

МІЖРЕГІОНАЛЬНА  
АКАДЕМІЯ УПРАВЛІННЯ ПЕРСОНАЛОМ



МАУП

Методичні матеріали  
щодо забезпечення самостійної роботи студентів  
з дисциплін

**“СТАТИСТИЧНЕ МОДЕЛЮВАННЯ”**  
**“АНАЛІЗ ДАНИХ”**  
(для бакалаврів)

МАУП

Київ  
ДП «Видавничий дім «Персонал»  
2010

Підготовлено доцентом кафедри прикладної математики та програмування  
*В. І. Панчуком*

Затверджено на засіданні кафедри прикладної математики  
та програмування (протокол № 8 від 03.04.08)

*Схвалено Вченою радою Міжрегіональної Академії управління персоналом*

**Панчук В. І.** Методичні матеріали щодо забезпечення самостійної роботи студентів з дисциплін “Статистичне моделювання” та “Аналіз даних” (для бакалаврів). — К.: ДП «Видавничий дім «Персонал», 2010. — 68 с.

Методична розробка містить пояснювальну записку, тематичний план, теоретичний мінімум, завдання для самостійної роботи і рекомендації до їх виконання, питання для самоконтролю, список літератури

- © Міжрегіональна Академія управління персоналом (МАУП), 2010
- © ДП «Видавничий дім «Персонал», 2010

## **ПОЯСНЮВАЛЬНА ЗАПИСКА**

Навчальний курс “Статистичне моделювання” є одним із курсів математичної освіти за фахом “Прикладна математика”. Він передбачає викладення засад методу статистичних випробувань (методу Монте-Карло) та побудови й аналізу найважливіших математичних моделей, оснований на цьому методі. Статистичне моделювання нерозривно пов’язано з іншим курсом — статистичного аналізу даних.

Метою вивчення обох курсів є актуалізація раніше вивченого слухачами матеріалу з різних предметів, зокрема з теорії ймовірностей та математичної статистики; формування у них наочного уявлення про можливості статистичних методів; подальше поглиблення ймовірнісної інтуїції мислення.

Основні завдання курсів — навчити студентів будувати статистичні моделі для аналізу складних систем, сприяти набуттю навичок практичного досвіду розв’язання задач методами статистичного моделювання, до яких застосування інших класичних методів неможливе або ускладнене та неефективне через наявність багатьох різномірних змінних тощо. Студенти повинні опанувати основні алгоритми моделювання рівномірно, нормально розподілених випадкових величин та випадкових величин з іншими законами розподілу; гауссівських процесів та процесів випадкового блукання; лінійної регресії; обчислення інтегралів за методом Монте-Карло; різновидів стохастичних систем тощо.

Універсальність і міждисциплінарність статистичного методу дозволяє використовувати його для розв’язання та аналізу ймовірнісних задач математичної фізики, техніки, екології, економіки тощо. Знання, які одержують студенти, можуть бути використані при вивченні та поглибленні різних дисциплін, зокрема таких, як “Дослідження операцій”, “Теорія масового обслуговування”, “Чисельні методи”.

Сучасний фахівець з прикладної математики повинен вміти самостійно підходити до вивчення будь-якого об’єкта природи за допомогою апарату математичного моделювання. Для цього йому потрібно опанувати методи побудови, структурної та параметричної ідентифікації моделей, а також методів їх дослідження. Останні найчастіше залежать від задачі, яка ставиться при дослідженні об’єкта.

**Метою дисципліни** є формування системи знань з методології та інструментарію статистичного чи імітаційного моделювання, побудови і дослідження моделей, їх аналізу та використання.

**Завданнями вивчення** дисципліни є:

- теоретико-практична підготовка студентів стосовно основ статистичного моделювання реальних процесів;
- опанування ними практичних засобів моделювання різних процесів.

### **ТЕОРЕТИЧНИЙ МІНІМУМ: ОСНОВИ СТАТИСТИЧНОГО МОДЕЛЮВАННЯ ТА АНАЛІЗУ ДАНИХ І МОДЕЛЕЙ**

Метою самостійної роботи є отримання студентами знань та навичок постановки задач статистичного моделювання, проведення статистичного аналізу даних та об'єкта моделювання. Побудовані моделі, з одного боку, мають відповідати вимогам максимальної простоти, а з іншого, — адекватно описувати процеси, за якими проводяться спостереження.

#### **Вимірювання випадкових величин на комп'ютері**

Дослідження властивостей математичних моделей випадкових величин пов'язане з проведенням статистичних експериментів, для яких потрібно мати наготові послідовності чисел, що певною мірою розглядаються як реалізація послідовності незалежних випадкових величин. Здебільшого необхідні для експериментів послідовності чисел одержують на комп'ютерах за допомогою різних програм, здатних виробляти детерміновані послідовності чисел достатньо складної нерегулярної структури, схожі на послідовності випадкових чисел. Оскільки такі послідовності чисел мають відмінності від “справжніх” випадкових послідовностей, зумовлені обмеженою машинною розмірністю чисел, то їх називають послідовностями псевдовипадкових чисел, а самі програми їх обчислення — датчиками псевдовипадкових чисел.

#### **Програмна реалізація датчиків псевдовипадкових чисел**

Обчислення псевдовипадкових чисел на комп'ютері відбувається, як правило, за рекурентними формулами:

$$\alpha_{i+1} = F(\alpha_i), \quad i = 0, 1, 2, \dots$$

які мають певний період  $L$ , зумовлений обмеженою машинною розмірністю чисел.

Перейдемо до простого способу побудови послідовних псевдовипадкових чисел  $v_0, v_1, \dots, v_n$  на відрізку  $[0;1]$  з розподілом, близьким до рівномірного, за рекурентною формулою

$$v_{n+1} = \{Kv_n\}, \quad n = 0, 1, \dots,$$

де  $\{x\}$  — дробова частина числа  $x = Kv_n$ ,  
 $K$  — ціле число, яке визначає властивості  $\{v_n\}$ .

Щоб дробова частина числа  $\{v_n\}$  не була занадто регулярною і мала достатньо великий період, доречно в її якості обирати ціле двозначне число, яке при діленні дає залишок  $\pm 3$ .

а) **Модельовання випадкової події  $A$  з можливістю настання  $P(A) = p$**

Позначимо через  $r_i$  числа, отримані при зверненні до генератора ПВЧ рівномірно розподілених на  $[0,1]$ .

А настає тоді, коли  $r \leq p$ ; у протилежному випадку відбувається подія  $\bar{A}$  з імовірністю  $r > p$ .

Дійсно

$$P(r < p) = \int_0^p f(r) dr = P(A) = p.$$

б) **Модельовання групи несумісних подій  $A_1, \dots, A_n$  з відомими ймовірностями  $P(A_i) = p_i, (p_0 = 0) \quad i = \overline{1, n}$**

Відкладемо ці ймовірності на відрізку  $[0,1]$ .

Якщо число  $r$  попало  $\sum_{n=0}^{k-1} p_i < r \leq \sum_{n=0}^k p_i$ , то подія  $A_k$  відбулась.

Цю процедуру називають визначенням результату випробування за жеребом. Вона ґрунтується на формулі:

$$P \left\{ \sum_{i=0}^{k-1} p_i < r \leq \sum_{i=0}^k p_i \right\} = p_k = P(A_k)$$

де  $p_0 = 0$ .

в) **Модельовання дискретної випадкової величини**

Дискретну випадкову величину  $X$

Можливе значення $X$	$x_1$	$x_2$	...	$x_n$
Ймовірність $p_i$	$p_1$	$p_2$	...	$p_i$

можна подати, як повну групу несумісних подій

$$A_1 = (X = x_1), \quad A_2 = (X = x_2), \dots, A_n = (X = x_n).$$

г) **Модельовання умовної події А при настанні події В з умовною ймовірністю P(A/B)**

Спочатку моделюємо подію  $B_1$ . Якщо вона настає, то моделюємо відбування події А. Якщо маємо  $\bar{B}$ , то відбування події А не моделюємо:

д) **Псевдовипадкові числа, рівномірно розподілені на відріжку [0;1]**

Згадаємо, що рівномірний розподіл на відріжку  $[a, b]$  має щільність

$$f(x) = p(x) = \begin{cases} c = \text{const}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

З умови нормування функції розподілу  $F(x) = \int_{-\infty}^{+\infty} p(x) dx = 1$  знаходимо

$$F(x) = \int_{-\infty}^{+\infty} c dx = c \int_a^b c dx = cx \Big|_a^b = c(b-a) = 1 \Rightarrow c = \frac{1}{b-a};$$

Отже,

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b], \end{cases}$$

$$MX = \int_{-\infty}^{+\infty} xp(x) dx = \dots = \frac{a+b}{2}; \quad DX = \int_{-\infty}^{+\infty} (x - \frac{a+b}{2})^2 p(x) dx = \dots = \frac{(b-a)^2}{12}.$$

Побудуємо функцію рівномірно розподіленої величини

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

Очевидно, коли  $a=0, b=1$ , то функція рівномірного розподілу набуває вигляду

$$P(x) = \begin{cases} 1 & x=0; \\ 0 & x \neq 0; \end{cases} \quad MX = \frac{1}{2}; \quad DX = \frac{1}{12}.$$

Нехай  $a=-h, b=h$ , тоді

$$P_h(x) = \begin{cases} \frac{1}{2h}, x \in [-h, h] \\ 0, x \notin [-h, h] \end{cases}$$

При  $h \rightarrow 0$  маємо

$$\lim_{h \rightarrow 0} P(x) = \begin{cases} \infty & x = 0; \\ 0 & x \neq 0; \end{cases} \equiv \delta(x)$$

З умови нормування  $\int_{-\infty}^{\infty} \delta(x) dx = 1$

“Функція”  $\delta$  не є функцією в класичному розумінні. Це так звана узагальнена функція (функція Дірака), яка має застосування при дослідженнях у квантовій механіці та різних розділах фізики, механіки.

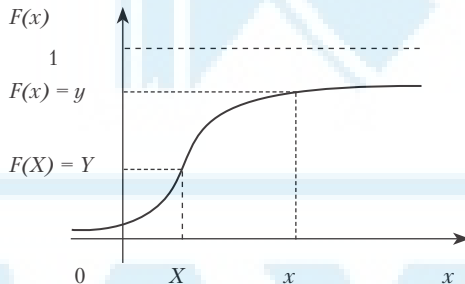
### Моделювання неперервних випадкових величин

Для моделювання неперервних випадкових величин використовується метод оберненої функції.

Нехай є деяка функція розподілу  $y = F(x)$ . На осі ординат розіграємо точку  $y \in [0, 1]$ , використовуючи функцію  $y = F(x)$ .

Можна довести твердження: якщо на осі ординат взяти випадкове число  $Y \in [0; 1]$  і знайти значення  $x$ , при якому  $F(x) = y$ , то випадкова величина  $x = F^{-1}(y)$  буде мати функцію розподілу  $F(x)$ .

Знайдемо  $F(x)$  випадкової величини  $X$ . За означенням вона дорівнює ймовірності  $P(X < x)$ .



З рис. випливає, що

$$P(X < x) = P(Y < y = F(x)) = \int_0^{F(x)} f(y) dy = F(x)$$

Отже, послідовність  $\{y_1, y_2, y_3, \dots\} \in Y[0, 1]$  перетворюється на послідовність  $\{x_1, x_2, x_3, \dots\}$  із функцією щільності  $f(x)$ .

**а) Моделювання рівномірно розподіленої випадкової величини на відрізку  $[a, b]$**

Для моделювання рівномірно розподіленої випадкової величини використаємо метод оберненої функції, за яким

$$y = F(x) = \int_a^x \frac{dt}{b-a} = \frac{t}{b-a} \Big|_a^x = \frac{x-a}{b-a} \Rightarrow x = y(b-a) + a$$

**б) Моделювання експоненціального розподілу випадкової величини**

Функція щільності

$$f(x) = \begin{cases} 0, & x \geq 0, \\ \lambda e^{-\lambda x}, & x, 0 \end{cases}$$

$$y = F(x) = \lambda \int_0^x e^{-\lambda t} dt = -e^{-\lambda t} \Big|_0^x = 1 - e^{-\lambda x} \Rightarrow y = 1 - e^{-\lambda x} \Rightarrow e^{-\lambda x} = 1 - y \Rightarrow$$

$$-\lambda x = \ln(1 - y) \Rightarrow x = -\frac{1}{\lambda} \ln(1 - R)$$

Величина  $1-R$  розподілена так само, як і  $R$ . Тоді, зробивши заміну  $1-R = R$ , знайдемо, що  $x = -\frac{1}{\lambda} \ln R$

**в) Моделювання нормального закону розподілу випадкових величин**

При моделюванні цього закону безпосередньо не можна скористатися методом оберненої функції. Тут використовують центральну граничну теорему (ЦГТ). Пояснимо, як саме.

Нехай  $X \sim N(m_x, \delta_x)$ , а  $Z \sim N(0, 1)$ . Тоді  $x = \delta_x Z + M_x$ . ЦГТ творить:

Якщо  $\{X_i\}_1^n$  — послідовність незалежних випадкових величин із середнім значенням  $MX_i = a$  і  $DX_i = \delta^2$ ,  $i = \overline{1, n}$ , то при  $n \rightarrow \infty$  функція розподілу таких випадкових величин

$$\bar{X}(n) = \frac{\frac{1}{n}(x_1 + \dots + x_n) - a}{\delta/\sqrt{n}} = \frac{(\bar{X}(n) - a)\sqrt{n}}{\delta} \quad (*)$$



прямує до стандартного розподілу  $\Phi(z)$ ,  
тобто

$$F_{\bar{X}(n)} \rightarrow \Phi(z) ,$$

$$\text{де } \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{y^2}{2}} dy$$

Тепер для одержання нормального закону розподілу випадкових величин достатньо підсумувати декілька значень розподілу випадкових величин, отриманих за допомогою генератора псевдовипадкових чисел (достатньо шести значень) і пронормувати отримані значення так, щоб одержати  $Z$  і за формулою (\*) перейти до  $X$ .

Зазвичай підсумовують 12 випадкових величин

$$Z = \sum_{i=1}^{12} R_i ,$$

щоб дисперсія дорівнювала 1.

### Розв'язування диференціальних рівнянь

Розглянемо спочатку простий приклад розв'язування крайової задачі

$$\frac{\partial^2 y}{\partial x^2} = 0; y(0)=2, y(1) = 3. \quad (1)$$

Апроксимуючи другу похідну симетричним різницеvim різностороннім співвідношенням, замість крайової задачі (1) отримуємо різницеву задачу

$$\frac{y(x-h) - 2y(x) + y(x+h)}{h^2} = 0; y(0) = 2, y(1) = 3, \quad (2)$$

де  $h$  — крок у напрямку  $x$ .

Виберемо  $h = 0,2$  і знайдемо розв'язок системи (2) в деякій сітковій точці проміжку  $[0, 1]$ , наприклад, у точці  $x = 0,6$ , за так званим методом Монте-Карло.

Нижче наведемо алгоритм знаходження розв'язку за таким методом.

Промодельюємо  $n$  раз процес блукання фіктивної частинки, вважаючи, що частинка може знаходитись тільки в одній із шести сіткових точок  $x = 0; x = 0,2; x = 0,4; x = 0,6; x = 0,8; x = 1$ . При цьому кожне блукання починатимемо з точки  $x=0,6$  і закінчуватимемо кожен раз при умові попадання частинки в одну з граничних точок  $x = 0$  і  $x = 1$ .

Переміщення частинки в сусідні точки зліва і справа є рівноймовірні (тобто виконуються з вірогідністю 0,5). Після такого процесу шукане наближення різницевої задачі (2), а, отже, даної крайової диференціальної задачі (1), виходить як середнє арифметичне з найдених граничних значень.

Практичне моделювання траєкторії блукання частинки в кожному випробуванні можна виконати, наприклад, шляхом підкидання правильної монети. При цьому потрібно заздалегідь домовитися, що випадання, наприклад герба, відповідає переміщенню частинки вправо, а випадання цифри — вліво.

Нижче наведемо результат такого моделювання для траєкторій 20 частинок ( $n = 20$ ):

$$y(0,6) \approx \frac{1}{20}(2+3+3+3+2+3+2+2+2+3+3+3+3+2+3+3+3+2+3) = 2,65$$

Для порівняння наведемо точний розв'язок задачі (1):

$$y = x + 2 \Big|_{x=0,6} = 2,60$$

Загальна ідея наближеного розв'язування звичайних диференціальних рівнянь, розглянута нами на простому прикладі, лежить в основі використання методу Монте-Карло для наближеного розв'язування диференціальних рівнянь з частинними похідними. При цьому як основний математичний апарат використовуються також і ланцюги Маркова.

Нехай, наприклад, потрібно розв'язати крайову задачу.

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0; \quad (x, y) \in G; \quad (3)$$

$$u = \varphi, \quad (x, y) \in \Gamma; \quad (4)$$

Відомо, що коли область  $G$ , в котрій шукається розв'язок задачі (3), (4), відрізняється від прямокутника чи круга, то точний аналітичний розв'язок такої задачі практично знайти неможливо. Разом з цим на практиці часто виникає необхідність у розв'язуванні задачі (3), (4) для довільних областей інтегрування  $G$ . В тих випадках задачу розв'язують наближено. При цьому найчастіше, а при використуванні комп'ютерів майже завжди, використовують метод скінчених різниць.

Покриємо область  $G$  квадратною сіткою з кроком  $h$  і всі вузли цієї сітки розіб'ємо на два класи: внутрішні і граничні вузли. Вузол називається внутрішнім, якщо всі його чотири сусідні вузли, знаходячись на відстані  $h$ , належать області  $\bar{G}$ . В протилежному випадку вузол називається граничним. Позначмо множину внутрішніх вузлів сітки через  $G_h$ , а граничних – через  $\Gamma_h$ .

Згідно з методом скінчених різниць диференціальну задачу (3),(4) апроксимуємо наступною різницевою задачею:

$$\frac{U_{A_1} - 2U_A + U_{A_3}}{n^2} + \frac{U_{A_2} - 2U_A + U_{A_4}}{n^2}, \quad A \in G_h; \quad (5)$$

$$U_B = \varphi_B, \quad (6)$$

де, наприклад,  $U_A$  – значення функції  $U$  в точці  $A$ , а  $A_1, A_2, A_3, A_4$  – найближчі до вузла  $A$  (рис. 1.). Значення функції  $\varphi$  в граничних вузлах  $B$  покладемо рівними  $\varphi_{B'}$ , де  $B'$  – найближча до вузла  $B$  точка на границі  $\Gamma$ .

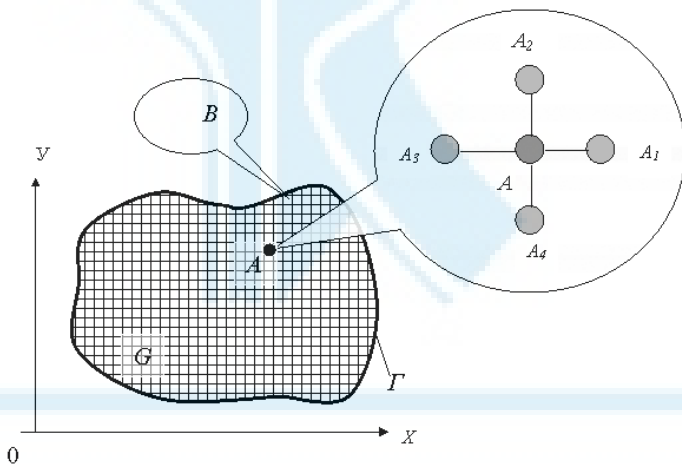


Рис. 1.

Відомо, що при деяких умовах розв'язування задачі (5),(6) при  $h \rightarrow 0$  рівномірно збігається в вузлах сітки до розв'язку задачі (3),(4).

Основна складність, котра виникає на практиці при використанні методу кінцевих різниць, полягає в фактичному розв'язуванні різни-

цевих задач. Зазвичай їх розв'язують яким-небудь детерміністичним методом, найчастіше методом ітерацій. Однак для цієї цілі можна успішно використовувати метод Монте-Карло, що зводиться в цьому випадку до простежування достатньо великої кількості траєкторій блукаючої фіктивної частинки вузлами вибраної квадратної сітки.

Позначуємо через  $A \rightarrow B$  подію, яка полягає в тому, що частинка, вийшовши з вузла  $A$ , прийде в вузол  $B$ , через  $A \rightarrow A_j \rightarrow B$  подію, що полягає в тому, що частинка, вийшовши через вузол  $A$ , прийде в вузол  $B$  і при цьому неодмінно побуває в вузлі  $A_j$ .

Зрозуміло, що настання події  $A \rightarrow B$  рівносильне настанню одного з декількох чотирьох неспільних подій:

$$A \rightarrow A_1 \rightarrow B, A \rightarrow A_2 \rightarrow B, A \rightarrow A_3 \rightarrow B, A \rightarrow A_4 \rightarrow B.$$

Ось чому, використовуючи теорему складання ймовірностей неумісних подій, отримуємо

$$P(A \rightarrow B) = \sum_{j=1}^u P(A \rightarrow A_j \rightarrow B). \quad (7)$$

Але події  $A \rightarrow A_j$  і  $A_j \rightarrow B$  можна вважати незалежними і, отже, по теоремі множення ймовірностей матимемо

$$P(A \rightarrow A_j \rightarrow B) = P(A \rightarrow A_j)P(A_j \rightarrow B), \quad i=1,2,3,4.$$

Підставимо ці вирази в формулу (7)

$$P(A \rightarrow B) = \sum_{j=1}^u P(A \rightarrow A_j)P(A_j \rightarrow B). \quad (8)$$

У найпростішому випадку блукання частинки ймовірності подій  $A \rightarrow A_1$ ,  $A \rightarrow A_2$ ,  $A \rightarrow A_3$ ,  $A \rightarrow A_4$  однакові і рівні  $\frac{1}{4}$ . Отже, формулу (8) можна переписати у вигляді

$$P(A \rightarrow B) = \sum_{j=1}^u P(A_j \rightarrow B). \quad (9)$$

Будемо припускати, що при попаданні частинки в граничний вузол, процес його блукання закінчується. Тоді можна написати

$$P(B \rightarrow B) = 1, \quad P(B \rightarrow \bar{B}) = 0, \quad (10)$$

де  $B$  і  $\bar{B}$  — довільні граничні вузли ( $\bar{B} \neq B$ ).

Нехай границя  $\Gamma_n$  складається з  $k$  граничних вузлів  $B_1, B_2, \dots, B_k$ . Позначуємо через  $U_A$  дискретну випадкову величину, закон розподілення ймовірностей котрої має вид

$$P(U_A = \varphi_{B_i}) = P(A \rightarrow B_i), \quad i = 1, 2, \dots, k.$$

Буква А в позначенні  $U_A$  показує, що випадкова величина  $U_A$  залежить від початкового вузла А процесу блукання частинки.

Покажемо, що математичне сподівання

$$MU_A = \sum_{i=1}^K P(A \rightarrow B_i) \varphi_{B_i} \quad (11)$$

цієї випадкової величини задовольняє різницевому рівнянню (5) і крайовій умові (6).

Справді, покладаючи в рівності (9)  $B = B_j$ , помножуючи його на  $\varphi_{B_j}$  і сумуючи результат по  $i=1, 2, \dots, k$ , отримуємо

$$MU_A = \frac{1}{4} \sum_{j=1}^K MU_{A_j}.$$

Але це рівняння можна представити у вигляді

$$\frac{MU_{A_1} - 2MU_A + MU_{A_3}}{h^2} + \frac{MU_{A_2} - 2MU_A + MU_{A_4}}{h^2} = 0,$$

тобто у вигляді рівняння (5). Далі, покладаючи у рівності (11)  $A = B$ , з урахуванням формул (10) знаходимо  $MU_B = \varphi_B$ , тобто крайова умова (6).

Таким чином, рівність

$$U_A = MU_A \quad (12)$$

можна розглядати у як імовірнісну модель задачі (5),(6).

Але згідно із законом великих чисел при достатньому великому

числі блукань  $n$  матимемо  $P(A \rightarrow B_j) \approx \frac{m_j}{n}$ ,

де  $m_j$  число блукань частинок, які стартували у внутрішньому вузлі А і фінішували у граничному вузлі  $B_j$ .

Тому в силу формул (11),(12) можна написати

$$U_A = MU_A = \sum_{j=1}^k P(A \rightarrow B_j) \varphi_{B_j} \approx \frac{1}{n} \sum_{j=1}^k m_j \varphi_{B_j}.$$

Зазвичай цю формулу записують у вигляді

$$U_A \approx \frac{1}{n} \sum_{j=1}^k \varphi_{B_j}, \quad (13)$$

Припускаючи, що деякі доданки в сумі можливо співпадають між собою. Формула (13) означає, що за приближене значення розв'язку

системи (5),(6) у вузлі  $A$  можна взяти середнє арифметичне з достатньо великого числа отриманих граничних значень.

Практично кожне з  $n$  блукань частинок можна здійснити таким чином. У вузлі  $A$ , для котрого потрібно знайти розв'язок системи (5),(6), за допомогою таблиці випадкових чисел визначають напрямки першого переміщення частинки. Наприклад, якщо чергове число з таблиці двозначних чисел виявиться в котрій із областей 00–24, 25–49, 50–74, 75–99, то вважатимемо, що частинка перемістилася відповідно направо, вверху, вліво, вниз. Потім цей процес спостереження за траєкторією блукання фіктивної частинки повторює доти, поки частинка не попаде в якийсь граничний вузол. При цьому немає потреби запам'ятовувати проміжні значення місцезнаходження частинки в процесі її блукання. Важливий тільки старт – фіксований вузол  $A$  і фініш – випадковий граничний вузол  $B_j$ . Ця обставина має велике значення при використанні комп'ютерів, оскільки у цьому випадку не потрібно великого обсягу запам'ятовуючого пристрою обчислювальної машини.

Розглянутий метод розв'язування системи (5),(6) за допомогою статистичного моделювання, зрозуміло, дозволяє знаходити приблизне значення функції  $U$  тільки в одному довільному вузлі  $A$ . Ця властивість методу Монте-Карло може виявитися корисною в тих випадках, коли потрібно знайти розв'язок крайової задачі не в усій області інтегрування, а тільки в деякій її частці.

Іншою особливістю методу Монте-Карло є можливість одночасного моделювання процесу блукань декількох частинок, які стартували з даного вузла  $A$ , що дозволяє в деяких випадках суттєво зменшити час розв'язування цим методом.

Чудовою особливістю методу Монте-Карло є також можливість використання одних і тих же блукань частинок для розв'язування крайової задачі з різними граничними умовами (з різною функцією  $\varphi$ ). У цьому відношенні є деяка аналогія з методом розв'язування крайових задач за допомогою функції Гріна. Більше того, одні й ті ж блукання можна використовувати для розв'язування того ж диференційного рівняння, але область інтегрування котрого міститься у середині даної області інтегрування.

Це безпосередньо слідує з факту перетину границі внутрішньої області траєкторіями блукань частинок, які стартують в цій області і фінішують на границі зовнішньої області.

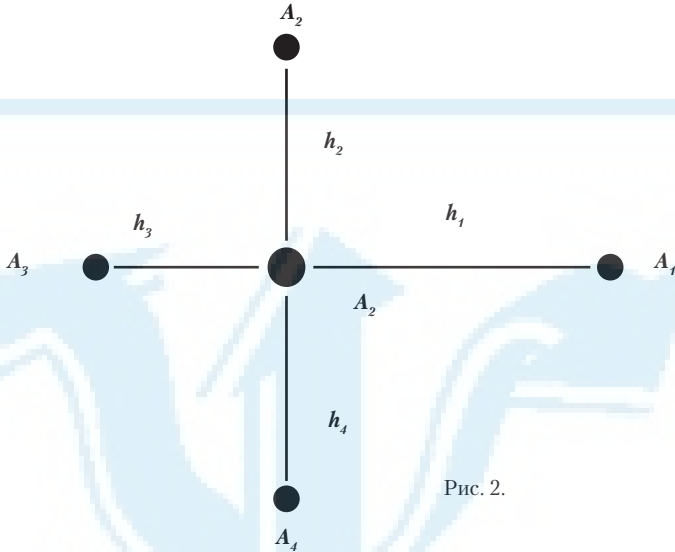


Рис. 2.

Якщо використовувати нерівномірну сітку відповідно до рис.2, то, використовуючи наближені формули

$$\frac{\partial^2 U_A}{\partial x^2} \approx \frac{\frac{U_{A_1} - U_A}{h_1} - \frac{U_A - U_{A_3}}{h_2}}{\frac{h_1 + h_2}{2}} =$$

$$= \frac{2}{h_1(h_1 + h_3)} U_{A_1} + \frac{2}{h_3(h_1 + h_3)} U_{A_3} - \frac{2}{h_1 + h_3} \left( \frac{1}{h_1} + \frac{1}{h_3} \right) U_A;$$

$$\frac{\partial^2 U_A}{\partial y^2} = \frac{2}{h_2(h_2 + h_4)} U_{A_2} + \frac{2}{h_4(h_2 + h_4)} U_{A_4} - \frac{2}{h_2 + h_4} \left( \frac{1}{h_2} + \frac{1}{h_4} \right) U_A,$$

отримуємо різницеве рівняння, яке апроксимує диференціальне рівняння (3)

$$U_A = \sum_{j=1}^4 P_j U_{A_j},$$

$$\text{в якому } P_1 = \frac{h_2 h_3 h_4}{(h_1 + h_2)(h_1 h_3 + h_2 h_4)}; \quad P_2 = \frac{h_1 h_3 h_4}{(h_2 + h_4)(h_1 h_3 + h_2 h_4)};$$

$$P_3 = \frac{h_1 h_2 h_4}{(h_1 + h_2)(h_1 h_3 + h_2 h_4)}; \quad P_2 = \frac{h_1 h_2 h_4}{(h_2 + h_4)(h_1 h_3 + h_2 h_4)};$$

причому  $\sum_{j=1}^4 P_j = 1$ .

У цьому випадку, вважаємо, що частинка переміщається із вузла  $A$  у сусідні вузли  $A_1, A_2, A_3, A_4$  відповідно з ймовірностями  $P_1, P_2, P_3, P_4$ , тобто

$$P(A \rightarrow A_j) = P_j \quad j = 1, 2, 3, 4.$$

При цьому рівність (8) набуває вигляду

$$P(A \rightarrow B) = \sum_{j=1}^4 P_j P(A_j \rightarrow A).$$

Практично моделювання першого кроку цього блукання здійснюється так. Нехай  $\rho$ -значення рівномірно розподіленої в інтервалі  $(0, 1)$  випадкової величини. Тоді,

якщо  $0 < \rho < P_1$ , то зміщення в  $A_1$ ;

якщо  $P_1 \leq \rho < P_1 + P_2$ , то зміщення в  $A_2$ ;

якщо  $P_1 + P_2 \leq \rho < P_1 + P_2 + P_3$ , то зміщення в  $A_3$ ;

якщо  $P_1 + P_2 + P_3 \leq \rho < 1$ , то зміщення в  $A_4$ ;

Послідуючі кроки моделювання блукання фіктивної частинки виконуються аналогічно.

Метод Монте-Карло можна використовувати також і для приближеного розв'язування диференційних рівнянь із змінними коефіцієнтами. Однак у цьому випадку ймовірності переходу частинки з вузла в сусідні вузли залежить від координат цього вузла.

При розв'язуванні крайових задач методом Монте-Карло важливо мати уявлення про тривалість одного випробування (блукання). Для цієї мети зручно використовувати математичне сподівання  $M_A$  числа кроків, які необхідно зробити блукаючій частинці, котра вийшла з внутрішнього вузла  $A$ , щоб досягти якого-небудь граничного вузла.

Формально величину  $M_A$  можна знайти так. За визначенням математичного сподівання маємо

$$M_A = \sum_{k=1}^{\infty} kP(A, k), \quad (14)$$

де  $P(A, k)$  — ймовірність того, що частинка, вийшовши з внутрішнього вузла  $A$ , досягне якого-небудь граничного вузла рівно за  $k$  кроків.



Міркуючи таким же чином, як і при виводі рівняння (9), отримуємо

$$P(A, k) = \frac{1}{4} \sum_{j=1}^4 P(A_j, k-1), \quad (15)$$

Тут  $P(A_j, k-1)$  — ймовірність того, що частинка, вийшовши з внутрішнього вузла  $A_j$ , досягне якого-небудь граничного вузла рівно за  $k-1$  кроків. Помножуючи рівність (15) на  $k-1$  і додаючи результати від  $k=1$  до  $k=\infty$ , будемо мати

$$\sum_{k=1}^{\infty} (k-1)P(A, k) = \frac{1}{4} \sum_{k=1}^{\infty} \sum_{j=1}^4 (k-1)P(A_j, k-1), \quad (16)$$

Природно припустити, що блукаюча частинка, вийшовши з вузла  $A$ , рано чи пізно попаде в якийсь граничний вузол, тобто ймовірність нескінченного блукання частинки внутрішніми вузлами дорівнює нулю.

Дійсно, позначимо  $P_i(k)$  ймовірність того, що частинка, вийшовши з внутрішнього вузла  $A_i$ , досягне якого-небудь граничного вузла за  $k$  або менше кроків. Очевидно, що  $P_i(k) > 0$ , якщо тільки  $k \geq k_0$ , де  $k_0$  — деяке число. Переходячи до протилежної події, напишемо

$$q_i(k) = 1 - p_i(k) \leq 1 - p,$$

де  $q_i(k)$  — ймовірність того, що частинка, яка вийшла з внутрішнього вузла  $A_i$ , не досягне жодного граничного вузла за  $k$  чи менше кроків, а  $p = \min_i p_i(k)$ .

Нехай за більше або рівно  $k$  кроків частинка перемістилася з вузла  $A_i$  у внутрішній вузол  $A_j$ . Оскільки  $p_j(k) \geq \min_i p_i(k) = p$ , то  $q_j(k) = 1 - p_j(k) \leq 1 - p$ .

Отже, за теоремою множення ймовірностей незалежних подій для ймовірності того, що частинка, яка вийшла з внутрішнього вузла  $A_i$ , не досягне жодного граничного вузла за  $2k$  чи менш кроків, маємо

$$q_i(k) \quad q_j(k) \leq (1-p)^2.$$

Якщо знову за більше або рівно  $k$  кроків частинка перемістилася в деякий вузол  $A_e$ , то для ймовірності того, що частинка, яка вийшла з початкового внутрішнього вузла  $A_i$ , не досягне жодного граничного вузла за  $3k$  чи менше кроків, аналогічно знаходимо

$$q_i(k) \quad q_j(k) \quad q_e(k) \leq (1-p)^3.$$

Продовжуючи так міркувати, в силу нерівності  $1-p < 1$  робимо висновок, що ймовірність того, що частинка, вийшовши з випадко-

вого внутрішнього вузла, не досягне жодного граничного вузла при достатній великій кількості кроків, як завжди мала.

Таким чином, можна написати,  $\sum_{k=1}^{\infty} P(A, k) = 1$ , звідки з урахуванням формул (14) і (16) отримаємо

$$M_A = \frac{1}{4} \sum_{j=1}^{\infty} M_{A_j} + 1. \quad (17)$$

Оскільки блукання частинки закінчується на границі, то природно покласти

$$M_B = 0, \quad (18)$$

де  $B$  – граничний вузол.

Таким чином, математичне сподівання числа кроків блукання частинки, необхідне для досягнення нею якого-небудь граничного вузла, можна знайти, розв'язавши систему різницьових рівнянь (17) з граничною умовою (18).

#### Зауваження.

1. Нехай  $G$  – квадрат  $0 \leq x, y \leq a$  зі стороною довжини  $a$  і  $x = sh$ ,

$y = th$  – координати вузла  $A$ , де  $h = \frac{a}{N}$ .

Тоді можна показати, що

$$M_A = -\frac{2}{N^2} \sum_{k=1}^{N-1} \sum_{e=1}^{N-1} \frac{\frac{k\pi t}{N} \sin \frac{e\pi s}{N} + \sin \frac{k\pi s}{N} \sin \frac{e\pi t}{N}}{\cos l\pi \sin \frac{k\pi}{2N}} \sin \frac{k\pi}{2} \sin \frac{(N-1)k\pi}{2N} \sin \frac{l\pi}{N} * \\ * \sin^2 \frac{l\pi}{2} \left(1 - \frac{\cos \frac{l\pi}{N} + \frac{k\pi}{N}}{2}\right)^2$$

де сумування виконується тільки за непарними індексами  $k$  і  $l$ . Ця формула дозволяє оцінити знизу і зверху величину  $M_A$  для довільних плоских не занадто витягнутих областей інтегрування  $G$ .

2. Для грубої прикидки середнього числа кроків, котрі робить частинка в одному блуканні, у випадку довільної області  $G$  іноді використовують наближену формулу

$$M_A \approx \frac{R^2}{2h^2}.$$

Тут  $R$  – середня відстань даного вузла  $A$  від усіх граничних вузлів;  $h$  – крок квадратної сітки.

3. Можна показати, що математичне сподівання числа кроків у кожному блуканні слабо залежить від розмірності області інтегрування  $G$ . Так, якщо областю інтегрування  $G \in k$ -вимірна сфера, то

$$\max_A M_A = \frac{c(k)v^{\frac{k}{2}}}{h^2}(1 + \varepsilon(k)),$$

де  $v$  — об'єм сфери;  $\lim_{k \rightarrow \infty} \varepsilon(k) = 0$ ;  $c(k)$  — постійна, наприклад,

$$c(1) = 0,25; c(2) = 0,312; c(3) = 0,385.$$

Важливість цього факту зрозуміла, якщо взяти до уваги, що число блукань не залежить від розмірності області інтегрування.

### **АНАЛІЗ ДАНИХ МЕТОДАМИ МАТЕМАТИЧНОЇ СТАТИСТИКИ**

Методи математичної статистики використовуються для визначення достовірності та оцінювання можливих похибок при вимірюванні й аналізі випадкових даних.

Нехай  $x(k)$  — деяка випадкова величина. Тоді для будь-якого фіксованого значення  $x$  випадкова подія  $x(k) \leq x$  визначається як множина усіх можливих наслідків  $k$  таких, що  $x(k) \leq x$ . Можливі наслідки деякого експерименту (або вимірювання) являють собою множину точок, яка називається вибірковим простором.

Якщо область значень випадкової величини неперервна, то **щільність імовірності**  $p(x)$  визначається диференціальним співвідношенням

$$p(x) = \lim_{\Delta x \rightarrow 0} \left[ \frac{\text{Prob}[x < x(k) \leq x + \Delta x]}{\Delta x} \right] \quad (1)$$

Тоді буде

$$p(x) \geq 0; \quad (2)$$

$$\int_{-\infty}^{\infty} p(x) dx = 1; \quad (3)$$

$$p(x) = \int_{-\infty}^{\infty} p(x) dx; \quad \frac{dP(x)}{dx} = p(x); \quad (4)$$

Розглянемо випадкову величину  $x$ . Основними характеристиками цієї величини є середнє значення  $p_x$  та дисперсія  $\sigma_x^2$ , що опису-

ють її центр розсіювання та величину цього розсіювання. Вони визначаються за формулами

$$\mu_x = M[x] = \int_{-\infty}^{\infty} xf(x)dx \quad (5)$$

$$\sigma_x^2 = M[(x - \mu_x)^2] = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x)dx, \quad (6)$$

де  $f(x)$  – щільність імовірності випадкової величини  $x$ . На практиці ці дві характеристики величини  $x$  визначити точно не вдається, бо, як правило, щільність імовірності сповна не відома. У зв'язку з цим доводиться задовольнятися оцінками середнього значення та дисперсії, що одержані за результатами обмеженої кількості спостережень.

Тоді замість (1.5), (1.6) будемо мати

$$\bar{x} = \hat{\mu}_x = \frac{1}{N} \sum_{i=1}^N x_i; \quad (7)$$

$$\bar{s}_b^2 = \bar{\sigma}_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2. \quad (8)$$

$\bar{x}, \bar{s}_b^2$  – вибіркові середнє та дисперсія відповідно. Індекс  $b$  означає, що це – зміщена оцінка.

Для визначення якості чи “правильності” оцінки використовують три основних властивості.

1. Бажано, щоб математичне сподівання (середнє значення) оцінки дорівнювало параметрові, що оцінюється, тобто

$$M[\hat{\phi}] = \phi, \quad (9)$$

де  $\hat{\phi}$  – оцінка параметра  $\phi$ . Якщо ця властивість має місце, то оцінка називається **незміщеною**.

2. Бажано також, щоб середньоквадратична похибка цієї оцінки була найменшою серед усіх можливих оцінок, тобто

$$M[(\hat{\phi}_1 - \phi)^2] \leq M[(\hat{\phi}_2 - \phi)^2], \quad (10)$$

де  $\hat{\phi}_1$  – оцінка, що досліджується,  $\hat{\phi}_2$  – будь-яка інша оцінка. Якщо ця властивість має місце, то оцінка називається **ефективною**.

3. Бажано, щоб оцінка збігалась до параметра, що оцінюється, із імовірністю, що прямує до одиниці при збільшенні розміру вибірки, тобто для будь-якого  $\epsilon > 0$

$$\lim_{N \rightarrow \infty} P\{|\hat{\phi} - \phi| \geq \epsilon\} = 0. \quad (11)$$

У цьому випадку оцінка називається **обґрунтованою**.

## Функція розподілу

Найважливішою з функцій розподілу є гауссова (нормальна) функція розподілу. Поширене застосування у математичній статистиці мають також інші функції розподілу, що тісно пов'язані із нормально розподіленими випадковими величинами.

### Нормальний розподіл

Щільність імовірності та функція розподілу визначаються за формулами

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[ -\frac{(x - \mu_x)^2}{2\sigma_x^2} \right]; \quad (12)$$

$$F(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \int_{-\infty}^x \exp \left[ -\frac{(\varepsilon - \mu_x)^2}{2\sigma_x^2} \right] d\varepsilon; \quad (13)$$

Зручніше користуватись стандартною нормально розподіленою випадковою величиною  $z = \frac{x - \mu_x}{\sigma_x}$ .

Якщо цей вираз підставити у (12), (13), то одержимо стандартну щільність та функцію розподілу з нульовим середнім значенням та дисперсію, що дорівнює одиниці ( $(\mu_x = 0; \sigma_x = 1)$ ):

$$f(x) = (\sqrt{2\pi})^{-1} \exp \left[ -z^2 / 2 \right]; \quad (14)$$

$$F(x) = (\sqrt{2\pi})^{-1} \int_{-\infty}^x \exp \left[ -\varepsilon^2 / 2 \right] d\varepsilon. \quad (15)$$

Позначуємо (для зручності) через  $z_\alpha$  значення  $z$ , що відповідає імовірності

$$F(z_\alpha) = \int_{-\infty}^{z_\alpha} f(z) dz = P\{z \leq z_\alpha\} = 1 - \alpha, \quad (16a)$$

$$\text{або } 1 - F(z_\alpha) = \int_{z_\alpha}^{\infty} f(z) dz = P\{z > z_\alpha\} = \alpha. \quad (16б)$$

Значення  $z_\alpha$ , яке задовольняє рівняння (12), має назву  $100\alpha$  – відсоткової точки нормального розподілу. Стандартна нормальна щільність імовірності унімодальна, монотонно змінюється з обох боків моди й симетрична із точками перегину  $\pm 1$ .

Модою називається максимальне значення (точка максимуму) функції розподілу.

$x^2$ -розподіл.

Нехай  $z_1, z_2, \dots, z_n \in n$  незалежних випадкових величин, кожна з яких має нормальний розподіл з нульовим середнім значенням та одиничною дисперсією. Визначимо нову випадкову величину виду

$$x_n^2 = z_1^2 + z_2^2 + \dots + z_n^2. \quad (17)$$

Випадкова величина  $x^2$  називається  $\chi^2$ -квадрат випадковою величиною з  $n$  — ступенями свободи. Число ступенів свободи  $n$  визначає число незалежних, або “вільних” квадратів, що містяться у цій сумі.

Щільність імовірності  $x^2$  має вигляд

$$f(x^2) = [2^{n/2} \Gamma(n/2)]^{-1} (x^2)^{n/2-1} e^{-x^2/2}, x^2 \geq 0, \quad (18)$$

де  $\Gamma(n/2)$  — гамма-функція.

Відповідна функція розподілу  $x_n^2$ , що дорівнює інтегралу від  $-\infty$  до  $x_n^2$ , називається  $\chi^2$ -квадрат розподілом з  $n$  ступенями свободи.  $100\alpha$  — відсоткові точки  $x^2$ -розподілу позначимо через  $x_{n,\alpha}^2$ , тобто

$$\int_{x_n^2}^{\infty} f(x^2) dx^2 = P\{x_n^2 x_{n,\alpha}^2\} = \alpha; \quad (19)$$

$$x^2 = n; \quad \sigma_{x^2}^2 = 2n. \quad (20)$$

Відзначимо кілька особливостей  $\chi^2$ -квадрат розподілу.

1.  $\chi^2$ -квадрат розподіл фактично є частковим випадком більш загального гамма-розподілу.
2. Важливе значення має розподіл випадкової величини, яка дорівнює квадратному кореню з випадкової величини, що має  $\chi^2$ -квадрат розподіл з двома ступенями свободи ( $x_2$ ), і називається **розподілом Релея**. Розподіл Релея має поширене застосування для розв'язку задач, що пов'язані з влучанням в двовимірну мішень; крім того, він виконує роль граничного розподілу екстремальних значень вузько смугастого гауссового випадкового сигналу, коли ширина сигналу прямує до нуля.
3. Із  $\chi^2$ -квадрат розподілом пов'язаний ще один важливий розподіл, який відповідає випадковій величині, що дорівнює квадратному кореню із  $\chi^2$ -квадрат розподілу з трьома ступенями свободи ( $x_3$ ) й має назву **розподілу Максвелла**.
4.  $\chi^2$ -квадрат розподіл наближається до нормального при збільшенні числа ступенів свободи. Зокрема, для  $n > 30$  величина

$\sqrt{2x_n^2}$  наближено нормальна з середнім  $\mu = \sqrt{2n-1}$  та дисперсією  $\sigma^2 = 1$ .

### ***t*-розподіл Ст'юдента**

Нехай  $y$  і  $z$  – незалежні випадкові величини такі, що  $y$  має  $x_n^2$  – розподіл, а  $z$  – нормальний розподіл з нульовим середнім та одиничною дисперсією. Визначимо нову випадкову величину

$$t_n = \frac{z}{\sqrt{y/n}}. \quad (21)$$

Випадкова величина  $t_n$  має  $t$ -розподіл Ст'юдента з  $n$  ступенями свободи. Відомо, що щільність імовірності випадкової величини  $t_n$  має вигляд

$$f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n} \Gamma(n/2)} \left[1 + \frac{t^2}{n}\right]^{-(n+1)/2} \quad (22)$$

Відповідна функція розподілу обчислюється інтегруванням щільності (18) від  $-\infty$  до  $x_n^2$  й має назву  $t$ -розподілу з  $n$  ступенями свободи.  $100\alpha$  – відсоткова точка  $t$ -розподілу позначається  $t_{n;\alpha}$  та

$$\int_{t_{n;\alpha}}^{\infty} f(t) dt = P\{t_n > t_{n;\alpha}\} = \alpha. \quad (23)$$

Середнє значення та дисперсія дорівнюють

$$M[t] = t = 0; \quad n > 1; \quad M\left[(t_n - \mu_t)^2\right] = \sigma_t^2 = \frac{n}{n-2}. \quad n > 2. \quad (24)$$

Із збільшенням числа ступенів свободи  $t$ -розподіл прямує до нормального розподілу.

### ***F*-розподіл**

Нехай  $y_1$  і  $y_2$  – незалежні випадкові величини такі, що  $y_1$  має  $x_n^2$ -розподіл з  $n_1$  ступенями свободи, а  $y_2$  має  $x_n^2$ -розподіл з  $n_2$  ступенями свободи. Визначимо нову випадкову величину

$$F_{n_1, n_2} = \frac{y_2/n_1}{y_2/n_2} = \frac{y_2 n_2}{y_2 n_1}. \quad (25)$$

Випадкова величина  $F_{n_1, n_2}$  – це  $F$ -розподілена випадкова величина з  $n_1$  і  $n_2$  степенями свободи. Відомо, що щільність імовірності  $F_{n_1, n_2}$  дорівнює

$$f(F) = \frac{\Gamma[(n_1 + n_2)/2] (n_1/n_2)^{n_1/2} F^{n_1/2-1}}{\Gamma(n_1/2) \Gamma(n_2/2) [1 + (n_1 F/n_2)]^{(n_1+n_2)/2}}. \quad (26)$$

Відповідна функція розподілу  $F_{n_1, n_2}$  обчислюється інтегруванням щільності (22) від  $-\infty$  до  $F_{n_1, n_2}$  й називається  $F$ -розподілом з  $n_1$  та  $n_2$  ступенями свободи.  $100\alpha$  -відсоткова точка  $F$ -розподілу позначається  $F_{n_1, n_2; \alpha}$

$$\int_{F_{n_1, n_2; \alpha}}^{\infty} f(F) dF = P\{F_{n_1, n_2} > F_{n_1, n_2; \alpha}\} = \alpha. \quad (27)$$

Середнє значення та дисперсія  $F_{n_1, n_2}$  дорівнюють відповідно

$$M[F_{n_1, n_2}] = \mu F = \frac{n_2}{n_2 - 2}; n_2 > 2; \quad (28)$$

$$M\left[(F_{n_1, n_2} - \mu F)^2\right] = \sigma_F^2 = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}, \quad n_2 > 4. \quad (29)$$

### **Функція правдоподібності**

Нехай маємо вибірку з  $N$  спостережень  $z$ , яку будемо вважати реалізацією  $N$ -вимірної випадкової величини, для якої відомий розподіл імовірності  $p(z/y)$  залежить від невідомого параметра  $y$ . До появи даних  $p(z/y)$  сполучає щільність імовірності із кожним конкретним виходом  $z$  для фіксованого  $y$ . Після появи даних настає момент для того, щоб розглянути різні можливі значення  $y$ , які могли б привести до заданої множини  $z$  фактично одержаних спостережень. Для цієї мети підходить **функція правдоподібності**  $L(y/z)$ , яка має ту ж саму форму, що й  $p(z/y)$ , але у якій тепер  $z$  фіксований, а  $y$  змінний.

Часто зручніше мати справу із логарифмічною функцією правдоподібності  $\ln L(y/z) = I(y/z)$ , яка містить довільну адитивну сталу. Однією із причин, завдяки яким функція правдоподібності має фундаментальне значення у теорії оцінювання, є її зв'язок з “принципом правдоподібності”, згідно з яким усе, що можна дізнатися про параметри моделі із даних, міститься у функції правдоподібності, усі ж інші аспекти даних не мають відношення до справи.

Значення параметрів, що доставляють максимум функції правдоподібності, називаються оцінками максимальної правдоподібності (МП). Другі похідні логарифмічної функції правдоподібності дають міру “розтягнутості” функції правдоподібності й можуть використовуватись для обчислення наближених стандартних похибок оцінок. Граничні властивості оцінок максимальної правдоподібності зазви-



чай доводяться для незалежних спостережень, але можуть використовуватись й для оцінок параметрів стаціонарних часових рядів.

### Вибіркові розподіли

Розглянемо випадкову величину  $x$  з функцією розподілу  $F(x)$ . Нехай  $x_1, x_2, \dots, x_N$  — вибірка із  $N$  спостережень  $x$ . Будь-яка величина, яка обчислена за цими вибірковими даними, також буде випадковою. Розглянемо, наприклад, середнє значення  $\bar{x}$ . Якщо взяти декілька вибірок розміром  $N$  з однієї й тієї ж випадкової величини  $x$ , то значення  $\bar{x}$ , що обчислені за різними вибірками, будуть, взагалі кажучи, різними. Таким чином,  $\bar{x}$  також випадкова величина. Функція розподілу цієї величини називається вибірковим розподілом  $\bar{x}$ .

#### 1. Розподіл вибіркового середнього з відомою дисперсією

Розглянемо вибіркове середнє вибірки з  $N$  незалежних спостережень випадкової величини  $x$ :

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (30)$$

Нехай спочатку випадкова величина  $x$  розподілена нормально з середнім  $\mu_x$  та відомою дисперсією  $\sigma_x^2$ . Вибірковий розподіл вибіркового середнього  $\bar{x}$  також нормальний. Середнє значення вибіркового розподілу величини  $\bar{x}$  дорівнює  $\bar{x} = \mu_x$ . Дисперсія  $\sigma_{\bar{x}}^2 = \sigma_x^2 / N$ . У зв'язку з цим можна стверджувати, що

$$P \left\{ \bar{x} > \left( \frac{\sigma_x^2 z_\alpha}{\sqrt{N}} + \mu_x \right) \right\} = \alpha. \quad (31)$$

Нехай тепер розподіл випадкової величини  $x$  відрізняється від нормального. Згідно з центральною граничною теоремою маємо наступний результат. Із збільшенням розміру вибірки  $N$  вибірковий розподіл вибіркового середнього  $\bar{x}$  прямує до нормального розподілу незалежно від розподілу початкової величини  $x$ .

Практично у багатьох випадках вибірковий розподіл випадкової величини  $\bar{x}$  вже для  $N > 4$  можна вважати нормальним.

Розглянемо дисперсію вибірки з  $N$  незалежних спостережень випадкової величини  $x$ :

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2. \quad (32)$$

Якщо випадкова величина  $x$  розподілена нормально з середнім  $\mu_x$  та дисперсію  $\sigma_x^2$ , то розподіли правої та лівої частин співвідношення

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sigma_x^2 x_n^2, \quad n = N - 1$$

збігаються. Значить, вибірковий розподіл вибіркової дисперсії  $s^2$  визначається із співвідношення

$$\frac{ns^2}{\sigma_x^2} = x_n^2, \quad n = N - 1. \quad (33)$$

Тому відносно можливих значень вибіркової дисперсії  $s^2$  можна стверджувати, що

$$P \left\{ s^2 > \frac{\sigma_x^2 x_n^2}{n} \right\} = \alpha. \quad (34)$$

## 2. Розподіл вибіркового середнього з невідомою дисперсією

Розглянемо вибіркове середнє вибірки з  $N$  незалежних спостережень випадкової величини  $x$ , що визначається формулою (1.26). Якщо  $x$  розподілена нормально з середнім  $\mu_x$  та невідомою дисперсією, то із співвідношень (1.17), (1.29) випливає, що

$$\frac{(\bar{x} - \mu_x)}{s / \sqrt{N}} = \frac{\sigma_x z / \sqrt{N}}{\sqrt{\sigma_x^2 x_n^2 / n} / \sqrt{N}} = \frac{z}{\sqrt{x_n^2 / n}} = t_n,$$

де  $t_n$  має  $t$ -розподіл з  $n = N - 1$  ступенями свободи. Значить, вибірковий розподіл вибіркового середнього  $\bar{x}$  з невідомою  $\sigma_x^2$  визначається співвідношенням

$$\frac{(\bar{x} - \mu_x) \sqrt{N}}{s} = t_n, \quad n = N - 1. \quad (35)$$

Тоді відносно можливих значень вибіркового середнього  $\bar{x}$  можна стверджувати, що

$$P \left\{ \bar{x} > \left( \frac{st_{n;x}}{\sqrt{N}} + \mu_x \right) \right\} = \alpha. \quad (36)$$

## 3. Довірчі інтервали

Використання вибіркових значень як оцінки параметрів випадкових величин дозволяє одержати тільки оцінки параметрів у окремих точках і не дозволяє робити висновки відносно ступеня близькості

вибіркових значень до параметрів, що оцінюються. Більш змістовними є процедури оцінювання, що пов'язані з одержанням інтервалу, який накриває параметр, що оцінюється, з відомим ступенем достовірності.

Такий інтервал називається довірчим. Степінь довіри, який відповідає твердженню, що є довірчим, називається **рівнем довіри**.

При оцінюванні середнього значення  $\mu_x$  довірчий інтервал можна побудувати за вибірковим значенням  $\bar{x}$ , якщо перегрупувати члени у формулі

$$P \left\{ z_{1-\alpha/2} < \frac{(\bar{x} - \mu_x) \sqrt{N}}{\sigma_x} \leq z_{\alpha/2} \right\} = \begin{cases} 0 \\ 1 \end{cases} \quad (37)$$

наступним чином

$$\left[ \bar{x} - \frac{\sigma_x z_{\alpha/2}}{\sqrt{N}} \leq \mu_x < \bar{x} + \frac{\sigma_x z_{\alpha/2}}{\sqrt{N}} \right]. \quad (38)$$

Якщо  $\sigma_x$  невідома, то довірчий інтервал для  $\mu_x$  можна побудувати за вибірковим значенням  $\bar{x}$  і  $s$ , якщо перегрупувати у формулі (31):

$$\left[ \bar{x} - \frac{st_{n;\alpha/2}}{\sqrt{N}} \leq \mu_x < \bar{x} + \frac{st_{n;\alpha/2}}{\sqrt{N}} \right], \quad n = N - 1. \quad (39)$$

У формулах (34), (35) використано властивості  $z_{1-\alpha/2} = -z_{\alpha/2}$  та  $t_{n;1-\alpha/2} = -t_{n;\alpha/2}$ . Цим інтервалам відповідає рівень довіри  $1-\alpha$ . Довірче твердження формулюється так: істинне значення  $\mu_x$  опиняється у зазначеному інтервалі із довірчою імовірністю  $1-\alpha$  або із довірчою імовірністю  $100(1-\alpha)\%$ . Подібні твердження можна робити відносно будь-яких оцінок параметрів, аби були відомі відповідні вибіркові розподіли.

**Приклад 1.** Нехай вибірка містить  $N = 30$  незалежних спостережень нормально розподіленої випадкової величини  $x$ .

60 61 47 56 61 63  
 65 69 54 59 43 61  
 55 61 56 48 67 65  
 60 58 57 62 57 58  
 53 59 58 61 67 62

Знайдемо 90 % — довірчий інтервал для середнього значення та дисперсії випадкової величини  $x$ . Згідно з формулою (35) довірчий

інтервал з рівнем довіри  $1 - \alpha$  для  $\mu_x$  будується за вибірковою середньою  $\bar{x}$  та дисперсією  $s^2$  при розмірі вибірки  $N = 30$ :

$$\left[ \bar{x} - \frac{st_{29;\alpha/2}}{\sqrt{30}} \leq \mu_x < \bar{x} + \frac{st_{29;\alpha/2}}{\sqrt{30}} \right]. \quad (40)$$

Для  $\alpha = 0,1$  з таблиці В.2.4 знаходимо  $t_{30;\alpha/2} = t_{30;0,05} = 1,697$ , звідси маємо

$$[(\bar{x} - 0,3048s) \leq \mu_x < (\bar{x} + 0,3048s)].$$

Згідно з (1.36) довірчий інтервал для дисперсії  $\sigma_x^2$  з рівнем довіри  $1 - \alpha$  будується за вибірковою дисперсією  $s^2$

$$\left[ \frac{29s^2}{x_{29;\alpha/2}^2} \leq \sigma_x^2 < \frac{29s^2}{x_{29;1-\alpha/2}^2} \right].$$

З таблиці для  $\alpha = 0,1$  знаходимо  $x_{29;\alpha/2}^2 = x_{29;0,05}^2 = 43,77$  та  $x_{29;1-\alpha/2}^2 = x_{29;0,95}^2 = 18,49$ . Звідси одержуємо інтервал

$$[0,6854s^2 \leq \sigma_x^2 < 1,622s^2].$$

Обчислимо  $\bar{x} = \frac{1}{N-1} \sum_{i=1}^N x_i = 58,61$ ;

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N-1} \left\{ \sum_{i=1}^N x_i^2 - N(\bar{x})^2 \right\} = 33,49.$$

Таким чином, довірчі інтервали з рівнем довіри 90 % для середнього значення та дисперсії випадкової величини  $x$  дорівнює

$$[56,85 \leq \mu_x < 60,37]; [22,91 \leq \sigma_x^2 < 54,22].$$

### Перевірка гіпотез

Припустимо, що ми маємо деяку оцінку  $\hat{\phi}$ , яку побудовано за вибіркою з  $N$  незалежних спостережень випадкової величини  $x$ . Припустимо також, що є підстави вважати, що істинне значення параметра  $\phi = \phi_0$ , вибіркові значення  $\hat{\phi}$ , імовірно, не будуть в точності дорівнювати  $\phi_0$  внаслідок вибіркової мінливості, що притаманна  $\hat{\phi}$ . У зв'язку з цим виникає питання. Якщо припустити, що  $\phi = \phi_0$ , то при якому відхиленні  $\hat{\phi}$  від  $\phi_0$  ця гіпотеза має бути відхилена як необґрунтована? Відповідь на це питання можна одержати у статистичних термінах, якщо обчислити імовірність будь-якого значущого відхилення  $\hat{\phi}$  від  $\phi_0$  за вибірковою розподілом  $\hat{\phi}$ . Якщо імовірність

такого відхилення маленька, то відмінність належить вважати значущою й гіпотеза  $\varphi = \varphi_0$  має бути відхилена.

Це міркування являє собою найпростіший вид статистичної процедури, що має назву перевірка гіпотез.

Припустимо, що вибіркове значення  $\hat{\varphi}$  має щільність імовірності  $f(\hat{\varphi})$ . Якщо гіпотеза  $\varphi = \varphi_0$  правильна, то  $f(\hat{\varphi})$  повинна мати середнє значення  $\varphi_0$  (рис. 1). Імовірність того, що  $\hat{\varphi}$  виявиться менше за нижню межу  $\varphi_{1-\alpha/2}$ , дорівнює

$$P\{\hat{\varphi} \leq \varphi_{1-\alpha/2}\} = \int_{-\infty}^{\varphi_{1-\alpha/2}} f(\hat{\varphi}) d\hat{\varphi} = \frac{\alpha}{2}, \quad (41)$$

а ймовірність того, що  $\hat{\varphi}$  перевищить  $\varphi_{\alpha/2}$ , дорівнює

$$P\{\hat{\varphi} > \varphi_{\alpha/2}\} = \int_{\varphi_{\alpha/2}}^{\infty} f(\hat{\varphi}) d\hat{\varphi} = \frac{\alpha}{2}. \quad (42)$$

Значить, імовірність того, що  $\hat{\varphi}$  не попаде у інтервал, обмежений значеннями  $\varphi_{1-\alpha/2}$  та  $\varphi_{\alpha/2}$ ,  $\in \alpha$ . Нехай тепер  $\alpha$  настільки мале, що вихід  $\hat{\varphi}$  за межі інтервалу, обмеженого значеннями  $\varphi_{1-\alpha/2}$  та  $\varphi_{\alpha/2}$ , уявляється вкрай неправдоподібним. У такому випадку нема ніяких поважних причин ставити під сумнів істинність початкової гіпотези.

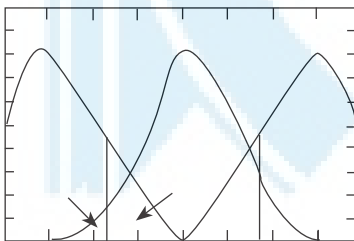


Рис.1. Визначення похибки другого роду

Імовірність  $\alpha$ , що використовується при випробовуванні гіпотез, називається рівнем значущості критерію. Область значень  $\hat{\varphi}$ , при яких гіпотеза має бути відхилена, називається критичною областю, або областю відхилення. Область значень  $\hat{\varphi}$ , при яких гіпотеза приймається, називається областю прийняття гіпотези.

При перевірці гіпотез можливі два типи похибок. По-перше, гіпотеза може бути відхилена, хоч вона фактично правильна. Така похибка зветься похибкою першого роду. З рис. 1.1 впливає, що ймо-

вірність похибки першого роду дорівнює  $\alpha$ . Навпаки, гіпотеза може бути прийнята, хоч вона фактично неправильна. Така похибка зветься похибкою другого роду. Ця похибка виникає внаслідок відхилення параметра  $\hat{\varphi}$ , що поступається гіпотезою, від істинного значення параметра  $\varphi$ , тобто  $\varphi = \varphi_0 \pm d$ . Тоді ймовірність того, що  $\hat{\varphi}$  попаде в область прийняття гіпотези, дорівнює  $\beta$ . Значить, ймовірність похибки другого роду дорівнює  $\beta$  при виявленні відхилення  $\pm d$  від гіпотетичного значення  $\varphi_0$ .

Імовірність  $1-\beta$  називається потужністю критерію. Очевидно, що при будь-якому даному розмірі вибірки  $N$  ймовірність похибки першого роду може бути зроблена як завгодно малою за рахунок зменшення рівня значущості  $\alpha$ . Але ж при цьому зростає ймовірність  $\beta$  похибки другого роду (зменшується потужність критерію). Єдиним засобом одночасного зменшення  $\alpha$  та  $\beta$  є збільшення розміру вибірки  $N$ . Ці міркування лежать в основі вибору необхідного розміру вибірки при статистичних експериментах.

**Приклад 2.** Будемо вважати, що для випадкової величини  $x$  середнє  $\bar{x} = 10$ , дисперсія  $x$  відома й дорівнює  $\sigma_x^2 = 4$ . Знайдемо розмір вибірки, яка дозволяє побудувати критерій перевірки гіпотези  $\bar{x} = 10$  з 5% рівнем значущості й 5% похибкою другого роду для виявлення 10% відхилень від гіпотетичного значення. Побудуємо для даного критерію також область прийняття гіпотези.

Вибіркове середнє  $\bar{x}$ , що визначається за формулою (3), є незміщеною оцінкою  $\mu_x$ . Відповідний вибірковий розподіл  $\bar{x}$  визначається за співвідношенням (31):

$$\bar{x} \sim \frac{\sigma_x}{\sqrt{N}} z + \mu_x,$$

де  $z$  — нормальна випадкова величина із нульовим середнім та одиничною дисперсією. Зауважимо, що такий вибірковий розподіл  $\bar{x}$  буде точним, якщо  $x$  розподілене нормально, й наближеним, якщо розподіл  $x$  відрізняється від нормального. Верхня та нижня межі області прийняття гіпотези у даному критерію відповідно дорівнюють

$$\frac{\sigma_x}{\sqrt{N}} z_{\alpha/2} + \mu_x; \quad \frac{\sigma_x}{\sqrt{N}} z_{1-\alpha/2} + \mu_x.$$

Якщо тепер істинне середнє значення дорівнює насправді  $\mu'_x = \mu \pm d$ , то з ймовірністю  $\beta$  з'явиться похибка другого роду, коли вибіркове значення  $\bar{x}$  виявиться меншим за верхню межу або біль-

шим за нижню межу. В термінах вибіркового розподілу  $\bar{x}$  маємо відповідно для верхньої та нижньої меж

$$\frac{\sigma_x}{\sqrt{N}} z_{1-\beta} + \mu_x + d ; \frac{\sigma_x}{\sqrt{N}} z_{\beta} + \mu_x - d .$$

Отже, справедливі такі рівності.

$$\frac{\sigma_x}{\sqrt{N}} z_{\alpha/2} + \mu_x = \frac{\sigma_x}{\sqrt{N}} z_{1-\beta} + \mu_x + d$$

$$z_{\alpha/2} = z_{1-\beta} + \frac{\sqrt{N}}{\sigma_x} d = -z_{\beta} + \frac{\sqrt{N}}{\sigma_x} d .$$

Звідси знаходимо необхідний розмір вибірки

$$N = \left[ \frac{\sigma_x (z_{\alpha/2} + z_{\beta})}{d} \right]^2 .$$

Для конкретних значень прикладу, що розглядається ( $\sigma_x = 2, z_{\alpha/2} = 1,96, z_{\beta} = 1,645, d = 0,1 * 10 = 1$ ), необхідний розмір вибірки дорівнює  $N = 52$ . Область прийняття гіпотези у даному критерію визначається відповідно верхньою та нижньою межами:

$$\frac{\sigma_x}{\sqrt{N}} z_{\alpha/2} + \mu_x = 10,54 ; \frac{\sigma_x}{\sqrt{N}} z_{1-\alpha/2} + \mu_x = 9,46 .$$

### Критерій згоди $\chi^2$ -квадрат

Для перевірки еквівалентності щільності ймовірності вибірових даних деякій гіпотетичній щільності часто використовується так званий критерій згоди  $\chi^2$ -квадрат. Загальна ідея критерію полягає у використанні як міри розбіжності щільності ймовірності, що спостерігаються, та гіпотетичної щільності деякої статистики, яка наближено підлягає розподілу  $\chi^2$ -квадрат ( $\chi^2$ -розподіл). Потім гіпотеза відносно їхньої еквівалентності перевіряється шляхом вивчення вибіркового розподілу цієї статистики.

Нехай дана вибірка з  $N$  незалежних спостережень випадкової величини  $x$  із щільністю  $f(x)$ . Згрупуємо  $N$  спостережень по  $K$  інтервалах, які в сукупності утворюють гістограму частот. Кількість спостережень, що опинились у  $i$ -му інтервалі, називається частотою  $i$ -го інтервалу, що спостерігається; позначимо її  $f_i$ . Кількість спостережень, що могли опинитися у  $i$ -му інтервалі, якби істинною щільністю  $x$  була б  $f(x)$ , називається частотою  $i$ -го інтервалу, що очікується;

позначимо її  $F_i$ . Розбіжність між ними у кожному інтервалі дорівнює  $f_i - F_i$ . Для вимірювання загальної розбіжності відповідними частотами, що очікуються, й просумуємо їх. Одержимо вибіркву статистику

$$x^2 = \sum_{i=1}^K \frac{(f_i - F_i)^2}{F_i}. \quad (43)$$

Цей розподіл наближено збігаються з  $x^2$ -розподілом (17). Число ступенів свободи у даному випадку дорівнює  $n = K$  мінус число деяких незалежних лінійних обмежень на спостереження. Одним з таких є обмеження, яке пов'язане з тим, що часто в останньому інтервалі групування повністю визначається частотами усіх попередніх інтервалів. Крім того, якщо, наприклад, гіпотетична щільність — нормальна з невідомими середнім та дисперсією, то з'являються два додаткових обмеження, оскільки для підбору нормальної щільності необхідно оцінити два параметри (середнє значення та дисперсію). Значить, у тому звичайному випадку, коли критерій згоди  $\chi^2$ -квадрат використовується для перевірки нормальності, число ступенів свободи в  $x^2$  дорівнює  $n = K - 3$ .

Після вибору числа ступенів свободи величини  $x^2$  перевірка гіпотези виконується таким чином. Припустимо, що випадкова величина  $x$  має щільність  $f(x)$ . Згрупуємо вибіркві значення по  $K$  інтервалів й обчислимо очікувану частоту для кожного з інтервалів при припущенні, що  $f(x) = f_0(x)$ . За формулою (1.39) визначуємо  $x^2$ . Оскільки будь-яке відхилення  $f(x)$  від  $f_0(x)$  збільшує  $x^2$ , використовуємо однібічний критерій (по верхній межі). Область прийняття гіпотези має вигляд

$$x^2 \leq x_{n;\alpha}^2 \quad (44)$$

де значення  $x_{n;\alpha}^2$  береться з таблиці. Якщо нерівність (40) задовольняється, то гіпотеза приймається з рівнем значущості  $\alpha$ .

Відомі два основні засоби застосування критерію згоди  $\chi^2$ -квадрат. Перший полягає у виборі таких інтервалів групування, що забезпечують рівність очікуваних частот для усіх інтервалів. За винятком гіпотези про рівномірність розподілу такий спосіб призводить до формування інтервалів різної довжини. Другий спосіб полягає у виборі інтервалів однакової довжини. У цьому випадку різними виявляться частоти влучання у той чи інший інтервал (за винятком рівномірного



розподілу). Якщо стандартне відхилення вибірових даних дорівнює  $s$ , то часто інтервали групування беруть довжиною  $\Delta x = 0,4s$ . Більш суттєва вимога полягає у тому, що очікувані частоти для кожного інтервалу групування мають бути досить великими з метою забезпечення прийнятної точності наближення  $\chi^2$  за формулою (39). Зазвичай рекомендується досягти виконання умови  $F_i > 3$ .

### Перевірка нормальності

**Приклад 3.** Нехай маємо вибірку з  $N = 200$  незалежних спостережень виходу генератора шуму, які перетворені у числовий вид. При довжині інтервалу  $\Delta x = 0,4s$  межі інтервалів групування, що визначені за стандартним нормальним розподілом, позначені як  $z_\alpha$ . Ці дані містяться у другому стовбці таблиці, що наведена нижче. У наступному стовбці наведені ті ж величини у вольтах. Наступний стовбець містить імовірність влучання вибірових значень у кожний з інтервалів групування. Добуток  $NP$  дає очікувані частоти для кожного з інтервалів. Необхідно перевірити вихід генератора шуму на нормальність за допомогою критерію  $\chi^2$  з рівнем значущості  $\alpha = 0,05$  (див. табл.)

Таблиця 1

$k$	$z_\alpha$	$x = sz_\alpha + \bar{x}$	$p$	$F = NP$	$f$	$ F - f $	$(F - f)^2 / F$
1	-2,0	-6,36	0,0228	4,5	4	0,5	0,06
2	-1,6	-5,04	0,0320	6,4	8	1,6	0,40
3	-1,2	-3,72	0,0603	12,1	10	2,1	0,36
4	-0,8	-2,40	0,0968	19,4	21	1,6	0,13
5	-0,4	-1,08	0,1327	26,5	29	2,5	0,24
6	0,0	0,24	0,1554	31,1	31	0,1	0,00
7	0,4	1,56	0,1554	31,1	27	4,1	0,54
8	0,8	2,88	0,1327	26,5	25	1,5	0,08
9	1,2	4,20	0,0968	19,4	20	0,6	0,02
10	1,6	5,52	0,0603	12,1	13	0,9	0,07
11	2,0	6,84	0,0320	6,4	6	0,4	0,03
12	$\infty$	$\infty$	0,0228	4,5	6	1,5	0,50
			1,000	200	200		2,43

$$N = 200; \bar{x} = 0,24; s = 3,30; n = K - 3 = 9; \chi^2 = 2,43.$$

За таблицею згідно з  $n = 9$  одержуємо  $\chi_{9,0,05}^2 = 16,92$ . Значить, гіпотеза про нормальність закону розподілу приймаються з рівнем значущості  $\alpha = 0,05$ .

## Статистичні похибки при оцінюванні основних параметрів

### Поняття про статистичні похибки

Точність оцінки деякого параметра випадкового процесу, що одержаний на основі вибірки, характеризується середнім квадратом похибки

$$M[(\hat{\phi} - \phi)^2] = M[(\hat{\phi} - M[(\hat{\phi})])^2] + M[(M[\hat{\phi}] - \phi)^2]. \quad (45)$$

Видно, що середній квадрат похибки складається з дисперсії оцінки, що характеризує долю “випадковості” у величині похибки

$$Var[\hat{\phi}] = M[(\hat{\phi} - M[(\hat{\phi})])^2] = M[(\hat{\phi}^2) - M^2[\hat{\phi}], \quad (46)$$

та з квадрату зміщення оцінки, який характеризує її систематичне відхилення:

$$b^2[\hat{\phi}] = M[b^2[\hat{\phi}]] = M[(M[\hat{\phi}] - \phi)^2]. \quad (47)$$

Отже, середній квадрат похибки дорівнює сумі дисперсії оцінки та квадрату зміщення оцінки:

$$M[(\hat{\phi} - \phi)^2] = Var[\hat{\phi}] + b^2[\hat{\phi}]. \quad (48)$$

Для зручності бажано представити похибку в долях від параметра, що оцінюється. При  $\phi \neq 0$  нормовані стандартна похибка, похибка зміщення та середньоквадратична похибка виражаються формулами

$$\varepsilon_r = \frac{\sigma[\hat{\phi}]}{\phi} = \frac{\sqrt{M[\hat{\phi}^2] - M^2[\hat{\phi}]}}{\phi}; \quad (49a)$$

$$\varepsilon_r = \frac{M[\hat{\phi}]}{\phi} - 1; \quad (49б)$$

$$\varepsilon = \frac{\sqrt{\sigma^2[\hat{\phi}] + b^2[\hat{\phi}]}}{\phi} = \frac{\sqrt{M[(\hat{\phi} - \phi)^2]}}{\phi}. \quad (49c)$$

Нормовану стандартну похибку  $\varepsilon_r$  часто називають коефіцієнтом варіації. Якщо  $\varepsilon_r$  мала, то можна вважати, що  $\hat{\phi}^2 = \phi^2(1 \pm \varepsilon_r)$ , так що

$$\hat{\phi} = \phi \sqrt{1 \pm \varepsilon_r} = \phi \left(1 \pm \frac{\varepsilon_r}{2}\right);$$

тоді маємо

$$\varepsilon_r[\hat{\phi}] \approx 2\varepsilon_r[\phi]. \quad (50)$$

Якщо зміщення оцінки  $\hat{\phi}$  мале, тобто  $b[\hat{\phi}] = 0$ , й нормована середньоквадратична похибка також мала, приміром,  $\varepsilon \leq 0,2$ , то щільність імовірності цієї оцінки можна наближено вважати гауссовою із середнім значенням  $M[\hat{\phi}] = \phi$  та середньоквадратичним відхиленням  $\sigma[\hat{\phi}] = \varepsilon\phi$ :

$$f(\hat{\phi}) = \frac{1}{\varepsilon\phi\sqrt{2\pi}} \exp\left[-\frac{(\hat{\phi} - \phi)^2}{2(\varepsilon\phi)^2}\right]. \quad (51)$$

Звідси випливають такі твердження відносно ймовірності одержання інших значень того ж параметра.

$$P\{\phi(1 - \varepsilon) \leq \hat{\phi} \leq \phi(1 + \varepsilon)\} \approx 0,68;$$

$$P\{\phi(1 - 2\varepsilon) \leq \hat{\phi} \leq \phi(1 + 2\varepsilon)\} \approx 0,95.$$

При цьому довірчий інтервал для невідомого істинного значення параметра  $\phi$ , що визначиться за будь-якою індивідуальною оцінкою  $\hat{\phi}$ , набуває вигляду

$$\left[\frac{\hat{\phi}}{1 + \varepsilon} \leq \phi \leq \frac{\hat{\phi}}{1 - \varepsilon}\right] \text{ з довірчою імовірністю } 68\%;$$

$$\left[\frac{\hat{\phi}}{1 + 2\varepsilon} \leq \phi \leq \frac{\hat{\phi}}{1 - 2\varepsilon}\right] \text{ з довірчою імовірністю } 95\%.$$

### **Оцінка середнього значення**

Припустимо, що окрема реалізація  $x(t)$ , що належить до стаціонарного (ергодичного) випадкового процесу  $[x(t)]$ , визначена на скінченному інтервалі часу довжиною  $T$ . Вибіркове значення можна обчислити за формулою

$$\hat{\mu}_x = \frac{1}{T} \int_0^T x(t) dt. \quad (52)$$

Істинне середнє значення

$$\mu_x = M[x(t)] \quad (53)$$

не залежить від часу  $t$ , якщо процес  $[x(t)]$  є стаціонарним.

$$M[\hat{\mu}] = M\left[\frac{1}{T} \int_0^T x(t) dt\right] = \frac{1}{T} \int_0^T [x(t)] dt = \frac{1}{T} \int_0^T \mu_x dt = \mu_x. \quad (54)$$

Значить, незалежно від довжини реалізації  $\hat{\mu}_x$  буде незміщеною оцінкою величини  $\mu_x$ . Тоді середній квадрат похибки оцінки  $\hat{\mu}_x$  дорівнює дисперсії.

$$\text{Var}[\hat{\mu}_x] = M\left[(\hat{\mu}_x - \mu_x)^2\right] = M[\hat{\mu}_x^2] - \mu_x^2. \quad (55)$$

де із формули (1.48) випливає, що

$$M[\hat{\mu}_x^2] = \frac{1}{T} \int_0^T \int_0^T M[x(t)x(\varepsilon)] dt d\varepsilon. \quad (56)$$

Розглянемо функцію

$$R_{xx}(t) = M[x(\tau)x(t+\tau)]. \quad (57)$$

Ця функція має назву коваріаційної функції стаціонарного процесу  $[x(t)]$ . Розглянемо також функцію

$$C_{xx}(\tau) = M[x(\tau)x(t+\tau)] - \mu_x = R_{xx} - \mu_x. \quad (58)$$

Якщо  $\mu_x \neq 0$ , то зручніше користуватися функцією  $C_{xx}$ . Дисперсію оцінки, що визначається формулою (51) або (52), можна виразити через коваріаційну функцію  $C_{xx}$ :

$$\text{Var}[\hat{\mu}_x] = \frac{1}{T^2} \int_0^T \int_{-\tau}^{T-\tau} C_{xx}(\eta) d\eta d\tau = \frac{1}{T} \int_{-T}^T \left(1 - \frac{|\tau|}{T}\right) C_{xx}(\tau) d\tau. \quad (59)$$

При великих значеннях  $T$ , коли  $|\tau| \ll T$ , дисперсія дорівнює

$$\text{Var}[\hat{\mu}_x] = \frac{1}{T} \int_{-\infty}^{\infty} C_{xx}(\tau) d\tau. \quad (60)$$

Розглянемо важливий частковий випадок, коли процес  $[x(t)]$  є обмеженим по частоті білим шумом, що має середнє значення  $\mu_x \neq 0$  та дисперсію  $\sigma_x^2$ .

$$G_{xx}(f) = \begin{cases} \frac{\sigma_x^2}{B} + \mu_x^2 \delta(f), & 0 \leq f \leq B. \\ 0, & \end{cases} \quad (61)$$

в інших випадках,

де  $B$  — ширина смуги частот. Відповідна коваріаційна функція дорівнює

$$C_{xx}(\tau) = \int_0^{\infty} G_{xx}(f) \cos 2\pi f \tau d\tau - \mu_x^2 = \sigma_x^2 \left( \frac{\sin 2\pi B \tau}{2\pi B \tau} \right). \quad (62)$$

Зауважимо, що  $C_{xx}(0) = \sigma_x^2$  та  $C_{xx}(r) = 0$  при  $r = n/(2B)$ , де  $n$  - ціле число. Таким чином, значення процесу у точках, що розділяються проміжками  $1/(2B)$ , некорельовані. У випадку гауссового процесу вони будуть статично незалежними. Тоді при  $BT \geq 5$  має місце наближена формула

$$\text{Var}[\hat{\mu}_x] = \frac{\sigma_x^2}{2BT}. \quad (63)$$

При  $\mu_x \neq 0$  нормована середньоквадратична похибка дорівнює

$$\varepsilon[\hat{\mu}_x] = \frac{1}{\sqrt{2BT}} \left( \frac{\sigma_x^2}{\mu_x} \right). \quad (64)$$

Нормована середньоквадратична похибка оцінки дорівнює

$$\varepsilon_x[\hat{\mu}_x^2] \approx 2\varepsilon_x[\hat{\mu}_x] \approx \frac{\sqrt{2}}{\sqrt{BT}} \left( \frac{\sigma_x}{\mu_x} \right). \quad (65)$$

### **Оцінка середнього квадрата**

Нехай  $x(t)$  є окремою реалізацією стаціонарного ергодичного випадкового процесу  $[x(t)]$ . Середній квадрат процесу  $[x(t)]$  можна знайти шляхом осереднення

$$\hat{\psi}_x^2 = \frac{1}{T} \int_0^T x^2(t) dt. \quad (66)$$

Істинний середній квадрат

$$\hat{\psi}_x^2 = M[x^2(t)] \quad (67)$$

не залежить від  $t$ , бо процес  $[x(t)]$  стаціонарний.

$$M[\hat{\psi}_x^2] = \frac{1}{T} \int_0^T M[x^2(t)] dt = \frac{1}{T} \int_0^T \psi_x^2 dt = \psi_x^2. \quad (68)$$

Значить, незалежно від довжини реалізації  $T$  величина  $\hat{\psi}$  є незміщеною оцінкою величини  $\hat{\psi}$

Середній квадрат похибки цієї оцінки дорівнює дисперсії

$$\begin{aligned} \text{Var}[\hat{\psi}_x^2] &= M\left[\left(\hat{\psi}_x^2 - \psi_x^2\right)^2\right] = M[\hat{\psi}_x^4] - \psi_x^4 = \\ &= \frac{1}{T^2} \iint_0^T \left( M[x^2(\varepsilon)x^2(\eta)] - \psi_x^4 \right) d\eta d\varepsilon \end{aligned} \quad (69)$$

Для гауссового процесу із середнім значенням  $\mu_x \neq 0$

$$M[x^2(\varepsilon)x^2(\eta)] = 2(R_{xx}^2(\eta - \varepsilon) - \hat{\mu}_x^4) + \hat{\mu}_x^4.$$

Таким чином

$$\text{Var}[\hat{\psi}_x^2] = \frac{2}{T} \int_{-T}^T \left(1 - \frac{|\tau|}{T}\right) (C_{xx}^2(\tau) + 2\mu_x^2 C_{xx}(\tau)) d\tau. \quad (70)$$

При великих  $T$ , коли  $|\tau| \ll T$ , дисперсія дорівнює

$$\text{Var}[\hat{\psi}_x^2] = \frac{2}{T} \int_{-\infty}^{\infty} (C_{xx}^2(\tau) + 2\mu_x^2 C_{xx}(\tau)) d\tau. \quad (71)$$

Значить, величина  $\hat{\psi}$  – обґрунтована оцінка параметру  $\psi_x^2$ , бо дисперсія  $\text{Var}[\hat{\psi}_x^2]$  прямує до нуля при  $T \rightarrow \infty$  й при умові, що функції  $C_{xx}(\tau)$  та  $C_{xx}^2(\tau)$  абсолютно інтегровані на інтервалі  $(-\infty, \infty)$ .

У випадках важливо для практичних застосувань обмеженого за частотою гауссового білого шуму ( $B$  – повна спектральна ширина процесу), що визначається формулою (57), маємо

$$\text{Згідно з (67)} \quad C_{xx}(\tau) = \sigma_x^2 \left( \frac{\sin 2\tau B\pi}{2\pi B\tau} \right).$$

$$\text{Var}[\hat{\psi}_x^2] \approx \frac{\sigma_x^4}{BT} + \frac{2}{BT} \mu_x^2 \sigma_x^2. \quad (72)$$

У загальному випадку нормована середньоквадратична похибка дорівнює

$$\varepsilon[\hat{\psi}_x^2] = \frac{1}{\sqrt{BT}} \left( \frac{\sigma_x}{\psi_x} \right)^2 + \frac{\sqrt{2}}{\sqrt{BT}} \left( \frac{\mu_x \sigma_x}{\psi_x^2} \right). \quad (73)$$

При  $\mu_x = 0$  величина  $\psi_x^2 = \sigma_x^2$  й остання рівність набуває вигляду

$$\varepsilon[\hat{\psi}_x^2] = 1/\sqrt{BT}. \quad (74)$$

### Оцінка дисперсії

Оцінка дисперсії може бути одержана як

$$\hat{\sigma}_x^2 = \hat{\psi}_x^2 - \hat{\mu}_x^2. \quad (75)$$

Вочевидь, що

$$\text{Var}[\hat{\sigma}_x^2] = M[\hat{\sigma}_x^4] - (M[\hat{\sigma}_x^2])^2,$$

де  $M[\hat{\sigma}_x^2] = M[\hat{\psi}_x^2] - M[\hat{\mu}_x^2]$ .

Звідси

$$\text{Var}[\hat{\sigma}_x^2] = \text{Var}[\hat{\psi}_x^2] - \text{Var}[\hat{\mu}_x^2] - 2\left(M[\hat{\sigma}_x^2\hat{\mu}_x^2] - M[\hat{\psi}_x^2]M[\hat{\mu}_x^2]\right). \quad (76)$$

На відміну від оцінок середнього значення та середнього квадрату оцінки дисперсії, що одержані у (1.71), будуть зміщені. В окремому випадку обмеженого по частоті гауссового білого шуму буде

$$M[\hat{\sigma}_x^2] \approx \psi_x^2 - \left(\frac{\sigma_x^2}{2BT} + \mu_x^2\right) \approx \sigma_x^2 \left(1 - \frac{1}{2BT}\right). \quad (77)$$

Значить, систематична похибка зміщення дорівнює

$$b[\hat{\sigma}_x^2] = M[\hat{\sigma}_x^2] - \sigma_x^2 \approx -\frac{\sigma_x^2}{2BT}. \quad (78)$$

Формула (1.72) у цьому випадку набуває вигляду

$$\text{Var}[\hat{\sigma}_x^2] \approx \frac{\sigma_x^4}{BT}. \quad (79)$$

Нормована випадкова похибка є

$$\varepsilon_r[\hat{\sigma}_x^2] \approx \frac{1}{BT}. \quad (80)$$

На відміну від (1.72) цей результат справедливий і для  $\mu_x \neq 0$ .

## **РЕГРЕСІЙНИЙ АНАЛІЗ**

Регресійний аналіз застосовується для побудови моделей зв'язку вихідних величин процесу із вхідними величинами, що є випадковими, з метою передбачення поведінки вихідних величин процесу залежно від конкретних значень вхідних величин.

Нехай модель процесу представляється рівнянням виду:

$$y = f(x)a \quad (1)$$

де:  $a = [a_0, a_1, \dots, a_m]^T$  — вектор невідомих параметрів моделі  
 $x = [x_1, x_2, \dots, x_n]^T$  — вектор значень незалежних змінних  
 $f = [f_0(x), f_1(x), \dots, f_n(x)]$  — вектор заданих функцій.

За результатами експерименту необхідно знайти оцінку  $\hat{a}$  параметрів  $a$ . Тоді оцінка  $\hat{y}$  для функції  $y$  обчислюється за формулою:

$$\hat{y} = f(x)\hat{a} \quad (2)$$

Нехай експеримент проводиться у  $N$  точках  $x^1, x^2, \dots, x^N$ . Позначимо вектор, координатами якого є результати його спостережень  $\hat{y}$  точках  $x^k$ :

$$\hat{y} = [\bar{y}^1, \bar{y}^2, \dots, \bar{y}^N]^T.$$

Власне кажучи, у кожній  $x^k$  точці може бути виконано  $v$  експериментів результатами яких будуть спостереження:

$$y^{-k_1}, y^{-k_2}, \dots, y^{-k_v}; \quad k = \overline{1, N}.$$

Тоді:

$$y^{-k} = \frac{1}{v}(y^{-k_1}, y^{-k_2}, \dots, y^{-k_v});$$

Задача — на підставі результатів спостережень  $\bar{y}^k$  знайти найкращі у деякому сенсі оцінки  $\hat{a}$  та  $\hat{y}$ .

### ***Теоретичні основи методу***

За основу методу побудови моделі покладено метод найменших квадратів.

Позначимо

$$\bar{y}^k = \frac{1}{v}(\bar{y}^{k_1}, \bar{y}^{k_2}, \dots, \bar{y}^{k_v})^T$$

— вектор значень, який обчислено у відповідності з (2).

$$\bar{y}^k = f^T \hat{a}; \quad k = \overline{1, N}. \quad (3)$$

або у матричному вигляді:

$$\hat{y} = F \hat{a}, \quad (4)$$

$$F = [f_j(x^k)] = \begin{bmatrix} f_0(x^1) & f_1(x^1) & f_m(x^1) \\ f_0(x^2) & f_1(x^2) & f_m(x^2) \\ \dots & \dots & \dots \\ f_0(x^N) & f_1(x^N) & f_m(x^N) \end{bmatrix}$$

Будемо вважати, що результати експерименту  $y^{-k}$  розподілені нормально, є випадковими величинами, що досліджуються, тобто математичне сподівання величини  $y^{-k}$  відповідає дійсному значенню:

$$M[\hat{y}] = y = Fa.$$



Нехай також результат спостережень у точці  $x^k$  не залежить від результату спостережень у точці  $x^j$  й дисперсія результатів спостережень  $y^{-k}$  однакова у всіх точках  $x^k$ . Останні дві умови виконуються, якщо:

$$M[(\hat{y} - y)(\hat{y} - y)^T] = \sigma^2 E.$$

Крім того, припустимо, що оцінка  $\hat{a}$ , що являє собою випадковий вектор, є незміщеною, тобто  $M[\hat{a}] = a$ , а дисперсія  $\sigma_k^2$  оцінки  $\hat{a}_k$ , має бути мінімальною

$$\sigma_k^2 = M[(\hat{a}_1 - a_k)^2], \quad k = \overline{0, n}$$

Оцінка  $\hat{a}$  задовольняє цим умовам, якщо:

$$S = \sum_{k=1}^N (y^{-k} - y^k)^2 = |y^{-k} - y^k|^2 \quad (5)$$

є мінімальною.

З урахуванням (3) маємо:

$$S = S(a) = \tilde{y}^T y + a^T F^T F a - 2\tilde{y}^T F a.$$

Якщо матриця  $F^T F$  не вироджена, тобто якщо матриця  $F$  має ранг  $m + 1$ , то сума квадратів (4) має один мінімум при:

$$\hat{a} = C F^T \hat{y}.$$

Матриця  $C = [F^T F]^{-1} = [c_{jk}]^m$  має назву дисперсійної.

### **Статистичний аналіз точності**

Похибка оцінки  $\hat{a}$ , що обчислена згідно з (5) за методом найменших квадратів, тим більша, чим більша дисперсія похибок спостережень. Точність оцінки  $\hat{a}_k$  та величини  $\hat{y}$  характеризується відповідно дисперсіями  $\sigma_{a_k}^2$  та  $\sigma_y^2$ , що залежить від матриці  $F$ . З метою визначення цієї залежності необхідно знайти вираз для коваріаційної матриці  $V(\hat{a})$ :

$$V(\hat{a}) = M[(\hat{a} - a)(\hat{a} - a)^T] = C \sigma^2.$$

Звідси для дисперсії  $\sigma_{a_k}^2$  оцінки  $\hat{a}_k$  маємо:

$$\sigma_{a_k}^2 = c_{kk} \sigma^2$$

$$\sigma_y^2 = f^T(x) C f(x) \sigma^2.$$

Оцінка  $\hat{y}$  визначається за формулою (3):

Далі, згідно з попереднім розділом, визначаються довірчі інтервали для оцінок параметрів  $\hat{a}$ .

### **Перевірка адекватності моделі**

Розглянемо величину:

$$S_3 = \sum_{k=1}^N v(y^{-k} - y^k)^2 = vS_1$$

що має  $K_3 = N - m - 1$  ступенів свободи.

Частка від ділення дисперсії неадекватності на оцінку дисперсії похибки одиночного спостереження

$$F = \frac{S_3 / K_3}{S_2 / K_2} \quad (7)$$

у випадку, коли модель адекватна, є випадковою величиною, що задовольняє  $\tau$ -розподілу Фанера з числом ступенів свободи  $K_3$  та  $K_2$ . Для визначеного рівняння значущості  $\alpha$  перевірки гіпотези про адекватність моделі можна визначити значення  $\tau_{kp}$ , що відповідає умові

$$P\{\tau > \tau_{kp}\} = \alpha = 1 - p.$$

### **Реалізація побудови моделі**

1. Вибір моделі:

$$y(x) = a_0 f_0(x) + a_1 f_1(x) + \dots + a_m f_m(x).$$

2. Для  $N > m + 1$  точок:

$$x^k = [x_1^k, x_2^k, \dots, x_n^k]^T \quad k = \overline{1, N}.$$

Обчислюється матриця:

$$F = [f_j(x^k)] = \begin{bmatrix} f_0(x^1) & f_1(x^1) & \dots & f_m(x^1) \\ f_0(x^2) & f_1(x^2) & \dots & f_m(x^2) \\ \dots & \dots & \dots & \dots \\ f_0(x^N) & f_1(x^N) & \dots & f_m(x^N) \end{bmatrix}$$

3. Обчислюється дисперсійна матриця  $C = [F^T F]^{-1}$

4. У кожній точці  $x^k$  здійснюється  $v$  випробувань, за результатами яких  $y^{-k_1}, y^{-k_2}, \dots, y^{-k_v}$  обчислюється середнє значення

$$y^{-k} = \frac{1}{v} \sum_{j=1}^v y^{-kj}$$

із дисперсією  $\sigma^2$ .

5. Обчислюються оцінки  $\hat{a}$  коефіцієнтів моделі за методом найменших квадратів:

$$\hat{a} = CF^T \hat{y}.$$

6. Обчислення дисперсії оцінок

$$\sigma_{a_k}^2 = c_{kk} \sigma^2.$$

7. Визначається оцінка функції  $y$  за формулою (3).

8. Обчислюється оцінка дисперсії похибок спостережень  $s$ :

$$s^2 = S_2 / (vK_2); \quad K_2 = N(v-1);$$

$$S_2 = \sum_{k=1}^N \sum_{j=1}^N (y^{-kj} - y^{-k})^2; \quad s_k^2 = c_{kk} s^2.$$

9. Обчислюється:

$$S_3 = \sum_{k=1}^N v(y^{-k} - \hat{y}^k)^2$$

10. Оцінюється адекватність моделі:

$$\tau = \frac{S_3 / K_3}{S_2 / K_2} \quad K_3 = N - m - 1$$

Для заданої надійності  $P\{\tau > \tau_{kp}\} = \alpha = 1 - p$  визначається  $\tau_{kp}$ . Якщо  $\tau < \tau_{kp}$ , то модель вважається адекватною процесу.

11. Якщо дисперсія спостережень  $\sigma^2$  відома й модель адекватна, то довірчий інтервал із довірчою ймовірністю  $p$  визначається за нерівністю:

$$|\hat{a}_k - a_k| < \sigma_k \varepsilon.$$

Значення  $\varepsilon$  визначаємо із таблиці за умови  $\Phi(\varepsilon) = p/2$ . Якщо дисперсія спостережень  $\sigma^2$  невідома і в кожній точці плану проводиться один експеримент, то:

$$|\hat{a}_k - a_k| < \sigma_k \varepsilon. \quad s_k = \sqrt{c_{kk} s}; k = \overline{0, m}.$$

$$s^2 = S_1 / K_1 = \sum_{k=1}^N (y^{-k} - \hat{y}^k)^2 / K_1,$$

$\varepsilon$  знаходимо із таблиці при  $K_1 = N - m - 1$  ступенях свободи та заданій довірчій ймовірності  $p$ .

12. Перевіряються коефіцієнти  $a_k$  на значущість. Якщо:

$$|\hat{a}_k| > t_{kp} s_k,$$

то коефіцієнт  $a_k$  вважається значущим;  $t_{kp}$  – критичне значення розподілу Ст'юдента для заданого рівня значущості  $\alpha$  та  $K_1$  ступенів свободи.  $t_{kp}$  визначається із таблиці за умови  $P\{|t| > t_{kp}\} = \alpha$ . Коефіцієнти, які виявились незначущими, відкидаються.

**Приклад 1.** Розглянемо процес, вихід якого залежить від двох змінних  $x_1, x_2$ . Виконуються у трьох точках два експерименти (див. таблицю).

К	$x_1^k$	$x_2^k$	Результати спроб		
			$y^{k1}$	$y^{k2}$	$y^{-k} = (y^{-k_1} + y^{-k_2})/2$
1	1	1	45,4	46,1	45,75
2	1	-1	38,0	38,4	38,2
3	-1	-1	34,5	34,8	34,65
4	-1	1	42,1	42,7	42,4

Побудуємо лінійну модель

$$y = a_0 + a_1x_1 + a_2x_2 + 2.$$

Матриця

$$F = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \end{bmatrix}$$

$$F^T F = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} = 4E_3$$

$$C = [F^T F]^{-1} = 1/4E_3$$

Визначимо параметри моделі

$$\hat{a} = CF^T \hat{y} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 45 & 75 \\ 38 & 20 \\ 34 & 65 \\ 42 & 4 \end{bmatrix} = \begin{bmatrix} 40 & 25 \\ 1 & 725 \\ 3 & 825 \end{bmatrix}$$

Таким чином, одержуємо:

$$\hat{y}F\hat{a} = 40,25 + 1,725x_1 + 3,825x_2 = [45,80 \quad 38,15 \quad 34,70 \quad 42,35]^T$$

Оскільки дисперсія спостережень  $\sigma^2$  невідома, то обчислюємо  $S_2$  та  $S_3$ .

$$S_3 = \sum_{k=1}^4 2(y^{-k} - \hat{y}^k)^2 = 0,004;$$

$$S_2 = (45,4 - 45,75)^2 + (46,1 - 45,75)^2 + (38,0 - 38,15)^2 + (38,4 - 38,15)^2 + (34,5 - 34,7)^2 + (34,8 - 34,7)^2 + (42,1 - 42,35)^2 + (42,7 - 42,35)^2 = 0,55;$$

$$K_2 = 4(2 - 1) = 4;$$

$$s^2 = \frac{0,55}{2 \cdot 4} = 0,069;$$

$$s_j = \sqrt{1/4s} = 1,138;$$

Для 95 % надійності з таблиці при  $N = 4$  одержуємо  $\epsilon = 2,776$  і

$$|\hat{a}_j - a_j| < 0,138 - 2,776 = 0,383.$$

З метою вивчення адекватності:

$$\tau = \frac{0,004/1}{0,55/4} = 0,03;$$

З таблиці для 99 % надійності одержуємо:

$$\tau_{kp} = 21,2; \quad \tau = 0,03 < \tau_{kp}.$$

Значить, обрана модель адекватна процесу.

Оцінимо значущість параметрів моделі. Для обраного рівня значущості  $p = 0,99$  таблиці знаходимо  $\epsilon = 13,277$ . Одержуємо:

$$|\hat{a}_0| = 40,25 > 13,277 - 0,138 = 1,832; \quad |\hat{a}_1| = 1,725 < 1,832, \quad |\hat{a}_2| = 3,825 > 1,832.$$

Таким чином, модель, що адекватна процесу, має вигляд:

$$\hat{y} = 40,25 + 3,825x_2.$$

### **ЗАВДАННЯ ДЛЯ САМОСТІЙНИХ РОБІТ ТА РЕКОМЕНДАЦІЇ ЩОДО ЇХНЬОГО ВИКОНАННЯ**

Кожна самостійна робота оформлюється у вигляді рукопису, який містить:

- постановку задачі;
- обґрунтування обраного методу розв'язування поставленого завдання, основні робочі формули;
- блок-схему алгоритму розв'язання задачі;

- текст програми;
- одержані на ЕОМ результати;
- аналіз результатів, оцінку похибок.

Для успішного розв'язання задачі необхідно знати теоретичний матеріал, сформульований у контрольних запитаннях, які містяться в кінці кожного завдання для самостійної роботи.

### ***Завдання 1. Програмна реалізація датчиків псевдовипадкових чисел***

Теорія ймовірностей та математична статистика вивчають властивості математичних моделей випадкових явищ. Існування тих чи інших властивостей таких моделей обґрунтовується, як і в інших областях математики, строгими дедуктивними міркуваннями. Однак пошук цих властивостей може відбуватися шляхом експериментів з математичними моделями. Для проведення таких експериментів необхідно мати наготові послідовності чисел, які, з того чи іншого погляду, можна розглядати як реалізації послідовності незалежних випадкових величин. Іншою важливою областю застосування таких послідовностей є методи Монте-Карло[4].

У більшості випадків необхідні для статичних експериментів послідовності чисел одержуються за допомогою ЕОМ. Існує багато різних програм, які виробляють детерміновані послідовності чисел, що мають достатньо складну нерегулярну структуру і тому здебільшого схожі на послідовності випадкових чисел. Щоб підкреслити принципову відмінність цих детермінованих послідовностей від “справжніх” випадкових послідовностей, такі програми називаються датчиками псевдовипадкових чисел [1].

Отже, для розв'язання прикладних задач методом Монте-Карло студентам необхідно одержувати на ЕОМ послідовність вибірових значень випадкової величини із заданим законом розподілу. Випадкову величину (ВВ) моделюють за допомогою перетворень одного чи декількох незалежних значень ВВ  $\alpha$ , яка є рівномірно розподіленою на інтервалі (0,1). Одним з методів одержання випадкової величини є метод псевдовипадкових чисел.

Псевдовипадкові числа, як правило, будуть обчислюватися студентами на ЕОМ за рекурентною формулою вигляду

$$\alpha_{k+1} = F(\alpha_k), \quad k = 0, 1, 2, \dots \quad (1)$$

Яка є періодичною з періодом  $L$ .

Метою самостійної роботи є засвоєння студентами методів моделювання на ЕОМ випадкових величин.

1. Псевдовипадкові числа з рівномірним розподілом на відрізьку [0,1]

Одним з найпростіших способів побудови послідовності  $x_1, x_2, \dots$  псевдовипадкових чисел з відрізьку [0,1] з розподілом близьким до рівномірного, є побудова рекурентної послідовності

$$x_{n+1} = \{Kx_n\}, n = 0, 1, \dots$$

де  $\{x\}$  – дробова частина числа  $x$ .  $K$  – ціле число, яке визначає властивості послідовності  $\{x_n\}$ . Щоб послідовність  $\{x_n\}$  мала достатньо великий період та не була занадто регулярною, доречно вибирати  $K$  як ціле двозначне число, яке при діленні на 8 дає залишок  $\pm 3$ .

Зокрема, при  $v_0 = 0.1357913$  та  $K = 37$  одержимо, що  $v_1 = 0.0242781$ ,  $v_2 = 0.8982897$ ,  $v_3 = 0.758603$ , ... [1], [4].

Якщо за рекурентну формулу обрати вираз

$$v_{n+1} = \{11v_n + \pi\}, n = 0, 1, \dots$$

то одержимо при  $v_0 = 0.1357913$  наступну послідовність псевдовипадкових чисел: 0.6352969; 0.129859; 0.5700416; 0.4120502; ...

2. Псевдовипадкові числа з рівномірним розподілом на відрізьку  $[a, b]$  з рівномірним розподілом на множині  $\{1, 2, \dots, n\}$ , показниковим розподілом.

Датчик псевдовипадкових чисел, які мають рівномірний розподіл на відрізьку [0,1], дозволяє одержати послідовність псевдовипадкових чисел, які мають задану функцію розподілу  $F(x)$ , якщо обернена до неї функція  $F_{-1}(y) = \sup\{x : F(x) \leq y\}$ ,  $0 < y < 1$  обчислюється достатньо просто.

Наприклад, якщо  $F(x)$  – функція рівномірного розподілу на відрізьку  $[a, b]$ , то

$$F_{-1}(y) = a + (b - a)y;$$

якщо  $F(x)$  – функція рівномірного розподілу на множині  $\{1, 2, \dots, n\}$ , то

$$F_{-1}(y) = 1 + [ny];$$

якщо  $F(x) = 1 - e^{-\alpha x}$  функція показникового розподілу з параметром  $\alpha$ , то

$$F_{-1}(y) = -\frac{1}{\alpha} \ln y.$$

### Псевдовипадкові числа з нормальним розподілом

Для побудови послідовності псевдовипадкових чисел з нормальним розподілом можна використати твердження [4]:

Випадкові величини  $\xi_1$  та  $\xi_2$  незалежні та мають стандартний нормальний розподіл тоді й тільки тоді, коли незалежні випадкові величини

$$\rho^2 = \xi_1^2 + \xi_2^2; \quad \varphi = \arg(\xi_1 + i\xi_2).$$

причому

$$P\{0 \leq \rho^2 \leq x\} = 1 - e^{-x/2}; \quad P\{0 \leq \varphi \leq x\} = \min\{1, \frac{x}{2\pi}\}, x \geq 0.$$

Тому, якщо випадкові величини  $v_1$  та  $v_2$  незалежні й мають рівномірний розподіл на відрізок  $[0, 1]$ , то випадкові величини

$$\xi_1 = \cos(2\pi v_1) \sqrt{-2 \ln v_2};$$

$$\xi_2 = \sin(2\pi v_1) \sqrt{-2 \ln v_2};$$

незалежні й мають стандартний нормальний розподіл з  $M\xi_1 = M\xi_2 = 0$ ,  $D\xi_1 = D\xi_2 = 1$ . Перехід від випадкової величини  $\xi$ , яка має стандартний нормальний розподіл до випадкової величини  $\eta$ , яка має нормальний розподіл з параметрами  $M\eta = a$ ;  $D\eta = \sigma^2$ , здійснюється за формулою

$$\eta = a + \sigma\xi.$$

### 3. Оцінка точності моделювання випадкових величин

Нехай деякий алгоритм здійснює моделювання випадкової величини  $\xi$  на ЕОМ. Після  $n$ -кратного звертання до цього алгоритму моделюється випадкова вибірка  $X = \{x_1, \dots, x_n\}$ . Необхідно за допомогою  $X$  перевірити гіпотезу  $H_0$  про те, що функція розподілу випадкової величини  $\xi$

$$F_\xi(x) \equiv F_0(x).$$

де  $F_0(x)$  — деяка фіксована функція розподілу. Конкуруюча гіпотеза  $H_1$ ,

$$F_\xi(x) \neq F_0(x).$$

Гіпотеза  $H_0$  називається гіпотезою узгодження. Для перевірки  $H_0$  відомо декілька статистичних критеріїв, які називаються критеріями узгодження. Розглянемо  $\chi^2$  — критерій Пірсона.

У  $X$  знаходимо

$$x_- = \min\{x_i\}, \quad x_+ = \max\{x_i\}$$

Та здійснюємо розбиття числової прямої на  $k > 1$  комірок:

$$(-\infty, x_- + h), (x_- + h, x_- + 2h) \dots (x_+ - 2h, x_+ - h), (x_+ - h, +\infty),$$



де  $h = \frac{x_+ - x_-}{k}$ . Бажано  $k$  обирати таким чином, щоб  $\min_{i=1, \dots, k} np_i \geq 5$ .

Обчислимо теоретичну ймовірність влучання  $\xi$  до  $i$ -ї комірки, якщо справедлива  $H_0$ .

$$p_i = \begin{cases} F_0(x_- + h), & i = 1, \\ F_0(x_- + ih) - F_0(x_- + (i-1)h), & 1 < i < k, \\ 1 - F_0(x_+ - h), & i = k \end{cases}$$

Число  $n_i$  вибіркових значень із  $X$ , які влучили в  $i$ -у комірку. Відзначимо, що  $\sum_{i=1}^k p_i = 1$ ,  $\sum_{i=1}^k n_i = n$ .

Обчислимо  $\chi^2$ -статистику:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}.$$

Тоді приймаємо гіпотезу

$$\begin{cases} H_0, \chi^2 < x^2, \\ H_1, \chi^2 \geq x^2, \end{cases}$$

де  $x^2$  обирається таким чином, щоб імовірність помилки першого роду дорівнювала заданому рівню значущості  $\alpha \ll 1$ :

$$P\{\chi^2 \geq x^2 / H_0\} = \alpha.$$

Виявляється, що при правильній гіпотезі  $H_0$  та  $n \rightarrow \infty$  випадкова величина  $\chi^2$  має  $\chi^2$ -розподіл з  $(k-1)$  ступенем вільності, який не залежить від  $F_0(\cdot)$ . Функцією цього розподілу є

$$F_{\chi^2_{k-1}}(y) = \frac{1}{2\Gamma\left(\frac{k-1}{2}\right)} \int_0^y \left(\frac{z}{2}\right)^{\frac{k-1}{2}-1} e^{-\frac{z}{2}} dz, y \geq 0.$$

Тоді  $x^2$  можна визначити з наближеної рівності:

$$1 - F_{\chi^2_{k-1}}(x^2) = \alpha,$$

Звідки

$$x^2 = F_{\chi^2_{k-1}}^{-1}(1 - \alpha).$$

Значення  $x^2$  залежні від  $k$  та  $\alpha$  наведено в табл.

$k$	$\alpha = 0.01$	$\alpha = 0.00025$	$\alpha = 0.05$	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.95$
1	6.6	6.0	3.8	0.0039	0.00098	0.00016
2	9.2	7.4	6.0	0.103	0.051	0.020

3	11.3	9.4	7.8	0.352	0.216	0.115
4	13.3	11.1	9.5	0.711	0.484	0.297
5	15.1	12.8	11.1	1.15	0.831	0.554
6	16.8	14.4	12.6	1.64	1.24	0.872
7	18.5	16.0	14.1	2.17	1.69	1.24
8	20.1	17.5	15.5	2.73	2.18	1.65
9	21.7	19.0	16.9	3.33	2.70	2.09
10	23.2	20.5	18.3	3.94	3.25	2.56
11	24.7	21.9	19.7	4.57	3.82	3.05
12	26.2	23.3	21.0	5.23	4.40	3.57
13	27.7	24.7	22.4	5.89	5.01	4.11
14	29.1	26.1	23.7	6.57	5.63	4.66
15	30.6	27.5	25.0	7.26	6.26	5.23
16	32.0	28.8	26.3	7.96	6.91	5.81

### Студентам необхідно:

1. На ЕОМ одержати (за допомогою методів з 1.2.3)  $N$  псевдовипадкових чисел  $\alpha_1, \alpha_2, \dots, \alpha_N$ .
2. Обчислити вибірково середнє та дисперсію

$$\bar{\alpha} = \frac{1}{N} \sum_{i=1}^N \alpha_i; \quad s^2 = \frac{1}{N-1} \sum_{i=1}^N (\alpha_i - \bar{\alpha})^2.$$

3. Надрукувати послідовність  $\{\alpha_i\}, \bar{\alpha}, s^2$ .
4. Провести перевірку гіпотези про відповідний закон розподілу одержаної послідовності псевдовипадкових чисел за допомогою  $\chi^2$ -критерію Пірсона. ( $N$  та конкретний метод побудови послідовності обираються викладачем).

### Завдання 2. Моделювання дискретних випадкових величин

Нехай випадкова величина  $\eta$  задана розподілом

$y$	$y_1$	$y_2$	...	$y_N$
$P$	$P_1$	$P_2$	...	$P_N$

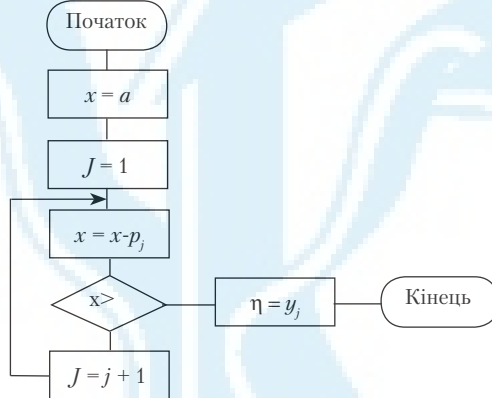
де  $p_j = P(\eta = y_j), j = \overline{1, N}$ .

Загальний метод моделювання дискретної випадкової величини базується на наступному алгоритмі. Оскільки  $\sum_{i=1}^N p_i = 1$ , то розіб'ємо відрізок  $[0, 1]$  на частини  $I_i$  таким чином, щоб  $|I_i| = p_i$ .



Будемо випадковим чином кидати на відрізок  $[0,1]$  точку та, якщо вона влучила в інтервал  $I_j$ , то випадковій величині  $\eta$  будемо ставити у відповідність значення  $y_j$ . Імовірність влучання в інтервал  $I_j$  дорівнює  $p_j$ . Отже побудована випадкова величина  $\eta$  має заданий розподіл.

Нижче наведено блок-схему даного алгоритму.



Псевдовипадкове число  $\alpha$  з інтервалу  $(0,1)$  студенти можуть одержати за допомогою стандартної підпрограми на відповідній мові програмування.

Таблиця 2

Номер	Закон розподілу випадкової величини $\eta$							
	1	$\eta$	7.5	12	14	-5.2		
	$P$	0.1	0.2	0.3	0.4			
2	$\eta$	10	7	3	1			
	$P$	0.5	0.3	0.1	0.1			
3	$\eta$	-5	-3.7	-2.1	-1.3	0	2.5	3.7
	$P$	0.3	0.1	0.15	0.15	0.1	0.1	0.1
4	$\eta$	1.34	2.15	3.17	9.2	10.3	11.15	13
	$P$	0.15	0.15	0.19	0.2	0.15	0.01	0.15
5	$\eta$	1	2	3	4	5	6	7
	$P$	0.01	0.09	0.15	0.25	0.1	0.2	0.2
6	$P\{\eta = k\} = C_n^k p^k (1-p)^{n-k}; p = 0.3; n = 10; k = \overline{0,10}$							

7	$P\{\eta = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \lambda = 3$							
8	$P\{\eta = k\} = p(1-p)^k, p=0.7$							
9	$P\{\eta = k\} = \frac{1}{n}, n=20$							
10	$\eta$	-5	-3	-1	0	1	3	5
	$P$	0.1	0.2	0.09	0.01	0.3	0.2	0.1

### Студентам необхідно:

1. Змодельовати на ЕОМ дискретну випадкову величину, задану розподілом у наведеній таблиці 2.
2. Надрукувати  $N$  одержаних значень  $\eta_1, \dots, \eta_N$ , обчислити

$$M\eta = \sum_{i=1}^N p_i \xi_i; \quad D\xi = \sum_{i=1}^N p_i \xi_i^2 - (M\xi)^2.$$

3. Обчислити ефективність алгоритму за формулою:

$$s = \frac{1}{1 + M\xi}$$

### Завдання 3. Моделювання неперервних випадкових величин

#### 1. Моделювання абсолютно неперервних випадкових величин

Будь-яка випадкова величина  $\xi$  повністю визначається функцією розподілу  $F\xi(x)$ :

$$F\xi(x) = P(\xi > x). \quad (5)$$

Абсолютно неперервною випадковою величиною називають випадкову величину, яка має всюди неперервну функцію розподілу  $F\xi(x)$ . Позначимо  $F^{-1}$  функцію, обернену до  $F\xi(x)$ . Така функція визначена на відрізку  $[0,1]$ , оскільки  $F\xi(x)$  монотонно зростає. Доведено [4], що якщо випадкова величина  $\alpha$  розподілена рівномірно на відрізку  $[0,1]$ , то випадкова величина  $\xi = F^{-1}(\alpha)$  має функцію розподілу  $F\xi(x)$ . Із цієї формули одержують стандартний спосіб моделювання абсолютно неперервної випадкової величини із заданою функцією розподілу  $F\xi(x)$ :

1. За відомою функцією  $F(x)$  аналітично або чисельно будемо обернену до неї функцію  $F^{-1}(\alpha)$ .
2. Генеруємо чергове значення випадкової величини  $\alpha$ , рівномірно розподіленої на відрізку  $[0,1]$ .

3. За формулою  $x = F^{-1}(\alpha)$  знаходимо чергове значення випадкової величини  $\xi$ , яка має функцію розподілу  $F(x)$ .

Зокрема, у випадку моделювання випадкової величини, яка має показниковий розподіл з параметром  $\lambda$ , функція розподілу має вигляд:

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

Для такої функції легко знайти обернену:

$$F^{-1}(x) = -\frac{1}{\lambda} \ln(x)$$

Однак стандартний метод моделювання зрідка використовується на практиці, оскільки для більшості найважливіших розподілів функції  $F^{-1}(\alpha)$  не виражаються через елементарні функції. Табулювання ж функції  $F^{-1}(\alpha)$  істотно ускладнює процес розв'язування.

## 2. Моделювання нормально розподіленої випадкової величини

Нехай  $\xi_1, \xi_2, \dots, \xi_N$  – незалежні випадкові величини, які рівномірно розподілені на відрізку  $[0, 1]$ .  $M\xi_i = 0.5$ .  $D\xi_i = 1/12$ . Складемо з них нормовану величину.

$$\eta = \sum_{i=1}^n \frac{\xi_i - 0.5}{\sqrt{n/12}} \quad (6)$$

Згідно з центральною граничною теоремою при  $n \rightarrow \infty$  випадкова величина  $\eta$  достатньо швидко прямує до стандартної нормально розподіленої випадкової величини з математичним сподіванням, яке дорівнює 0, та дисперсією, рівною 1.

Звичайно беруть  $n=12$ , тоді

$$\eta = \sum_{i=1}^{12} \xi_i - 6.$$

Моделювання випадкової величини  $\xi = N(a, \sigma)$ , яка має нормальний розподіл з параметрами  $a$  та  $\sigma$ , базується на формулі

$$\xi = a + \sigma\eta,$$

де  $\eta = N(0, 1)$ .

Блок-схему алгоритму моделювання випадкової величини  $\xi = N(a, \sigma)$  наведено нижче. Тут  $\alpha$  знаходимо за допомогою стандартної підпрограми реалізації псевдовипадкового числа  $(0, 1)$ .

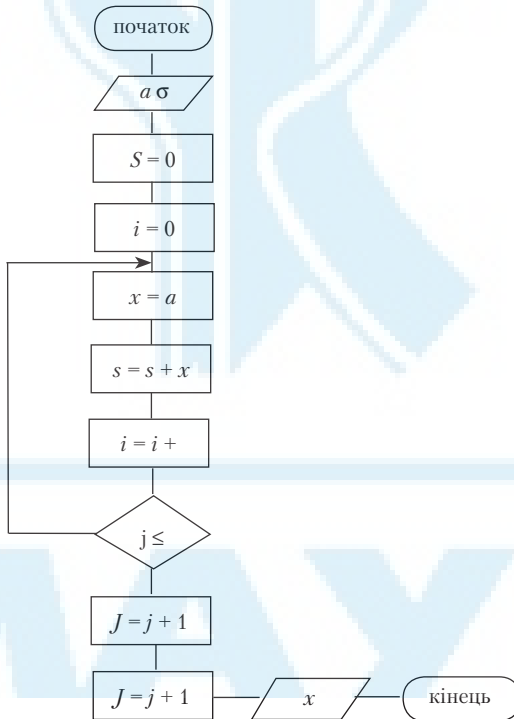
### 3. Метод Монте-Карло оцінювання $M\xi$ та $D\xi$

Оцінимо математичне сподівання випадкової величини  $\xi$ , яка є складною функцією декількох незалежних випадкових величин із заданими законами розподілу. Найчастіше побудувати функцію  $F_\xi(x)$  у явному вигляді неможливо, але величину  $M\xi$  можна оцінити методом Монте-Карло.

Отже, нехай  $M\xi$  існує. Оберемо  $N$  реалізації  $\xi_1, \xi_2, \dots, \xi_N$  випадкової величини  $\xi$  й обчислимо середнє арифметичне

$$\bar{a}_N = \frac{1}{N} \sum_{i=1}^N \xi_i. \quad (7)$$

Оскільки послідовність однаково розподілених незалежних випадкових величин, у котрих  $a$  математичне сподівання, задовольняє закон великих чисел, то величина  $a_N$  збігається за ймовірністю до математичного сподівання при  $N \rightarrow \infty$ , тобто  $a_N \rightarrow a$ .



Отже, при великих  $N$  величина  $a_N \approx a$ . Для оцінки похибки припустимо, що величина  $\xi$  має дисперсію  $D\xi$ . Якщо величини  $\xi_1, \xi_2, \dots, \xi_N$  незалежні за сукупністю, то завдяки центральній граничній теоремі випадкова величина

$$\frac{a_N - a}{\sqrt{D\xi / N}}$$

розподілена асимптотично нормально [3].

Задамо довірчу ймовірність  $\beta$ , наприклад  $\beta = 0.95$ . Для нормально розподіленої випадкової величини  $\beta P(|\beta| < 1.96) = 0.95$ . Отже, при великих  $N$  з ймовірністю, близькою до одиниці, виконується нерівність

$$|\bar{a}_N - a| < 1.96 \sqrt{D\xi / N} \quad (8)$$

Однак, коли ми починаємо знаходити  $M\xi$ , значення  $D\xi$ , як правило, невідоме. Але в більшості задач величину  $D\xi$  можна оцінити емпірично за формулою

$$D\xi \cong \frac{1}{N} \sum_{i=1}^N \xi_i^2 - \left[ \frac{1}{N} \sum_{i=1}^N \xi_i \right]^2. \quad (9)$$

Отже, обчислення потрібно скінчити, якщо

$$1.96 \sqrt{\frac{D\xi}{N}} < \varepsilon, \quad (10)$$

де  $\varepsilon$  — задана точність.

### Студентам необхідно:

1. Знайти 95 %-довірчий інтервал шириною  $2\varepsilon$  для математичного сподівання випадкової величини  $F(\xi, \eta, \zeta, \psi)$ , де  $\xi, \eta, \zeta, \psi$  — випадкові величини з відомими законами розподілу.

$\xi$  — розподілена нормально з параметрами  $a, \sigma$ ;

$\eta$  — показниковий розподіл з параметром  $\lambda$ ;

$\zeta$  — біноміальний розподіл з параметрами  $n$  і  $p$ ;

$\psi$  — розподіл за законом Пуассона з параметром  $\lambda$ .

2. Дані для обчислень брати з таблиці.

Номер	$a$	$\sigma$	$\lambda$	$n$	$p$	$F(\xi, \eta, \zeta, \psi)$
1	2	3	4	5	6	7
1	1	0.1	2	3	0.1	$(\xi^2 + \eta^3) / (1 + \zeta)$

1	2	3	4	5	6	7
2	2	0.5	1			$\arctg\xi / (1 + \psi^2)$
3			1	5	0.2	$\ln(1 + \psi) / (\arctg\xi^2 + 1)$
4	0,5	1	0.2			$\cos\xi + \eta$
5	2	0.1	2			$\xi + e^{-\psi}$
6	1	0.2		3	0.1	$1 / (\ln(\xi^2 + \zeta^2))$
7	2	0.1				$\sin\xi$
8	2	0.5	2			$\cos(\eta + \cos^3\xi)$
9	1	0.1	1			$sh(\cos\xi) / (1 + \eta^2)$
10	2	0.5	5			$\xi^2 e^{-\eta^2}$
11			2			$\ln(\cos\psi + 1)$
12	1.5	1.0	3			$\xi + \eta$
13			2			$\cos(\psi^2 + 1)$
14	0	2.0	1	5	0.4	$\xi^2 + \eta^2 + \zeta$
15	0	1.0	1	3	0.8	$(\cos\xi + \eta) / (2 + tg^2\zeta)$
16	0	1.0	1	3	0.8	$(\cos\eta + \xi) / (2 + tg^2\zeta)$
17	2	1.0	1			$e^{-(\xi+1/(1+\eta))}$
18	2	0.4	3			$\sqrt{\xi^2 + \eta^2}$
19	-2	0.2	0.5			$\sqrt{\xi^2 \sin^2 \eta}$
20	2	0.2	0.8			$\sqrt{\xi \sin^2 \eta^2}$



## Контрольні запитання до завдань 1–3

1. Що таке псевдовипадкове число?
2. Методи одержання послідовності псевдовипадкових чисел, розподілених рівномірно на  $(a, b)$ .
3. Як проводиться моделювання дискретної випадкової величини?
4. У чому полягає стандартний метод моделювання абсолютно неперервної випадкової величини?
5. Яка особливість є у моделюванні нормально розподіленої випадкової величини?
6. Як оцінюється математичне сподівання випадкової величини  $\xi$ ?
7. У чому полягає метод Монте-Карло?

### ***Завдання 4. Моделювання гауссівського випадкового процесу за допомогою моделі авторегресії та ковзаючого середнього***

Нехай  $\{\xi_t^* \equiv \xi^*(t, \omega)\} \subset R^1$  – випадковий процес, який має нульове математичне сподівання, кореляційну функцію  $R(t)$  та дробово-раціональну спектральну щільність  $S(\lambda)$ . Змоделюємо цей процес на проміжку  $[0, T]$  з кроком дискретизації  $\Delta$ .

Нехай  $t = m\Delta$  ( $m = 0, \pm 1, \pm 2, \dots$ ) – дискретний час,  $\eta = \eta_t(t\omega)$  – дискретний “білий шум”, тобто послідовність незалежних гауссівських випадкових величин, які розподілені за стандартним законом  $N_1(0, 1)$ . Випадковий процес  $\xi_t$ , який визначається лінійним рекурентним рівнянням

$$\varepsilon_{m\Delta} = \sum_{k=0}^q a_k \eta_{(m-k)\Delta} - \sum_{i=1}^p b_i \xi_{(m-i)\Delta},$$

називається процесом авторегресії та ковзаючого середнього порядку  $(p, q)$ .

З теорії часових рядів [9] відомо, що за вказаних умов  $\xi_{m\Delta}$  є гауссівською послідовністю з нульовим математичним сподіванням та спектральною щільністю

$$S_{\xi}(z) = \frac{|A(z)|^2}{|B(z)|^2} = \frac{A(z)A(z^{-1})}{B(z)B(z^{-1})}, \quad z = e^{-j\lambda\Delta}. \quad (1)$$

$$A(z) = \sum_{k=1}^q h_k z^k, \quad B(z) = 1 + \sum_{i=1}^p h_i z^i.$$

Підберемо параметри  $p$ ,  $q$ ,  $\{a_k\}$ ,  $\{b_i\}$  таким чином, щоб  $S_\xi(z)$  відповідала спектральній щільності  $S_{\xi^*}(z)$  послідовності  $\varepsilon_{m\Delta}$ , яка моделюється.

Обчислимо  $S_{\xi^*}(z)$  за заданою кореляційною функцією  $R(m\Delta)$ :

$$S_{\xi^*}(z) = \sum_{m=-\infty}^{+\infty} R(m\Delta) z^m = F(z) + F\left(\frac{1}{z}\right) - R(0), \quad (2)$$

де

$$F(z) = \sum_{m=0}^{\infty} R(m\Delta) z^m \quad (3)$$

є  $z$ -перетворення [9] послідовності  $\{R(0), R(\Delta), \dots\}$ .

За умовою  $S(\lambda)$  — дробово-раціональна функція, тому кореляційна функція може бути подана у вигляді скінченної суми:

$$R(m\Delta) = \sum_{r,s} A_{r,s} m^r e^{\lambda_s m}, \quad \lambda_s = \lambda_s^{(1)} + j\lambda_s^{(2)}. \quad \lambda_s^{(2)} > 0, \quad (4)$$

де  $A_{r,s}$  — деякі коефіцієнти, які не залежать від  $m$ . Використовуючи вираз (4) у виразі (3) з урахуванням властивостей  $z$ -перетворення, одержуємо:

$$F(z) = \sum_{r,s} A_{r,s} F_{r,s}(z) \quad (5)$$

$$F_{r,s}(z) = \begin{cases} (1 - ze^{\lambda_s})^{-1}, & r = 0, \\ ze^{\lambda_s} (1 - ze^{\lambda_s}) Q_{r-1}(ze^{\lambda_s}), & r \geq 1, \end{cases}$$

де  $Q_r(z)$  - визначений многочлен ступеня  $r$ :

$$Q_0(z) = 1, \quad Q_1(z) = 1 + z, \quad Q_2(z) = 1 + 4z + z^2, \quad Q_3(z) = 1 + 11z + 11z^2 + z^3, \\ Q_4(z) = 1 + 26z + 66z^2 + 26z^3 + z^4.$$

Підставляючи (5) у (2), елементарними перетвореннями приводимо  $S_{\xi^*}(z)$  до вигляду (1) та з умови  $S_{\xi^*}(z) \equiv S_\xi(z)$  визначаємо  $p$ ,  $q$ ,  $\{a_k\}$ ,  $\{b_i\}$ .

Зокрема, якщо  $R(r) = e^{-\omega_* |r|}$ ,  $\omega_* > 0$ , то за формулами (4), (5), (2), (1) одержуємо

$$R(m\Delta) = e^{-m\lambda_1}, \quad \lambda_1 = -\gamma_*, \quad \gamma_* = \omega_* \Delta, \quad A_{01} = 1;$$

$$F(z) = \frac{1}{1-z\rho} + \frac{1}{1-z^{-1}\rho} = \frac{1}{(1-z\rho)(1-z^{-1}\rho)}, \quad \rho = e^{-\gamma}; \quad (6)$$

$$A(z) = \sqrt{1-z^2}, \quad B(z) = 1-z\rho.$$

Отже,  $q = 0$ ,  $p = 1$ ,  $a_0 = \sqrt{1-\rho^2}$ ,  $b_1 = -\rho$ , і моделювання  $\xi_t^*$  здійснюємо за допомогою процесу авторегресії та ковзаючого середнього порядку (1,0) (чистого авторегресійного процесу):

$$\xi_{m\Delta} = a_0 \eta_{m\Delta} - b_1 \xi_{(m-1)\Delta}, \quad m = 0, 1, 2, \dots \quad (7)$$

### Студентам необхідно:

1. Змоделювати на комп'ютері гауссівський випадковий процес, який описується формулами (6), (7),.
2. Одержати графіки  $N$  реалізацій даної моделі випадкового процесу.
3. Дані для обчислень:

Номер	$\omega$	$T$	$\Delta$	$N$ реалізацій
1	$20\pi$	5	0,02	5
2	$10\pi$	3	0,01	4
3	$20\pi$	4	0,03	3
4	$15\pi$	5	0,01	3
5	$17\pi$	10	0,10	5
6	$13\pi$	7	0,10	4
7	$9\pi$	3	0,01	3
8	$5\pi$	2	0,01	5
9	$7\pi$	1	0,005	3
10	$18\pi$	3	0,01	4

### Завдання 5. Моделювання процесів випадкового блукання

Процес випадкового блукання відноситься до класу процесів з незалежними приростами і досить часто зустрічається у прикладних задачах з економіки, статистичної фізики, техніки.

Визначимо процес  $\xi_1 \in R^1(t \in [0, \infty))$  випадкового блукання “частинки” (за яку може бути прийнято, наприклад, броунівську частинку, деякий стохастичний об'єкт чи систему, на які випадковим чином діє зовнішнє середовище) із дискретним часом  $t = m\Delta$

$m = 0, 1, \dots$  та дискретним фазовим простором станів:

$$\xi_1 \in \{r : r = l\delta, l = 0 \pm 1 \pm 2, \dots\}.$$

Тут  $\Delta > 0$  та  $\delta > 0$  — кроки дискретизації часу й фазового простору відповідно.

Нехай у момент  $t$  частинка знаходиться у стані  $r \in X : \xi_t = x$ . Тоді у наступний момент  $t + \Delta$  вона переходить у стан  $x \pm \delta$  з імовірністю  $p_{\pm}$ , а з імовірністю  $1 - p_+ - p_-$  залишається у тому ж самому стані  $x$  ( $p_{\pm} > 0, 0 < p_+ + p_- \leq 1$ ). Початковий стан  $\xi_0 = 0$ .

Для моделювання  $\xi_t$  визначимо послідовність незалежних дискретних випадкових величин  $\eta_0, \eta_1, \dots$ :

$$\eta_t \in \{-\delta, 0, \delta\}. P\{\eta_t = \pm\delta\} = p_{\pm}. P\{\eta_t = 0\} = 1 - p_+ - p_-.$$

Моделювання  $\{\eta_t\}$  здійснюється за таким методом:

1. Звернення до стандартної підпрограми та одержання псевдовипадкового числа з інтервалу  $(0, 1)$  з рівномірним законом розподілу
2. Порівняння цього числа  $A$  з величинами  $\{s_i\}$  та визначення номера  $k$ , для якого  $s_{k-1} \leq A \leq s_k$ , де  $k = \overline{1, 3}$

$$s_0 = 0; s_1 = p_-; s_2 = 1 - p_+; s_3 = 1.$$

3. Прийняття рішення про те, що реалізацією  $\eta_t$  буде:

$$-\delta, k = 1;$$

$$0, k = 2;$$

$$\delta, k = 3.$$

Тоді процес  $\xi_t$  породжується процесом  $\eta_t$  за допомогою очевидного співвідношення

$$\xi_{t+\Delta} = \xi_t + \eta_t, \quad t = 0, \Delta, 2\Delta, \dots \quad (18)$$

Важливе теоретичне та прикладне значення має асимптотична поведінка  $\xi_t$  при  $\Delta \rightarrow 0$ . Щоб запобігти виродженню цього процесу, параметри  $\delta, p_+, p_-$  повинні задовольняти асимптотичним співвідношенням

$$\delta = b\sqrt{\Delta} + o(\sqrt{\Delta}), p_{\pm} = \frac{1}{2} + \frac{a}{2b}\sqrt{\Delta} + o(\sqrt{\Delta}),$$

Де  $b > 0, a \in R^1$  — деякі константи. Роблячи граничний перехід при  $\Delta \rightarrow 0$ , одержуємо в результаті дифузійний випадковий процес з ко-

ефіцієнтом переносу  $a$  та коефіцієнтом дифузії  $b$ , а при  $a = 0, b = 1$  – стандартний вінерівський процес  $\{\omega_1 \equiv \omega(t, \omega)\}$ .

Треба зауважити, що існує “трубка” з межами

$$\int_{\pm} (t) = at \pm y_{1-\frac{1}{2}} b \sqrt{t}. \quad (19)$$

В яку при кожному  $t$  дифузійний процес із параметрами  $(a, b)$  повинен попасти з імовірністю 1-.

Тут

$$y_{1-\frac{1}{2}} = \phi^{-1}\left(1 - \frac{1}{2}\right),$$

де  $\phi^{-1}(z)$  – функція обернена до функції розподілу стандартного нормального закону:

$$\phi^{-1}(z) \approx \frac{2.30753 + 0.27061\theta}{1 + 0.99229\theta + 0.0481\theta^2} - \theta.$$

$$\theta = \sqrt{2 \ln z} \quad 0,5 \leq z \leq 1$$

#### Студентам необхідно:

1. Одержати  $N$  реалізацій процесу випадкового блукання.
2. Зобразити їх графічно.
3. На цьому ж малюнку зобразити “трубку” (19).
4. Дані для обчислень:

Номер	$p_+$	$p_-$	$\delta$	$T$	$\Delta$	$\varepsilon$	$N$
1	0,5	0,5	$\sqrt{\Delta}$	100	0,1	0,1	3
2	0,5	0,5	1	1000	1	0,1	4
3	$0,5 + 0,25 \sqrt{\Delta}$	$0,5 - 0,25 \sqrt{\Delta}$	$\sqrt{\Delta}$	100	0,1	0,1	5
4	0,3	0,2	$\sqrt{\Delta}$	100	0,1	0,01	3
5	0,2	0,3	$\sqrt{\Delta}$	100	0,1	0,05	4
6	0,4	0,4	1	1000	1	0,01	5
7	$0,5 + 0,25 \sqrt{\Delta}$	$0,5 - 0,25 \sqrt{\Delta}$	1	100	1	0,01	3
8	$0,5 + 0,25 \sqrt{\Delta}$	$0,5 - 0,25 \sqrt{\Delta}$	$\sqrt{\Delta}$	1000	0,1	0,01	4

9	$0,35+0,3\sqrt{\Delta}$	$0,35-0,3\sqrt{\Delta}$	$\sqrt{\Delta}$	100	0,1	0,1	3
10	$0,4+0,25\sqrt{\Delta}$	$0,4-0,25\sqrt{\Delta}$	$\sqrt{\Delta}$	100	0,1	0,1	3

### Контрольні запитання до завдань 4, 5

1. Що таке гауссівський випадковий процес?
2. Дайте означення кореляційної функції  $R(\tau)$  випадкового процесу.
3. Дайте означення спектральної щільності випадкового процесу.
4. Яка відмінність між траєкторією і реалізацією випадкового процесу?
5. Що таке модель авторегресії та ковзаючого середнього?
6. Дайте означення процесу броунівського руху, вінерівського процесу.
7. Як здійснюється моделювання дискретних випадкових величин  $\eta_t$ ?
8. Як зображається функція  $\phi(z)$  та “трубка”, в яку влучає дифузійний процес з параметрами  $(a, b)$ ?

### ***Завдання 6. Обчислення інтегралів методом Монте-Карло***

При обчисленні багатовимірних інтегралів виникають дві проблеми: по-перше, дуже важко одержати замкнені вирази для гарантованих оцінок похибки; по-друге, навіть якщо така оцінка одержана, то для ряду класів функцій вона є практично непотрібною, оскільки для досягнення заданої гарантованої похибки квадратурна формула повинна містити настільки багато вузлів, що виконати комп'ютерні обчислення практично неможливо.

Одним з методів обчислення інтеграла, при якому похибка оцінюється не гарантовано, а лише деяким ступенем достовірності, є метод Монте-Карло.

Нехай необхідно обчислити найближче значення інтеграла

$$I(f) = \iiint_{\Omega} \dots \int f(P) dx_1 dx_2 \dots dx_n, \quad (20)$$

де  $P = (x_1, x_2, x_n)$  – точка  $n$ - вимірного простору.

Без обмеження загальності можна вважати, що область

$$\Omega = \{(r_1, r_2, \dots, r_n) \mid x_i \in [0, 1], \quad i = \overline{1, n}\}$$

$n$  – вимірний одиничний куб.

Якщо  $P_j$  – послідовність випадкових попарно незалежних точок, рівномірно розподілених у кубі  $\Omega$ , то випадкові величини  $s_j = f(P_j)$  попарно незалежні й однаково розподілені, причому

$$M(s_j) = \iint_{\Omega} \dots \int f^2(P) dx_1 \dots dx_n - \left[ \iint_{\Omega} \dots \int f(P) dx_1 \dots dx_n \right]^2 \equiv D.$$

Тобто математичне сподівання даних величин збігається з шуканим значенням інтеграла. Розглянемо випадкову величину

$$S_N = \frac{1}{N} \sum_{j=1}^N s_j.$$

Завдяки попарній незалежності випадкових величин  $s_j$  одержимо:

$$M(S_N) = \frac{1}{N} \sum_{j=1}^N M(s_j) = \iint_{\Omega} \dots \int f(P) dx_1 \dots dx_n,$$
$$D(S_N) = \frac{1}{N^2} \sum_{j=1}^N D(s_j) = \frac{1}{N} D.$$

Якщо точки  $P_j$  незалежні за сукупністю, то завдяки центральній граничній теоремі випадкова величина  $(S_N - I(f)) / \sqrt{D/N}$  розподілена асимптотично нормально. Задамо довірчу ймовірність  $\beta = 0.95$ . Для нормально розподіленої випадкової величини  $\epsilon$

$$P(|\xi| < 1.96) = 0.95$$

Таким чином, при великих  $N$  з імовірністю близькою до 0,95 виконується нерівність

$$|S_N - I(f)| \leq 1.96 \sqrt{\frac{D}{N}}.$$

Величина  $D$  у правій частині апіорі є невідомою, але її можна емпірично оцінити за формулою

$$D \equiv \frac{1}{N} \sum_{i=1}^N s_i^2 - \left( \frac{1}{N} \sum_{i=1}^N s_i \right)^2.$$

Наведемо схему алгоритму для обчислення інтеграла за методом Монте-Карло для обраної ймовірності  $\beta = 0.95$ .

Послідовно для  $n = 1, 2, \dots$  одержуємо координати випадкових точок  $P_n$  та обчислюємо величини  $z_n, S_n, d_n, D_n$ , використовуючи рекурентні співвідношення

$$\begin{aligned} z_n &= z_{n-1} + f(P_n), \\ S_n &= z_n / n, \end{aligned} \quad (21)$$

$$d_n = d_{n-1} + \frac{n}{n-1} (f(P_n) - S_n)^2.$$

$$D_n = \frac{d_{n-1}}{n-1}$$

Початкові умови для рекурентних формул мають вигляд

$$z_1 = S_1 = f(P_1), \quad d_1 = D_1 = 0.$$

Паралельно обчислюємо величину

$$\omega_n = 1.96 \sqrt{\frac{D_n}{n}}.$$

Обчислення припиняємо якщо  $\omega_n \leq \varepsilon$ . У цьому випадку вважаємо, що шукане значення інтеграла  $I(f) \cong S_n$ , а нерівність

$$|I(f) - S_n| \leq \varepsilon \quad (22)$$

Виконується з ймовірністю 0,95.

### Студентам необхідно:

1. Методом Монте-Карло обчислити інтеграл за одиничним  $n$ -вимірним кубом.
2. Дані брати з таблиці

номер	$n$	$f(x)$	номер	$n$	$f(x)$
1	5	$\sin\left(\sum_{i=1}^n x_i^3\right)$	8	5	$\text{arctg}\left(\sum_{i=1}^n x_i^2\right)$
2	6	$\cos\left(\sum_{i=1}^n x_i^3\right)$	9	7	$\text{arctg}\left(\sum_{i=1}^n x_i^2\right)$
3	6	$1/\left(1 + \sum_{i=1}^n x_i\right)$	10	6	$\text{ch}\left(\prod_{i=1}^n x_i\right)$



4	5	$1/\left(1+\sum_{i=1}^n x_i^2\right)$	11	5	$\frac{\sin x_i^2 \cos x_i^3}{1+\sum_{i=1}^n \sin^2 x_i}$
5	4	$\sum_{i=1}^n \cos x_i^2$	12	8	$\arctg\left(\prod_{i=1}^n x_i\right)$
6	7	$1/\sqrt{1+2\sum_{i=1}^n x_i^2}$	13	7	$1/\left(1+\arctg\left(\sum_{i=1}^n x_i\right)\right)$
7	6	$\frac{\cos\left(\sum_{i=1}^4 x_i\right)}{\cos\left(\sum_{i=1}^6 x_i\right)}$	14	8	$\sin\left(\sum_{i=1}^n x_i\right)/\left(1+\sum_{i=1}^n \cos x_i^2\right)$

### Контрольні запитання

1. У чому полягають труднощі обчислення багатократних інтегралів?
2. Як слід знаходити багатократні інтеграли за методом Монте-Карло?
3. Як оцінюється наближене значення інтеграла, отримане методом статистичних випробувань?.
4. Методом статистичних випробувань знайдіть наближений розв'язок означеного інтеграла від вибраної вами функції
5. Яким чином можна зменшити трудомісткість статистичного оцінювання значення інтеграла
6. Наведіть порівняльну характеристику різних методів наближеного обчислення інтегралів

### СПИСОК ЛІТЕРАТУРИ

#### Основна

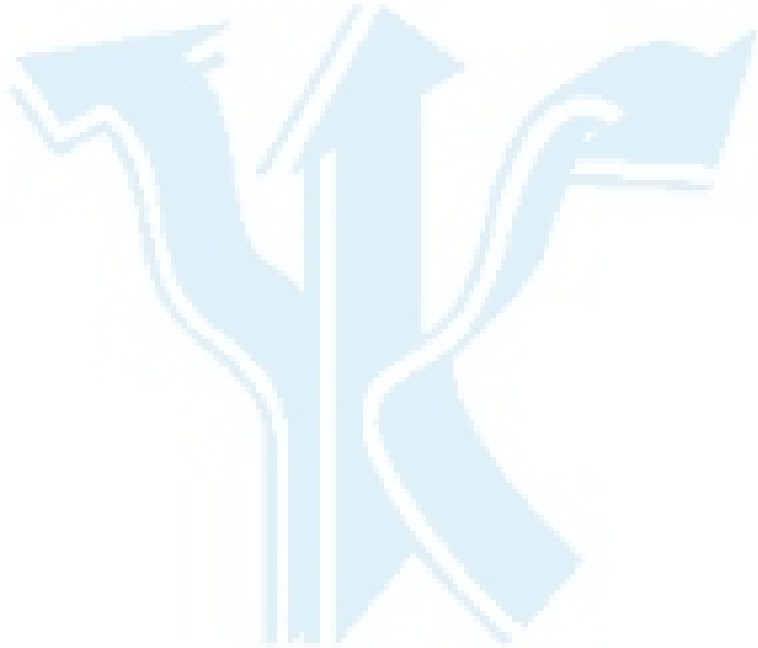
1. Харин Ю. С., Степанова М. Д. Практикум на ЕВМ по математической статистике. — Минск, 1987. — 304 с.
2. Плис А. И., Сливина Н. А. Лабораторный практикум по высшей математике. — М.: Высш. шк., 1994. — 410 с.
3. Гихман И. И., Скороход А. В., Ядренко М. И. Теория вероятностей и математическая статистика. — К.: Высш. шк., 1979. — 408 с.

4. *Ермаков С. М., Михайлов Г. А.* Статистическое моделирование. — 2-е изд. — М.: Наука, 1982. — 296 с.
5. *Розанов Ю. А.* Теория вероятностей, случайные процессы и математическая статистика. — М.: Наука, 1985. — 320 с.
6. *Боровков А. А.* Курс теории вероятностей. — М.: Наука, 1976. — 288 с.
7. *Боровков А. А.* Математическая статистика: Оценка параметров. Проверка гипотез. — М.: Наука, 1984. — 472 с.
8. *Бусленко Н. П.* Моделирование сложных систем. — М.: Наука, 1978. — 399 с.
9. *Справочник по теории вероятностей и математической статистике / В. С. Королюк, Н. И. Портенко, А. В. Скорокод, А. Ф. Турбин.* — М.: Наука, 1985. — 640 с.
10. *Маликов В. Т., Кветный Г. Н.* Вычислительные методы и применение ЭВМ. — К.: Высш. шк., 1989. — 213 с.
11. *Н. Матвеев В. Ф., Ушаков В. Г.* Системы массового обслуживания. — М., 1984. — 230 с.
12. *Коваленко И. Н., Гнеденко Б. В.* Теория вероятностей. — К.: Высш. шк., 1990. — 328 с.
13. *Адмиан Дж.* Стохастические системы. — М.: Мир, 1987. — 376 с.
14. *Гихман И. И., Скороход А. В.* Введение в теорию случайных процессов. — М.: Наука, 1977. — 568 с.
15. *Валтер Я.* Стохастические модели в экономике. — М.: Статистика, 1970.

*Додаткова*

16. *Соболь И. М.* Метод Монте-Карло. — М.: Наука, 1985. — 78 с.
17. *Советов Б. Я., Яковлев С. А.* Моделирование систем: Учебник. — М.: Высш. шк., 1985.
18. *Зеленський К. Х., Ігнатенко В. М., Коц О. П.* Комп'ютерні методи прикладної математики. — К.: Академперіодика, 2002. — 480 с.
19. *Ясинський В. К., Юрченко І. В.* Статистичне моделювання. Методичні вказівки до лабораторних робіт. — Чернівці: Рута, 1997. — 45 с.
20. *Томашевский В. М., Жданова О. Г., Жолдаков О. О.* Вирішення практичних завдань методами комп'ютерного моделювання. — К., 2001. — с 267 с.

21. *Михайлов Г. А., Войтишек А. В.* Численное статистическое моделирование. — М.: Академия, 2006.
22. *Соболь И. М.* Численные методы Монте-Карло. — М.: Наука, 1973.
23. *Харин Ю. С. и др.* Основы имитационного и статистического моделирования. — Минск, 1997.



МАУП

## **ЗМІСТ**

Пояснювальна записка .....	3
Теоретичний мінімум:основи статистичного моделювання та аналізу даних і моделей .....	4
Аналіз даних методами математичної статистики.....	19
Регресійний аналіз .....	39
Завдання для самостійних робіт та рекомендації щодо їх виконання.....	45
Список літератури.....	65

Відповідальний за випуск *А. Д. Вегеренко*  
Редактор *О. М. Коваленко*  
Комп'ютерне верстання *І. О. Музика*

Зам. № ВКЦ-4242

Формат 60×84/16. Папір офсетний. Друк ротатійний, трафаретний.  
Тираж 30 пр.

Міжрегіональна Академія управління персоналом (МАУП)  
03039 Київ-39, вул. Фрометівська, 2, МАУП

ДП "Видавничий дім "Персонал"  
03039 Київ-39, просп. Червонозоряний, 119, літ. XX

*Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи ДК № 3262 від 26.08.2008 р.*