ЛЕКЦІЯ 5(А).
Неперервні випадкові величини.

1. *Визначення випадкової величини неперервного типу*. 2. *Деякі властивості щільності $f_\xi(x)$*. 3. *Сингулярні розподіли.*

1. Визначення випадкової величини неперервного типу.

Припустимо, подібно як і у випадку дискретних випадкових величин, що метою дослідження є встановлення *розподілу* деякої кількісної ознаки X для окремих елементів в межах заданої генеральної популяції. Іншими словами, необхідно вивчити:

- Які значення може приймати ця ознака та як часто окремі значення в цій популяції зустрічаються.

Якщо виявиться, що:

- ➢ X може приймати *будь-які значення з деякого інтервалу на прямій*,

тобто кількість їх – незлічена, то поняття та методи дослідження, запропоновані для ознак дискретного типу, будуть недостатніми для їх вивчення. Ось чому описова статистика виділяє два типи кількісних ознак:

- ➢ *дискретні* та *неперервні*:

Описова статистика пропонує наступне визначення:

Визначення (описове). Кажемо, що маємо справу з вимірювальною ознакою X *неперервного* типу, якщо ця ознака може приймати *будь-які значення* з деякого *інтервалу* на прямій: $X \in \Omega_X = (a, b)$. Не виключаємо при цьому, що інтервал цей може бути необмежений, тобто $a = -\infty$, або $b = +\infty$, або ж і $a = -\infty$, і $b = +\infty$.

Припустимо тепер, що намагаючись формально описати схему *випадкового* вибору при статистичному дослідженні кількісної властивості *неперервного* типу $X \in (a, b)$, хочемо поставити їй у відповідність деяку випадкову величину (ξ):

$$X \Leftrightarrow \xi.$$

Тоді, подібно до X, випадкова величина ξ теж буде приймати *будь-які значення* з цього інтервалу: $\xi \in (a, b)$. І якщо $(\Omega, \Im, P)$ – деякий ймовірнісний простір, то множина можливих значень функції $\xi = \xi(\omega)$, $\omega \in \Omega$, буде незлічена. Тому не маємо можливості, як це було зроблено у випадку дискретної випадкової величини:

- Приписати кожному значенню x додатну ймовірність:

$$p(x) = P\{\xi = x\},\ x \in (a, b).$$

Тому, подібно до того, як описова статистика виділяє два типи кількісних ознак: *дискретні* та *неперервні*, теорія ймовірностей крім дискретних, окремо визначає також *неперервні* випадкові величини.

Визначення. Кажемо, що ξ є *неперервною* випадковою величиною з *щільністю* $f_\xi(x)$, якщо її функцію розподілу $F_\xi(x) = P\{\xi < x\}$ для будь-якого дійсного числа x можна записати у вигляді:

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(u)du, -\infty < x < \infty.$$

Зауваження. З цього визначення випливає, що функція розподілу неперервної випадкової величини теж буде неперервною. Якщо крім цього $F_\xi(x)$ – диференційована функція, то щільність $f_\xi(x)$, $-\infty < x < \infty$, визначається, як похідна функції розподілу $F_\xi(x)$:

$$f_\xi(x) = (F_\xi(x))' = \frac{dF_\xi(x)}{dx}.$$

2. Деякі властивості щільності $f_\xi(x)$.

Приведемо спочатку основні, або «*визначальні*» властивості щільності.

Властивість 1. Оскільки $F_\xi(x)$ – не спадна функція, то $f_\xi(x) \ge 0$, $-\infty < x < \infty$.

Властивість 2. Оскільки $F_\xi(\infty) = P\{\xi < \infty\} = 1$, то $\int_{-\infty}^{\infty} f_\xi(x)dx = 1$.

Термін «*визначальні*» в даному випадку необхідно інтерпретувати в сенсі властивостей «*необхідних та достатніх*».

- З одного боку вони безпосередньо випливають з визначення щільності.
- З іншого боку – якщо функція $f_\xi(x)$ має такі властивості, то $F_\xi(x)$ буде функцію розподілу деякої випадкової величини, а отже $f_\xi(x)$ буде її щільністю, а сама випадкова величина буде *неперервною*.

Зауваження. Якщо порівнювати *дискретні* та *неперервні* випадкові величини, то щільність $f_\xi(x)$ слід розглядати, як *неперервний аналог* розподілу $\{(x_i, p_i), i = 1, 2, \ldots, n, \ldots\}$ дискретної величини. Нагадаємо, що відповідні «*визначальні*» властивості розподілу виглядають наступним чином:

1) $p_i > 0$, $i = 1, 2, \ldots, n, \ldots$;
2) $p_1 + p_2 + \ldots + p_n + \ldots = 1$.

Сформулюємо та обґрунтуємо ще декілька корисних з практичної точки зору властивостей *неперервних* випадкових величин та їх щільності.

Властивість 3. Нехай $[c, C]$ – деякий інтервал на числовій прямій. Тоді ймовірність випадкової події $P\{c \le \xi < C\}$ дорівнює площі фігури, утвореної графіком щільності $f_\xi(x)$ на цьому інтервалі.

Дійсно, використовуючи визначення неперервної випадкової величини та властивість 5, встановлену для функції розподілу $F_\xi(x)$, отримаємо:

$$P\{c \le \xi < C\} = F(C) - F(c) = \int_{-\infty}^{C} f_\xi(u)du - \int_{-\infty}^{c} f_\xi(u)du = \int_{c}^{C} f_\xi(x)dx.$$

Властивість 4. Якщо ξ – неперервна випадкова величина, то для довільного дійсного числа c має місце рівність: $P\{\xi = c\} = 0$.

Дійсно, оскільки функція розподілу є неперервною *зліва*, то спрямовуючи ($C \to c$) отримаємо наступну рівність:

$$P\{\xi = c\} = \lim_{C\to c} P\{c \le \xi < C\} = P\{c \le \xi < c_{+0}\} = F_\xi(c_{+0}) - F_\xi(c).$$

Варто зауважити, що рівність ця справедлива як для дискретних, так і для неперервних випадкових величин. У випадку неперервної величини функція $F_\xi(x)$ буде неперервною, а отже для довільного c $F_\xi(c_{+0}) = F_\xi(c)$.

На практиці властивість 4 означає, що *немає сенсу* говорити про *точне* значення неперервної випадкової величини. Виправданим є тільки оцінювання можливості для її значень потрапити до *певного числового інтервалу* ненульової довжини. Іншими словами:

- Для кожного *конкретного* числа c має місце рівність $P\{\xi = c\} = 0$, тоді, як *додатну* ймовірність можуть мати лише випадкові події типу:

$$P\{\xi \in (c, C)\}, \text{ при умові, що } C - c > 0.$$

Тому дамо оцінку для ймовірностей подібних подій з використанням щільності $f_\xi(x)$, та принагідно встановимо, звідки походить термін «*щільність розподілу*».

Властивість 5. Якщо ξ – неперервна випадкова величина з *щільністю* $f_\xi(x)$, то для будь-якого дійсного числа $-\infty < x < \infty$, має місце наступна наближена рівність:

$$P\{x \le \xi < x + \Delta x\} \approx f_\xi(x) \cdot \Delta x, \ (\Delta x > 0),$$

тобто ймовірність для випадкової величини ξ потрапити до числового інтервалу $[x, x + \Delta x]$ *прямо пропорційна* значенню щільності $f_\xi(x)$.

Нехай $[x, x + \Delta x]$ – довільний числовий інтервал довжини Δx. Використовуючи визначення похідної функції, а також властивості функції розподілу отримаємо:

$$f_\xi(x) = (F_\xi(x))' = \lim_{\Delta x\to 0} \frac{F_\xi(x+\Delta x) - F_\xi(x)}{\Delta x} = \lim_{\Delta x\to 0} \frac{P(x \le \xi < x + \Delta x)}{\Delta x}.$$

Тобто маємо наступну наближену рівність:

$$f_\xi(x) \approx \frac{P(x \le \xi < x + \Delta x)}{\Delta x},$$

яка одночасно доводить властивість 4) та пояснює походження терміну «*щільність розподілу*:

- Функція $f_\xi(x)$ визначає «*величину ймовірності*», яка «*припадає*» на одиницю довжини інтервалу $[x, x + \Delta x]$.

Доведену наближену рівності можна також трактувати наступним чином:

- Ймовірність того, що значення неперервної випадкової величини буде знаходитись «*поблизу*» дійсного числа x *приблизно* дорівнює ($f_\xi(x)\cdot\Delta x$).

Використовуючи цю інтерпретацію можемо зробити кілька дуже корисних з практичної точки зору висновків.

Властивість 6. Неперервна випадкова величина з щільністю $f_\xi(x)$ може приймати тільки такі значення x, для яких $f_\xi(x) > 0$.

Властивість 7. Чим більша величина щільності $f_\xi(x)$, тим більш *ймовірними* буде для неперервної випадкової величини ξ значення x.

Властивість 8. Справедлива наступна *формальна* рівність:

$$P\{\xi = x\} = f_\xi(x) \cdot dx.$$

Тому з формальної точки зору визначення:

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(u)du = P\{\xi < x\}$$

означає, що $F_\xi(x)$ дорівнює «*сумі ймовірностей* $P\{\xi = u\}$», причому сума ця поширюється на всі можливі значення ($u < x$) величини ξ менші від x.

Точно такий зв'язок існує між функцією розподілу $F_\eta(x)$ та розподілом $p_i = P\{\eta = x_i\}$, $i = 1, 2, \ldots$ дискретної випадкової величини η:

$$F_\eta(x) = \sum_{i:(x_i<x)} p_i\ , -\infty < x < \infty,$$

де сума поширюється на всі можливі значення (x_i) величини η, строго менші від x ($x_i < x$).

3. Сингулярні розподіли.

Окрім розглянутих вже *дискретних* та *неперервних* розподілів, існують ще розподіли третього типу: *сингулярні розподіли*.

Визначення. Припустимо, що $F_\xi(x)$ функцію розподілу. Точка x називається *точкою росту* (або *зростання*) функції $F_\xi(x)$, якщо будь-якого додатного числа $\varepsilon > 0$ виконується нерівність:

$$F_\xi(x + \varepsilon) - F_\xi(x - \varepsilon) > 0.$$

Так, наприклад:

- Для *дискретного* розподілу $\{(x_k, p_k), k = 1, 2, ... \}$ точками зростання функції розподілу є точки x_k, $k = 1, 2, ...$.
- Для *неперервного* розподілу, зосередженого в інтервалі $[a, b]$, точками зростання функції розподілу будуть всі точки інтервалу $[a, b]$.

Визначення. Кажемо, що функція $F_\xi(x)$ визначає розподіл *сингулярного типу*, якщо $F_\xi(x)$ – неперервна функція, множина точок росту якої має *нульову лебегову* міру.

Нагадаємо, що лебегівська міра для інтервалу $[a, b)$ числової прямої дорівнює довжині: $\Delta = b - a$ цього інтервалу.

Прикладом сингулярної функції може бути *крива Кантора* $K(x)$, яка визначається наступним чином.

- Нехай $K(x) = 0$ для $x < 0$; $K(x) = 1$ для $x > 1$;
- $K(x) = ½$ для $1/3 \le x \le 2/3$;
- $K(x) = ¼$ для $1/9 \le x \le 2/9$; $K(x) = ¾$ для $7/9 \le x \le 8/9$;
- $K(x) = 1/8$ для $1/27 \le x \le 2/27$; $K(x) = 3/8$ для $7/27 \le x \le 8/27$; $K(x) = 5/8$ для $19/27 \le x \le 20/27$; $K(x) = 7/8$ для $25/27 \le x \le 26/27$;

Продовжуючи цю процедуру, а потім до-визначаючи побудовану функцію таким чином, щоб зробити її неперервною, ми отримаємо розподіл, для якого загальна довжина інтервалів сталості становить:

$$\frac{1}{3}+\frac{2}{9}+\frac{4}{27}+\ldots+\frac{2^k}{3^{k+1}}+\ldots=1.$$

Таким чином множина точок росту *кривої Кантора* $K(x)$ має *нульову лебегову* міру.

Наведений вище приклад показує, що розподіли *сингулярного типу* існують. Але вони, скоріше, представляють теоретичний інтерес, з формальної точки зору. В практичних застосуваннях статистичних методів, принаймні в економічних моделях, ми, як правило, не зустрічаємо розподілів сингулярного типу. Тому далі будемо розглядати тільки два основних типи розподілів: дискретні та неперервні.

З теоретичної точки зору проблему існуючих типів розподілів випадкових величин остаточно розв'язує наступна, фундаментальна для теорії ймовірностей, теорема А. Лебега.

Теорема (Лебега). Кожну функцію розподілу $F_\xi(x)$ можна *єдиним способом* розкладена у вигляді суми:

$$F_\xi(x) = F_\xi^{(д)}(x) + F_\xi^{(н)}(x) + F_\xi^{(с)}(x),$$

трьох *неспадних* функцій $F_\xi^{(д)}(x)$, $F_\xi^{(н)}(x)$, $F_\xi^{(с)}(x)$.

При цьому:

1. Функція $F_\xi^{(д)}(x)$ характеризує *дискретну* складову розподілу $F_\xi(x)$ і має вигляд:
$$F_\xi^{(д)}(x) = \sum_{x_i<x} p(x_i),\ p(x_i) \ge 0,\ \sum_i p(x_i) \le 1.$$
2. Функція $F_\xi^{(н)}(x)$ характеризує *неперервну* складову розподілу $F_\xi(x)$ і має вигляд:
$$F_\xi^{(н)}(x) = \int_{-\infty}^{x} f_\xi(u)du,\ f_\xi(x) \ge 0,\ \int_{-\infty}^{\infty} f_\xi(x)dx \le 1.$$
3. Функція $F_\xi^{(с)}(x)$ характеризує *сингулярну* складову розподілу $F_\xi(x)$ і є неперервною функцією: $0 \le F_\xi^{(с)}(x) \le 1$, множина *точок росту* якої має *нульову лебегову* міру.