ПРАКТИЧНЕ ЗАНЯТТЯ № 5.
Головні задачі математичної статистики.
Непараметричне оцінювання.

Вказівка: Припустимо, що предметом статистичного дослідження є генеральна популяція з функцією розподілу $F(x)$. Маємо в розпорядженні просту випадкову вибірку $(\xi_1, \xi_2, \dots, \xi_n)$ чисельності n з цієї популяції.
Емпірична функція розподілу, побудована на підставі вибірки $(\xi_1, \xi_2, \dots, \xi_n)$, визначається рівністю:

$$\hat{F}_n(x) = \frac{\nu_n}{n},$$

де $\nu_n(x)$ означає кількість елементів вибірки менших від x.

Емпірична функція розподілу $\hat{F}_n(x)$ є статистичною оцінкою функції розподілу $F(x)$ генеральної популяції:

$$\hat{F}_n(x) \approx F(x).$$

Приклад 1. Маємо дві урни A та B в яких є як *білі*, так і *чорні* кулі.
Повторюється $n = 25$ разів наступний експеримент:

- *з кожної урни випадковим чином вибирається по одній кулі, а потім в отриманій таким чином парі підраховується кількість білих куль.*

Після цього кулі повертаються до своїх урн (початковий склад урн не змінюється).
Результатом цього експерименту є наступна вибірка (y_i – кількість білих куль в парі i-того по порядку експерименту (i = 1, 2, …, 25)):

$$y_1 = 2;\ y_2 = 0;\ y_3 = 2;\ y_4 = 1;\ y_5 = 1;\ y_6 = 1;\ y_7 = 1;\ y_8 = 2;\ y_9 = 1;\ y_{10} = 0;$$
$$y_{11}=2;\ y_{12}=1;\ y_{13}=1,\ y_{14} = 0;\ y_{15} = 1;\ y_{16} = 0;\ y_{17} = 1;\ y_{18}=2;\ y_{19} = 0\ y_{20} = 1;$$
$$y_{21} = 1;\ y_{22} = 0;\ y_{23} = 1;\ y_{24} = 2;\ y_{25}=1.$$

Завдання 1. Збудувати *точковий ряд*: $\{(x_1, n_1), (x_2, n_2), \dots, (x_k, n_k)\}$, тобто визначити множину різних значень $\{x_1, x_2, \dots, x_k\}$, що зустрічаються у вибірці, та підрахувати, скільки разів (n_i) кожне з них (x_i) зустрічаються у вибірці (i = 1, 2, …, k).

Відповідь: {(0; 6), (1; 13), (2; 6)}.

Завдання 2. На підставі точкового ряду: $\{(x_1, n_1), (x_2, n_2), \dots, (x_k, n_k)\}$, збудувати емпіричний розподіл для випадкової величини ξ, яку представляє вибірка $\{y_1, y_2, \dots, y_n\}$: $\{(x_1, \hat{p}_1), (x_2, \hat{p}_2), \dots, (x_k, \hat{p}_k)\}$, де $\hat{p}_i = n_i/n$ – частота появи значення x_i у вибірці (i = 1, 2, ..., k).

Вказівка: Емпіричний розподіл $\{(x_1, \hat{p}_1), (x_2, \hat{p}_2), \dots, (x_k, \hat{p}_k)\}$ є статистичною оцінкою ймовірнісного розподілу $\{(x_1, p_1), (x_2, p_2), \dots, (x_k, p_k)\}$.

Відповідь: {(0; 0,24), (1; 0,52), (2; 0,24)}.

Завдання 3. Збудувати емпіричну функцію розподілу $\hat{F}_n(x)$, та знайти її значення в точках x = 0,3; x = 1; x = 1,5.

Відповідь:

x_i	0	0.3	1	1,5	2	2,2
p_i	0	0,24	0,24	0,76	0,76	1

Завдання 4. Побудувати графік емпіричної функцію розподілу $\hat{F}_n(x)$.

Приклад 2. З генеральної популяції з функцією розподілу $F(x)$ вибрано наступну просту випадкову вибірку:

$$\mathbf{x_1 = 2{,}1;\ x_2 = 1{,}5;\ x_3 = 3{,}2;\ x_4 = 0{,}3;\ x_5 = 2{,}1;}$$
$$\mathbf{x_6 = 2{,}5;\ x_7 = 1{,}7,\ x_8 = 4{,}3;\ x_9 = 3{,}2;\ x_{10} = 0{,}7.}$$

Завдання 1. На підставі випадкової вибірки побудувати *варіаційний* ряд:

$$x^*_1 \le x^*_2 \le \ldots \le x^*_{10},$$

порядкуючи елементи вибірки.

Відповідь: $x^*_1 = 0{,}3$; $x^*_2 = 0{,}7$; $x^*_3 = 1{,}5$; $x^*_4 = 1{,}7$; $x^*_5 = 2{,}1$; $x^*_6 = 2{,}1$; $x^*_7 = 2{,}5$; $x^*_8 = 3{,}2$; $x^*_9 = 3{,}2$; $x^*_{10} = 4{,}3$.

Завдання 2. Використовуючи визначення емпіричної функції розподілу $\hat{F}_n(x)$ знайти її значення в точках $[x_3, x_5, x_8]$ випадкової вибірки.

Відповідь: $\hat{F}_n(3{,}2) = 0{,}7$; $\hat{F}_n(2{,}1) = 0{,}4$; $\hat{F}_n(4{,}3) = 0{,}9$.

Завдання 3. Побудувати графік емпіричної функцію розподілу $\hat{F}_n(x)$.

Емпіричний розподіл.

Припустимо, що ми маємо справу зі статистичною ознакою дискретного типу, яка має наступний розподіл:

$$\{(x_1, p_1), (x_2, p_2), \ldots, (x_k, p_k)\},$$

Тобто $\xi \in \{x_1, x_2, \ldots, x_k\}$, при цьому

$$p_i = P\{\xi = x_i\},\ i = 1, 2, \ldots, k.$$

У цьому випадку проблему непараметричного оцінювання вирішує *емпіричний розподіл*, який визначається наступним чином:

Це набір пар

$$\{(x_1, \hat{p}_1), (x_2, \hat{p}_2), \ldots, (x_k, \hat{p}_k)\},$$

де числа $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_k$, визначають частоту появи відповідних значень $x_1, x_2, \ldots, x_k$ у випадковій вибірці $(\xi_1, \xi_2, \ldots, \xi_n)$. Іншими словами:

$$\hat{p}_i = \frac{n_i}{n},\ i = 1, 2, \ldots, k,$$

Приклад 3. Предметом статистичного дослідження є кількісна властивість (X) дискретного типу. Підставою до дослідження є наступна проста випадкова вибірка розмірності $n = 50$:

7	7	4	7	10	2	7	4	7	2
7	7	7	7	10	2	10	10	10	10
7	7	2	10	2	2	2	10	7	10
7	10	7	2	2	4	7	7	7	4
10	2	2	10	7	7	7	7	10	2

Завдання 1. Збудувати *точковий ряд*: $\{(x_1, n_1), (x_2, n_2), \ldots, (x_k, n_k)\}$, тобто визначити множину різних значень $\{x_1, x_2, \ldots, x_k\}$, що зустрічаються у вибірці, та підрахувати, скільки разів (n_i) кожне з них (x_i) зустрічаються у вибірці ($i = 1, 2, \ldots, k$).

Завдання 2. На підставі точкового ряду: $\{(x_1, n_1), (x_2, n_2), ..., (x_k, n_k)\}$, збудувати емпіричний розподіл $\{(x_1, \hat{p}_1), (x_2, \hat{p}_2), ..., (x_k, \hat{p}_k)\}$, для випадкової величини ξ, яку представляє вибірка $\{y_1, y_2, \dots, y_n\}$. $\hat{p}_i$ означає частоту появи значення x_i у вибірці:

$$\hat{p}_i = \frac{n_i}{n} \quad i = 1, 2, ..., k.$$

Завдання 3. Збудувати емпіричну функцію розподілу $\hat{F}_n(x)$, та намалювати її графік.

Розв'язок:

Завдання 1. Аналізуючи дані, які містяться в таблиці, приходимо до висновку, що статистична ознака може приймати наступні значення:

$$x_1 = 2;\ x_2 = 4;\ x_3 = 7;\ x_4 = 10;\ (k = 4).$$

Точковий ряд, побудований на основі аналізованої вибірки, представимо у вигляді таблиці:

$x_1 = 2$	$x_2 = 4$	$x_2 = 7$	$x_2 = 10$
$n_1 = 12$	$n_2 = 4$	$n_1 = 21$	$n_1 = 13$

Завдання 2. Провівши необхідні розрахунки, отримаємо:

$$\hat{p}_1 = \frac{12}{50} = 0{,}24;\ \hat{p}_2 = \frac{4}{50} = 0{,}08;\ \hat{p}_3 = \frac{21}{50} = 0{,}42;\ \hat{p}_4 = \frac{13}{50} = 0{,}26.$$

Отриманий емпіричний розподіл статистичної ознаки, побудований на основі аналізованої вибірки, представимо у вигляді таблиці:

x_i	$x_1 = 2$	$x_2 = 7$	$x_2 = 7$	$x_2 = 10$
$\hat{p}_i$	$\hat{p}_1 = 0{,}24$	$\hat{p}_2 = 0{,}08$	$\hat{p}_3 = 0{,}42$	$\hat{p}_4 = 0{,}26$

$$\hat{p}_i = \frac{n_i}{n},\ i = 1, 2, ..., k,$$

Приклад 4. Припустимо, що випадкова величина ξ має розподіл B(6; 0,25), тобто біноміальний, з параметрами ($n = 6$, $p = 0{,}25$).

Завдання 1. Використовуючи таблицю ймовірностей $Q(k, n, p) = P\{B(n, p) \geq k\}$ для біноміального розподілу, обчисліть ймовірності p_A та p_B наступних випадкових подій: $A = \{\xi = 2\}$; $B = \{\xi = 3\}$.

Завдання 2. З генеральної сукупності, що має біноміальний розподіл із параметрами $n = 6$, $p = 0{,}25$, (B(6; 0,25)), було взято наступну вибірку об'ємом $N = 60$:

1	2	5	1	0	0	1	0	3	2
2	2	1	1	1	2	0	2	1	2
0	1	0	0	1	0	1	0	1	1
0	2	0	0	2	1	2	1	1	0
1	1	2	3	2	2	0	0	2	3
0	0	2	1	2	2	3	0	1	2

Використовуючи задану вибірку, оцініть p_A та p_B, тобто обчисліть наближені значення $p*_A$ та $p*_B$ ймовірності випадкових подій A і B.

Завдання 3. Використовуючи визначення функції розподілу $F_\xi(x)$ випадкової величини ξ та таблицю ймовірностей $Q(k, n, p)$, обчислити значення $F_\xi(2,2)$; $F_\xi(3,7)$ функції розподілу в точках: (x = 2,2); (x = 3,7).

Завдання 4. Використовуючи визначення емпіричної функції розподілу $\hat{F}_n(x)$ і задану вибірку, обчислити значення $\hat{F}_{60}(2{,}2)$ та $\hat{F}_{60}(3{,}7)$ емпіричної функції розподілу в точках: (x = 2,2); (x = 3,7).

Розв'язок:

Завдання 1. На основі визначення біноміального розподілу:

$$P\{A\} = P\{\xi = 2\} = P\{B(6;\, 0{,}25) = 2\} = C_6^2 \cdot 0{,}25^2 \cdot 0{,}75^4 .$$

Таблиця для біноміального розподілу містить ймовірності $Q(k, n, p)$, задані формулою:

$$Q(k, n, p) = P\{B(n, p) \geq k\}.$$

Враховуючи наступну рівність:

$$P\{B(6;\, 0{,}25) = 2\} = P\{B(6;\, 0{,}25) \geq 2\} - P\{B(6;\, 0{,}25) \geq 3\},$$

приходимо до висновку, що:

$$P\{A\} = P\{\xi = 2\} = Q(2, 6, 0{,}25) - Q(3, 6, 0{,}25).$$

З таблиці знаходимо:

$$Q(2, 6, 0{,}25) = 0{,}466;\ Q(3, 6, 0{,}25) = 0{,}169.$$

Тому:

$$P\{A\} = P\{\xi = 2\} = 0{,}466 - 0{,}169 = 0{,}297.$$

Аналогічно отримаємо:

$$P\{B\} = P\{\xi = 3\} = Q(3, 6, 0{,}25) - Q(4, 6, 0{,}25) = 0{,}169 - 0{,}037 = 0{,}132.$$

Відповідь: [$P\{A\}$ = 0,297; $P\{B\}$ = 0,132].

Завдання 2. На основі випадкової вибірки, представленої в розглянутому прикладі, побудуємо *точковий ряд*:

$x_1 = 0$	$x_2 = 1$	$x_3 = 2$	$x_4 = 3$	$x_5 = 4$	$x_6 = 5$	$x_7 = 6$
$n_1 = 18$	$n_2 = 19$	$n_3 = 18$	$n_4 = 4$	$n_5 = 0$	$n_6 = 1$	$n_7 = 0$

У випадку статистичної ознаки дискретного типу проблему непараметричного оцінювання вирішує *емпіричний розподіл*. Випадкова подія A безпосередньо пов'язана з розподілом ознаки ξ, тобто:

$$A = \{\xi = 2\} = \{\xi = x_3\}.\ P\{A\} = P\{\xi = x_3\} = p_3.$$

Тому наближенням для ймовірності p_A буде частота $\hat{p}_A$ появи випадкової події A у вибірці, розрахована за формулою:

$$\hat{p}_A = \frac{n_3}{n} = \frac{18}{60} = 0{,}3.$$

Розмірковуючи подібним чином, у випадку випадкової події *B* ми отримаємо:

$$\hat{p}_B = \frac{n_4}{n} = \frac{4}{60} = 0{,}06(6).$$

Відповідь: [$\hat{p}_A$ = 0,3; $\hat{p}_B$ = 0,06(6)].

Завдання 3. Виходячи з визначення функції розподілу $F_\xi(x)$ випадкової величини ξ отримуємо:

$$F_\xi(2{,}2) = P\{\xi < 2{,}2\} = 1 - P\{\xi \ge 2{,}2\}.$$

Оскільки ознака ξ може приймати значення тільки з множини

$$\xi \in \{0, 1, 2, 3, 4, 5, 6\},$$

то випадкові події $\{\xi \ge 2{,}2\}$ та $\{\xi \ge 3\}$ еквівалентні. Отже, використовуючи таблицю $Q(k, n, p)$ біноміального розподілу отримуємо:

$$F_\xi(2{,}2) = 1 - P\{\xi \ge 3\} = 1 - Q(3, 6, 0{,}25) = 1 - 0{,}169 = 0{,}831.$$

Розмірковуючи подібним чином, у випадку значення $x = 3{,}7$ отримаємо:

$$F_\xi(3{,}7) = P\{\xi < 3{,}7\} = 1 - P\{\xi \ge 3{,}7\} = 1 - P\{\xi \ge 4\} =$$
$$= 1 - Q(4, 6, 0{,}25) = 1 - 0{,}037 = 0{,}963.$$

Відповідь: [$F_\xi(2{,}2)$ = 0,831; $F_\xi(3{,}7)$ = 0,963].

Завдання 4. З урахуванням визначення емпіричної функції розподілу значення $\hat{F}_{60}(2{,}2)$ розраховується як відношення:

$$\hat{F}_N(2{,}2) = \hat{F}_{60}(2{,}2) = \frac{\nu_{60}(2{,}2)}{60},$$

де $N = 60$ – обсяг вибірки, $\nu_{60}(2{,}2)$ – кількість елементів цієї вибірки, що менші ніж 2,2. Використовуючи точковий ряд розподілу, побудований на основі випадкової вибірки, отримуємо:

$$\nu_{60}(2{,}2) = n_1 + n_2 + n_3 = 18 + 19 + 18 = 55.$$

$$\hat{F}_{60}(2{,}2) = \frac{55}{60} = 0{,}916(6).$$

Розмірковуючи подібним чином, у випадку значення $x = 3{,}7$ отримаємо:

$$\nu_{60}(3{,}7) = n_1 + n_2 + n_3 + n_4 = 18 + 19 + 18 + 4 = 59.$$

$$\hat{F}_{60}(3{,}7) = \frac{59}{60} = 0{,}983(3).$$

Відповідь: [$\hat{F}_{60}(2{,}2)$ = 0,916(6); $\hat{F}_{60}(3{,}7)$ = 0,983(3)].