ЛАБОРАТОРНА РОБОТА № 1

МОДЕЛЮВАННЯ ПРОЦЕСУ ВІДНОСНО-ЧАСТОТНОГО ВИЗНАЧЕННЯ ЙМОВІРНОСТІ

Основні поняття та визначення

Випадковим процесом називається випадковий дослід або експеримент, при якому можливі різні результати, які неможливо заздалегідь передбачити. Величина *X*, що є результатом випадкового процесу, називається *випадковою величиною*. Непостійність результату такого досліду може бути пов'язана з наявністю випадкових помилок вимірів або зі статистичною природою самої вимірюваної величини (наприклад, процес розпаду радіоактивної речовини). Випадкові величини зазвичай позначають великими літерами латинського алфавіту *X, Y, Z,...*, а їх можливі значення – малими *x, y, z..*.

Випадкові величини бувають дискретні і неперервні, одновимірні (залежні від однієї змінної) або багатовимірні (залежні від двох і більше змінних).

Дискретною випадковою величиною називається така величина, можливі значення якої рівні одному із значень із скінченої, або нескінченої множини, елементи якої можуть бути пронумеровані.

Неперервною випадковою величиною називається така величина, можливі значення якої безперервно заповнюють деякий інтервал (скінченний або нескінченний) числової осі.

Повною характеристикою випадкової величини *X* з імовірнісної точки зору є її *закон розподілу*, тобто заданий тою чи іншою мірою зв'язок між можливими значеннями випадкової величини та ймовірностями їх появи.

Загальною формою закону розподілу випадкових величин є *функція розподілу ймовірностей* - це така функція $F(x)$, значення якої в точці x рівне ймовірності того, що при проведенні досліду значення випадкової величини X виявиться менше, ніж x:

$$F(x) = P(X < x) \quad (1)$$

Як випливає із визначення, функція розподілу є невід’ємною неспадною функцією, значення якої лежать на відрізку $0 \le F(x) \le 1$. Мають місце граничні рівності $F(-\infty) = 0$ і $F(\infty) = 1$. Крім того функція $F(x)$ для дискретної випадкової величини ступінчата розривна, а для неперервної випадкової величини – неперервна.

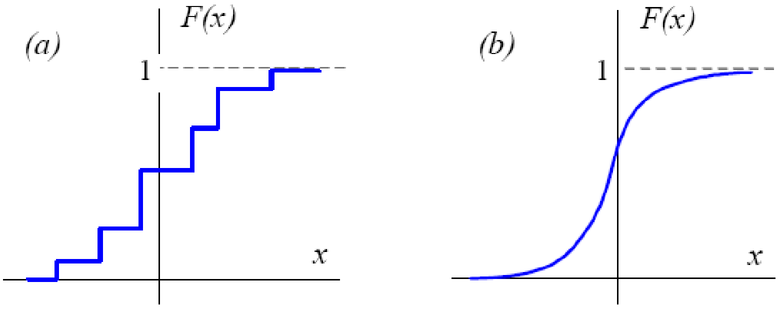


Рис. 1. Графічне зображення функції розподілу ймовірностей:
(а) – для дискретної випадкової величини, (б) – для неперервної випадкової величини.

Похідна від функції розподілу

$$f(x) = \frac{dF(x)}{dx}$$

називається *густиною розподілу ймовірності*, або просто *густиною ймовірності* даної випадкової величини. Очевидно, що:

$$f(x)dx = P(x < X \leq x + dx),$$

тобто величина $f(x)dx$ є ймовірністю потрапляння випадкової величини X в напівінтервал $(x,\ x + dx]$

Для неперервної випадкової величини X густина ймовірності $f(x)$ є неперервною функцією. Для дискретної випадкової величини $X\{x_i\}$, що приймає фіксовані значення $\{x_1, x_2, ..., x_n, ...\}$ з ймовірностями $\{p_1, p_2, ..., p_n, ...\}$, густина ймовірності виражається сумою дельта-функцій:

$$f(x) = \sum_i p_i \cdot \delta(x - x_i)$$

В обох випадках густина ймовірності $f(x) \geq 0,\ \forall x$ і задовольняє умову нормування

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Щоб у стислій формі виразити найбільш суттєві особливості того чи іншого розподілу, використовують особливі числові характеристики випадкової величини, які називають її *моментами n-го порядку*:

$$M\{X^n\} \equiv \overline{x^n} = \int_{-\infty}^{\infty} x^n f(x)dx.$$

Зауважимо, що $\overline{x^n}$ є не випадковою, а певною, детермінованою величиною.

Однією з важливих характеристик є момент 1-го порядку, який називають *математичним сподіванням* або *середнім значенням* випадкової величини:

$$M\{X\} \equiv \bar{x} = \int_{-\infty}^{\infty} xf(x)dx.$$

Момент 2-го порядку

$$M\{X^2\} \equiv \overline{x^2} = \int_{-\infty}^{\infty} x^2 f(x)dx$$

називають *середнім квадратом* випадкової величини.

Для дискретної випадкової величини $X\{x_i\}$, з розподілом $P(X = x_i) = p_i$, $\sum_i p_i = 1$, математичне сподівання і середній квадрат будуть відповідно:

$$M\{X\}\equiv\overline{x}=\sum_i x_i p_i\,,$$

$$M\{X^2\}\equiv\overline{x^2}=\sum_i x_i^2 p_i\,.$$

Іншою важливою характеристикою випадкової величини є *дисперсія*, яка визначається як математичне сподівання квадрата відхилення випадкової величини X від свого середнього значення:

$$D\{X\}\equiv\sigma^2=M\left\{\left(X-M\{X\}\right)^2\right\}.$$

Неважко показати, що має місце рівність:

$$\sigma^2=M\{X^2\}-\left(M\{X\}\right)^2=\overline{x^2}-(\overline{x})^2,$$

тобто дисперсія є різницею між середнім квадратом випадкової величини X і квадратом її середнього значення.

Величина σ, тобто додатній квадратний корінь з дисперсії називається *стандартним* або *середньоквадратичним відхиленням*. Середньоквадратичне відхилення кількісно показує, наскільки сильно значення випадкової величини X розкидані навколо середнього значення $\overline{x}$.

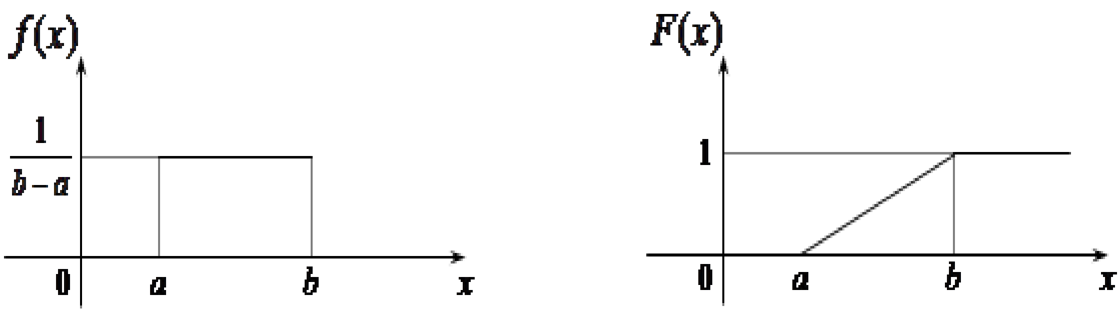
Неперервна випадкова величина X має *рівномірний розподіл* на відрізку $[a,b]$, де $a,b\in\mathbb{R}$, якщо її густина імовірності має вигляд:

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

Інтегральна функція розподілу:

$$F(x)=\int_{-\infty}^{x} f(\xi)d\xi=\begin{cases}0, & x<a\\ \dfrac{x-a}{b-a}, & x\in[a,b]\\ 1, & x>b\end{cases}$$

На рис. 2 схематично показано густину імовірності та інтегральну функцію розподілу для рівномірно розподіленої випадкової величини X.



a) b)

Рис. 2. Графічне зображення рівномірно розподіленої випадкової величини X.

Для рівномірно розподіленої випадкової величини X легко обчислити її числові характеристики

$$\bar{x} = \int_{-\infty}^{\infty} x f(x) dx = \frac{a+b}{2}$$

$$\overline{x^2} = \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{a^2+ab+b^2}{3}$$

$$\sigma^2 = \overline{x^2} - (\bar{x})^2 = \frac{(b-a)^2}{12}$$

Коли неперервна випадкова величина X має рівномірний розподіл на відрізку $[a,b]$, то це записують так: $X \in U[a,b]$. Якщо $a=0,\ b=1$, тобто $X \in U[0,1]$, то такий неперервний рівномірний розподіл називають *стандартним*. Для стандартного рівномірного розподілу числові характеристики:

$$\bar{x} = \frac{1}{2}, \quad \overline{x^2} = \frac{1}{3}, \quad \sigma^2 = \frac{1}{12}$$

Має місце твердження: якщо випадкова величина $X \in U[0,1]$, а випадкова величина $Y = a + (b-a)X$, то $Y \in U[a,b]$. Тобто, маючи генератор випадкових чисел, рівномірно розподілених на інтервалі [0,1], можна легко побудувати генератор випадкових чисел, рівномірно розподілених на заданому інтервалі $[a,b]$.

Дискретна випадкова величина X має *дискретний рівномірний розподіл* на відрізку $[a,b]$, якщо вона на цьому відрізку приймає скінченне число значень $\{x_1, x_2, ..., x_n\}$ з однаковими ймовірностями $p_i = p = \frac{1}{n}$.

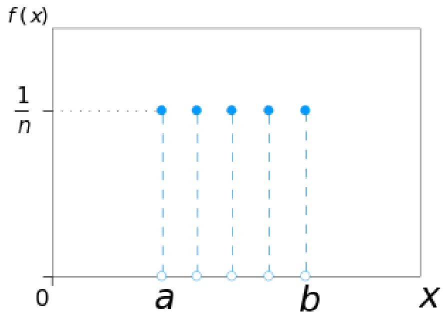


Рис. 3. Густина ймовірності дискретної випадкової величини з рівномірним розподілом

Числові характеристики дискретної випадкової величини з рівномірним розподілом:

$$\bar{x} = \sum_{i=1}^{n} x_i p_i = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\overline{x^2} = \sum_{i=1}^{n} x_i^2 p_i = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 p_i = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 = \overline{x^2} - (\overline{x})^2$$

Простим прикладом дискретного випадкового процесу з рівномірним розподілом є підкидання монетки, або кидання кубика.

Нехай у нас є гральний кубик. Випадіння кожної з його граней є *рівноймовірною* подією, і така ймовірність, зрозуміло, є $P = 1/n = 1/6$.

Тоді *математичне сподівання* при підкиданні кубика:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} i == \frac{1}{6}(1+2+3+4+5+6) = 3,5;$$

середній квадрат

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} i^2 = \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) = 15,17;$$

дисперсія:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (i - \overline{x})^2 = \overline{x^2} - (\overline{x})^2 = 15,17 - 3,5^2 = 2,92.$$

середньоквадратичне відхилення:

$$\sigma = \sqrt{\sigma^2} = 1,7.$$

Розглянемо випадкову величину *X*, яка є результатом експерименту з багатократного підкидання кубика. Нехай у результаті експерименту отримана скінченна послідовність $\{x_i\}$, i=1,2,…,N змодельованих значень випадкової величини *X*. Таку скінченну послідовність називають *вибіркою* з випадкового процесу.

Нехай у вибірці кожне число k=1,2,3,4,5,6 з'являється N_k разів. Уведемо частоту випадання числа k:

$$\nu_k = \frac{N_k}{N}.$$

Якщо повторити ще раз ту саму кількість N кидань, то отримаємо, взагалі кажучи, інше значення N_k і ν_k. Але при $N \to \infty$, $\quad \nu_k \to P_k = P = \frac{1}{6}$.

Для вибірки вводяться поняття *вибіркового математичного сподівання* $\overline{x}_N$, вибіркового *середнього квадрату* $\overline{x_N^2}$, *вибіркової дисперсії* σ_N^2 і *вибіркового середньоквадратичного відхилення* σ_N:

$$\overline{x}_N = \frac{1}{N}\sum_{i=1}^{N} x_i,$$

$$\overline{x^2} = \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

$$\sigma_N^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} \left(x_i - \overline{x}_N \right)^2 = \overline{x_N^2} - (\overline{x}_N)^2 .$$

$$\sigma_N = \sqrt{\sigma_N^2}$$

Значення $\overline{x}_N$, σ_N^2, σ_N при відносно великих N, можна прийняти в якості оцінок математичного сподівання $\overline{x}$, дисперсії σ^2 і середньоквадратичного відхилення σ величини X, тобто, $\overline{x} \cong \overline{x}_N$, $\sigma^2 \cong \sigma_N^2$, $\sigma \cong \sigma_N$. Наближені рівності стають точними в межі, коли $N \to \infty$.

Завдання

1) Згенерувати вибірку $\{x_i\}$, $i=1,2,\ldots,N$ з цілих випадкових чисел (варіант 1 – 1÷4; варіант 2 – 1÷6; варіант 3 – 1÷8; варіант 4 – 1÷12; варіант 5 – 1÷20 або інший запропонований викладачем варіант). Для цього використати генератор псевдовипадкових чисел обраної мови програмування.
2) Побудувати залежність частоти випадання k-го числа від номера k. Графік представити у вигляді стовпців (побудувати гістограму).
3) Обчислити для згенерованого масиву чисел вибіркове математичне сподівання $\overline{x}_N$, вибіркову дисперсію σ_N^2, вибіркове середньоквадратичне відхилення σ_N. Порівняти отримані значення із теоретичними.
4) Виконати п.1 – п.3 для N=10, N=100, N=1000, N=10000.
5) Зобразити графічно залежності $\overline{x}_N$ та σ_N від $\lg N$. Проаналізувати отримані результати.
6) Зробити висновки. Оформити звіт.