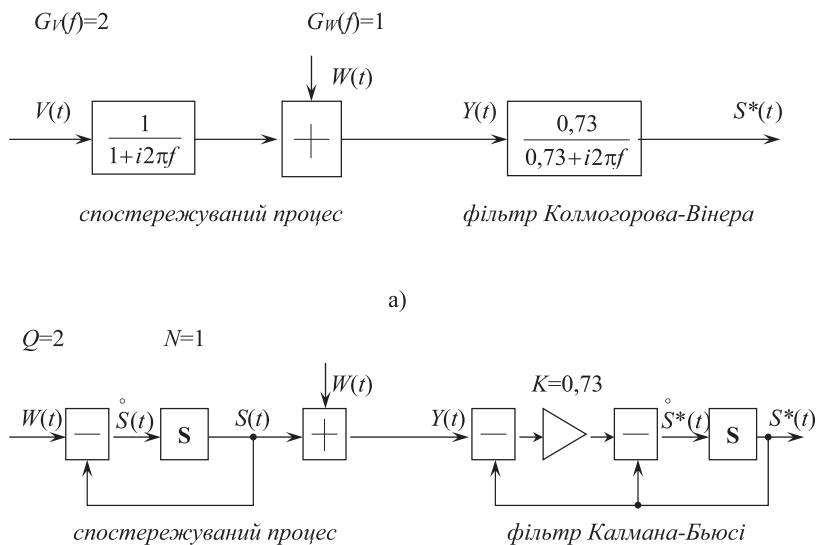


Рис. 5.9. – Спостережуваний процес і фільтр Колмогорова-Вінера (Калмана-Бьюсі): а) частотне зображення, б) зображення у просторі фазових змінних; «i»-уявна одиниця

5.4.3. Оптимальна нелінійна фільтрація

При застосуванні нелінійних методів модуляції переданий сигнал є нелінійною функцією повідомлення. Тому спостережуваний сигнал описується нелінійним диференціальним рівнянням. Це приводить до необхідності використовувати нелінійну фільтрацію.

Загальна постановка задачі збігається з попереднім випадком лінійної фільтрації нестационарних процесів. Тобто з фізичної точки зору на вхід приймача системи зв'язку надходить сигнал $Y(t)$, в якому закодоване (недоступне безпосередньому спостереженню) повідомлення, спотворене завадами. Для сигналу $Y(t)$ формується відповідна математична модель у вигляді спостережуваної системи.

Розглянемо спочатку деякий частинний випадок, коли повідомлення $S(t)$ являє собою марківський процес.

Запишемо рівняння спостереження у вигляді

$$Y(t) = x(t, S) + \Xi(t), \quad 0 \leq t \leq T, \quad (5.59)$$

де $x(t, S)$ – нелінійна функція повідомлення $S(t)$; $\xi(t)$ – білий гаусів шум нульовим середнім значенням та спектральною щільністю N_0 .

Рівняння стану, як і у попередньому випадку, задамо у вигляді

$$\frac{dS}{dt} = -\alpha S(t) + \alpha \Xi(t),$$

де покладено $N_1 = 4/\alpha$. Тоді $S(t)$ – гаусів марковський дифузійний процес з одичиною дисперсією. Для цього процесу щільність імовірності визначається рівнянням Фоккера-Планка-Колмогорова, яке у даному випадку має вигляд

$$\frac{\partial}{\partial t} p(s, t) = -\frac{\partial}{\partial s} [\alpha s p(s, t)] + \frac{1}{4} N_1 \frac{\partial^2}{\partial s^2} p(s, t). \quad (5.60)$$

Вибираючи за критерій оптимальності критерій мінімуму середньоквадратичної похибки

$$\min ME^2(t) = \min M[S(t) - S^*(t)]^2,$$

можна одержати оптимальну оцінку, якою є математичне сподівання апостеріорного розподілу

$$S^*(t) = M[p(s|y)] = \int_S s p(s|y) ds.$$

При цьому похибка оцінювання характеризується апостеріорною дисперсією

$$\sigma_S^2 = \int_S (s^* - s)^2 p(\bar{s}|\bar{y}) ds.$$

Звідси витікає, що для знаходження оптимальної оцінки та її похибки необхідно мати щільність апостеріорного розподілу $p(s|y)$, або, що еквівалентно, $p(s)$ та $p(y|s)$. Функція $p(s)$ визначається з рівняння (5.60), а $p(y|s)$ – може бути знайдена з рівняння (5.59). Враховуючи, що $x(t, s)$ є відомою функцією аргументів t та s , а завада – гаусова бачимо, що розподіл $p(y|s)$ також має гаусів розподіл.

У цих умовах оцінка $S^*(t)$ є оптимальною не тільки за критерієм мінімуму середньоквадратичної похибки, а й за критерієм максимуму апостеріорної ймовірності.

Наведемо деякі конкретні співвідношення.

Відомо, що апостеріорна щільність ймовірності для реалізації $S(t)$ повідомлення (у скінчену мить часу спостереження) може бути знайдена з нелінійного диференціального рівняння

$$\frac{\partial}{\partial t} p(s|y) = \frac{\partial}{\partial s} [A_1(t,s)p(s|y)] + \frac{1}{2} \frac{\partial^2}{\partial s^2} [A_2(t,s)p(s|y)] + [F(t,s) - \overline{F(t,s)}] p(s|y). \quad (5.61)$$

Тут A_1 , A_2 – відповідно, коефіцієнти зносу та дифузії,

$$F(t,s) = \frac{d}{dt} \ln p(y|s).$$

З точністю до сталих величин функція $F(t,s)$ у загальному випадку має вигляд

$$F(t,s) = -\frac{1}{N_0} [y(t) - x(t,s)]^2,$$

а для неенергетичних параметрів сигналу – вигляд

$$F(t,s) = \frac{2}{N_0} y(t)x(t,s).$$

Для великих значеннях відношення сигнал-шум та значному часі спостереження апостеріорна щільність імовірності може бути апроксимована гаусівською

$$p(s|y) = \frac{2}{\sqrt{2\pi\sigma_s^2(t)}} \exp\left[-\frac{[s(t) - s^*(t)]^2}{2\sigma_s^2(t)}\right]. \quad (5.62)$$

Це гаусівське наближення у теорії нелінійної фільтрації. При цьому оцінкою переданого повідомлення $s(t)$ є середнє значення $s^*(t)$, а похибка фільтрації визначається дисперсією $\sigma_s^2(t)$.

Повернемось тепер до обговорення принципових результатів у теорії нелінійної фільтрації. Розглянемо векторне стохастичне диференціальне рівняння

$$d\vec{X}(t) = \vec{f}(\vec{X}, t)dt + V(\vec{X}, t)d\vec{B}(t), \quad t \geq 0, \quad (5.63)$$

де $\vec{X}(t)$ – векторний процес стану, недоступний спостереженню. Спостерігається векторний процес $\vec{Y}(t)$, що залежить від $\vec{X}(t)$ та похибки спостереження $W(t)d\vec{\Gamma}(t)$:

$$d\vec{Y}(t) = \vec{\eta}(\vec{X}, t)dt + W(t)d\vec{\Gamma}(t), \quad t \geq 0. \quad (5.64)$$

Тут \vec{X} , \vec{f} – n -вимірні, \vec{Y} , $\vec{\eta}$ – m -вимірні векторні процеси; $\vec{B}(t)$ та $\vec{\Gamma}(t)$ – складаються із взаємно незалежних стандартних вінерових процесів, тобто

$$M d\vec{B} d\vec{B}^{tr} = Idt;$$

$$M d\vec{\Gamma} d\vec{\Gamma}^{tr} = Idt.$$

Припускається також, що матриця W не залежить від \vec{X} , вона має ранг m ; кореляційна матриця похибок спостереження є симетричною невід'ємно визначеною матрицею розміром $m \times m$;

$$W(t)W^{tr}(t) = P(t).$$

Ставиться задача оцінювання умовній щільності ймовірності $p^*(\vec{x}, t | \vec{y}_{[0, t]})$ величини $\vec{X}(t)$ при заданому значенні $\vec{y}_{[0, t]}$ всієї реалізації $\vec{y}(\tau)$ на інтервалі $0 \leq \tau \leq t$. Справджується таке твердження.

Теорема 5.4 (Кушнера)

Нехай процеси $\vec{X}(t)$ та $\vec{Y}(t)$ задовольняють відповідно рівняння (5.63) та (5.64) за умов $VV^{tr} = Q$ та $WW^{tr} = P$. Тоді

$$dp^* = p^*(\vec{x}, t + dt | \vec{y}_{[0, t+dt]}) - p^*(\vec{x}, t | \vec{y}_{[0, t]}) = - \sum_{i=1}^n \frac{\partial (f_i p^*)}{\partial x_i} dt + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 (g_{ij} p^*)}{\partial x_i \partial y_j} dt + p^* \{ \vec{\eta}(\vec{X}, t) - \vec{\eta}(t) \}^{tr} p^{-1}(t) \{ d\vec{X}(t) - \vec{\eta}^*(t) dt \}. \quad (5.65)$$

Тут $\vec{\eta}^*$ – умовне математичне сподівання $\vec{\eta}$ при заданому $\vec{y}_{[0, t]}$;
 f_i – i -та компонента вектора $\vec{f}(\vec{x}, t)$;

g_{ij} – елемент матриці $G(\vec{x}, t) = V(\vec{x}, t)V^{tr}(\vec{x}, t)$.

Стохастичне диференціальне рівняння (5.63) повністю розв'язує задачу нелінійної фільтрації. Зауважимо, що розв'язання цієї задачі суттєво пов'язано з марківськими процесами та рівнянням Фоккера-Планка-Колмогорова, тому що перші два члени у (5.63) – це і є багатовимірне рівняння Фоккера-Планка-Колмогорова). Зауважимо також, що для лінійних систем при $\vec{f} = A\vec{X}$, $\vec{\eta} = C\vec{X}$ та Q , що не залежить від \vec{X} , одержується фільтр Калмана-Бьюсі та диференціальне рівняння Ріккати.

Приклад 5.2

Розглянемо нелінійну фільтрацію у гаусовому наближенні при деяких додаткових спрощуючих припущеннях.

Припустимо, що $A_1(s,t)$ залежить лінійно від $S(t)$, а коефіцієнт $A_2(s,t) = \text{const}$. Тоді, підставляючи (5.62) у (5.61), можна одержати систему нелінійних диференціальних рівнянь

$$\begin{cases} \frac{d}{dt} S^*(t) = A_1(t, s^*) + K(t) \frac{\partial F(t, s^*)}{\partial s^*}; \\ \frac{d}{dt} K(t) = 2K(t) \frac{\partial A_1(t, s^*)}{\partial s^*} + K^2(t) \frac{\partial^2 F(t, s^*)}{\partial s^{*2}} + A_2, \end{cases} \quad (5.66)$$

де $K(t) = \sigma_s^2(t)$.

Для гаусова марківського повідомлення, заданого у вигляді

$$\frac{dS}{dt} = -\alpha S(t) + \alpha \Xi_1(t)$$

маємо $A_1(t, s^*) = -\alpha s(t)$, $A_2(t, s^*) = \frac{1}{2} N_1$.

При цьому рівняння (5.66) далі можуть бути спрощені і приведені до вигляду

$$\begin{cases} \frac{d}{dt} S^*(t) = -\alpha S^*(t) + k(t) \frac{\partial F(t, s^*)}{\partial s^*}; \\ \frac{d}{dt} k(t) = -2\alpha k(t) + \frac{1}{2} N_1 + k^2(t) \frac{\partial^2 F(t, s^*)}{\partial s^{*2}} \end{cases} \quad (5.67)$$

Якщо $X(t, s) = S(t)f(t)$, маємо

$$F(t, s) = \frac{1}{2} N_0 [Y(t) - f(t)S(t)]^2$$

і рівняння (5.67) приводяться до вигляду

$$\begin{cases} \frac{d}{dt} S^*(t) = -\alpha S^*(t) + \frac{2k}{N_0} f(t) [Y(t) - f(t)S^*(t)]; \\ \frac{dk}{dt} = \frac{N_1}{2} - 2\alpha k - \frac{2}{N_0} f^2 k^2, \end{cases} \quad (5.68)$$

що з точністю до позначень збігають з рівнянням (5.57)-(5.58) фільтра Калмана-Бьюсі.

Наведені рівняння дають відповідний нелінійний (або лінійний — у останньому випадку) фільтр, який відновлює передане повідомлення $S(t)$ з мінімальною середньоквадратичною похибкою.

Висновки

1. Теорія статистичних рішень з точки зору розв'язання основних задач приймання сигналів у торії електрозв'язку охоплює три розділи — перевірка гіпотез, оцінювання параметрів розподілів та фільтрація повідомлень.

З формальної точки зору кожна з названих задач може розглядатися як частинний випадок наступної — перевірка гіпотез є частинним випадком задачі оцінювання параметрів, які приймають дискретні значення; оцінювання параметрів є частинним випадком фільтрації повідомлень.

2. При математичній постановці задачі вибору гіпотез у загальному випадку задаються функції правдоподібності, ймовірності гіпотез, матриця втрат. Обсяг необхідної апріорної інформації залежить від обраного показника якості рішень та критерію їх оптимальності.

Найбільш поширені у теорії електрозв'язку такі критерії: байєсівський, мінімум середньої ймовірності похибки, Неймана-Пірсона, максимум правдоподібності. Найбільш широким є клас байєсівських вирішуючих правил, куди входять вирішуючі правила, одержані і за іншими важливими критеріями оптимальності.

Найпростіша задача у теорії перевірки гіпотез є задача перевірки простої гіпотези проти простих альтернатив. Задача перевірки гіпотез із заважаючими параметрами за рахунок усереднення по заважаючих параметрах може бути зведена до попереднього випадку.

У задачах перевірки складної гіпотези проти складних альтернатив в загальному випадку вирішуючі правила вже не мають структури байєсівських.

3. Задача оцінювання параметрів розподілів узагальнює задачу перевірки гіпотез; тут параметри звичайно набирають значення з континуальної множини.

З принципіальної точки зору постановки задач оцінювання параметрів збігається з відповідними постановками здач вибору гіпотез.

На відміну від задач вибору гіпотез у теорії оцінювання параметрів розглядаються якісні характеристики одержаних оцінок, основними з них є: незсуненість, ефективність, обгрунтованість.

Для порівняння якості різних оцінок параметрів важлива нижня границя дисперсії несунених оцінок, яка задається нерівністю Крамера-Рао. Відповідні узагальнення існують і для оцінок векторних параметрів.

4. Задача фільтрації повідомлення узагальнює задачу оцінювання параметрів – її можна розглядати як узагальнення задачі оцінювання параметрів, які змінюються у часі.

Серед задач фільтрації повідомлення важливі три випадки: лінійна фільтрація стаціонарних випадкових процесів, лінійна фільтрація нестаціонарних випадкових процесів та нелінійна фільтрація.

Лінійна фільтрація стаціонарних у широкому розумінні випадкових процесів забезпечується фільтром Колмогорова-Вінера, який мінімізує середньоквадратичну похибку відновлення.

Частотна характеристика оптимального фільтра обчислюється через спектральні щільності повідомлення та шуму.

Лінійна фільтрація нестаціонарних випадкових процесів забезпечується фільтром Калмана-Бьюсі, який також мінімізує середньоквадратичну похибку. Рівняння фільтра Калмана-Бьюсі визначають алгоритм знаходження оцінок повідомлення (інакше кажучи – структурну схему фільтра) та похибку фільтрації. На відміну від фільтрації стаціонарних процесів у даному випадку при постановці задачі задаються рівняння спостереження та рівняння стану.

В задачах нелінійної фільтрації передане повідомлення нелінійно «закодоване» у прийнятому сигналі.

Задача фільтрації формулюється у термінах методу змінних стану – задається рівняння спостереження та рівняння стану. Найчастіше за критерій оптимальності вибирається критерій мінімуму середньоквадратичної похибки. При цьому для марківського повідомлення оптимальною оцінкою повідомлення є математичне сподівання апостеріорного розподілу, а похибкою оцінювання є апостеріорна дисперсія. У загальному випадку суттєве значення у теорії нелінійної фільтрації має гаусове наближення, яке вірне для великих значень відношення сигнал-шум та значному часі спостереження.

Запитання, задачі й вправи

1. Як пов'язані об'єктивна реальність, теорія ймовірностей та математична статистика?

2. Поясніть фізичну суть перевірки гіпотез, оцінювання параметрів та фільтрації.

3. Що таке показник якості рішень та критерій їх оптимальності? Наведіть приклади показників якості та критеріїв оптимальності рішень для різних задач.

4. Сформулюйте постановку задачі перевірки гіпотез.

5. Сформулюйте постановку задачі оцінювання параметрів розділів.

6. Сформулюйте задачу фільтрації повідомлень.

7. Наведіть вирішуюче правило перевірки гіпотези проти однієї альтернативи, оптимальне за критерієм ідеального спостерігача. У чому полягає його узагальнення у багатоальтернативній задачі?

8. У чому полягає узагальнення вирішуючих правил з питання 7 у задачах перевірки гіпотез із заважаючими параметрами?

9. Що таке максимально правдоподібна оцінка параметра розподілу? Наведіть рівняння правдоподібності.

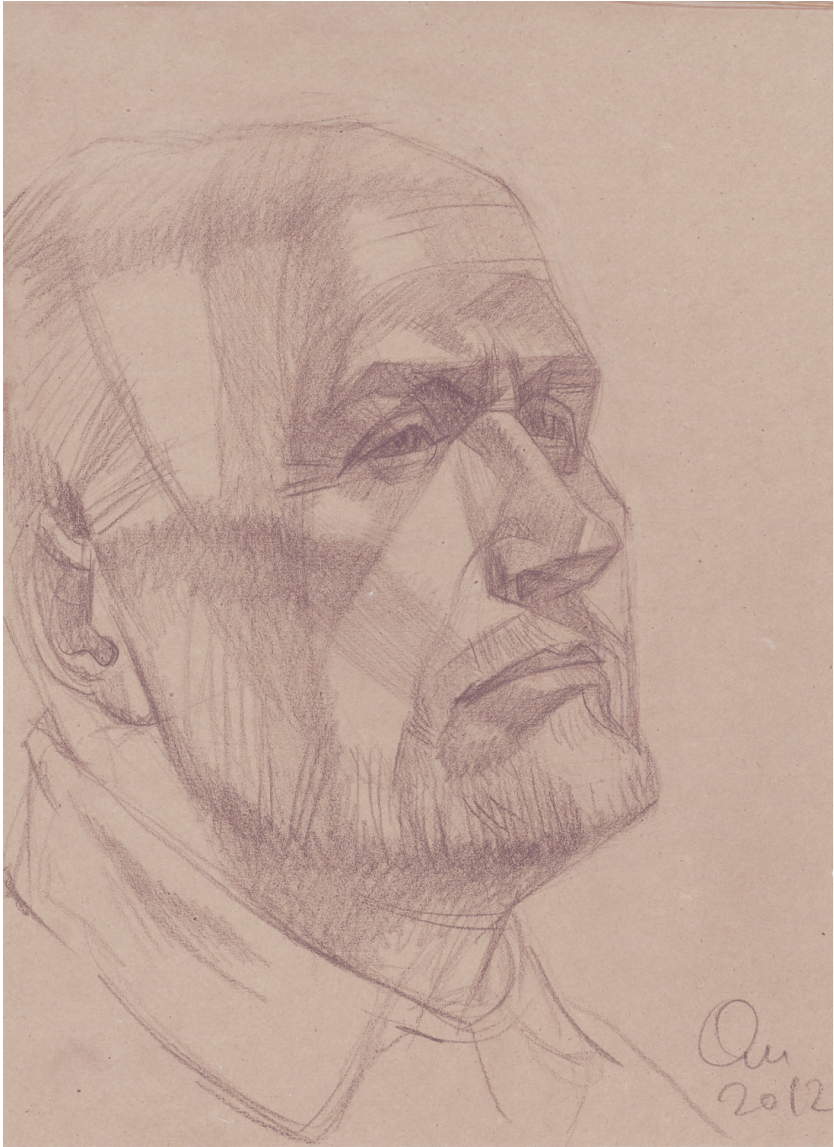
10. Що таке нерівність Крамера-Рао для оцінки скалярного параметра? Як вона узагальнюється для векторного параметра?

11. Що таке послідовні алгоритми вибору гіпотез та оцінювання параметрів?

12. Як ставиться та розв'язується задача оптимальної лінійної фільтрації стаціонарних випадкових процесів?

13. У чому полягає особливість фільтрації фільтром Калмана-Бьюсі? Що таке рівняння спостереження і стану, а також рівняння фільтра Калмана-Бьюсі?

14. У чому полягає особливість нелінійної фільтрації? Наведіть приклад рівняння спостереження, стану та розв'язання задачі для марківського повідомлення.



Віктор Олександрович Омельченко у пенсійний період

ПІСЛЯМОВА

Прошло багато часу з тих пір, коли читався цей курс. Однак, математика не старіє. Більш того, актуальність цих матеріалів у нашому випадку зростає.

Вища школа почала деградувати ще в останній період існування Радянського Союзу. Після його розвалу у 1991-му році, в Україні і далі руйнувалась освіта. Викладачам і науковим робітникам по півроку не видавалась заробітна плата. Деякі викладачі Вищої школі навіть почали брати плату за іспити, заліки тощо. До чого це привело? У 2018-му році ректор Харківського національного університету радіоелектроніки у газеті «Самопоміч» писав, що те, що: « в нашій країні називається «вищою світою» фактично в масі своїй є просто розподілом дипломів за гроші».

Ситуація далі ще більше ускладнювалась. Спочатку пандемія коронавірусу примусила, принаймні частково, перейти на систему роботи «on line». Широкомасштабне вторгнення Росії в Україну і систематичне на протязі двох років руйнування інфраструктури і житлового фонду в країні і далі заводило Вищу освіту майже «у глухий кут».

Але близьке майбутнє України – це членство у Євросоюзі. Після закінчення війни з Росією і перемоги Вища школа не тільки буде повинна відбудовуватися, але й перебудовуватися, щоб піднятися до найкращих зразків у Західній Європі та США. При розбудові такої Системи освіти ця книга стане невеликою відповідною цеглинкою у будові.

*25 Лютого 2024 р.
Віктор Омельченко*

СПИСОК ЛІТЕРАТУРИ

1. *Марченко Б.Г., Омельченко В.А.* Вероятностные модели случайных сигналов и полей в прикладной статистической радиофизике. – К.:УМКВО, 1988.
2. Імовірнісні моделі випадкових сигналів та полів у прикладах і задачах. Навчальний посібник / за редакцією *В.О. Омельченка*. – К.:ІСДО. 1996
3. *Омельченко В.О. Санніков В.Г.* Теорія електричного зв'язку. ч.1 – 1994, ч.2 – 1995. ч.3 – 1997. Підручник – К.:МОУ.
4. *Колмогоров А.Н., Фомин С.В.* Элементы теории функций и функционального анализа. – М.:Наука, 1972.
5. *Вентцель.Е.С.* Теория вероятностей. - М.:Гос. изд. ф.-м. лит. – 1962.
6. *Гнеденко Б.В.* Курс теории вероятностей. – М.:Наука,1988.
7. Прикладна теорія випадкових сигналів і полів. Колективна монографія / за ред. *Я.П. Драгана В.О. Омельченка*, – Харків, Львів, Тернопіль: АНУ, 1993.
8. *Марченко Б.Г.* Метод стохастических интегральных представлений и его приложение в радиотехнике. – К.:Наука. Думка,1973.
9. *Драган Я.П.* Структура и представление стохастических сигналов. – К.:Наука. думка, 1980.
10. *Левин Б.Р., Шварц В.* Вероятностные модели и методы в системах связи и управления. М.:Радио и связь, 1985.
11. *Колесник В.Д., Полтырев Г.Ш.* Курс теории информации. – М.:Наука. 1982.
12. *Гихман И.И., Скороход А.В., Ядренко М.И.* Теория вероятностей и математическая статистика. – К.:Вища школа.1988.
13. *Тихонов В.И.* Оптимальный прием сигналов. – М.:Радио и связь, 1983.
14. *Браммер К., Зифлинг.* Фильтр Калмана-Бьюси. – М.:Наука, 1982.

Рецензия на книгу В.А. Омельченко «Теорія ймовірностей у електров'язку»

В книге Виктора Александровича Омельченко с единых теоретико-множественных позиций изложены различные ветви математики, или ветви теории вероятностей в широком смысле.

В первом разделе описывается математический аппарат современной теории вероятностей и ее ветвей: элементы теории множеств, мера, измеримые функции, интеграл по мере, функциональные пространства.

Второй раздел посвящен положениям классической теории вероятностей. Выбранный — идеологический — метод изложения позволил автору при сравнительно небольшом объеме материала дойти до описания современных результатов теории вероятностей.

Третий раздел посвящен теории случайных процессов. Изложены классические и современные результаты этой теории. Особый интерес для прикладников представляет расширение идеи теоремы Винера-Хинчина-Бохнера о существовании энергетического спектра стационарного случайного процесса (относительно континуального базиса гармоник) на нестационарные случайные процессы с конечной энергией относительно любого счетного ортогонального базиса гильбертова пространства детерминированных функций.

Четвертый раздел описывает положения теории информации К. Шеннона. В нем использована техническая терминология, характерная для теории связи. Материал излагается методически весьма удачно — от простого к сложному. Рассмотрен, по сути, один и тот же круг вопросов: в дискретном канале связи без помех, в дискретном канале с помехами и в непрерывном канале связи (как для передачи дискретных, так и непрерывных сообщений).

Пятый раздел посвящен части математической статистики — теории статистических решений, применяемой при обработке сигналов. Математическим терминам выбор (проверка) гипотез, оценка параметров распределений, фильтрация придан физический смысл обнаружения сигналов, оценка их параметров и фильтрация сообщений. Обсуждена задача статистического синтеза алгоритмов обработки сигналов для указанных задач по различным критериям оптимальности. Изложены положения байесовской теории при различных частных случаях байесовского критерия. Рассмотрены вопросы линейной фильтрации аналоговых сообщений фильтром Колмогорова-Винера и фильтром Калмана-Бьюси.

Особенность книги — метод изложения теорий: сначала автор дает строгие определения, а далее, основываясь на них, приводит принципиальные результаты и дает их трактовку в прикладной области, прежде всего, в области теории связи. Такой метод изложения позволяет при сравнительно небольшом объеме разделов обсудить достаточно большой объем результатов — от классических до современных.

Материал удачно методически оформлен — в конце каждого раздела даются выводы и приводятся примеры и задачи, решение которых позволяет закрепить изучаемый материал. Достоинством книги является изложение материала на государственном языке.

Книга может быть опубликована в открытой печати. Она может быть полезна студентам, аспирантам и научным работникам в области связанных специальностей, а также радиотехнических, радиофизических и других смежных специальностей.

Рецензент:
Доктор физико-математических наук, профессор
А.С. Мазманишвили

ЗМІСТ

Передмова	5
1. Математичний апарат сучасної теорії ймовірностей та її гілок.....	6
1.1. Елементи теорії множин	6
1.2. Міра, вимірні функції, інтеграл за мірою.....	13
1.3. Функціональні простори.....	19
1.4. Аксиоматика А.М. Колмогорова теорії ймовірностей.....	29
Висновки	32
Запитання, задачі й вправи.....	32
2. Класична теорія ймовірностей	34
2.1. Випадкові події.....	34
2.2. Випадкові величини.....	42
2.3. Класичні закони розподілу ймовірностей.....	51
2.4. Випадкові вектори.....	61
Висновки	87
Запитання, задачі й вправи.....	89
3. Випадкові процеси.....	92
3.1. Основні означення та підходи до вивчення випадкових процесів	92
3.2. Інтегральне зображення випадкових процесів зі скінченними енергетичними характеристиками	107
3.3. Зображення рядами випадкових процесів зі скінченними енергетичними характеристиками	114
3.4. Приклади класів випадкових процесів	128
Висновки	146
Запитання, задачі й вправи.....	148
4. Теорія інформації.....	153
4.1. Передавання інформації по дискретному каналі без завад.....	153

4.2. Передавання інформації по дискретному каналу із завадами.....	170
4.3. Передавання інформації по неперервному каналу	183
Висновки	204
Запитання, задачі й вправи.....	205
5. Теорія статистичних рішень.....	208
5.1. Постановка задач статистичного синтезу	209
5.2. Перевірка (вибір) гіпотез.....	223
5.3. Оцінювання параметрів розподілів.....	235
5.4. Фільтрація повідомлень.....	244
Висновки	259
Запитання, задачі й вправи.....	260
Післямова.....	263
Список літератури	264
Рецензія на книгу В.А. Омельченко «Теорія ймовірностей у електров'язку»	265

Для нотаток

Для нотаток

Для нотаток

Навчальне видання

Омельченко Віктор Олександрович

Теорія ймовірностей у електровзв'язку

Підп. до друку 25.03.2024. Формат 60x84/16.

Папір офсетний. Ум. друк. арк. 17.

Замовл. № 03-25-01. Наклад 100 прим.

Видання та друк ВПП «Контраст»

Україна, 61166, м. Харків, пр. Науки, 40, оф. 504Б.

Св-во: ДК №1778 від 05.05.2004

тел. (050) 343-53-23, (098) 581-56-00.

www.kontrast.kh.ua

e-mail: kontrast.main@gmail.com

ISBN 978-617-7405-74-9



9 786177 405749