Статистичні характеристики

Стохастичними називають експерименти, результат яких неможливо передбачити до проведення експерименту, але який можна повторити в незалежний спосіб будь-яке число разів. Фіксований результат експерименту, який не можна виразити через сукупність інших результатів, називається елементарною подією. Частота появи випадкової події:

$$N(S) = \frac{m}{n} \ ,$$

де n – загальна кількість проведених експериментів, m – кількість ситуацій, що супроводжуються появою події S.

Ймовірність – це ненегативне число, яке є кількісною мірою **об'єктивної** можливості реалізації випадкової події S:

$$P(S)=\lim_{n\to\infty} \frac{m}{n} \ .$$

Повною групою подій називається така система подій, що в результаті проведеного експерименту неодмінно станеться одна і тільки одна з подій. Сума ймовірностей усіх подій повної групи дорівнює одиниці:

$$\sum_{i=1}^{n} P(S_i) = 1.$$

Дві події називаються сумісними, якщо поява однієї з них не виключає появу другої події в тому самому випробуванні. Якщо поява випадкової події *A* впливає на ймовірність появи іншої випадкової події *B*, то використовують поняття **умовної ймовірності**. Імовірність події *B* при умові, що мала місце подія *A*, позначається $P(B/A)$.

Імовірність **одночасної** появи двох подій *A* і *B* у припущенні, що перша подія вже відбулась

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

Якщо ці події незалежні, то

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

Імовірність появи ***хоча б однієї*** з двох сумісних подій

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

Формула Байєса

Нехай подія *A* може відбутись тільки разом з однією із попарно несумісних подій $H_1, H_2, \cdots, H_n$, які називаються *гіпотезами (hypothesis)* і утворюють повну групу, тобто $\sum_{i=1}^{n} P(H_i) = 1$.

Імовірність $P(H_i)$ гіпотези H_i, яка відома ще до початку випробувань, називається **апріорною ймовірністю.**

Імовірність події А визначається **формулою повної ймовірності**:

$$P(A) = \sum_{i=1}^{n} P(H_i) \cdot P(A|H_i).$$

Апостеріорна ймовірність $P(Hi|A)$ гіпотези Hi за умови, що відбулась подія A, обчислюється за формулою Байєса:

$$P(H_i|A) = \frac{P(H_i) \cdot P(A|H_i)}{P(A)}.$$

Випадкова величина *X* - це величина, яка в результаті випробувань може приймати певні значення (із сукупності своїх можливих значень x_1, x_2, … , x_i, …, x_n) з певною ймовірністю $P(x_\mathrm{i})$.

Випадкові величини поділяються на дискретні та неперервні.

Перелік всіх можливих значень дискретної випадкової величини і їх ймовірностей називають **законом розподілу**.

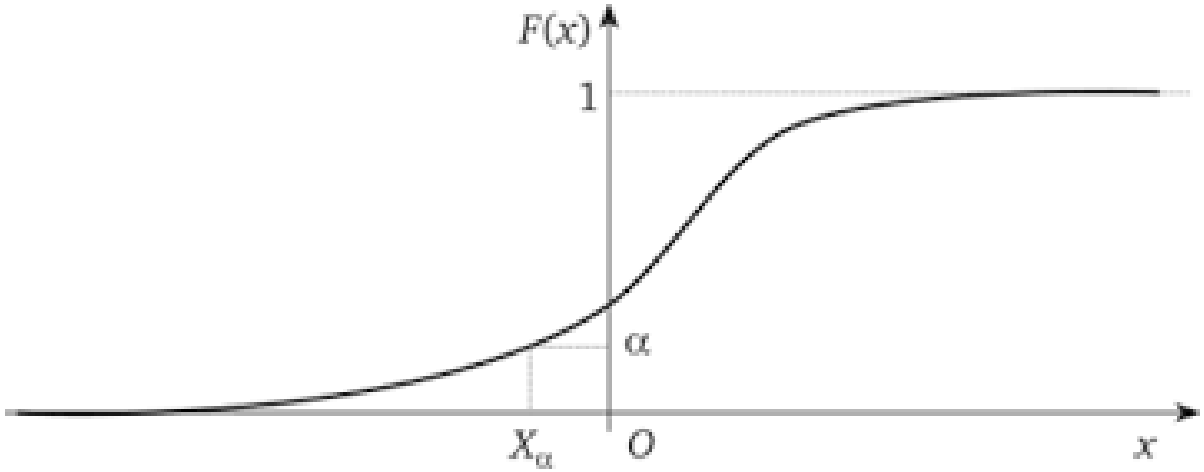
Закон розподілу може бути задано функціями:

а) функцією розподілу (інтегральна функція розподілу);

б) функцією густини розподілу (диференціальна функція розподілу).

Функцією розподілу випадкової величини *X* називають ймовірність того, що випадкова величина *X* прийме значення менше аргументу цієї функції *x*:

$F(x) = \mathrm{P}(X < x)$.



Функція розподілу має наступні властивості:

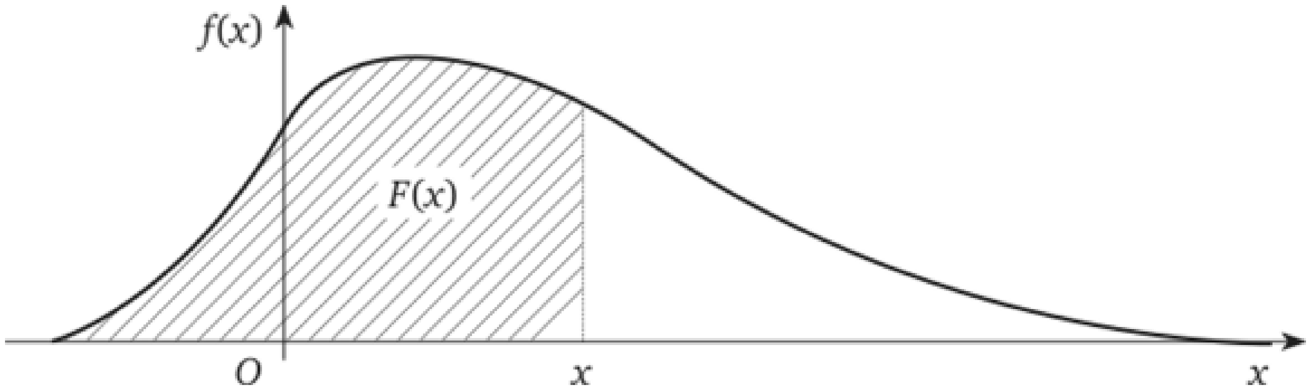
1) $0 \le F(x) \le 1$;

2) F(x) неперервно зростає на всій числовій осі;

3) ймовірність того, що випадкова величина X прийме значення з інтервалу (a;b), дорівнює приросту функції розподілу на цьому інтервалі: $P(a < X < b) = F(b) - F(a)$;

4) $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to +\infty} F(x) = 1$.

Диференціальною функцією неперервного розподілу *(густиною ймовірності,* або *густиною розподілу*) називають першу похідну від інтегральної функції $f(x) = F'(x)$.



Властивості:

1) $f(x) \geq 0$;

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

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

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

Генеральна сукупність — вся множина однорідних за певною ознакою подій, які можуть бути предметом дослідження.

Оскільки часто дослідити всю генеральну сукупність неможливо, вибирають підмножину, яка повинна представляти всю генеральну сукупність для статистичних висновків. Таку підмножину називають «**статистична вибірка**».

Вибірка представляє *n* експериментів, в яких можна виміряти одну і ту ж випадкову змінну**.**

Шляхом математичної обробки статистичної вибірки отримують статистичні характеристики випадкових величин.

Основні числові характеристики випадкових величин

Математичним очікуванням випадкової величини *X* називається сума добутків всіх можливих значень *X* на відповідні їм ймовірності:

$$M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i \;.$$

Математичне очікування характеризує середнє значення випадкової величини *X* із врахуванням ймовірностей її можливих значень.

Властивості математичного очікування:

1. Математичне очікування сталої величини С дорівнює цій же сталій:

$$M\ (C) = C\ ,\ C = \text{const}.$$

2. Сталий множник можна виносити за знак математичного очікування:

$$M\ (CX) = CM\ (X),\ C = \text{const}.$$

3. Математичне очікування суми (різниці) двох випадкових величин дорівнює сумі (різниці) їх математичних очікувань:

$$M\ (X \pm Y) = M\ (X) \pm M\ (Y).$$

Дисперсією $D(X)$ випадкової величини X називають математичне очікування квадрата відхилення X від її математичного очікування:

$$D(X) = M((X - M(X))^2) = \sum_{i=1}^{n} (x_i - M(X))^2 p_i \ .$$

Дисперсія має такі властивості:

1. Дисперсія сталої величини дорівнює нулю:

$$D(C) = 0 \ , \ C = \text{const.}$$

2. За знак дисперсії можна винести квадрат сталого множника:

$$D(CX) = C^2 D(X) \ , \ C = \text{const.}$$

3. Дисперсія суми та різниці двох незалежних випадкових величин дорівнює **сумі** їх дисперсій:

$$D(X \pm Y) = D(X) + D(Y) \ .$$

4. Дисперсія дорівнює різниці математичного очікування квадрата випадкової величини X і квадрата її математичного очікування:

$$D(X) = M(X^2) - M^2(X) \ .$$

Середнім квадратичним відхиленням (СКВ) σ(*x*) випадкової величини *X* називається корінь квадратний з її дисперсії:

$$\sigma(x)=\sqrt{D(X)}\,.$$

СКВ суми кінцевого числа незалежних випадкових величин дорівнює кореню квадратному із суми квадратів СКВ цих величин:

$$\sigma(X_1+X_2+\ldots+X_n)=\sqrt{\sigma^2(X_1)+\sigma^2(X_2)+\ldots+\sigma^2(X_n)}\,.$$

У випадку, коли дисперсії цих величин однакові:

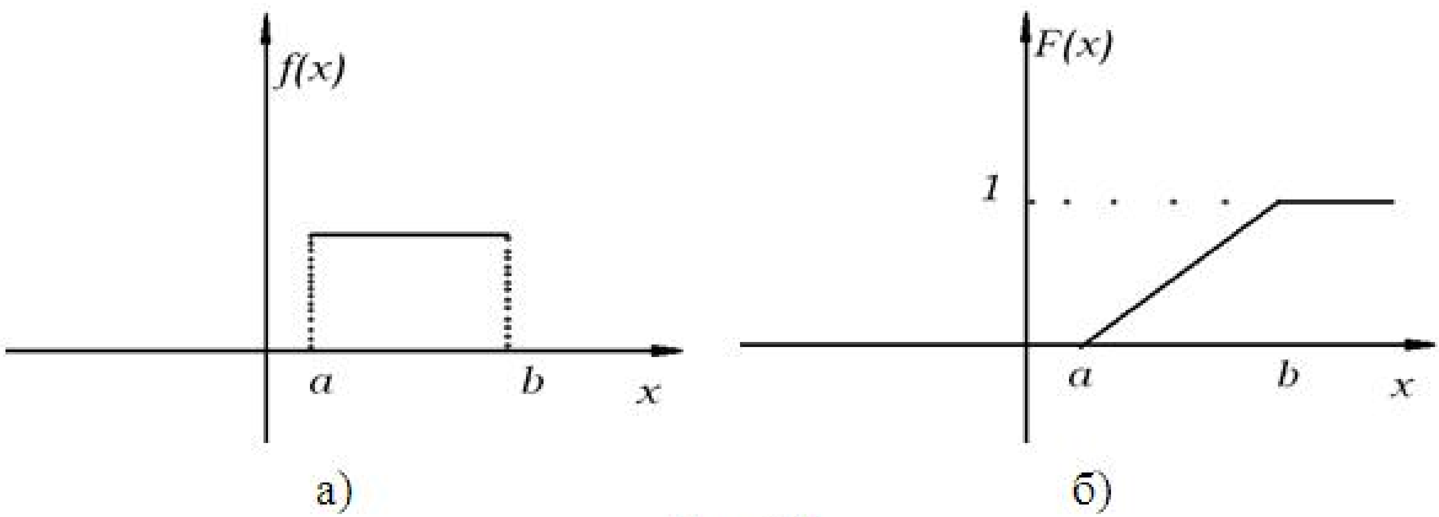
$$\sigma^2(X_i)=\sigma^2,\ \left(i=\overline{1,\ n}\right),\quad \sigma(X_1+X_2+\ldots+X_n)=\sigma\sqrt{n}\,.$$

Дисперсія та СКВ відносяться до характеристик розсіяння можливих значень випадкової величини навколо її математичного очікування (центра розподілу).

Найбільш вживані закони розподілу

Рівномірний:

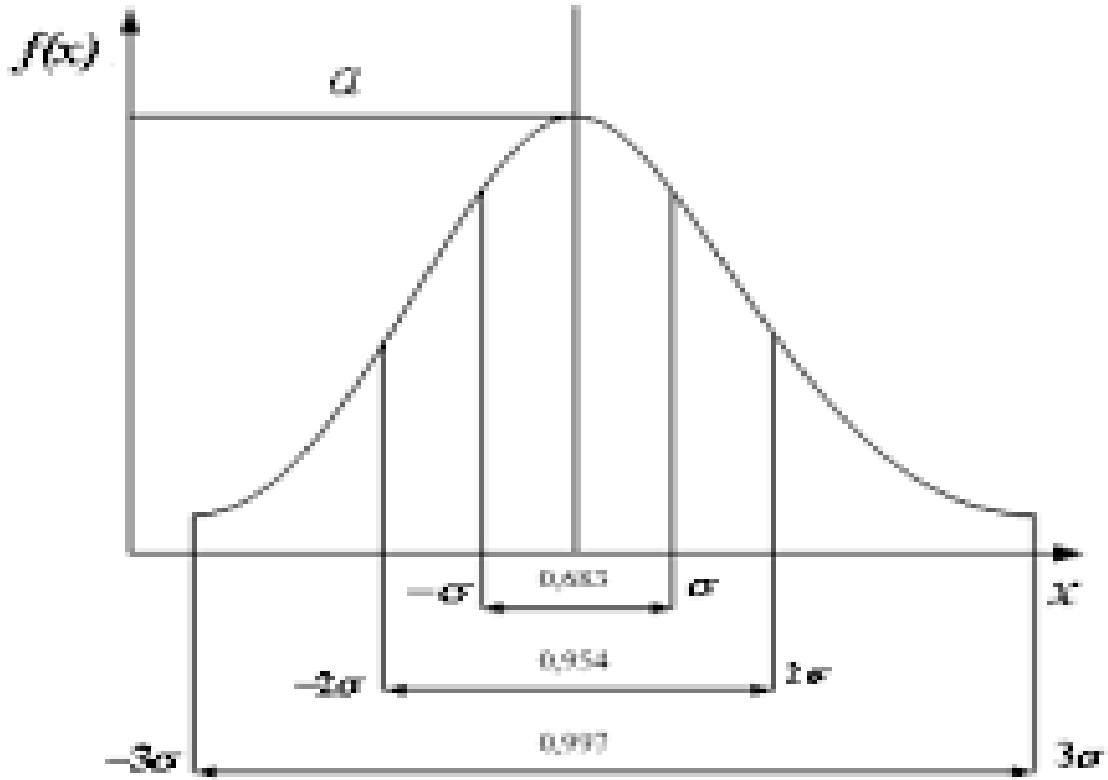
$$f(x)=\begin{cases}0, & при \quad x<a,\\ \dfrac{1}{b-a}, & при \quad a\le x\le b,\\ 0, & при \quad x>b.\end{cases}$$



Використовується при мінімальних відомостях про випадкову величину.

Нормальний (закон Гауса):

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}} \qquad F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-a)^2}{2\sigma^2}}dt$$



Він є граничним законом, до якого наближаються інші закони розподілу.

Оцінювання статистичних характеристик на основі статистичної вибірки

Основними характеристиками вибірки є:

- вибіркове математичне очікування (вибіркове середнє) :

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

– вибіркова дисперсія

$$D_X^* = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2,$$

– вибіркове середньоквадратичне (стандартне) відхилення

$$\sigma_X = \sqrt{D_X}\ .$$

Точковою оцінкою невідомого параметра розподілу θ називається наближене значення цього параметра, отримане за даними вибірки x_1, x_2, …, x_n.

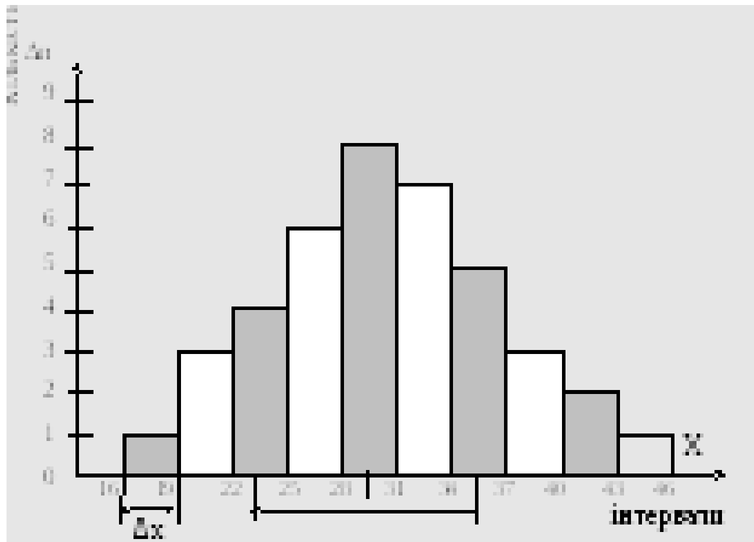
Іноді в статистичних розрахунках важливо не лише знайти оцінку параметра розподілу, але і охарактеризувати її точність. Для цього вводиться поняття про довірчий інтервал для параметра θ - це інтервал (θ_1, θ_2), що містить (накриває) істинне значення θ із заданою ймовіністю P = 1 – α, тобто

$$P=\{ \theta_1< \theta < \theta_2\}=1- \alpha.$$

Число P = 1 – α називають довірчою ймовірністю (чи надійністю), а малу величину α - рівнем значущості.

В цьому випадку кажуть про **інтервальну** оцінку параметра θ.

Гістограма – це ступінчастий графік, для побудови якого по осі абсцис відкладають значення вимірюваної величини, розбиті на інтервали (біни), а по осі ординат – частоти $\Delta n_i/n$ попадання у біни Δx_i. Ці частоти є точковими оцінками ймовірностей попадання значення випадкової величини у кожний бін.



Якщо обсяг вибірки і кількість бінів збільшувати до нескінченності, ширина стовпчика гістограми стає нескінченно малою величиною і гістограма перетворюється у криву диференціального закону розподілу.