

## Розділ 4

# Імовірнісне моделювання

- ◆ Метод статистичних випробувань
- ◆ Методи генерування випадкових чисел
- ◆ Програмні генератори випадкових чисел
- ◆ Моделювання випадкових величин із заданими законами розподілу
- ◆ Моделювання випадкових процесів і векторів
- ◆ Аналіз результатів моделювання

### 4.1. Метод статистичних випробувань

*Метод статистичних випробувань* — це числовий метод математичного моделювання випадкових величин, який передбачає безпосереднє включення випадкового фактора в процес моделювання і є його істотним елементом. Вплив випадкових факторів на систему моделюється за допомогою випадкових чисел. Результатом моделювання є випадкові процеси або величини, які характеризують систему, що моделюється. Щоб їх імовірнісні характеристики (імовірність деяких подій, математичне сподівання, дисперсія випадкових величин, імовірності попадання випадкової величини в задану область та ін.) співпадали з аналогічними параметрами реальної системи або процесу під час моделювання потрібно отримати велику кількість реалізацій випадкових величин або процесів. Таким чином, метод полягає в багатократному проведенні випробувань побудованої ймовірнісної моделі і подальшій статистичній обробці результатів моделювання з метою визначення шуканих характеристик розглядуваного процесу у вигляді оцінок його параметрів. Точність оцінок цих параметрів визначає ступінь наближення розв'язку задачі до ймовірносних характеристик.

На практиці метод статистичних випробувань доцільно використовувати в таких випадках, коли:

- ◆ розв'язувати задачу цим методом простіше, ніж будь-яким іншим;
- ◆ досліджується система, функціонування якої визначається багатьма ймовірнісними параметрами елементарних явищ;
- ◆ важко або неможливо побудувати аналітичну ймовірнісну модель системи.

Важливою властивістю цього методу є те, що для звичайних числових методів обсяг обчислень зростає в разі збільшення розмірності задачі приблизно як по-

казникова функція розмірності задачі, а для методу статистичних випробувань – лише як лінійна функція розмірності.

Незалежно від типу досліджуваної моделі системи, застосовуючи метод статистичних випробувань, необхідно виконати такі кроки.

1. Визначити, що являтиме собою кожне випробування і зазначити, яке випробування буде успішним, а яке – ні.
2. Обчислити кількість випробувань, які необхідно провести, щоб отримати результати із заданою точністю, і провести ці випробування.
3. Виконати статистичну обробку результатів випробувань та обчислити оцінки необхідних статистичних характеристик.
4. Проаналізувати точність отриманих статистичних характеристик.

Така послідовність кроків є обов'язковою під час розв'язування будь-якої задачі за допомогою методу статистичних випробувань. Однак конкретний зміст цих кроків залежить від поставленого завдання та типу досліджуваної системи. У цьому разі метод завжди потребує використання генераторів випадкових чисел із заданим законом розподілу.

У методі статистичних випробувань особливе значення відіграють випадкові числа, рівномірно розподілені в інтервалі  $[0, 1]$ . Найважливіша їх властивість полягає в тому, що за їх допомогою можна отримати вибіркові значення, які мають будь-який інший розподіл, або промодельювати випадковий процес з різними статистичними властивостями.

Отже, для використання методу статистичних випробувань необхідні певні можливості, а саме:

- ◆ генерувати випадкові числа, рівномірно розподілені в інтервалі  $[0, 1]$ ;
- ◆ описувати модельовані випадкові явища функціями розподілу ймовірностей та ймовірнісними процесами;
- ◆ мати методи отримання випадкових величин функцій розподілу ймовірностей (дискретних і неперервних), які базуються на випадкових числах, рівномірно розподілених у інтервалі  $[0, 1]$ ;
- ◆ оцінювати статистичні характеристики випадкових величин з отриманих за допомогою методу статистичних випробувань чисел вибіркової послідовності;
- ◆ визначати точність отриманих статистичних оцінок як функцій від числа випробувань.

Випадкові числа, рівномірно розподілені в інтервалі  $[0, 1]$ , мають дві основні властивості:

1. Якщо  $r_i$  ( $i = 1, 2, 3, \dots$ ) – випадкові числа, рівномірно розподілені в інтервалі  $[0, 1]$ , то їх кумулятивний розподіл  $F$  (за визначенням  $F(r_i) = P(r_i < r)$ ), задовольняє співвідношенням:

$$F(r_i) = \begin{cases} r_i, & \text{якщо } 0 \leq r_i \leq 1, \forall i, \\ 0, & \text{якщо } r_i < 0, \forall i, \\ 1, & \text{якщо } r_i > 1, \forall i. \end{cases}$$

Слід зауважити, що теоретично ці випадкові числа повинні бути вибірковими значеннями неперервної величини з функцією щільності, визначеною таким чином:

$$f(r) = \begin{cases} 1, & \text{для всіх } 0 \leq r \leq 1, \\ 0, & \text{для всіх інших значень.} \end{cases}$$

Насправді ж під час комп'ютерного моделювання використовуються тільки дискретні значення, в яких після десяткової коми є фіксована кількість десяткових знаків.

2. Випадкові числа  $r_1, r_2, \dots, r_n$  є незалежними, якщо їх сумісний кумулятивний розподіл  $G$  можна подати як добуток окремих функцій розподілу:

$$G(r_1, r_2, \dots, r_n) = F_1(r_1) F_2(r_2) \dots F_n(r_n),$$

або, враховуючи, що  $n$  випадкових чисел мають однакові розподіли, можна записати:

$$G(r_1, r_2, \dots, r_n) = F(r_1) F(r_2) \dots F(r_n),$$

Методи генерування випадкових чисел, рівномірно розподілених у інтервалі  $[0, 1]$ , буде описано нижче.

Розглянемо кілька задач, для розв'язування яких можна застосувати метод статистичних випробувань.

#### Приклад 4.1

Необхідно знайти площу фігури (рис. 4.1), обмежену функцією  $y = f(x)$  та осями координат  $Ox$  і  $Oy$ .

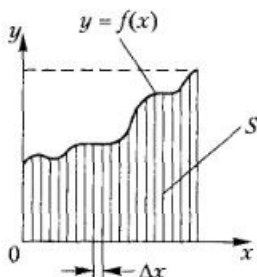


Рис. 4.1. Схема обчислення інтеграла

У числових методах для інтегрування використовується наближене зображення інтеграла у вигляді квадратурної формули. Одним із найпростіших є метод прямокутників. У разі використання методу прямокутників інтеграл апроксимується такою формулою:

$$S = \sum_{i=1}^n f(x_i) \Delta x.$$

Ця формула і є формулою числового інтегрування. Чим більша кількість інтервалів  $n$  і менший крок  $\Delta x$ , тим точніше можна обчислити площу  $S$ .

Тепер покажемо, як можна розв'язати цю задачу за допомогою методу статистичних випробувань. Спочатку пронормуємо функцію  $y = f(x)$  так, щоб уписати її в одиничний квадрат<sup>1</sup>. Припустимо, що  $\xi$  — деяка випадкова величина, рівномірно розподілена в інтервалі  $[0, 1]$ . Тоді ймовірність попадання значення  $\xi$  в будь-який відрізок  $[a, b] \in [0, 1]$  буде залежати тільки від довжини відрізка  $[a, b]$ , а не від місця його розташування в інтервалі  $[0, 1]$ , тобто ймовірність того, що вибіркоче значення випадкової величини  $\xi$  потрапить у деякий відрізок  $0 \leq a \leq b \leq 1$ , дорівнюватиме довжині цього відрізка:  $P(a \leq \xi \leq b) = \int_a^b d\xi = b - a$ .

Будемо використовувати одне значення випадкової величини  $\xi$  для визначення координати  $x_i$ , а друге — для визначення координати  $y_i$ . Таким чином, пара значень випадкової величини  $\xi$  задаватиме на площині точку з координатами  $(x_i, y_i)$ . Ймовірність попадання цієї точки в деяку область одиничного квадрата пропорційна площі цієї області та не залежить від місця розташування області в одиничному квадраті.

Проведемо  $N$  випробувань. Випробування будемо вважати успішним, якщо точка з координатами  $(x_i, y_i)$  потрапить в область під кривою  $y = f(x)$  або на неї. Підрахуємо кількість успішних випробувань, позначимо їх через  $m$  і визначимо частість успішних випробувань —  $m/N$ . На рис. 4.1 видно, що у разі збільшення кількості випробувань ця величина наближається до ймовірності попадання точки в заштриховану область  $P = S/S_{\text{од.кв}} = S \approx m/N$ , де  $S_{\text{од.кв}}$  — площа одиничного квадрата. Таким чином, згідно з теоремою Бернуллі

$$\lim_{N \rightarrow \infty} P\left(\left|\frac{m}{N} - p\right| < \varepsilon\right) = 1.$$

У разі прямування кількості випробувань  $N$  до нескінченності частість успішних випробувань буде відрізнятися від ймовірності  $p$  на нескінченно малу величину  $\varepsilon$ . Отже, можна вважати, що  $m/N$  — наближене значення шуканої площі  $S$ .

Цей приклад демонструє те, як метод статистичних випробувань може бути використано під час розв'язування детермінованих задач. На практиці такий підхід використовується для знаходження площ або об'ємів деяких багатовимірних фігур, які утворюються у випадку перетину різних геометричних тіл. У цьому разі число випробувань  $N$ , які необхідно провести для обчислення площі або об'єму, не залежить від кратності визначеного інтеграла.

#### Приклад 4.2

Припустимо, що чотири стрільці одночасно стріляють у рухому ціль. Ймовірність влучення в ціль кожним стрільцем дорівнює 0,5. Ціль вважається враженою, якщо в неї влучило два або більше стрільців. Потрібно знайти ймовірність ураження цілі.

Використовуючи методи теорії ймовірностей, цю задачу досить легко розв'язати аналітично. Дійсно, ймовірність ураження цілі одним пострілом  $p_{\text{вр}} = 1 - p_{\text{невр}}$ , де  $p_{\text{невр}}$  — ймовірність того, що ціль не буде вражена взагалі, визначається за формулою

$$p_{\text{невр}} = 0,5^4 + C_4^1 \cdot 0,5^1 \cdot 0,5^3 = 0,5^4 + 4 \cdot 0,5^1 \cdot 0,5^3 = 0,3125.$$

Звідси ймовірність ураження цілі

$$p_{\text{вр}} = 1 - p_{\text{невр}} = 1 - 0,3125 = 0,6875.$$

<sup>1</sup> Таке нормування необхідне для того, щоб спростити процедуру випробувань. Усі сучасні мови моделювання та програмування загального призначення мають вбудовані засоби генерування рівномірно розподілених незалежних випадкових величин у інтервалі  $[0, 1]$ .

Тепер покажемо, як розв'язати цю задачу за допомогою методу статистичних випробувань. Процедуру розіграшу можна реалізувати, одночасно підкидаючи чотири монети. Для моделювання підкидання однієї монети використовується одне значення  $r_i$ . Якщо  $r_i < p$ , вважаємо, що монета падає лицевим боком, і, таким чином, стрілець влучив у ціль. Інакше вважаємо, що стрілець промахнувся. Одне випробування — це підкидання чотирьох монет. Зробимо  $N$  випробувань і позначимо через  $m$  число успішних випробувань (дві або більше монет упали лицевим боком, що свідчить про те, що в ціль улучило два або більше стрільців). Тоді, згідно з теоремою Бернуллі,  $p_{\text{вр}} \approx m/N$ . У разі значного збільшення числа випробувань  $N$  і при будь-якому значенні  $\epsilon$  частість враження пілі буде збігатись до ймовірності  $p_{\text{вр}} = 0,6875$ .

### Приклад 4.3

Розглянемо більш складну задачу, яку розв'язати аналітично досить важко. Нехай є деяка ціль довільної форми загальною площею  $S$ , на яку бомбардувальники скидають  $n$  бомб. Площа враження кожної бомби — це круг з радіусом  $R$  (рис. 4.2). Ціль вважається враженою, якщо зруйновано  $K$  відсотків її площі  $S$ . Необхідно знайти ймовірність ураження цілі.



Рис. 4.2. Схематичне зображення ураження цілі бомбою

Для цього розглянемо область улучення бомб. За допомогою генератора випадкових чисел отримаємо координати падіння  $n$  бомб. Біля кожної точки падіння опишемо коло радіусом  $R$  (див. рис. 4.3) і визначимо площу враження, яку заштриховано на рисунку. Площу враження можна легко обчислити, використовуючи методи геометрії. Якщо площа враження становить  $K$  відсотків (або більше) загальної площі цілі  $S$ , то ціль вважається враженою, а випробування — успішним. Інакше ціль вважається неураженою, а випробування — неуспішним.

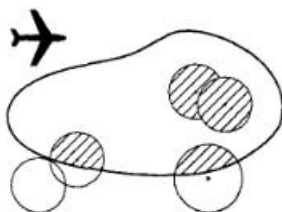


Рис. 4.3. Схематичне вирішення задачі

Проведемо  $N$  випробувань, моделюючи кожного разу координати  $n$  точок падіння бомб, і підрахуємо кількість випробувань, під час яких ціль була враженою. Тоді, згідно з теоремою Бернуллі, ймовірність враження цілі визначається за формулою  $p_{\text{вр}} \approx m/N$ , де  $m$  — кількість випробувань, за яких ціль була вражена. Оцінку математичного сподівання площі враження цілі можна визначити як

$$\bar{S} = \frac{1}{N} \sum_{i=1}^N S_i,$$

де  $S_i$  — площа враження під час  $i$ -го випробування. Згідно із законом великих чисел, якщо  $N \rightarrow \infty$ , то оцінка буде наближатись до математичного сподівання.

За допомогою методу статистичних випробувань можна обчислити будь-які характеристики випадкових величин і процесів. Крім того, цей метод можна застосовувати для розв'язування не тільки ймовірнісних, а й детермінованих задач. Але під час його застосування слід пам'ятати, що для отримання результату з наперед заданою точністю необхідно провести велику кількість випробувань, для чого потрібні довгі послідовності випадкових чисел.

## 4.2. Генератори випадкових чисел

Найбільше прикладів генерування випадкових чисел можна знайти в ігровому бізнесі. Це номери в спортивних лотереях, числа, які випадають на рулетці, варіанти розкладу карт тощо. Більшість комп'ютерних ігор теж базується на випадкових числах.

### 4.2.1. Типи генераторів

Без комп'ютера використання випадкових чисел, передбачене методом статистичних випробувань, не має сенсу, тому генератори випадкових чисел повинні бути безпосередньо з'єднані з комп'ютером. Це можна зробити за допомогою апаратних приставок до комп'ютера (апаратні методи) або спеціальних програм (програмні методи). Крім того, під час моделювання можна використати готові таблиці випадкових чисел, які слід розміщати в пам'яті комп'ютера або на зовнішньому накопичувачі.

*Апаратні методи* генерування випадкових чисел базуються на використанні деяких фізичних явищ (наприклад, шумів електронних приладів, радіоактивного випромінювання та ін.). Під час застосування апаратних генераторів випадковий електричний сигнал перетворюють у двійковий код, який уводиться в комп'ютер за допомогою спеціальних аналого-цифрових перетворювачів. Один з найбільш поширених методів – це використання шумів електронних приладів. Якщо на підсилювач не подавати ніякого сигналу та увімкнути його на повну потужність, то буде чути шипіння (шум). Це і є шум електронних елементів підсилювача, який є випадковим процесом. Цей неперервний сигнал можна перетворити в дискретний. Існують різні схеми перетворення випадкового сигналу в послідовність двійкових цифр [47]. У більшості випадків його підсилюють і встановлюють граничне значення напруги шумового сигналу, перевищення якого можна вважати значенням двійкової одиниці на деякому малому проміжку часу  $t$ . У протилежному випадку отримуємо двійковий нуль. Для отримання  $m$ -розрядного випадкового двійкового числа провадиться  $m$  вимірювань неперервного сигналу у фіксовані моменти часу  $t_1, t_2, \dots, t_m$ .

Вбудовані в комп'ютери апаратні генератори випадкових чисел останнім часом часто використовуються в системах захисту інформації. Прикладом застосування таких генераторів для забезпечення конфіденційності, цілісності та достовірності електронної інформації, яка зберігається в комп'ютері або передається по мережі, є пристрій для шифрування даних PadLock, інтегрований у деякі моделі



процесорів, розроблених компанією Intel. Пристрій має інтерфейс прикладного рівня, що дає змогу розробникам програмного забезпечення отримувати випадкові числа без використання програмних драйверів. Такий спосіб отримання високоякісних випадкових послідовностей простіший та ефективніший, ніж використання апаратно-програмної RNG (Random Number Generator) архітектури і суто програмних генераторів, що особливо важливо під час побудови захищених і криптографічних програм.

*Табличний метод.* У 1955 році корпорація «Ренд» опублікувала таблиці випадкових чисел, які мали мільйон значень. Для заповнення цих таблиць застосовувались апаратні методи. Дані цих таблиць можна використовувати під час моделювання систем за допомогою методу статистичних випробувань. У сучасних комп'ютерах ці таблиці можна зберігати на зовнішніх носіях або навіть в основній пам'яті. Головним недоліком табличного методу є те, що під час його використання витрачаються значні об'єми основної пам'яті комп'ютера.

Найбільш розповсюдженими на практиці є *програмні генератори*, які дають змогу отримувати послідовності *випадкових чисел* за рекурентними формулами. Якщо бути абсолютно точним, то числа, які виробляють програмні генератори, насправді є *псевдовипадковими* («псевдо» у перекладі з грецької – *нібито*). Так їх називають тому, що алгоритми їх отримання завжди є детермінованими.

Загалом же програмні генератори повинні задовольняти таким вимогам:

- ◆ генерувати статистично незалежні випадкові числа, рівномірно розподілені в інтервалі  $[0, 1]$ ;
- ◆ мати можливість відтворювати задані послідовності випадкових чисел;
- ◆ затрати ресурсів процесора на роботу генератора повинні бути мінімальними;
- ◆ легко створювати незалежні послідовності випадкових чисел (поток).

Слід звернути увагу на те, що більшість програмних генераторів виробляють випадкові числа, рівномірно розподілені в інтервалі  $[0, 1]$ . Необхідність моделювання таких чисел обумовлена тим, що на їх основі можна отримати випадкові числа практично будь-яких розподілів. Потрібно також мати на увазі, що випадкові числа, які виробляють програмні генератори, є *квазірівномірно* розподіленими («квазі» у перекладі з латинської – *майже*). Причина в тому, що вони створюються комп'ютером, кількість двійкових розрядів якого обмежена, і за його допомогою можна зобразити тільки дискретні (а не неперервні) значення з діапазону від 0 до 1.

Якість роботи генераторів визначається статистичними властивостями послідовностей випадкових чисел, які він виробляє, – незалежністю і випадковістю. Властивості послідовностей перевіряються за статистичними критеріями, детально описаними нижче.

Здатність відтворювання послідовності випадкових чисел полягає в тому, що за однакових початкових умов і параметрів генератор повинен відтворювати одні й ті ж послідовності псевдовипадкових чисел. Ідентичні послідовності випадкових чисел рекомендується використовувати у випадку, коли потрібно порівняти альтернативні варіанти систем, що моделюються, і налагодити програми. Однак можливість відтворення не завжди бажана під час моделювання систем і в комп'ютерних іграх (як це було в перших версіях відомої гри «Тетріс», коли кожна гра

починалась з тієї ж послідовності фігур). Для усунення такого недоліку початкові значення величин, необхідних для запуску програмного генератора, рекомендується брати з таймера комп'ютера.

Під час дослідження складних систем виникає необхідність у моделюванні послідовностей випадкових чисел великої довжини. Для їх створення потрібні швидкодіючі алгоритми генерування з мінімальними вимогами до ресурсів комп'ютера. Інколи дослідники з невеликим досвідом роботи використовують під час моделювання складних систем один і той же генератор, звертаючись до нього з різних місць програми. Однак це часто призводить до того, що процес моделювання швидко вироджується у зв'язку з виходом псевдовипадкової послідовності за межі аперіодичності. Тому для моделювання різних випадкових факторів бажано мати окремі послідовності (набори значень), які відтворювались би одним і тим же генератором, але за різних значень параметрів.

У більшості генераторів псевдовипадкових чисел  $x_i$  використовується рекурентна процедура  $x_{i+1} = f(x_i)$ . Найпростішим та найдавнішим серед таких генераторів є генератор фон Неймана та Метрополіса, робота якого базувалась на методі середин квадратів. Пояснимо суть цього методу.

Зобразимо умовно довільне чотирирозрядне десяткове число як  $x\ x\ x\ x$ . Піднесемо це число до квадрату і в отриманому результаті відкинемо по дві цифри зліва і справа:

$$\begin{array}{r} \times \quad x \ x \ x \ x \\ \quad \quad \underline{x \ x \ x \ x} \\ x \ x \mid \times \quad x \ x \ x \ x \mid x \ x \\ \quad \quad \underline{x \ x \ x \ x} \end{array}$$

Чотири цифри, які залишилися, і є новим випадковим числом. Якщо результатом множення є число з кількістю цифр менше восьми, зліва дописуються додаткові нулі. Реалізувати програмно цей генератор дуже просто, але він має ряд вад:

- ◆ якщо початкове число парне, то може відбутися виродження послідовності, тобто починаючи з деякого значення всі наступні дорівнюватимуть нулю (спробуйте взяти як початкове число 4500);
- ◆ числа, які виробляє генератор, є сильно корельованими.

#### 4.2.2. Лінійні конгруентні генератори

Зважаючи на ці вади, на практиці використовують більш складні програмні генератори. У більшості сучасних програмних генераторів використовується властивість *конгруентності*, яка полягає в тому, що два цілих числа  $A$  і  $B$  є конгруентними за модулем  $m$ , якщо їх різниця  $(A - B)$  є числом, яке ділиться на  $m$  без остачі (тобто є кратним  $m$ ). Записується це так:

$$A = B \pmod{m}.$$

Наприклад, щоб знайти число, конгруентне з числом 134 за модулем 10, необхідно знайти цілочислову *остачу* від ділення 134 на 10, яка дорівнює 4.



Наведемо кілька прикладів обчислення конгруентних значень для різних  $m$ :

$$12 \equiv 5 \pmod{7}; 35 \equiv 5 \pmod{10}; 125 \equiv 5 \pmod{10}.$$

Серед методів генерування випадкових чисел найбільш поширеним є лінійний мультиплікативний конгруентний метод:

$$x_{i+1} = (ax_i + c) \pmod{m}, \quad (4.1)$$

де  $i = 1, 2, \dots$ ;  $a, c$  і  $m$  – цілі константи. Щоб отримати нове число, необхідно взяти псевдовипадкове число  $x_i$  (або задати вихідне  $x_0$ ), помножити його на коефіцієнт  $a$ , додати константу  $c$  і взяти модуль отриманого числа за  $m$ , тобто розділити на  $m$ , і отримати остачу. Ця остача і буде наступним псевдовипадковим числом  $x_{i+1}$ . У разі правильного підбору параметрів цей генератор повертає випадкові числа від 0 до  $m - 1$ .

Отримані за формулою (4.1) значення  $x_{i+1}$  належать до діапазону  $0 < x_{i+1} < m - 1$  і мають рівномірний дискретний розподіл. Для того щоб отримати випадкове значення  $r_{i+1}$  з інтервалу  $[0, 1]$ , необхідно число  $x_{i+1}$  розділити на  $m$ . У цьому разі всі значення  $m, c, a, x_0$  повинні бути додатними й задовольняти умовам:  $0 < m; a < m; c < m; x_0 < m$ . Отримана за формулою (4.1) послідовність називається *лінійною конгруентною послідовністю*.

Однією із вад лінійних конгруентних генераторів є те, що отримані випадкові числа  $x_{i+1}$  суттєво залежать від значень  $m, c, a, x_0$  і обчислюються за однією й тією ж формулою (4.1), тобто не є абсолютно випадковими. Але незважаючи на те що алгоритм їх отримання є детермінованим, за умови відповідного вибору констант  $m, c, a$  послідовність чисел  $x_{i+1}$ , на основі яких отримують значення  $r_{i+1}$ , повністю задовольнятиме більшості статистичних критеріїв.

Ще одна вада цих генераторів стосується того, що випадкові числа  $r_{i+1}$ , отримані за допомогою генератора, можуть приймати тільки дробово-раціональні значення –  $0; 1/m; 2/m; \dots; (m-1)/m$ . Більше того, числа  $r_{i+1}$  можуть приймати лише деякі з указаних значень залежно від вибраних параметрів  $m, c, a$  і  $x_0$ , а також від того, як реалізується операція ділення чисел з плаваючою комою на число  $m$  у комп'ютері, тобто залежно від типу комп'ютера і системи програмування. Наприклад, якщо  $m = 10, x_0 = a = c = 7$ , то отримаємо послідовність  $7; 6; 9; 0; 7; 6; 9; 0, \dots$ , яка не є випадковою. Це свідчить про важливість правильного вибору значень констант  $m, c, a$  і  $x_0$ . Правильно підібрані значення іноді називають *магічними числами*.

Наведений приклад ілюструє й те, що конгруентна послідовність завжди є циклічною, тобто вона починає повторюватися через певну кількість випадкових чисел. Кількість значень, після яких випадкові числа починають повторюватися, називається *повним періодом* генератора і є основним його параметром. Значення повного періоду залежать від розрядності комп'ютера, а також від значень  $m, c, a$  і  $x_0$ . Існує теорема, яка визначає умови існування повного періоду генератора, а саме:

- ◆ числа  $c$  і  $m$  повинні бути взаємно простими, тобто мати взаємний дільник 1;
- ◆ значення  $b = a - 1$  має бути кратним  $q$  для кожного простого  $q$ , бути дільником  $m$ ;
- ◆ значення  $b$  має бути кратним 4, якщо  $m$  кратне 4.

Достатність цих умов уперше було доведено Халлом (Hull) і Добеллом (Dobell). Повне доведення теореми можна знайти в літературі [23].

Якщо  $c > 0$ , то генератор називається *мішаним*, а якщо  $c = 0$  – *мультиплікативним*.

Розглянемо, як потрібно вибирати параметри лінійного конгруентного генератора, щоб отримати послідовність з повним періодом. Для отримання такої послідовності необхідно вибирати значення  $m = 2^g - 1$ , де  $g$  – довжина розрядної сітки комп'ютера. Для 32-розрядного комп'ютера  $m$  – найбільше ціле число, яке може бути відтворене в ньому, дане число дорівнює  $2^{31} - 1 = 2147483647$  (один розряд відводиться під знак числа). У цьому разі ділення  $x_{i+1}/m$  виконувати не обов'язково. Якщо в результаті роботи генератора буде отримане число  $x_{i+1}$ , яке більше ніж те, що може бути відтворене в комп'ютері, виникне переповнення розрядної сітки. Це призведе до втрати крайніх лівих двійкових знаків цілого числа, які перевищили допустимий розмір. Однак розряди, що залишились, саме і є значеннями  $x_{i+1} \pmod{2^g}$ . Таким чином, під час генерування замість операції ділення можна скористатись переповненням розрядної сітки.

Стосовно константи  $c$  теорема стверджує, що для отримання послідовності з повним періодом генератора значення  $c$  повинне бути непарним, і, крім того,  $a - 1$  має ділитися на 4. Для такого генератора початкове значення  $x_0$  може бути довільним і лежати в діапазоні від 0 до  $m - 1$ . Якщо  $c = 0$ , то отримуємо мультиплікативний конгруентний метод, який передбачає використання таких рекурентних виразів:

$$x_i = ax_i \pmod{m}. \quad (4.2)$$

Цей метод більш швидкодіючий, ніж попередній, але він не дає послідовності з повним періодом. Дійсно, з виразу (4.2) видно, що значення  $x_{i+1} = 0$  може з'явитись тільки в тому випадку, якщо послідовність вироджується в нуль. Взагалі, якщо  $d$  – будь-який дільник  $m$  і  $x_i$  кратне  $d$ , всі наступні елементи мультиплікативної послідовності  $x_{i+1}, x_{i+2}, \dots$  будуть кратними  $d$ . Таким чином, якщо  $c = 0$ , потрібно, щоб  $x_i$  і  $m$  були взаємно простими числами з  $m$  і знаходились між 0 і  $m$ .

Що стосується вибору  $a$ , то у випадку, коли  $m = 2^g$ , де  $g \geq 4$ , викладені в праці [23] умови приводять до єдиної вимоги стосовно значення  $a$ :

$$a \equiv 3 \pmod{8} \quad \text{або} \quad a \equiv 5 \pmod{8}.$$

У цьому випадку четверта частина всіх можливих значень множників дає довжину періоду, що дорівнює  $m/4$ , яка й буде максимальним періодом генератора.

Наведемо одну з найкращих програм генерування послідовностей рівномірно розподілених випадкових чисел.

```
/* Програма мовою C для TT800: версія від 8 липня 1996 р. */
/* M. Matsumoto, email: matumoto@math.keio.ac.jp */
/* genrand() генерує одне псевдовипадкове число */
/* з подвійною точністю */
/* рівномірно розподілене в інтервалі [0, 1] */
/* для кожного запиту. Можна вибирати будь-які */
/* 25 початкових значень */
/* окрім всіх нулів. */
```

```

/* Див.: ACM Transactions on Modelling and Computer Simulation, */
/* Vol. 4, No. 3, 1994, pages 254-266. */

#include
#define N 25
#define M 7

double
genrand()
{
    unsigned long y;
    static int k = 0;
    static unsigned long x[N]={ /* Початкові 25 значень за бажанням */
        0x95f24dab, 0x0b685215, 0xe76ccae7, 0xaf3ec239, 0x715fad23,
        0x24a590ad, 0x69e4b5ef, 0xbf456141, 0x96bc1b7b, 0xa7bdf825,
        0xc1de75b7, 0x8858a9c9, 0x2da87693, 0xb657f9dd, 0xffdc8a9f,
        0x8121da71, 0x8b823ecb, 0x885d05f5, 0x4e20cd47, 0x5a9ad5d9,
        0x512c0c03, 0xea857ccd, 0x4cc1d30f, 0x8891a8a1, 0xa6b7aadb
    };
    static unsigned long mag01[2]={
        0x0, 0x8ebfd028 /* Це чарівний вектор 'a', що не змінюється*/
    };
    if (k==N) { /* Генерування N слів одночасно */
        int kk;
        for (kk=0;kk> 1) ^ mag01[x[kk] % 2];
    }
    for (; kk> 1) ^ mag01[x[kk] % 2];
    }
    k=0;
    }
    y = x[k];
    y ^= (y << 7) & 0x2b5b2500; /* s i b, чарівні вектори*/
    y ^= (y << 15) & 0xdb8b0000; /* t i c, чарівні вектори */
    y &= 0xffffffff; /* вилучить цей рядок, якщо довжина слова = 32 */
    /* Наступний рядок був доданий в версії 1996 р. для зменшення кореляції. */
    y ^= (y >> 16); /* Додати для версії 1994 р. */
    k++;
    return( (double) y / (unsigned long) 0xffffffff);
}
/* main() виводить перші 50 згенерованих чисел */
main()
{ int j;
  for (j=0; j<50; j++) {
    printf("%5f ", genrand());
    if (j%8==7) printf("\n");
  }
  printf("\n");
}

```

Існують й інші конгруентні методи генерування випадкових чисел, серед яких слід відзначити адитивний.

Найпростіший генератор, послідовність якого  $x_{i+1}$  залежить більше ніж від одного з попередніх значень, — це генератор, що використовує числа Фібоначчі:

$$x_{i+1} \equiv (x_i + x_{i-1}) \pmod{m}.$$

Цей генератор широко використовувався в 50-ті роки ХХ століття, але, як показали подальші дослідження, статистичні властивості його досить низькі. Більш складні методи генерування випадкових чисел можна знайти в праці [23].

Розглянуті в цьому розділі методи генерування випадкових чисел не стосуються дослідників, які використовують мови або пакети моделювання. Ці засоби містять вбудовані генератори випадкових чисел, якість яких залежить від розробників цих програмних засобів. На жаль, є ще багато пакетів моделювання, в яких застосовуються генератори досить низької якості [18]. Що стосується мов програмування загального призначення, то засоби генерування випадкових чисел, вбудовані в них, взагалі не задовольняють будь-яким статистичним критеріям. Тому їх не слід використовувати під час проведення відповідальних досліджень.

### 4.3. Перевірка послідовностей випадкових чисел

Статистичні властивості всіх послідовностей випадкових чисел, які будуть використовуватись під час проведення досліджень, потрібно ретельно перевіряти. Для цього застосовують емпіричні та теоретичні критерії.

*Емпіричні* критерії – це звичайні тести, в яких під час обчислення статистичних даних використовують вибіркові значення  $r_i$ , що виробляються генератором.

*Теоретичні* критерії не є тестами в тому розумінні, в якому вони передбачаються в математичній статистиці. У разі їх використання не потрібна послідовність випадкових значень  $r_i$ . Глобальна оцінка властивостей генератора формується на основі числових значень його параметрів. Теоретичні критерії визначають характеристики послідовності за допомогою методів, які ґрунтуються на рекурентних правилах створення послідовності.

Відомо одинадцять емпіричних критеріїв [23], що застосовуються для перевірки статистичних властивостей послідовностей дійсних чисел  $r_i$ ,  $i = 1, 2, \dots$ , які вважаються незалежними та рівномірно розподіленими в інтервалі  $[0, 1]$ .

Для оцінювання наближеності отриманого розподілу до рівномірного застосовують чотири типи тестів:

- ◆ *частотний* – з використанням або критерію Колмогорова–Смирнова, або критерію  $\chi^2$ ;
- ◆ *автокореляційний* – з вимірюванням кореляції між  $x_n$  і  $x_{n+k}$ , де  $k$  – зсув по послідовності ( $k = 1, 2, 3, \dots$ );
- ◆ *серіальний* – з фіксацією частоти появи всіх можливих комбінацій чисел (по 2, по 3, по 4 рази) і виконанням оцінювання за критерієм  $\chi^2$ ;
- ◆ *циклічний* – з перевіркою кількості циклів більше і менше деякої константи, за яку береться значення математичного сподівання із підрахунком істинного числа циклів різної довжини, що порівнюється за критерієм  $\chi^2$  з очікуваним числом циклів.

Серед інших критеріїв важливу роль відіграє *спектральний критерій*, який застосовується для перевірки конгруентних генераторів випадкових чисел. Вважається, що цей тест найбільш потужний. Спектральний тест використовують

для перевірки гіпотези про рівність сумісних розподілів  $t$  послідовних елементів випадкової послідовності. Якщо задано послідовність  $\{r_i\}$  з періодом  $m$ , то для перевірки за цим тестом необхідно проаналізувати множину всіх  $m$  точок

$$\{r_i, r_{i+2}, \dots, r_{i+t-1}\}, \quad 0 \leq i < m.$$

у  $t$ -вимірному просторі.

## 4.4. Моделювання випадкових подій та дискретних величин

У разі дослідження складних систем методом статистичних випробувань необхідно мати можливість отримувати за допомогою комп'ютера вибіркові значення випадкових величин, які мають різні закони розподілу. Випадкові величини зазвичай моделюють за допомогою перетворення одного або кількох незалежних значень випадкової величини  $R$ , рівномірно розподіленої в інтервалі  $[0, 1]$ , що позначаються як  $r_i$ ,  $i = 1, 2, 3, \dots$  ( $r_i \in [0, 1]$ ). Значення  $r_i$  генерують, як звичайно, за допомогою програмних генераторів випадкових чисел.

### 4.4.1. Незалежні випадкові події

Припустимо, що ймовірність настання деякої елементарної випадкової події  $A$  в одному випробуванні дорівнює  $P(A) = p$ . Вважається, що умови проведення кожного випробування однакові і його можна повторити нескінченну кількість разів. Якщо  $r_i$  – це значення рівномірно розподіленої в інтервалі  $[0, 1]$  величини, то можна стверджувати, що за умови  $r_i \leq p$  (рис. 4.4) настане подія  $A$ , а якщо  $r_i > p$ , то відбудеться подія  $\bar{A}$ .

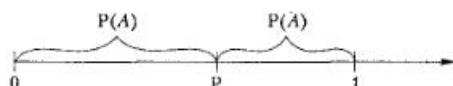


Рис. 4.4. Моделювання настання випадкових подій

Дійсно, якщо  $f(r)$  – функція щільності рівномірно розподіленої випадкової величини  $r$ , то

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

Ця модель добре описує такі події, як обслуговування вимоги в пристрої СМО, що може бути вільним або зайнятим, успішну або ні спробу виконання деякого завдання, влучення або ні в ціль, розгалуження потоків інформації у двох і більше напрямках. У деяких мовах для моделювання випадкової події використовується спеціальний блок (наприклад, у мові GPSS – блок TRANSFER, який працює в статистичному режимі [68]).

### 4.4.2. Група несумісних подій

Нехай є група несумісних подій  $A_1, A_2, \dots, A_k$ , настання яких необхідно дослідити. Відомі ймовірності настання цих подій  $p_1 = P(A_1), p_2 = P(A_2), p_3 = P(A_3), \dots, p_k = P(A_k)$ . Якщо події несумісні, то  $\sum_{i=1}^k p_i = 1$ . Припустимо, що  $p_0 = 0$ . На відрізьку  $[0, 1]$  числової осі відкладемо значення цих імовірностей (рис. 4.5).

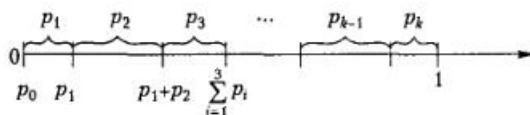


Рис. 4.5. Моделювання групи несумісних подій

Якщо отримане від генератора випадкових чисел значення  $r_i$  потрапляє в інтервал від  $\sum_{k=0}^{i-1} p_k$  до  $\sum_{k=0}^i p_k$ , вважаємо, що відбулася подія  $A_i$ . Таку процедуру називають визначенням результату випробування за жеребом. Вона ґрунтується на формулі

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

де  $p_0 = 0$ .

Ця модель часто використовується в теорії прийняття рішень і добре відтворює процеси вибору однієї з багатьох альтернатив у комп'ютерних іграх, розгалуження потоків інформації у вузлах мережі в кількох напрямках, вибір одного з багатьох пристроїв для обслуговування в СМО і т. ін.

### 4.4.3. Умовна подія

Умовна подія  $A$  – це подія, яка відбувається з імовірністю  $P(A/B)$  тільки за умови, що настала подія  $B$  (рис. 4.6). У цьому разі має бути задана ймовірність  $P(B)$  настання події  $B$ . Моделювання настання умовної події  $A$  провадиться таким чином. Спочатку випадкове число  $r_1$ , отримане від генератора випадкових чисел, використовується для моделювання настання події  $B$ . Подія  $B$  настає в тому випадку, якщо справджується нерівність  $r_1 \leq P(B)$ . Настання події  $A$  моделюється за допомогою числа  $r_2$ . Для цього перевіряється умова  $r_2 \leq P(A)$ , за виконання якої приймається рішення, що подія  $A$  відбулася. Якщо ж подія  $B$  не відбулася (тобто настає подія  $\bar{B}$ ), то настання події  $A$  моделювати не потрібно. Таким чином, можна скоротити загальну кількість випробувань.

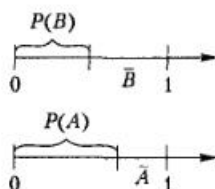


Рис. 4.6. Моделювання настання умовної події



#### 4.4.4. Випадкова дискретна величина

Одне з основних понять теорії ймовірностей – дискретна випадкова величина  $X$ , яка набуває конкретних значень  $x_i$  з імовірністю  $p_i$ . Ці випадкові величини називають цілочисловими. Якщо можливі значення випадкової величини становлять скінченну послідовність, то розподіл імовірностей випадкової величини визначають, задаючи значення  $x_1, x_2, \dots, x_n$  і відповідних їм імовірностей  $p_1, p_2, \dots, p_n$ . Моделювання випадкової дискретної величини виконується аналогічно моделюванню групи несумісних подій, тобто випадкову величину  $X$  подають як повну групу подій:

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

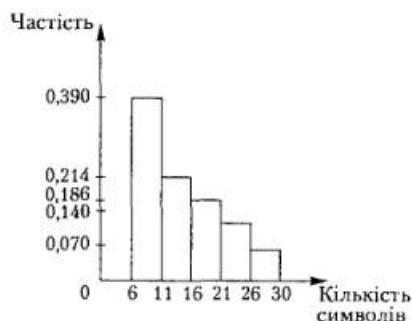
Для моделювання дискретної випадкової величини  $X$  зручно використовувати дискретну кумулятивну функцію. Для цього аналізують можливі значення випадкової величини  $X$  і будують гістограму розподілу можливих значень.

Побудову і використання кумулятивної функції розглянемо на прикладі моделювання процесу введення даних під час роботи текстового терміналу. В табл. 4.1 наведено результати, які відображають результати спостереження за об'ємом інформації, яка вводиться з терміналу під час обробки одного повідомлення.

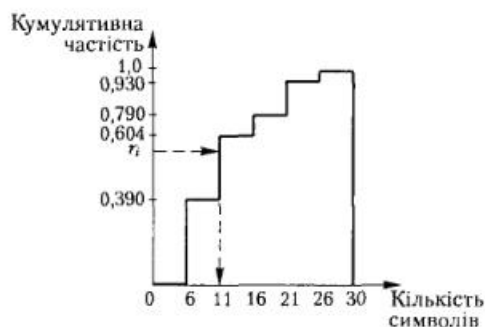
**Таблиця 4.1.** Результати спостереження за об'ємом введеної з терміналу інформації

Кількість символів	Розподіл (частка повідомлень зазначеної довжини)	Кумулятивний розподіл (частка повідомлень зазначеної або меншої довжини)
Менше 6	Відсутній	Відсутній
6–10	0,390	0,390
11–15	0,214	0,604
16–20	0,186	0,790
21–25	0,140	0,930
26–30	0,070	1,000
Більше 30	Відсутній	1,000

На рис. 4.7 і 4.8 зображено відповідно гістограму та кумулятивну функцію розподілу наведених у табл. 4.1 даних.



**Рис. 4.7.** Гістограма розподілу довжини повідомлень



**Рис. 4.8.** Кумулятивна функція розподілу довжини повідомлень

Слід звернути увагу, що висота кумулятивної функції за заданих значень кількості символів дорівнює сумі значень, наведених на рис. 4.7. Для того щоб під час імітаційного моделювання роботи терміналу відтворити кількість символів, які вводяться з клавіатури, необхідно згенерувати випадкове число з діапазону від 0 до 1 (значення по вертикальній осі), а потім на горизонтальній осі визначити кількість уведених символів, які відповідають цьому числу. Наприклад, якщо випадкове число дорівнює 0,578, (див. рис. 4.8), то кількість символів, уведених з терміналу, можна прийняти таким, що дорівнює 11. Цей підхід ілюструє метод оберненої функції, згідно з яким спочатку генерується випадкове рівномірно розподілене число  $r_i$ , що задає значення кумулятивної функції розподілу, за яким потім визначається значення аргументу функції  $x_i = F^{-1}(r_i)$ ,  $i = 1, 2, \dots, n$ , де  $F^{-1}$  – обернена до  $F$  функція.

На практиці часто застосовують дискретні випадкові величини, що набувають лише невід'ємних значень  $j = 0, 1, 2, \dots, k, \dots, n$  з ймовірностями  $p_0, p_1, p_2, \dots, p_k, \dots, p_n$ , тобто функція розподілу дискретної величини  $x$  має вигляд

$$F(x) = \sum_{j=0}^k p(j).$$

У цьому випадку обернену функцію можна записати як

$$x_k = j \quad \text{для} \quad F(j-1) < r_k < F(j), \quad (4.3)$$

де згідно з умовою  $F(-1) = 0$ .

#### 4.4.5. Геометричний розподіл

Для моделювання випадкової величини  $X$  з геометричним розподілом необхідно задати таблицю її значень та їх ймовірність (табл. 4.2).

**Таблиця 4.2.** Значення ймовірності геометричного розподілу випадкової величини

Значення $X$	0	1	...	$n$
Ймовірність	$p$	$(1-p)p$		$(1-p)^n p$

Прикладом випадкової величини з таким розподілом може бути загальна кількість випробувань, які потрібно провести до першого успішного випробування, наприклад кількість пострілів, які потрібно виконати до першого влучення в ціль.

Загалом, ймовірність того, що випадкова величина приймає значення  $k$ , визначається за формулою

$$p_k = p(1-p)^k, \quad k = 0, 1, \dots, n.$$

Для моделювання випадкової величини з геометричним розподілом можна скористатися табл. 4.2 або методом оберненої функції (4.3), але за великих  $n$  такі підходи потребують багато комп'ютерного часу. З цієї причини для отримання значення випадкової величини з геометричним розподілом використовують таку формулу [15]:

$$x_i = \left\lfloor \frac{\ln r_i}{\ln(1-p)} \right\rfloor, \quad 0 < p < 1.$$

У вищенаведеному виразі дужки  $\lfloor \ ]$  означають цілу частину виразу. Справді,

$$\begin{aligned} P(X = x) &= P[k \leq \ln r / \ln(1-p) < k+1] = \\ &= P[-k \ln(1-p) \leq -\ln r < -(k+1) \ln(1-p)] = \\ &= P[(1-p)^{k+1} \leq r < (1-p)^k] = (1-p)^k - (1-p)^{k+1} = p(1-p)^k, \end{aligned}$$

так як випадкова величина  $r$  розподілена рівномірно в інтервалі  $[0, 1]$ .

#### 4.4.6. Біноміальний розподіл

Біноміальний розподіл, або розподіл Бернуллі, — це розподіл дискретної випадкової величини, яка приймає два і тільки два значення: 1 — «true», або «істина», та 0 — «false», або «хибність». Цей розподіл показує ймовірність настання деякої події за  $n$  незалежних повторних випробувань, у кожному з яких подія настає з імовірністю  $p$ , тобто ймовірність  $s$  успішних наслідків у  $n$  випробуваннях

$$f(s) = \frac{n!}{s!(n-s)!} p^s (1-p)^{n-s}.$$

Функція розподілу ймовірності має такий вигляд:

$$F(k) = \sum_{s=0}^k \frac{n!}{s!(n-s)!} p^s (1-p)^{n-s}, \quad k = 0, 1, \dots, n.$$

Залежно від значення  $n$  можна вибрати один із двох способів моделювання випадкової величини з біноміальним розподілом. За невеликих  $n$  значення випадкової біноміально розподіленої величини визначається як кількість чисел у послідовності  $\{r_i\}$  з  $n$  чисел, які не перевищують значення  $p$ . Припустимо, що потрібно отримати випадкову величину, яка належить біноміальному розподілу з параметрами  $n = 7$  і  $p = 0,3$ . Для цього спочатку генеруємо послідовність із семи значень  $r_i$ : 0,0234; 0,1234; 0,7459; 0,0341; 0,8451; 0,1905; 0,5302, а потім рахуємо ті з них, які менші ніж  $p$ . У даному випадку в послідовності тільки чотири значення менші, ніж 0,3. Таким чином, значення випадкової величини, розподіленої за біноміальним законом, дорівнює 4.

За великих значень  $n$  і малих  $p$  можна діяти таким чином. Генеруємо рівномірно розподілені випадкові числа  $r_i$  доти, доки не виконається умова

$$r_i \leq \sum_{j=0}^n u_j, \quad (4.4)$$

де  $u_0$  та  $u_{j+1}$  задаються виразами

$$u_0 = (1-p)^n \quad \text{і} \quad u_{j+1} = u_j \frac{n-1}{j+1} \frac{p}{1-p}.$$

Значення випадкової величини з біноміальним розподілом дорівнює кількості випробувань  $n$ , які необхідно провести, доки не буде справджуватись умова (4.4).

### 4.4.7. Розподіл Пуассона

Випадкову величину з розподілом Пуассона можна отримати, якщо припустити, що кількість незалежних випробувань  $n$  у біноміальному розподілі прямує до нескінченності, а ймовірність успішного випробування  $p$  — до нуля, причому добуток  $np$  є незмінним і дорівнює  $\lambda$ . Функція щільності розподілу Пуассона задається виразом

$$f(s) = \frac{\lambda^s}{s!} e^{-\lambda}.$$

Таким чином, розподіл Пуассона є граничним випадком біноміального та описує випадкові події, які мають місце дуже рідко. На практиці згідно з біноміальним законом розподілені кількість дефектів у готовому виробі та кількість аварій на транспорті за деякий тривалий проміжок часу, кількість дзвінків у телефонній мережі за одиницю часу та ін.

Щоб отримати випадкову величину  $s$  з розподілом Пуассона, генеруємо послідовність рівномірно розподілених випадкових чисел  $r_i$  і знаходимо їх добуток, перевіряючи нерівність

$$\prod_{i=1}^n r_i < e^{-\lambda}. \quad (4.5)$$

У разі виконання умови (4.5) число  $n - 1$  і є випадковою величиною, що належить сукупності, розподіленій за законом Пуассона з математичним сподіванням  $\lambda$ . Якщо умові (4.5) відповідає перше із чисел  $r_i$ , то значення випадкової величини  $s$  дорівнює 0.

## 4.5. Моделювання неперервних випадкових величин

Існує кілька методів моделювання значень неперервних випадкових величин з довільним законом розподілу на основі випадкових чисел, рівномірно розподілених у інтервалі  $[0, 1]$ : метод оберненої функції, метод відсіювання, наближені методи тощо.

### 4.5.1. Метод оберненої функції

Розглянемо метод моделювання випадкової величини, яка має функцію щільності ймовірностей  $f(x)$  і монотонно зростаючу функцію розподілу  $F(x)$  (рис. 4.9). Суть методу така. За допомогою генератора випадкових чисел генеруємо значення випадкової величини  $r_i$ , якому відповідає точка на осі ординат. Значення випадкової величини  $x_i$  з функцією розподілу  $F(x)$  можемо одержати з рівняння  $F(x_i) = r_i$ .

Дійсно, якщо на осі ординат відкласти значення  $r_i$  випадкової величини, розподіленої рівномірно в інтервалі  $[0, 1]$ , і на осі абсцис знайти значення  $x_i$  випадкової

величини (рис. 4.9), при якому  $F(x_i) = r_i$ , то випадкова величина  $X = F^{-1}(r)$  буде мати функцію розподілу  $F(x)$ . За визначенням функція розподілу  $F(x)$  випадкової величини  $X$  дорівнює ймовірності  $P(X < x)$ :

$$P(X < x) = P(r < F(X)) = \int_0^{F(x)} f(r) dr = F(x).$$

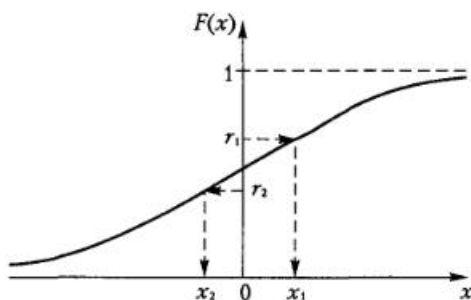


Рис. 4.9. Використання методу оберненої функції для генерування неперервної випадкової величини

Таким чином, послідовність випадкових чисел  $r_1, r_2, r_3, \dots$  перетворюється на послідовність  $x_1, x_2, x_3, \dots$ , яка має задану функцію щільності розподілу  $f(x)$ . Звідси випливає загальний алгоритм моделювання випадкових неперервних величин, що мають задану функцію розподілу ймовірностей:

- ◆ генерується випадкове число  $r_i \in [0, 1]$ ;
- ◆ обчислюється випадкове число  $x_i$ , яке є розв'язком рівняння

$$r_i = \int_{-\infty}^{x_i} f(x) dx.$$

Приклади застосування методу наведені нижче.

#### 4.5.2. Рівномірний розподіл

У загальному випадку випадкова величина  $X$  є рівномірно розподіленою на відрізьку  $[a, b]$ , якщо її щільність розподілу ймовірностей має вигляд

$$f(x) = \begin{cases} 0, & x < a, \\ \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & x > b. \end{cases}$$

Функцію розподілу ймовірностей можна знайти як

$$F(x) = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a},$$

тобто

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

Графіки функцій щільності  $f(x)$  та ймовірності  $F(x)$  зображено на рис. 4.10.

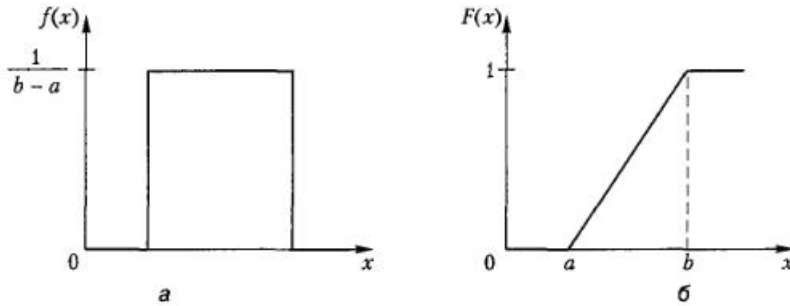


Рис. 4.10. Функції щільності (а) і розподілу (б) рівномірно розподіленої випадкової величини

Математичне сподівання та дисперсія випадкової величини  $X$  визначаються як

$$M(X) = \frac{a+b}{2}, \quad D(X) = \frac{b-a}{12}.$$

Для моделювання випадкової рівномірно розподіленої на відрізку  $[a, b]$  величини можна скористатись методом оберненої функції. Обчислимо функцію розподілу випадкової величини та прирівняємо її до значення  $r_i$ :

$$r_i = \int_a^{x_i} \frac{dx}{b-a} = \frac{x_i - a}{b-a}.$$

Звідси знаходимо значення випадкової величини з функцією розподілу  $f(x)$ :

$$x_i = (b-a)r_i + a.$$

Цю формулу також можна отримати, якщо виконати лінійне перетворення інтервалу  $[0, 1]$  у відрізок  $[a, b]$ . Для цього потрібно змінити масштаб функції рівномірного розподілу, помноживши її на  $(b-a)$ , а потім змістити її на величину  $a$ .

Функція рівномірного розподілу широко застосовується для моделювання випадкових величин, для яких функція розподілу невідома, а відоме лише її середнє значення. У такому випадку припускають, що відомими є середнє значення випадкової величини та деяке розсіювання  $\pm \Delta$  її значень відносно середнього. Це дає змогу стверджувати, що дана випадкова величина має рівномірний розподіл. У мові GPSS такий розподіл часто використовується в блоках ADVANCE для моделювання затримки проходження інформації або під час генерування потоків транзактів у блоках GENERATE. Наприклад, щоб згенерувати потік транзактів, які надходять у модель кожні  $5 \pm 2$  хв, використовується блок GENERATE 5.2.



Прикладами реальних задач, в яких виникає необхідність моделювання рівномірно розподілених випадкових величин, можуть бути аналіз помилок округлення під час проведення числових розрахунків (точність задається як кількість десяткових знаків), час переміщення головок у магнітних накопичувачів (мінімальний та максимальний час), відхилення від розкладу руху транспортних засобів (наприклад, метро).

### 4.5.3. Експоненціальний розподіл

Експоненціальний закон розподілу набув широкого використання в теорії надійності складних систем. Функція щільності експоненціального розподілу випадкової величини має вигляд

$$f(x) = \lambda e^{-\lambda x}.$$

Для її моделювання скористаємося методом оберненої функції. Маємо

$$r_i = \int_0^{x_i} f(x) dx = \int_0^{x_i} \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x_i}. \quad (4.6)$$

З виразу (4.6) знаходимо значення  $x_i$ :

$$x_i = -\frac{1}{\lambda} \ln(1 - r_i).$$

Можна показати, що випадкові величини  $(1 - r_i)$  мають такий самий розподіл, що і величини  $r_i$ . Тоді, замінивши  $1 - r_i$  на  $r_i$ , отримаємо

$$x_i = -\frac{1}{\lambda} \ln r_i.$$

Випадкові величини з експоненціальним розподілом широко застосовуються в задачах моделювання та аналізу СМО, наприклад під час моделювання процесів виходу з ладу та ремонту обладнання, які виникають у складних системах, у разі визначення інтервалів часу між послідовними викликами абонентів у телефонній мережі або замовлень від незалежних клієнтів у будь-якій мережі обслуговування (швидка допомога, служби ремонту, виклик таксі і т. ін.)

Покажемо ще один підхід до моделювання випадкової величини, розподіленої за експоненціальним законом, з використанням методу оберненої функції, який прийнятий у мові GPSS [68]. Цей підхід передбачає заміну функції розподілу ймовірностей кусково-лінійною апроксимуючою функцією.

Розглянемо найпростіший випадок, коли  $\lambda = 1$ . Апроксимуємо функцію експоненціального розподілу лінійними відрізками таким чином, щоб кожний відрізок можна було використовувати для моделювання за допомогою методу оберненої функції. У мові GPSS функція розподілу апроксимована 23 відрізками. Точки апроксимації  $x_i$  та значення функції  $F(x_i)$  у цих точках наведені в табл. 4.3.

Таблиця 4.3. Вузли та значення апроксимованої функції

$x_i$	$F(x_i)$	$x_i$	$F(x_i)$	$x_i$	$F(x_i)$	$x_i$	$F(x_i)$	$x_i$	$F(x_i)$	$x_i$	$F(x_i)$
0	0	0,4	0,509	0,75	1,38	2,3	0,9	3,2	0,96	5,3	0,995
0,1	0,1	0,5	0,69	0,8	1,6	2,52	0,92	3,5	0,97	6,2	0,998
0,2	0,222	0,6	0,915	0,84	1,83	2,81	0,94	3,9	0,98	7	0,999
0,3	0,355	0,7	1,2	0,88	2,12	2,99	0,95	4,6	0,99	8	0,9998

На рис 4.11 зображено лінійну апроксимацію експоненціальної функції розподілу  $F(x)$  з параметром  $\lambda = 1$ , а на рис. 4.12 – функцію, обернену до неї. Перша функція відображає задані в табл. 4.2 значення, а друга використовується під час моделювання випадкових величин з експоненціальним розподілом.

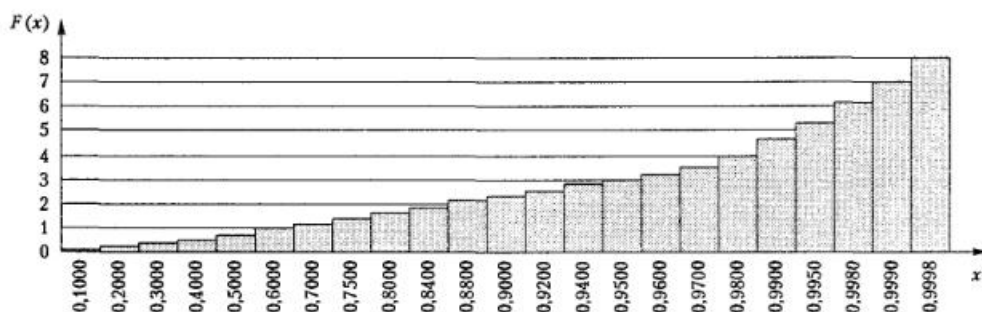


Рис. 4.11. Апроксимація експоненціальної функції розподілу

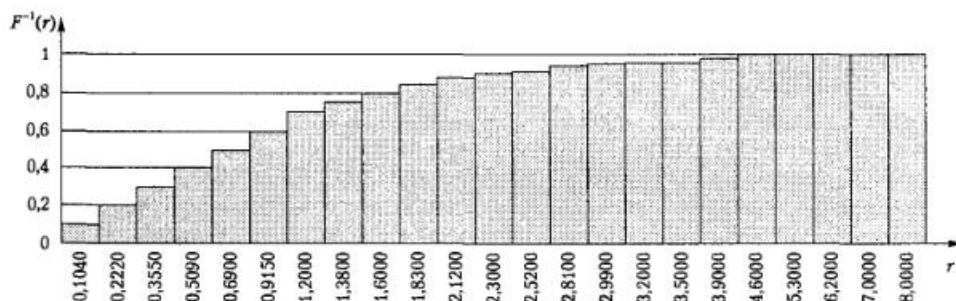


Рис. 4.12. Обернена функція до апроксимованої

За необхідності моделювання випадкової величини  $X$ , розподіленої за експоненціальним законом з математичним сподіванням  $\lambda_x \neq 1$ , діють таким чином:

- ◆ генерують значення випадкової величини  $r_b$ , яке використовують як аргумент оберненої функції (рис. 4.12) експоненціального розподілу з параметром  $\lambda = -1$ , і знаходять значення функції  $F(r)$  (для підвищення точності оцінювання параметрів моделювання функцію розподілу  $F(r)$  інтерполюють лінійними відрізками);
- ◆ знаходять добуток отриманого значення  $F(r)$  на  $1/\lambda_x$ .

Наприклад, для моделювання часу затримки транзактів, що має експоненціальний закон розподілу з параметром  $\lambda_x = 0,01$ , у мові GPSS використовують блок ADVANCE 100, FN\$XPDIS

У цьому операторі функція XPDIS (див. табл. 4.3) задає експоненціальний розподіл з інтенсивністю  $\lambda = 1$ :

```
XPDIS FUNCTION RN1,C24; exponential distribution function
0.0/.1..104/.2..222/.3..355/.4..509/.5..69/
.6..915/.7..1.2/.75.1.38/.8.1.6/.84.1.83/
.88.2.12/.9.2.3/.92.2.52/.94.2.81/.95.2.99/
.96.3.2.97.3.5/.98.3.9/.99.4.6/.995.5.3/
.998.6.2/.999.7/.9998.8
```

#### 4.5.4. Пуассонівський потік

Розглянемо моделювання пуассонівського потоку з інтенсивністю  $\lambda$ , основна властивість якого полягає в тому, що ймовірність надходження  $k$  вимог протягом інтервалу довжиною  $t$  становить

$$p_k(t) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}, \quad k = 1, 2, \dots$$

Для пуассонівського потоку інтервали часу між надходженням двох сусідніх вимог мають експоненціальний закон розподілу (див. розділ 2.1). Тому для його моделювання достатньо отримати ряд чисел з таким розподілом. Це можна реалізувати за допомогою методу оберненої функції, якщо ряд випадкових чисел  $r_j$ , рівномірно розподілених у інтервалі  $[0, 1]$ , перетворити згідно з функцією, оберненою до експоненціальної функції розподілу

$$t_j = F^{-1}(x) = -\bar{T} \ln(r_j),$$

де  $t_j - j$ -й проміжок часу між надходженнями двох сусідніх вимог;  $\bar{T} = 1/\lambda$  – середнє значення проміжку часу між надходженнями двох сусідніх вимог;  $r_j - j$ -е число в послідовності випадкових чисел з рівномірним розподілом у інтервалі  $[0, 1]$ .

У мові GPSS для моделювання пуассонівського потоку вимог з  $\bar{T} = 2$  год (одиниця часу в моделі дорівнює 1 хв) використовується блок GENERATE 120, FN\$XPDIS.

#### 4.5.5. Нормальний розподіл

Випадкова величина  $X$  має нормальний розподіл (розподіл Гаусса), якщо її щільність розподілу ймовірностей описується виразом

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

де  $m$  – математичне сподівання, а  $\sigma$  – середньоквадратичне відхилення.

Функція розподілу нормально розподіленої величини  $X$  має вигляд

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Графіки функцій щільності ймовірностей  $f(x)$  і розподілу  $F(x)$  зображено на рис. 4.13.

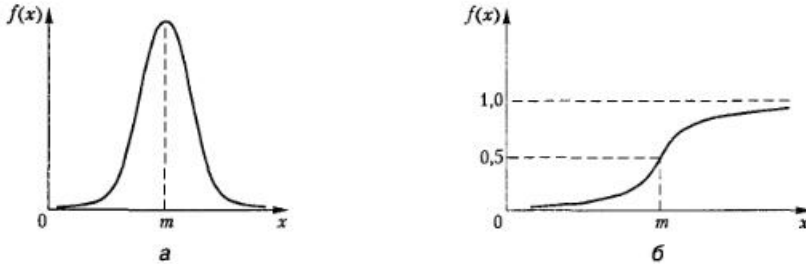


Рис. 4.13. Графіки функції щільності ймовірностей (а) та розподілу (б) випадкової величини

Для моделювання випадкової величини з нормальним законом розподілу безпосередньо скористатися методом оберненої функції не можна, оскільки неможливо аналітично виконати перетворення виду  $X = F^{-1}(r)$ . Тому для моделювання слід скористатися методом згорток.

Метод згорток базується на центральній граничній теоремі – одному із найбільш видатних результатів теорії ймовірностей: за широких припущень відносно розподілів суми великої кількості взаємно незалежних малих випадкових величин має місце розподіл, який близький до нормального. Метод згорток передбачає зображення випадкової величини як суми незалежних однаково розподілених випадкових величин зі скінченними математичним сподіванням і дисперсією.

Центральна гранична теорема формулюється таким чином.

Якщо  $X_1, \dots, X_n$  – послідовність незалежних випадкових величин із скінченним математичним сподіванням  $M[X_i] = a, i = \overline{1, n}$  і дисперсією  $D[X_i] = \sigma^2, i = \overline{1, n}$ , то у разі необмеженого збільшення значення  $n$  функція розподілу випадкової величини

$$\bar{X}^*(n) = \frac{1/n(X_1 + \dots + X_n) - a}{\sigma/\sqrt{n}} = \frac{(\bar{X}(n) - a)\sqrt{n}}{\sigma}$$

наближається до функції розподілу стандартного нормального закону  $\Phi(z)$  при всіх значеннях аргументу, тобто

$$F_{\bar{X}^*(n)} \xrightarrow{n \rightarrow \infty} \Phi(z),$$

де

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-u^2/2} du, \quad u = \frac{x-m}{\sigma}.$$

Функція  $\Phi(z)$  називається функцією Лапласа, для якої є детальні таблиці.

Найпростіший метод отримання значення випадкової величини, що має заданий нормальний розподіл, передбачає виконання таких кроків. Спочатку формують послідовність  $r_i$  ( $i = \overline{1, n}$ ) незалежних, рівномірно розподілених у інтервалі  $[0, 1]$

величин і обчислюють суму  $-Z = \sum_{i=1}^{12} r_i - 6$ . Величина  $n = 12$  є хорошим наближенням до нормально розподіленої випадкової величини з нульовим математичним сподіванням  $m_z = 0$  і одиничним середньоквадратичним відхиленням  $\sigma_z = 1$ . Нормальний розподіл з параметрами  $m_z = 0$  та  $\sigma_z = 1$  називається стандартним.

Перейти від випадкової величини  $Z$  з нульовим математичним сподіванням і одиничним середньоквадратичним відхиленням до випадкової величини  $X$ , яка має математичне сподівання  $m_x$  і середньоквадратичне відхилення  $\sigma_x$ , дає змогу виконати лінійне перетворення

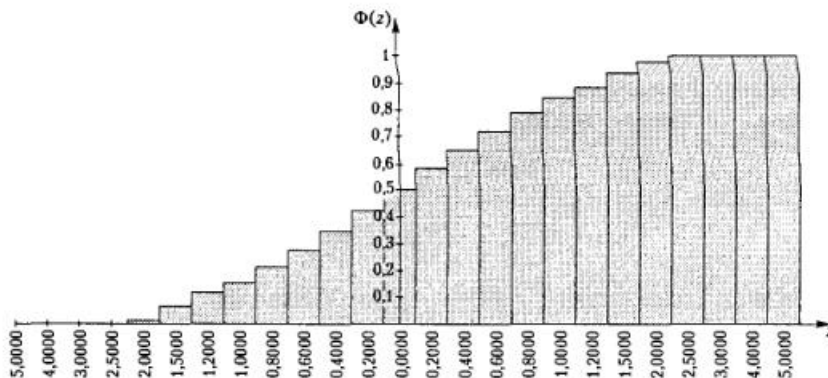
$$X = \sigma_x Z + m_x. \quad (4.7)$$

У системі моделювання GPSS [68] для моделювання випадкової величини з нормальним розподілом прийнято підхід, який базується на методі оберненої функції. За такого підходу функція нормального розподілу випадкової величини  $Z$  з параметрами  $m_z = 0$  і  $\sigma_z = 1$  наближається кусково-лінійною функцією. Як відзначили розробники інтерпретатора GPSS/PC [68], для цього достатньо 24 відрізки. У табл. 4.4 занесено відповідні значення аргументу  $z_i$  і функції  $\Phi(z)$ .

**Таблиця 4.4.** Вузли апроксимації і значення функції нормального розподілу

$z_i$	$\Phi(z)$	$z_i$	$\Phi(z)$	$z_i$	$\Phi(z)$	$z_i$	$\Phi(z)$	$z_i$	$\Phi(z)$
-5	0	-1,5	0,06681	-0,4	0,34458	0,6	0,72575	2	0,97725
-4	0,00003	-1,2	0,11507	-0,2	0,42074	0,8	0,78814	2,5	0,99379
-3	0,00135	-1	0,15866	0	0,5	1	0,84134	3	0,99865
-2,5	0,00621	-0,8	0,21186	0,2	0,57964	1,2	0,88493	4	0,99997
-2	0,02275	-0,6	0,27425	0,4	0,65542	1,5	0,93319	5	1

Для того щоб одержати нормально розподілену випадкову величину з математичним сподіванням  $m_x \neq 0$  і середньоквадратичним відхиленням  $\sigma_x \neq 1$ , необхідно виконати обчислення за формулою (4.7). На рис. 4.14 зображено графік функції, отриманої в результаті апроксимації функції нормального розподілу  $\Phi(z)$ , а на рис. 4.15 – графік функції  $\Phi^{-1}(r)$ , якою зручніше користуватися під час моделювання (як аргумент використовують значення  $r_i$  від генератора випадкових чисел і отримують значення функції).



**Рис. 4.14.** Кусково-лінійна апроксимація функції нормального розподілу

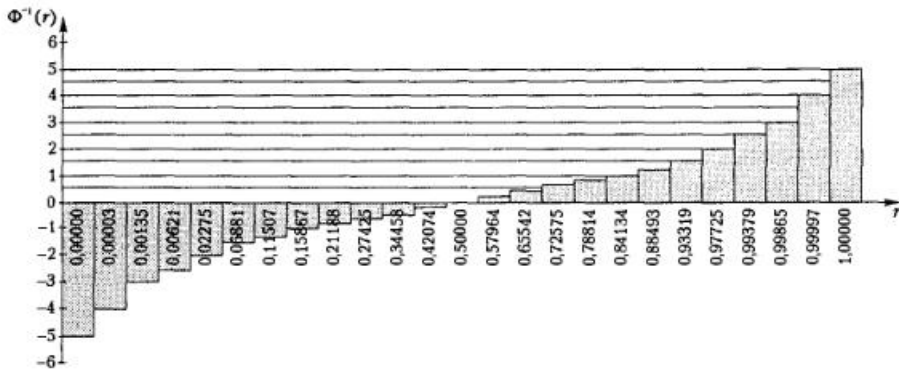


Рис. 4.15. Обернена функція до апроксимованої

У системі GPSS для моделювання нормально розподілених випадкових величин використовується функція NOR, яка апроксимує функцію стандартного нормального розподілу:

```
NOR FUNCTION RN1,C25
0,-5/.00003,-4/.00135,-3/.00621,-2.5/.02275,-2
.06681,-1.5/.11507,-1.2/.15866,-1/.21186,-.8/.27425,-.6
.34458,-.4/.42074,-.2/.5,0/.57926,.2/.65542,.4
.72575,.6/.78814,.8/.84134,1/.88493,1.2/.93319,1.5
.97725,2/.99379,2.5/.99865,3/.99997,4/1.5
```

Наприклад, щоб отримати значення випадкової величини з параметрами  $m_x = 60$  і  $\sigma_x = 10$ , потрібно задати змінну NORM1, яка викликає функцію NOR:

```
NORM1 FVARIABLE 60+10#FN$NOR
```

Якщо в моделі потрібно здійснити затримку транзактів на проміжок часу, що має нормальний розподіл (необхідно забезпечити невід'ємність значень, тобто потрібно, щоб виконувалась умова  $m_x \geq 5\sigma_x$ ), використовують блок ADVANCE V\$NORM1.

Недоліком розглянутих вище методів моделювання є те, що значення функції нормального розподілу, які лежать за межами  $m_x \pm \sigma_x$  суттєво відрізняються від точних значень. Щоб зменшити загальну похибку моделювання, треба використовувати більш точні методи отримання значень функції нормального розподілу. Ці методи базуються на такій властивості. Якщо  $X_1$  і  $X_2$  є незалежними нормально розподіленими випадковими величинами з нульовим математичним сподіванням і одиничним середньоквадратичним відхиленням, то величина кута між віссю абсцис і вершиною випадкового вектора з координатами  $(x_1, x_2)$  має рівномірний розподіл і не залежить від довжини вектора  $(\sqrt{x_1^2 + x_2^2})$  (рис. 4.16).

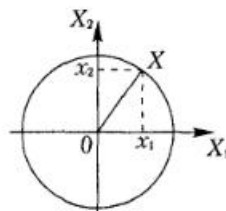


Рис. 4.16. Зображення вектора для моделювання нормального розподілу



Квадрат довжини вектора в цьому випадку має розподіл  $\chi^2$  з двома ступенями свободи і моделюється як окремий випадок показового розподілу з параметром  $\lambda = 1/2$ .

Існує два методи моделювання нормального розподілу, які використовують цю властивість:

**Метод Бокса–Мюллера (Box–Muller).** Генеруємо пару нормально розподілених чисел з  $m_x = 0$  і  $\sigma_x = 1$  за допомогою двох випадкових чисел  $r_1$  і  $r_2$ :

$$x_1 = -2 \ln r_1 \cos(2\pi r_2); \quad x_2 = -2 \ln r_2 \cos(2\pi r_1).$$

Таким чином, отримуємо два числа  $x_1$  і  $x_2$  з нормальним розподілом.

**Метод Марсальї–Брея (Marsaglia–Bray).** Існує більш швидка модифікація цього методу. Генерують два випадкових числа  $r_1$  і  $r_2$ , вважаючи, що  $v_1 = -1 + 2r_1$ ,  $v_2 = -1 + 2r_2$ , обчислюють суму  $s = v_1 + v_2$ . Якщо  $s \geq 1$ , то повторюють процедуру, якщо  $s < 1$ , то одержують два нормально розподілених числа:

$$x_1 = v_1 \sqrt{\frac{-2 \ln s}{s}}, \quad x_2 = v_2 \sqrt{\frac{-2 \ln s}{s}}.$$

Щоб одержати за цим методом 100 пар нормально розподілених чисел, потрібно генерувати 127 пар випадкових чисел. Це простий та швидкий метод, у разі його застосування більша частина часу роботи алгоритму витрачається на обчислення логарифму.

#### 4.5.6. Логарифмічно-нормальний розподіл

Логарифмічно-нормальний розподіл – це такий розподіл випадкової величини, в якій нормальний розподіл має натуральний логарифм її значень. Цей розподіл придатний для моделювання мультиплікативних процесів так само, як нормальний розподіл – для адитивних. Дійсно, використовуючи центральну граничну теорему, можна показати, що добуток незалежних додатних випадкових величин прямує до логарифмічно-нормального розподілу.

Логарифмічно-нормальна величина є результатом взаємодії великої кількості незалежних малих випадкових факторів. Внесок кожного фактора пропорційний уже досягнутому рівню досліджуваної величини, тобто характер впливу є мультиплікативним. Функція щільності логарифмічно-нормального розподілу має вигляд

$$f_{\eta}(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(\ln x - \ln a)^2}{2\sigma^2}}.$$

Щоб перевірити, чи мають емпіричні вибіркові дані логарифмічно-нормальний розподіл, потрібно обчислити логарифм від кожного елементу вибірки і перевірити її на нормальність, наприклад за допомогою критерію  $\chi^2$ . Якщо трансформований набір даних має нормальний розподіл, то вхідні дані є логарифмічно-нормально розподіленими.

На відміну від нормального розподілу, значення випадкової величини  $x$  з логарифмічно-нормальним розподілом завжди додатне і використовується під час моделювання економічних, фізичних, біологічних систем багатьох типів. Вони добре описують процеси, в яких значення змінної, що спостерігається, є випадковою часткою значення попереднього спостереження.

Випадковими величинами з цим розподілом можуть бути тривалість безвідмовної роботи виробу в режимі спрацювання та старіння, розмір банківського вкладу, довжини слів певної мови та переданих повідомлень у мережі, розміри файлів, що зберігаються в комп'ютері.

Як і нормальний розподіл, логарифмічно-нормальний також визначається двома параметрами  $a$  і  $\sigma$ . Графіки функції щільності логарифмічно-нормального розподілу для різноманітних значень параметрів  $a$  і  $\sigma$  зображено на рис. 4.17.

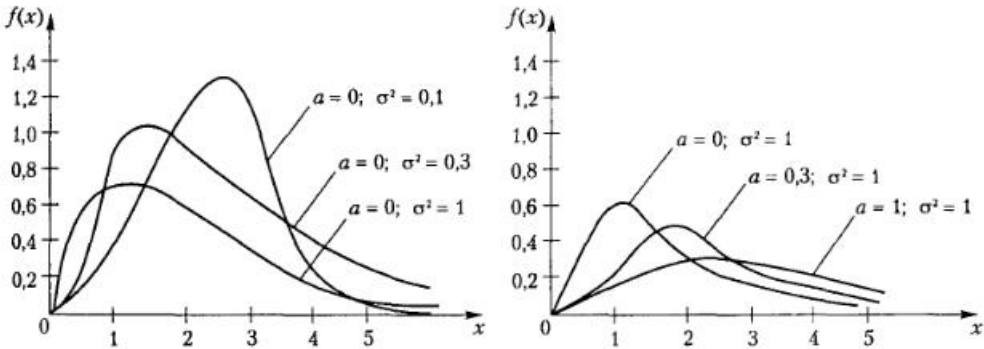


Рис. 4.17. Графіки функції щільності логарифмічно-нормального розподілу

Метод моделювання логарифмічно-нормального розподілу передбачає підставлення в рівняння  $L = e^N$  значень з вибірки  $N(m, \sigma^2)$ , які мають нормальний розподіл з математичним сподіванням  $m$  і дисперсією  $\sigma^2$ , де  $L$  – логарифмічно-нормальний розподіл. Математичне сподівання для цього розподілу  $M[x] = e^{m+\sigma^2/2}$  і дисперсія  $D[x] = e^{2m+\sigma^2}(e^{\sigma^2} - 1)$ .

#### 4.5.7. Розподіл і потоки Ерланга

Випадкові величини з експоненціальним розподілом не завжди адекватно описують деякі реальні процеси та події, наприклад час обслуговування і моменти надходження вимог до СМО. Для більш точного моделювання таких процесів доцільніше використовувати гамма-розподілені випадкові величини або ті, що мають розподіл Ерланга. Розподіл Ерланга є результатом підсумовування взаємно незалежних і однаково розподілених експоненціальних випадкових величин і є окремим випадком гамма-розподілу.

Функція щільності розподілу Ерланга  $k$ -го порядку з інтенсивністю  $\lambda$  має такий вигляд:

$$f_{\eta}(x) = \frac{\lambda(\lambda x)^{k-1}}{(k-1)!} e^{-\lambda x}, \quad x \geq 0.$$

Математичне сподівання і дисперсія розподілу Ерланга визначаються як  $M[x] = 1/k\lambda$ ,  $D[x] = 1/k\lambda^2$ . Графік функції щільності розподілу Ерланга зображено на рис. 4.18.

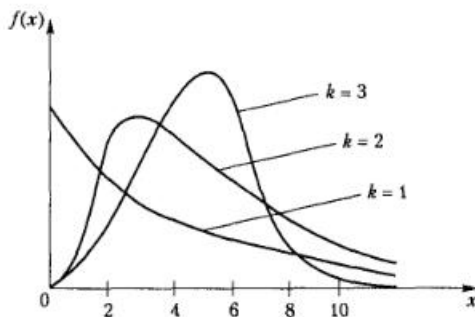


Рис. 4.18. Графік функції щільності розподілу Ерланга

Для моделювання розподілу Ерланга використовують метод згорток випадкових величин з експоненціальними функціями розподілу. Як зазначалось у розділі 2.1, для цього треба лише обчислити суму  $k$  експоненціально розподілених випадкових величин. Зі збільшенням  $k$  розподіл Ерланга наближається до нормального.

Мовою GPSS потік Ерланга другого порядку із середнім значенням часу надходження 180 можна задати таким чином:

```
GENERATE...1
SDFG ADVANCE 90, FN$EXPDIS
ADVANCE 90, FN$EXPDIS
SPLIT 1, SDFG
...
TERMINATE 1
```

Дамо пояснення до цієї програми. У початковий момент часу в моделі генерується один транзакт. Кожний блок ADVANCE імітує затримку транзакту на певний проміжок часу, який має експоненціальний закон розподілу. Блок SPLIT створює копію транзакту і направляє його до блока з міткою SDFG, транзакт-оригінал надходить до моделі і так далі.

Для дослідження властивостей розподілу Ерланга в системі GPSS можна скористатись таким кодом:

```
EXPDIS FUNCTION RN1,C24
0,0/.100,.104/.200,.222/.300,.355/.400,.509
.500,.690/.600,.915/.700,1.200/.750,1.380
.800,1.600/.840,1.830/.880,2.120/.900,2.300
.920,2.520/.940,2.810/.950,2.990/.960,3.200
.970,3.500/.980,3.900/.990,4.600/.995,5.300
.998,6.200/.999,7/1,8
TPTABLE X2,1,20,50 ;визначає таблицю
GENERATE ...1
SDFG ADVANCE 100, FN$EXPDIS
ADVANCE 100, FN$EXPDIS
ADVANCE 100, FN$EXPDIS
SPLIT 1, SDFG ;створює одну копію транзакту
SAVEVALUE 2, C1 ;запам'ятовує час появи транзакту
```

```

SAVEVALUE 2, X1 : визначає інтервал часу між транзактами
SAVEVALUE 1, C1 : запам'ятовує час появи попереднього
                : транзакту
TABULATE TP : буде гістограму
TERMINATE 1

```

Оператор TABLE, блоки SPLIT, SAVEVALUE і TABULATE необхідні для збирання статистичних даних про інтервали надходження транзактів до моделі [63].

Результати моделювання в разі використання оператора START 100000000 наведено на рис. 4.19. Пропонується дослідити властивості розподілу Ерланга при різних значеннях  $k$  шляхом зміни кількості блоків ADVANCE у наведеній програмі.

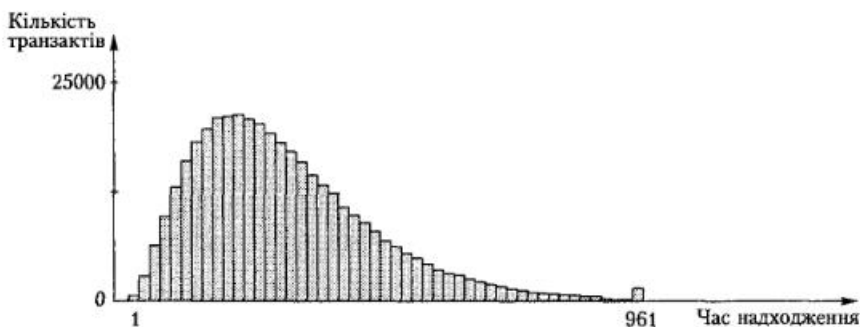


Рис. 4.19. Гістограма, отримана за результатами моделювання

#### 4.5.8. Гамма-розподіл

Випадкова величина має гамма-розподіл з параметрами  $\alpha$  та  $\beta$ , якщо її функція щільності має вигляд

$$f_{\gamma}(x) = \begin{cases} \frac{\beta^{-\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)}, & 0 \leq x < \infty; \\ 0, & x < 0, \end{cases} \quad (4.8)$$

де  $\alpha$  – параметр форми розподілу,  $\beta$  – масштабний коефіцієнт,  $\Gamma(\alpha)$  – гамма-функція або функція Ейлера, яка визначається як [18]

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx, \quad z > 0.$$

Нагадаємо деякі властивості гамма-функції:  $\Gamma(z+1) = z\Gamma(z)$  для  $z > 0$ ;  $\Gamma(k+1) = k!$  для від'ємних цілих  $k$ ;  $\Gamma(k+1/2) = \sqrt{\pi} \cdot 1 \cdot 3 \cdot 5 \dots (2k-1)/2^k$  для додатних цілих  $k$ ;  $\Gamma(1/2) = \sqrt{\pi}$ .

Математичне сподівання гамма-розподілу і дисперсія визначаються як  $M_{\gamma}[x] = \alpha\beta$  та  $D_{\gamma}[x] = \alpha\beta^2$ . Вигляд функції гамма-розподілу значною мірою залежить від значень параметрів. Ця властивість дає змогу використовувати випадкові величини з цим розподілом для моделювання різноманітних фізичних та економічних

явищ і процесів. Графіки функції розподілу при різних значеннях параметрів  $\alpha$  і  $\beta$  зображено на рис. 4.20.

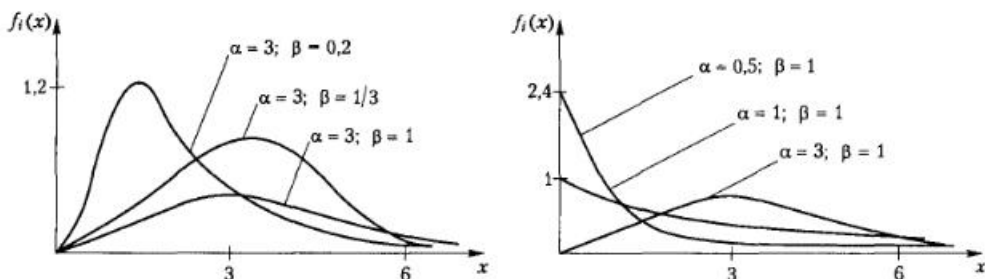


Рис. 4.20. Графіки функції щільності для гамма-розподілу

Гамма-розподіл є узагальненням розподілу Ерланга, коли кількість підсумованих експоненціальних величин не є цілим числом. Якщо  $\alpha = 0,5$  і  $\beta = 2$ , то випадкові величини з гамма-розподілом можна інтерпретувати також як суму квадратів нормально розподілених випадкових величин, тобто таких, які мають розподіл  $\chi^2$ . Таким чином, розподіл  $\chi^2$ , розподіл Ерланга та експоненціальний розподіл є окремими випадками гамма-розподілу.

Гамма-розподіл має важливу властивість. Сума будь-якої кількості незалежних гамма-розподілених випадкових величин  $m$  з однаковим значенням параметра  $\beta$  теж підпорядковується гамма-розподілу, але з параметрами  $(\alpha_1 + \alpha_2 + \dots + \alpha_m)$  і  $\beta$ .

Метод моделювання випадкової величини з гамма-розподілом залежить від значень параметрів  $\alpha$  та  $\beta$ . Якщо  $\alpha = 1$  і  $\beta = 1$ , гамма-розподіл перетворюється в експоненціальний розподіл, і тому для отримання випадкової величини з гамма-розподілом можна використати відповідні методи моделювання. Якщо  $\alpha$  має ціле значення можна перейти до моделювання розподілу Ерланга. Якщо  $\alpha = 0,5$ , гамма-розподіл перетворюється на розподіл  $\chi^2$ , тому для його моделювання досить підносити до квадрату вибіркві значення нормально розподілених випадкових величин.

Від випадкової величини  $X$ , яка має гамма-розподіл з будь-яким значенням параметра  $\alpha$  і значення  $\beta = 1$ , досить легко можна перейти до випадкової величини  $X'$  з параметрами  $\alpha$  і  $\beta > 1$ . Для цього використовується перетворення виду  $X' = \beta X$ . Причому ефективність і швидкодія методу зростає зі збільшенням значення  $\alpha$ .

Існує багато методів моделювання значень гамма-розподіленої випадкової величини [18, 46, 67]. Основна проблема, яка виникає під час її моделювання, – це обчислення гамма-функції. Справа в тому, що цей інтеграл обчислити аналітично неможливо. Тому для його обчислення зазвичай використовують числові методи.

Щоб отримати значення гамма-функції  $\Gamma(z)$  або  $1/\Gamma(z)$ , можна скористатися такою формулою [26, 59]:

$$\frac{1}{\Gamma(z)} \approx z(1+z)(1+a_1z+a_2z^2+\dots+a_{13}z^{13}), \quad z \in [-1, 1],$$

де

$$\begin{aligned} a_1 &= -0,422784335092; & a_2 &= -0,233093736365; & a_3 &= 0,191091101162; \\ a_4 &= -0,024552490887; & a_5 &= -0,017645242118; & a_6 &= 0,008023278113; \end{aligned}$$

$$a_7 = -0,000804341335; \quad a_8 = -0,000360851496; \quad a_9 = 0,000145624324;$$

$$a_{10} = -0,000017527917; \quad a_{11} = -0,000002625721; \quad a_{12} = 0,000001328554;$$

$$a_{13} = -0,00000018122.$$

Обчислення гамма-функції для різних значень  $x$  ускладнюється тим, що вона залежить від трьох аргументів –  $x$ ,  $\alpha$  та  $\beta$ . Тому на практиці під час моделювання у формулі функції щільності (4.8) використовується неповна гамма-функція

$$\Gamma(\alpha) = \int_0^x t^{\alpha-1} e^{-t} dt, \quad (4.9)$$

для обчислення якої за умови, що  $\alpha < 1$  можна скористатися таким виразом [59]:

$$\Gamma(\alpha) = x^\alpha \sum_{i=0}^{\infty} \frac{(-1)^i x^i}{i! (\alpha + i)}.$$

Для  $\alpha > 1$  інтеграл (4.9) можна легко обчислити за допомогою будь-яких формул числового інтегрування, наприклад квадратурних формул Ньютона–Котеса або Гаусса [59].

Отримана функція щільності гамма-розподілу використовується для перетворення випадкових незалежних рівномірно розподілених величин. Для цього область можливих значень випадкової величини  $X$  розбивається на  $n$  однакових інтервалів, кількість яких залежить від заданої точності апроксимації функції  $f(x)$ . Потім за допомогою значення  $r_i$  (методом розіграшу за жеребом) обирається один із  $n$  інтервалів, у якому отримують випадкові числа з функцією щільності розподілу  $f(x)$ .

Для оцінювання близькості функції щільності розподілу ймовірностей отриманих значень випадкової величини до функції щільності розподілу  $f(x)$  використовують метод найменших квадратів. Цей метод передбачає задання максимально допустимої похибки (наприклад,  $\varepsilon = 10^{-4}$ ).

Наведені вище формули та метод кускової апроксимації функції щільності можна використати, щоб задати таблицю значень функції гамма-розподілу при фіксованих значеннях параметрів  $\alpha$  і  $\beta$ , як це зроблено для мови GPSS у генераторі програм ISS 2000 [61].

Наведемо ще кілька алгоритмів моделювання випадкової величини, яка має гамма-розподіл, при різних значеннях параметра  $\alpha$  [46].

1.  $0 < \alpha < 1$ .

Генеруємо три числа:  $r_1$ ,  $r_2$  та  $r_3$ , які є незалежними реалізаціями випадкової величини, рівномірно розподіленої в інтервалі  $[0, 1]$ . Обчислимо значення:  $X = r_1^{1/\alpha}$ ,  $Y = r_2^{1/(1-\alpha)}$ .

Якщо  $X + Y \leq 1$ , обчислюємо  $W = X(X + Y)$  і розраховуємо значення  $\gamma_i$  випадкової величини за формулою

$$\gamma_i = W(-\ln(r_3))^\beta.$$

Або знаходимо нові значення  $X$  і  $Y$  і повторно перевіряємо умову  $X + Y \leq 1$ .



2.  $1 \leq \alpha < 5$ .

Позначимо через  $a = \lfloor \alpha \rfloor$  цілу частину від  $\alpha$  і через  $b = \alpha - \lfloor \alpha \rfloor$ . Обчислимо значення

$$X = \frac{\alpha}{a} \left[ -\ln \left( \prod_{i=1}^a r_i \right) \right].$$

Якщо  $r_{a+1} > \left( \frac{X}{\alpha} \right)^b e^{-\frac{X}{\alpha-1}}$ , то обчислюємо нове значення  $X$ .

Або розраховуємо значення  $\gamma_i$  випадкової величини за формулою

$$\gamma_i = \pm X \beta.$$

3.  $\alpha \geq 5$ .

У цьому випадку провадиться зважений відбір значень послідовності, що має розподіл Ерланга.

Якщо  $r_1 \geq \alpha - \lfloor \alpha \rfloor$ , то  $\gamma_i$  визначається як значення випадкової величини з розподілом Ерланга з параметрами  $\lfloor \alpha \rfloor, \beta$ .

Якщо  $r_1 < \alpha - \lfloor \alpha \rfloor$ , то  $\gamma_i$  обчислюється як значення випадкової величини, що має розподіл Ерланга з параметрами  $\lfloor \alpha \rfloor + 1, \beta$ .

#### 4.5.9. Бета-розподіл

Бета-розподіл визначений у скінченному інтервалі й при різних значеннях параметрів описується різноманітними кривими. Ці криві можуть бути симетричними, асиметричними, мати форму дзвону, U-подібну форму і т. ін. Одна із найпростіших різновидностей бета-розподілу – розподіл Парето, який часто використовується в економічних моделях для моделювання розподілу доходів або витрат.

Те, що бета-розподіл визначений лише в скінченному інтервалі, вносить обмеження на об'єкт моделювання (значення випадкової величини  $X$  лежить в інтервалі від 0 до 1). Прикладами можуть бути функції щільності оцінок імовірності, або частки чогось, експертні суб'єктивні ймовірності події, що нас цікавлять.

Функція щільності бета-розподілу має вигляд

$$f_{\beta}(x) = \begin{cases} \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} x^{\alpha_1-1} (1-x)^{\alpha_2-1}, & 0 \leq x \leq 1; \\ 0, & \text{в інших випадках,} \end{cases}$$

де  $0 < \alpha_1, \alpha_2 < \infty$ .

Математичне сподівання

$$M_{\beta}[x] = \frac{\alpha_1}{\alpha_1 + \alpha_2},$$

а дисперсія

$$D_{\beta}[x] = \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2 (\alpha_1 + \alpha_2 + 1)}.$$

Види функцій розподілу залежно від параметрів  $\alpha_1$  та  $\alpha_2$ , зображено на рис. 4.21.

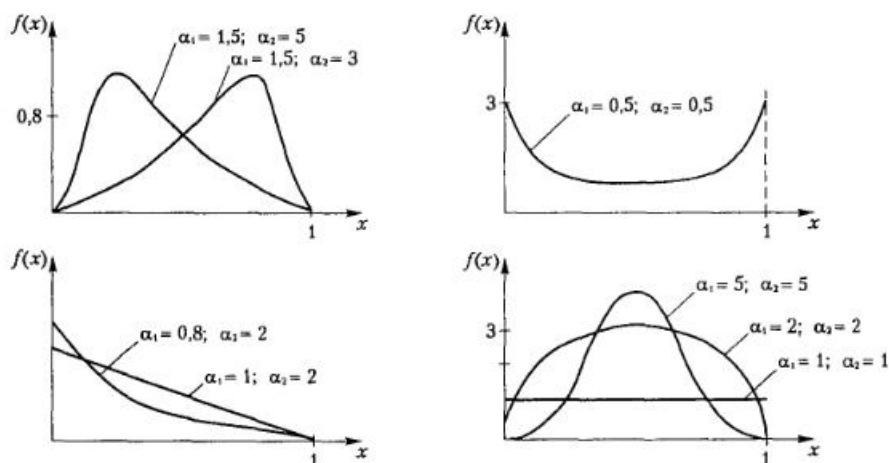


Рис. 4.21. Функції щільності бета-розподілу

Метод моделювання випадкової величини базується на такій властивості бета-розподілу: якщо  $\gamma_1$  і  $\gamma_2$  – дві незалежні гамма-розподілені випадкові величини з параметрами  $\alpha_1, \beta$  та  $\alpha_2, \beta$  відповідно, то значення  $\gamma_1/(\gamma_1 + \gamma_2)$  підпорядковується бета-розподілу з параметрами  $\alpha_1$  і  $\alpha_2$ . Тому метод моделювання передбачає перетворення випадкової величини за формулою

$$x_i = \frac{\gamma_1}{(\gamma_1 + \gamma_2)},$$

де  $\gamma_1$  і  $\gamma_2$  – дві незалежні гамма-розподілені випадкові величини з параметрами  $\alpha_1, 1$  та  $\alpha_2, 1$ .

#### 4.5.10. Розподіл Вейбулла

Розглянемо, якому розподілу підпорядковуються значення випадкової величини, які визначають тривалість безвідмовної роботи складної системи з кількох об'єктів, за умови, що з ладу можуть виходити окремі об'єкти. Через  $\xi$  позначимо тривалість безвідмовної роботи системи, через  $F_\xi(t) = P(\xi < t)$  – неперервну і диференційовану функцію розподілу випадкової величини, а через  $\lambda(t)$  – інтенсивність відмови елементів, що працювали до часу  $t$ .

$$\lambda(t) = -\frac{f_\xi(t)}{1 - F_\xi(t)} \approx \frac{n(t) - n(t + \Delta t)}{\Delta t n(t)}, \quad (4.10)$$

де  $n(t)$  – число об'єктів системи, які працювали безвідмовно до моменту часу  $t$ , а  $\Delta t$  – нескінченно малий відрізок часу. Інтенсивність відмов визначається як відношення кількості об'єктів системи, що вибули з ладу до моменту часу  $t$ , до

загальної кількості об'єктів системи, що працювали безвідмовно,  $n(t)$ . Розв'яжемо рівняння (4.10) відносно функції розподілу  $F_{\xi}(t)$ :

$$F_{\xi}(t) = 1 - e^{-\int_0^t \lambda(t) dt} \quad (4.11)$$

Із формули випливає, що конкретний вигляд функції  $F_{\xi}(t)$  залежить від функції  $\lambda(t)$ . На рис. 4.22 зображено графік життєвого циклу складного виробу. У ньому можна виділити три періоди, кожному з яких відповідають три окремих відрізки графіка. Для кожного відрізка існує своя функція  $\lambda(t)$  і, отже, свій закон розподілу часу безвідмовного функціонування системи  $F_{\xi}(t)$ . Для першого відрізка (період припрацювання) системи параметр  $\alpha < 1$ , для другого (період нормальної експлуатації)  $\alpha = 1$ , для третього (період старіння)  $\alpha > 1$ . На графіку видно, чому в період припрацювання не бажано продавати вироби або експлуатувати в критичних режимах складні системи.

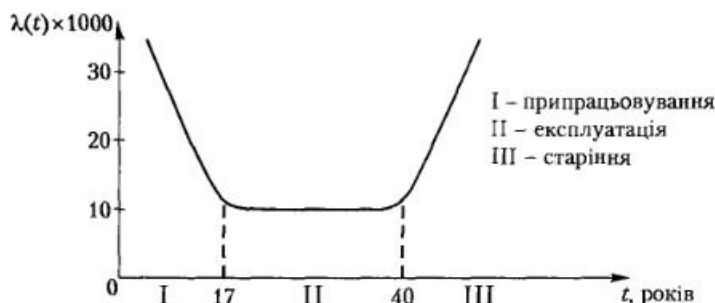


Рис. 4.22. Графік функції зміни інтенсивності відмов у часі для складної системи

Розглянемо випадок, коли функція  $\lambda(t)$  має вигляд

$$\lambda(t) = \lambda_0 \alpha t^{\alpha-1},$$

де  $\lambda_0, \alpha > 0$  – деякі числові параметри, які характеризують систему. Якщо підставити різні значення  $\alpha$  у формули (4.10) та (4.11), отримаємо

$$F_{\xi}(t) = 1 - e^{-\lambda_0 t^{\alpha}}, \quad t \geq 0.$$

Відповідно, функція щільності ймовірності такої випадкової величини має такий вигляд:

$$f_{\xi}(t) = \lambda_0 \alpha t^{\alpha-1} e^{-\lambda_0 t^{\alpha}}.$$

Це і є розподіл Вейбулла. Математичне сподівання і дисперсія задаються виразами

$$M_{\xi}[t] = \lambda_0^{-\frac{1}{\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right) \quad \text{та} \quad D_{\xi}[t] = \lambda_0^{-\frac{2}{\alpha}} \left[ \Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right].$$

Для моделювання випадкових величин  $w_i$ , розподілених по закону Вейбулла, використовуються незалежні рівномірно розподілені в інтервалі  $[0, 1]$  випадкові величини, що перетворюються за методом оберненої функції. Значення  $w_i$  отримують за формулою

$$w_i = \left( \frac{-\ln(r_i)}{\lambda_0} \right)^{1/\alpha},$$

де  $1/\lambda_0$  – масштабний параметр;  $\alpha$  – параметр крутизни.

#### 4.5.11. Гіпер- і гіпоекспоненціальні розподіли

Експоненціально розподілені випадкові величини використовуються і для моделювання випадкових величин з гіпер- і гіпоекспоненціальним розподілом. Ці розподіли дають змогу замінити неекспоненціальний розподіл випадкової величини сумою незалежних зважених експоненціальних розподілів (такий спосіб називається *методом суперпозиції*). Вони широко використовуються в теорії масового обслуговування. Це дає можливість застосовувати методи теорії марківських процесів з неперервним часом для розрахунків характеристик систем, в яких випадкові процеси підпорядковуються неекспоненціальним законам розподілу.

Якщо потрібно отримати випадкову величину з неекспоненціальним розподілом з коефіцієнтом варіації<sup>1</sup>  $C > 1$ , можна скористатись гіперекспоненціальним розподілом, з функцією розподілу ймовірностей

$$F_i(x) = P(X \leq x) = \sum_{i=1}^k \omega_i (1 - e^{-\mu_i x}); \quad \mu_i > 0, \quad \omega_i > 0; \quad \sum_{i=1}^k \omega_i = 1.$$

Математичне сподівання і дисперсія цієї випадкової величини задаються виразами

$$M[x] = \sum_{i=1}^k \frac{\omega_i}{\mu_i}, \quad D[x] = \sum_{i=1}^k \frac{2\omega_i}{\mu_i^2} - \left( \sum_{i=1}^k \frac{\omega_i}{\mu_i} \right)^2.$$

Легко показати, що коефіцієнт варіації визначається як

$$C = \frac{\sqrt{D(t)}}{\bar{x}} \geq 1.$$

Експоненціальний розподіл з коефіцієнтом варіації  $C = 1$  отримуємо за умови, що  $\mu_i = \mu$  для всіх  $i$ .

Розглянемо деякий обслуговуючий центр, обведений контуром на рис. 4.23, який має  $k$  паралельно з'єднаних пристроїв для обслуговування з імовірністю використання  $\omega_i$ . Припустимо також, що в довільний момент часу може бути зайнято не більше одного пристрою з  $k$ , тобто нова вимога надходить до обслуговуючого центру тільки після того, як закінчиться обслуговування попередньої вимоги

<sup>1</sup> Коефіцієнт варіації  $C$  — це відношення стандартного відхилення до математичного сподівання випадкової величини.

і вона залишить центр. Тоді, якщо час обслуговування вимог на кожному пристрої підпорядковується експоненціальному закону розподілу з інтенсивністю  $\mu_i$ , то час обслуговування вимог у центрі в цілому матиме гіперекспоненціальний розподіл. Схема з паралельними етапами обслуговування, наведена на рис. 4.23, використовується для моделювання гіперекспоненціального розподілу. Наприклад, під час моделювання обчислювальних систем такий розподіл добре описує функціонування центрального процесора комп'ютера [64].

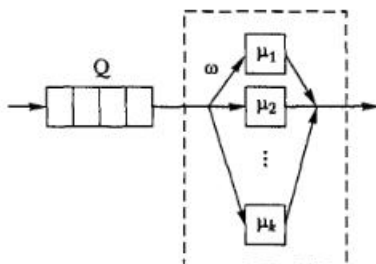


Рис. 4.23. Схема моделі для отримання гіперекспоненціального розподілу

Якщо необхідно отримати розподіл з коефіцієнтом варіації  $C < 1$ , можна скористатись гіпоекспоненціальним розподілом з функцією розподілу ймовірностей

$$F_t(x) = P(X \leq x) = 1 - \frac{\mu_2}{\mu_2 - \mu_1} e^{-\mu_1 x} - \frac{\mu_1}{\mu_1 - \mu_2} e^{-\mu_2 x}, \quad (\mu_1 \neq \mu_2).$$

Математичне сподівання, дисперсія і коефіцієнт варіації випадкової величини  $x$  задаються виразами

$$M[x] = \frac{1}{\mu_1} + \frac{1}{\mu_2}; \quad D[x] = \frac{1}{\mu_1^2 + \mu_2^2}; \quad C = \sqrt{1 - \frac{2\mu_1\mu_2}{(\mu_1 + \mu_2)^2}} < 1.$$

За умови, що всі коефіцієнти однакові ( $\mu_k = \mu$ ), час перебування вимоги в обслуговуючому центрі (обведений на рис. 4.24) буде мати  $k$ -розподіл Ерланга:

$$F_t(x) = 1 - e^{-k\mu x} \sum_{i=0}^{k-1} \frac{(k\mu x)^i}{i!}.$$

Для моделювання пристрою СМО, час обслуговування якого є випадковою величиною з гіпоекспоненціальним розподілом, необхідно послідовно з'єднати  $k$  пристроїв для обслуговування, у кожному з яких час обслуговування має експоненціальний розподіл. Під час моделювання слід урахувати, що в будь-який момент часу повинен бути зайнятим лише один пристрій (рис. 4.24), тобто нова вимога може надійти до першого пристрою для обслуговування тільки після того, як попередня вимога залишить останній пристрій.

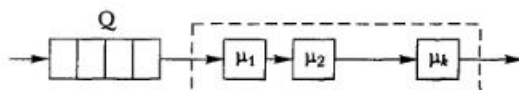


Рис. 4.24. Схема моделі для отримання гіпоекспоненціального розподілу

У разі моделювання обчислювальних систем випадкові величини з таким розподілом застосовують для визначення часу роботи пристроїв введення-виведення комп'ютера.

## 4.6. Моделювання випадкових векторів

Під час моделювання систем керування багатьох типів виникає необхідність генерувати багатовимірні випадкові вектори, які мають заданий сумісний розподіл або багатомірний розподіл. У цьому разі окремі компоненти вектора можуть бути незалежними.

Розглянемо моделювання неперервного випадкового вектора зі складовими  $X, Y$ . Нехай випадкові величини  $X, Y$  описуються спільною функцією щільності  $f(x, y)$ , яку може бути використано для визначення функції щільності  $f(x)$  випадкової величини  $X$ :

$$f_x(x) = \int_{-\infty}^{+\infty} f(x, y) dy.$$

Маючи функцію щільності  $f(x)$ , можна знайти випадкове число  $x_i$ , а потім, якщо  $x = x_i$ , знайти умовний розподіл випадкової величини  $Y$ :

$$f_y(y_i | x = x_i) = f(x, y) | f_x(x_i).$$

З цього виразу для функції щільності можна визначити випадкову величину  $y$ . Тоді пара чисел  $(x_i, y_i)$  буде реалізацією неперервного випадкового вектора  $(X, Y)$ . Такий спосіб реалізації двомірних векторів можна узагальнити і для моделювання багатомірних випадкових векторів. Однак слід мати на увазі: зі збільшенням числа компонентів вектора складність обчислень різко зростає, що перешкоджає широкому використанню цього методу на практиці.

## 4.7. Моделювання випадкових процесів

Випадковий процес – це процес (тобто зміна в часі стану деякої системи чи об'єкта), який розвивається під впливом якихось випадкових чинників і для якого задано ймовірнісні характеристики його протікання. До числа таких процесів можна віднести багато виробничих процесів, які супроводжуються випадковими флуктуаціями, а також процесів, з якими можна зустрітись у природничих науках, економіці, соціології тощо.

Моделювання будь-якого процесу, в тому числі і випадкового, полягає у відтворенні значень (величин) реалізації цього процесу. Випадковий стаціонарний процес задається значеннями математичного сподівання та автоковаріаційною або автокореляційною функцією. Для його моделювання скористаємось параметричними моделями авторегресії, які широко застосовуються для аналізу часових рядів [6].

Авторегресійний процес  $k$ -го порядку з постійними коефіцієнтами визначається рівнянням регресії

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_k y_{t-k} + e_t. \quad (4.12)$$

Значення процесу (4.12) у будь-який момент часу  $t$  визначається через попередні значення та випадкове збурення  $e_t$ . На практиці звичайно використовують авторегресійні моделі процесів першого і другого порядку (процес Маркова і Юла-Уокера), автокореляційна функція яких є згасаючою (рис. 4.25).

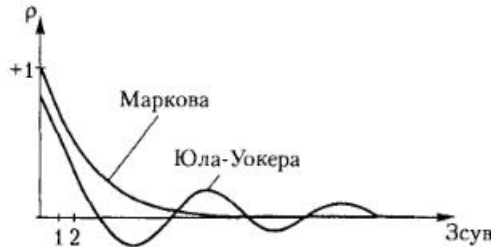


Рис. 4.25. Автокореляційна функція стаціонарного процесу

Параметри процесу  $a_1, \dots, a_k$  визначаються через коефіцієнти автокореляції. Так, для процесу Юла

$$a_0 = M[y_t], \quad a_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}, \quad a_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2},$$

де  $\rho_1, \rho_2$  — значення автокореляційної функції при зсувах 1 та 2.

Під час побудови рівняння авторегресії висуваються дві гіпотези. Перша — про стаціонарність процесу, друга — про те, що збурення  $e_t$  є випадковим процесом у широкому розумінні слова з нормальною функцією розподілу, нульовим математичним сподіванням і дисперсією  $\sigma^2$ . На рис. 4.26, а зображено графік випадкового процесу і нормальний розподіл збурення  $e_t$  (рис. 4.26, б).

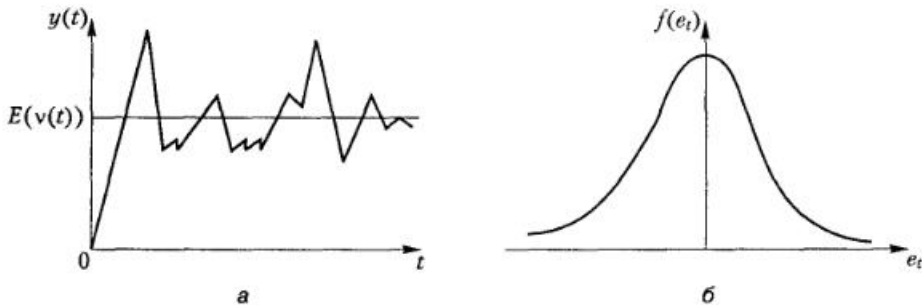


Рис. 4.26. Графіки випадкового процесу (а) і нормального розподілу збурення (б)

Процес  $y_t$  називається марківським, якщо для будь-яких моментів часу  $t_1 < t_2 < t_3 < \dots < t_n$  умовна ймовірність значення  $y_t$  залежатиме від  $y_{t-1}$  і не залежа-



тиме від того, в якому стані процес знаходився в попередні моменти часу. Під час моделювання марківського стаціонарного процесу з параметрами  $M[y_t]$ ,  $\rho_1$ ,  $\sigma^2$  діють таким чином. За початковий член ряду можна взяти будь-яке значення випадкового процесу (тому що будь-яка частина стаціонарного процесу є повноцінним представником усього процесу і має ті ж імовірнісні характеристики), наприклад  $y_{t-1} = 0$  або  $y_{t-1} = M[y_t]$ . За формулою (4.12) розраховуємо значення  $y_{t+1}$  при заданих  $a_0 = M[y_t]$  і  $a_1 = \rho_1$  без урахування збурення  $e_t$  і моделюємо значення нормально розподіленої випадкової величини з математичним сподіванням, що дорівнює нулю, і дисперсією  $\sigma^2$ . Отримане значення додаємо до  $y_{t-1}$  і таким чином отримуємо нове значення реалізації випадкового процесу. Повторюємо процедуру для обчислення інших значень за формулою (4.12), задаючи як початкове значення  $y_t$ , тобто моделюємо  $y_{t+1}$ , ...,  $y_{t+k}$ . Така методика дає змогу моделювати випадкові стаціонарні процеси з будь-якими автокореляційними та багатомірними функціями розподілів.

## 4.8. Статистична обробка результатів моделювання

Основою для обчислення статистичної оцінки параметра системи є реалізація випадкової величини, яка формується під час прогонів імовірнісної імітаційної моделі. Статистична оцінка також є функцією від випадкових величин, які дістають у результаті прогонів моделі, тому і згадана оцінка є випадковою величиною, закон розподілу якої залежить від закону розподілу досліджуваної випадкової величини та оцінюваного параметра. Чим більше реалізацій випадкової величини, тим точнішу статистичну оцінку параметра системи ми отримуємо. У таких умовах обробка результатів моделювання повинна провадитись лише з використанням методів і алгоритмів, які є оптимальними з погляду затрат часу та використання ресурсів комп'ютера. Під час вибору таких засобів необхідно враховувати, що всі статистичні оцінки мають бути ще й *незміщеними, ефективними і спрощеними*.

Розглянемо деякі методи обчислення основних статистичних оцінок.

### 4.8.1. Оцінювання ймовірності

Оцінкою ймовірності  $p$  настання деякої події  $A$  є її частість:

$$\hat{p} = \frac{m}{N},$$

де  $m$  – кількість випробувань, під час за яких випадкова подія спостерігалась;  $N$  – загальна кількість випробувань. Для її використання зазвичай на програмному рівні організовують два лічильники, один з яких призначено для підрахунку загальної кількості випробувань  $N$ , а другий – кількості успішних випробувань  $m$ .

### 4.8.2. Оцінювання розподілу випадкової величини

Для оцінювання функції розподілу випадкової величини, як звичайно, будується гістограма. Під час її побудови область можливих значень випадкової величини розбивають на  $n$  діапазонів і підраховують кількість попадання значень випадкової величини в конкретний інтервал –  $m_k$  ( $k = 1, \dots, n$ ). Оцінка ймовірності попадання випадкової величини в  $k$ -й інтервал має такий вигляд:

$$\hat{p}_k = \frac{m_k}{N}.$$

Цю величину називають відносною частістю. У процесі моделювання під час підрахунку значень  $m_k$  кожному інтервалу ставлять у відповідність окремий лічильник.

### 4.8.3. Оцінювання математичного сподівання

Для оцінювання математичного сподівання випадкової величини використовується формула

$$M(x) = \bar{x} = \frac{1}{N} \sum_{k=1}^n m_k x_k,$$

де  $x_k$  – значення випадкової величини, що належить  $k$ -му інтервалу;  $m_k$  – кількість попадань значень випадкової величини в інтервал;  $N$  – загальна кількість випробувань. У більш простому випадку для оцінювання математичного сподівання випадкової величини можна використати звичайне середнє арифметичне:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

Щоб запобігти непотрібному завантаженню пам'яті, суму доцільніше підраховувати шляхом поступового накопичення.

### 4.8.4. Оцінювання дисперсії

Для оцінювання дисперсії випадкової величини можна використати формулу

$$S^2 = \frac{1}{N-1} \sum_{k=1}^n (x_k - \bar{x})^2 m_k$$

або в простому випадку

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2,$$

де  $S^2$  – оцінка дисперсії випадкової величини  $x$ .

Безпосередньо використовувати ці формули в розрахунках дисперсії нераціонально, тому що зі збільшенням кількості значень  $x_i$  змінюється також середнє

значення, а для його обчислення потрібно запам'ятовувати всі  $N$  значень  $x_i$ . Тому доцільніше використовувати таку формулу:

$$S^2 = \hat{\sigma}^2 = \frac{1}{N-1} \left[ \sum_{i=1}^N x_i^2 - \frac{1}{N} \left( \sum_{i=1}^N x_i \right)^2 \right].$$

У цьому випадку достатньо накопичувати тільки суми двох послідовностей –  $x_i$  і  $x_i^2$ . Але і такий спосіб має недолік – його використання може призвести до переповнення розрядної сітки комп'ютера. Щоб запобігти цьому, потрібно змінити послідовність дій при обчисленнях, використовуючи формулу

$$S^2 = \hat{\sigma}^2 = \sum_{i=1}^N \left( \frac{x_i}{N-1} \right) x_i - \left( \sum_{i=1}^N x_i / N \right) \left( \sum_{i=1}^N x_i / n - 1 \right).$$

#### 4.8.5. Оцінювання кореляційного моменту

Для обчислення оцінки кореляційного моменту можна використовувати формулу

$$\hat{k}_{XY} = \frac{1}{N-1} \left[ \sum_{k=1}^N (x_k - \bar{x})(y_k - \bar{y}) \right]$$

або, більш зручну для обчислень,

$$\hat{k}_{XY} = \frac{1}{N-1} \left[ \sum_{k=1}^N x_k y_k - \frac{1}{N} \sum_{k=1}^N x_k \sum_{k=1}^N y_k \right].$$

При обчисленнях за цією формулою теж доцільно змінити послідовність дій.

### 4.9. Визначення кількості реалізацій під час моделювання випадкових величин

Точність оцінок параметрів системи, які отримують під час обробки результатів моделювання, у першу чергу залежить від кількості випробувань  $N$ . Слід врахувати, що обсяг вибірки  $N$  завжди обмежений, тому вищезгадані оцінки матимуть різні похибки і дисперсії.

Якщо треба оцінити значення деякого параметра  $a$  за результатами моделювання  $x_i$ , то за його оцінку слід брати величину  $\bar{x}$ , яка є функцією від усіх значень  $x_i$ . Статистична оцінка  $\bar{x}$  також є випадковою величиною, тому вона буде відрізнятися від  $a$ , тобто

$$|a - \bar{x}| < \varepsilon,$$

де  $\varepsilon$  – точність або похибка оцінки. Імовірність того, що ця нерівність виконується, позначимо через  $\alpha$ :

$$P(|a - \bar{x}| < \varepsilon) \geq \alpha. \quad (4.13)$$

У теорії ймовірностей  $\varepsilon$  – це *довірчий інтервал* для  $\alpha$ , довжина якого фактично дорівнює  $2\varepsilon$ , а  $\alpha$  – *довірчий рівень*, або *надійність оцінки*. Вираз (4.13) можна застосувати для визначення точності результатів статистичних випробувань.

### 4.9.1. Оцінювання ймовірності

Припустимо, що метою моделювання є оцінювання ймовірності настання деякої події  $A$ , яка визначає стан системи. У кожній з  $N$  реалізацій процесу настання події  $A$  є випадковою величиною  $\xi$ , що набуває значення  $x_1 = 1$  з імовірністю  $p$  і  $x_2 = 0$  з імовірністю  $1 - p$ . Тоді можна визначити математичне сподівання і дисперсію відповідно за формулами

$$M[\xi] = x_1 p + x_2 (1 - p) = p, \quad (4.14)$$

$$D[\xi] = (x_1 - M[\xi])^2 p + (x_2 - M[\xi])^2 (1 - p) = p(1 - p). \quad (4.15)$$

Як оцінку  $p$  використовують частість настання події  $A$ . Ця оцінка є незміщеною, спроможною та ефективною. За умови, що  $N$  задано, для отримання цієї оцінки достатньо накопичувати  $m$ :

$$\frac{m}{N} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (4.16)$$

де  $x_i$  – настання події  $A$  в реалізації  $i$ .

За формулами (4.14)–(4.16) визначимо вибіркоче математичне сподівання  $M[m/N] = p$  і дисперсію  $D[m/N] = p(1 - p)/(N - 1)$ .

Згідно з центральною граничною теоремою (у даному випадку її можна взяти у вигляді теореми Хінчина) випадкова величина  $m/N$  буде мати розподіл, близький до нормального (рис. 4.27). Тому для кожного рівня достовірності  $\alpha$  з таблиць нормального розподілу можна знайти таку величину  $t_\alpha$ , при якій точність обчислюватиметься за формулою

$$\varepsilon = t_\alpha \sqrt{D[m/N]}. \quad (4.17)$$

Якщо  $\alpha = 0,05$ , то  $t_\alpha = 1,96$ , а якщо  $\alpha = 0,003$ , то  $t_\alpha = 3$ .

Підставимо у формулу (4.17) вираз дисперсії:

$$\varepsilon = t_\alpha \sqrt{\frac{p(1-p)}{N-1}}.$$

Звідси

$$N = t_\alpha^2 \frac{p(1-p)}{\varepsilon^2} + 1. \quad (4.18)$$

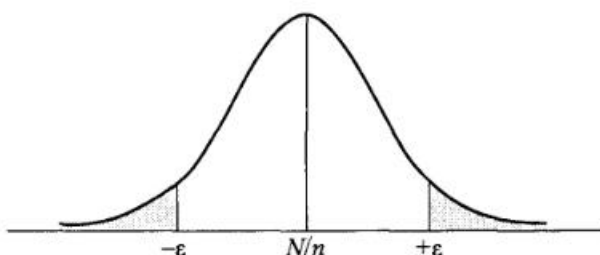


Рис. 4.27. Функція нормального розподілу для побудови довірчого інтервалу

З формули (4.18) видно, що при  $p = 1$  або  $p = 0$ , кількість реалізацій, які необхідно провести для підтвердження того, що подія  $A$  настає (або ні), дорівнює одиниці. Але оскільки ймовірність  $p$  заздалегідь невідома, провадять випробування ( $N = 50 \dots 100$ ), оцінюють частість  $m/N$  і підставляють її значення у вираз (4.18) замість  $p$ , після чого визначають остаточну кількість реалізацій. Графік залежності числа реалізацій для  $\alpha = 0,05$  і різних значень  $p$ , якщо  $\varepsilon = 0,05$ , наведено на рис. 4.29.

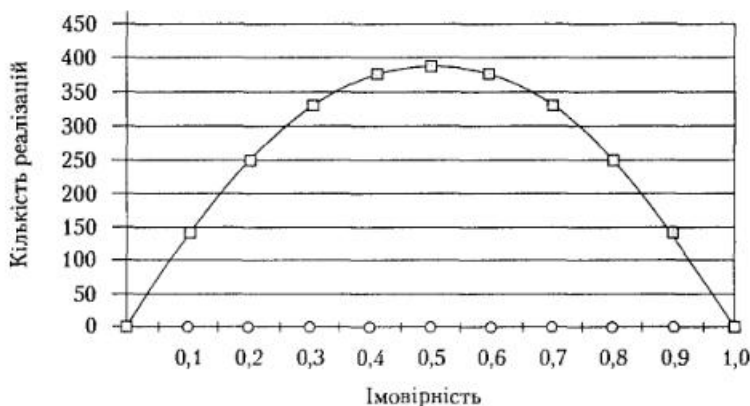


Рис. 4.28. Залежність числа реалізацій від значень ймовірності

#### 4.9.2. Оцінювання середнього значення

Нехай випадкова величина має математичне сподівання  $a$  і дисперсію  $\sigma^2$ . У  $i$ -й реалізації вона набуває значення  $x_i$ . Як оцінку математичного сподівання  $a$  використовуємо середнє арифметичне:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (4.19)$$

Згідно з центральною граничною теоремою при великих значеннях  $N$  середнє арифметичне (4.19) буде мати нормальний розподіл з математичним сподіванням  $a$  і дисперсією  $\sigma^2/(N-1)$ . Тоді

$$\varepsilon = t_{\alpha} \frac{\sigma}{\sqrt{N-1}}.$$

Звідси

$$N = t_{\alpha}^2 \sigma^2 / \varepsilon^2 + 1. \quad (4.20)$$

Оскільки дисперсія  $\sigma^2$  випадкової величини невідома, потрібно провести кілька десятків (50 ... 100) випробувань і знайти оцінку  $\hat{\sigma}^2$ , а потім отримане значення підставити у формулу (4.20), щоб визначити необхідну кількість реалізацій  $N$ . У цьому випадку замість нормально розподіленої величини необхідно скористатись  $t$ -розподілом Стьюдента з  $N - 1$  ступенями вільності для визначення  $t_{\alpha}$ . Зауважимо, що за збільшення ступенів вільності  $t$ -розподіл наближається до нормального. З практичного погляду, якщо  $N$  більше 30, користуються нормальним розподілом.

## Висновки

- ◆ Метод статистичних випробувань визначається як спосіб побудови і дослідження на комп'ютері моделі системи або процесу з використанням послідовностей випадкових чисел.
- ◆ Метод полягає в багатократному проведенні випробувань побудованої моделі й подальшій статистичній обробці результатів моделювання з метою визначення шуканих характеристик розглядуваного процесу у вигляді оцінок його параметрів.
- ◆ У методі статистичних випробувань особливе значення відіграють випадкові числа, рівномірно розподілені в інтервалі  $[0, 1]$ , за допомогою яких можна отримати вибіркові значення з будь-якими статистичними властивостями.
- ◆ Для генерування випадкових чисел використовуються апаратні, табличні та програмні методи.
- ◆ Апаратні методи генерування випадкових чисел базуються на використанні деяких фізичних явищ і процесів – випадковий електричний сигнал перетворюють у двійковий код, який уводиться в комп'ютер за допомогою спеціальних аналого-цифрових перетворювачів.
- ◆ У разі використання табличного методу випадкові числа можна зберігати на зовнішніх носіях або навіть в основній пам'яті комп'ютера.
- ◆ Програмні генератори дають змогу отримувати послідовності випадкових чисел за рекурентними формулами. Більшість програмних генераторів виробляють випадкові числа, рівномірно розподілені в інтервалі  $[0, 1]$ .
- ◆ Серед програмних генераторів найбільш розповсюдженим є лінійний мультиплікативний конгруентний генератор.
- ◆ Основним параметром програмного генератора є повний період – кількість чисел, після якої випадкові числа починають повторюватися.
- ◆ Для перевірки статистичних властивостей усіх послідовностей випадкових чисел, які будуть використовуватись під час проведення досліджень, застосовують емпіричні та теоретичні критерії.

- ◆ Емпіричні критерії – це звичайні статистичні тести, в яких використовують вибіркові значення, що виробляються генератором.
- ◆ Теоретичні критерії визначають характеристики послідовності за допомогою методів, які формуються на основі числових значень параметрів генератора.

## Контрольні запитання та завдання

1. Доведіть, що незалежні випадкові величини  $r_i$  та  $1 - r_i$ , отримані з рівномірного розподілу в інтервалі  $[0, 1]$ , мають однаковий розподіл.
2. Обчисліть імовірність того, що за 50 підкидань монети 10 разів випаде герб. Спочатку розв'яжіть задачу аналітично, а потім порівняйте результат з оцінкою, отриманою за допомогою методу статистичних випробувань.
3. Використовуючи метод статистичних випробувань, обчисліть площу круга. Зважаючи на те, що формула для визначення площі круга –  $S = \pi r^2$ , оцініть значення константи  $\pi$ .
4. Молодий пілот, виконуючи перший свій політ, не зміг здійснити вдалу посадку літака. Визначте число спроб зайти на посадку, які необхідно зробити пілоту, для того щоб політ благополучно закінчився. Ймовірність того, що за одну спробу стажист вдало завершить посадку літака, становить 0,1. Кількість палива в баках вважати необмеженою.
5. Skorиставшись методом статистичних випробувань, знайдіть оцінку інтеграла  $\int_0^1 x^{3/2} dx$  як площі під кривою  $x^{3/2}$  на відрізку  $0 \leq x \leq 1$ .
6. Імовірність одержання заліку студентом, який не відвідував лекційні і практичні заняття, становить 0,13. Skorиставшись методом статистичних випробувань, знайдіть оцінку ймовірності того, що студент одержить залік, якщо загальне число спроб здати залік не може перевищувати 3.
7. Знайдіть оцінку площі рівнобічного трикутника зі стороною 1 см. Визначіть рівняння прямих, які утворюють сторони трикутника і використайте ці вирази в процедурі оцінювання попадання «випадкової» точки в трикутник для методу статистичних випробувань.
8. Для одержання допуску до іспиту студенту необхідно отримати залік з лабораторних робіт (лабораторний курс складається з  $K$  лабораторних робіт). Припустимо, що ймовірність здавання однієї лабораторної роботи студентом становить 0,3 з однієї спроби. Оцініть кількість днів, потрібних студенту на одержання допуску до іспиту, якщо за один день він зможе здати не більше однієї лабораторної роботи (протягом семестру він не захистив жодної роботи).
9. Стрільцю необхідно вразити чотири мішені. Ймовірність ураження однієї мішені становить 0,14. Оцініть кількість набоїв, необхідних для того, щоб уразити всі чотири мішені.



10. Багдадський злодій ув'язнений у підземелля з трьома дверима [46]. Одні двері ведуть на волю, другі – у довгий тунель, а треті – у короткий. Потрапивши в один із тунелів, злодій знову опиняється в темниці. Він знову пробує вийти на волю, але не пам'ятає, в які двері входив минулого разу. Імовірність того, що злодій обере потрібні двері, дорівнює 0,3, ймовірність попадання в короткий тунель – 0,2; ймовірність попадання в довгий тунель – 0,5. Час перебування злодія в короткому тунелі – 3 хв, у довгому – 6 хв. Визначіть середній час пошуку шляху на волю. Побудуйте процедуру статистичних випробувань для розв'язування задачі.
11. Змоделюйте поведінку винищувача-бомбардувальника [67], який атакує об'єкт ракетами класу «повітря-земля». Кожна ракета наводиться індивідуально. Розміри об'єкта – 60×150 м. Заходження на атаку відбуваються в напрямі, який збігається з напрямом довгої осі цілі, точка прицілу – геометричний центр цілі. Фактичну точку влучення для кожної ракети можна визначити горизонтальним і вертикальним відхиленнями (рис. 4.29). Для відстані, з якої запускають ракети, обидва відхилення є незалежними, нормально розподіленими щодо точки прицілювання і мають нульове середнє значення.

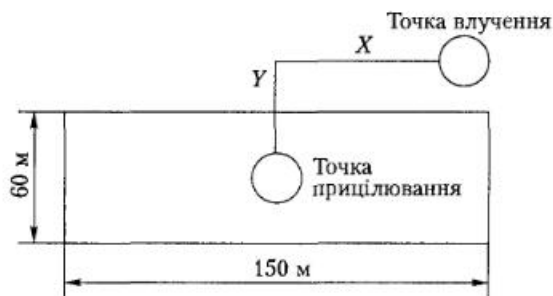


Рис. 4.29. Схема влучення ракети в ціль

Середньоквадратичне відхилення становить 60 м у напрямку  $X$  і 30 м – у напрямку  $Y$ . Бомбардувальник під час кожного заходження випускає шість ракет. Узнявши обсяг вибірки в 10 заходжень, знайдіть оцінку середнього числа влучень для кожної атаки.

12. Змоделюйте випадкову двомірну дискретну величину  $(X, Y)$ . Випадкова величина  $X$  може набувати значення 1, 2, 3, 4, а випадкова величина  $Y$  – 10, 20, 30, 40, у цьому разі кожній парі значень  $x_i, y_j$  відповідає ймовірність  $p_{ij}$  (табл. 4.5).

Таблиця 4.5. Ймовірність двомірної дискретної величини

Значення $x_i$	Значення $y_j$			
	10	20	30	40
1	0,02	0,05	0,1	0,05
2	0,03	0,1	0,05	0,03
3	0,01	0,15	0,1	0,02
4	0,04	0	0,15	0,1

13. Припустимо, що об'єм споживання води в місті є випадковою величиною, що має нормальний розподіл [67]. Знайдіть оцінку середнього споживання води в день так, щоб помилка не перевищувала  $\pm 6000$  л з імовірністю 0,95. Відомо, що розумна область розкиду споживання води становить 120 000 л за день. Який обсяг вибірки необхідний для цього дослідження?
14. Використовуючи результати 20 імітаційних прогонів для оцінювання часу перебування відвідувачів у системі [46], що наведені в дужках (1,1; 2,8; 3,7; 1,9; 4,9; 1,6; 0,4; 3,8; 1,5; 3,4; 1,9; 2,1; 3,8; 1,6; 3,2; 2,9; 3,7; 2,0; 4,2; 3,3), обчисліть оцінки для вибіркового середнього, дисперсії і коефіцієнта варіації. Побудуйте гістограму, яка містить п'ять інтервалів (довжина кожного інтервалу дорівнює одиниці, нижня межа першого інтервалу дорівнює нулю).
15. Дано випадкові некорельовані змінні  $A, B, C$ . Змінна  $A$  має нормальний розподіл з математичним сподіванням 100 і середньоквадратичним відхиленням 20. Змінна  $B$  також розподілена нормально з математичним сподіванням 20 і середньоквадратичним відхиленням 5. Розподіл змінної  $C$  задано в табл. 4.6.

**Таблиця 4.6.** Функція розподілу випадкової величини

Значення $C$	10	20	30	40
Імовірність	0,10	0,25	0,50	0,15

Застосовуючи метод статистичних випробувань, оцініть середнє значення нової змінної  $D$ , що визначається так:  $D = (A + B)/C$ . Використайте вибірку із 10 значень, яку необхідно отримати за допомогою розподілу, що демонструється на рис. 4.15.

16. Проаналізуйте модель наземної протиповітряної ракетної установки. Скільки необхідно виконати випробувань, кожне з яких моделює одну відсіч повітряної атаки, щоб оцінка частоти поразки цілі відрізнялася від істинної ймовірності не більше ніж на 0,02 з імовірністю не меншою за 0,85?