

## Лекція 2. Статистичні показники вибірки

### 2.1. Дискретний статистичний (незгрупований) розподіл

Дискретним статистичним розподілом вибірки називається відповідність між варіантами  $x_i$  варіаційного ряду та їх частотами  $n_i$  (або відносними частотами  $\omega_i = \frac{n_i}{n}$ ).

Загальний вигляд дискретного статистичного ряду:

Варіанти, $x_i, i = \overline{1, m}$	$x_1$	$x_2$	$x_3$	...	$x_m$
Частоти, $n_i, i = \overline{1, m}$	$n_1$	$n_2$	$n_3$	...	$n_m$

$$\sum_{i=1}^m n_i = n$$

Відносна частота дискретного статистичного ряду:

Відносні частоти, $\omega_i, i = \overline{1, m}$	$\omega_1 = \frac{n_1}{n}$	$\omega_2 = \frac{n_2}{n}$	$\omega_3 = \frac{n_3}{n}$	...	$\omega_m = \frac{n_m}{n}$
--	----------------------------	----------------------------	----------------------------	-----	----------------------------

$$\sum_{i=1}^m \omega_i = 1$$

### 2.2. Інтервальний (згрупований) статистичний розподіл

Інтервальним статистичним розподілом вибірки називають перелік часткових інтервалів і відповідних їм частот або відносних частот.

Загальний вигляд інтервального варіаційного ряду:

Інтервали $(x_i), i = \overline{1, k}$	$(x_1; x_2)$	$(x_2; x_3)$	$(x_3; x_4)$	...	$(x_{k-1}; x_k)$
Частоти $n_i^*, i = \overline{1, k}$	$n_1^*$	$n_2^*$	$n_3^*$	...	$n_k^*$

Для визначення оптимальної кількості інтервалів застосовують або формулу *Стерджеса*  $k \approx 3,322 \cdot \lg(n) + 1$  або формулу  $k \approx 4 \cdot \lg(n) + 1$  та у деяких випадках при дуже малих вибірках  $k \approx \sqrt{n}$ , де  $n$  - кількість елементів сукупності.

Ширину рівновеликих інтервалів визначають за формулою:

$$h = \frac{x_{\max} - x_{\min}}{k},$$

де  $x_{\max}$  та  $x_{\min}$  - відповідно найбільше та найменше значення варіанту ряду у вибірці.

Частота  $n_i^*$  варіант, які потрапили в кожний частинний інтервал  $(x_i; x_{i+1})$ , записуються в другий рядок таблиці.

Для визначення відносної частоти інтервального ряду обчислюють щільність відносних частот.

Щільність відносних частот інтервального ряду можна обчислити використавши частоту варіант, тобто:

$$\frac{\omega_i^*}{h} = \frac{n_i^*}{nh}.$$

Відносна частота інтервального статистичного ряду:

Відносні частоти, $\frac{\omega_i^*}{h}$ , $i = \overline{1, k}$	$\frac{\omega_1^*}{h} = \frac{n_1^*}{n \cdot h}$	$\frac{\omega_2^*}{h} = \frac{n_2^*}{n \cdot h}$	$\frac{\omega_3^*}{h} = \frac{n_3^*}{n \cdot h}$	...	$\frac{\omega_k^*}{h} = \frac{n_k^*}{n \cdot h}$
---	--	--	--	-----	--

**Наприклад 2.2.** У результаті вибіркового аналізу добової виручки авіакомпанії дістали вибірку обсягом  $n = 40$  (млн.грн):

0,87 0,94 0,99 0,90 0,90 0,87 0,85 0,87  
 0,90 0,94 0,87 0,87 0,82 0,90 0,94 0,90  
 0,85 0,85 0,87 0,94 0,81 0,82 0,87 0,97  
 0,90 0,94 0,85 0,81 0,87 0,85 0,90 0,82  
 0,99 0,90 0,94 0,82 0,97 0,81 0,85 0,87

Скласти: а) варіаційний ряд; б) дискретний статистичний розподіл вибірки; в) інтервальний розподіл, якщо кількості інтервалів  $k = 6$ . Обчислити відносні частоти для дискретного ряду та щільність відносних частот для інтервального ряду.

*Розв'язання:* а) Випишемо різні значення варіант, які потрапили у вибірку:

0,87; 0,94; 0,99; 0,90; 0,85; 0,82; 0,81; 0,97.

Розмістивши їх у порядку зростання, дістанемо дискретний варіаційний ряд:

0,81; 0,82; 0,85; 0,87; 0,90; 0,94; 0,97; 0,99.

б) Обчислимо частоту кожної варіанти із варіаційного ряду і складемо таблицю – дістанемо дискретний статистичний розподіл вибірки:

Табл. 2.2. Дискретний варіаційний ряд

$x_i$	0,81	0,82	0,85	0,87	0,90	0,94	0,97	0,99
$n_i$	3	4	6	9	8	6	2	2

Обчислимо відносні частоти згідно формули  $\omega_i = \frac{n_i}{n}$  для дискретного розподілу:  $\omega_1 = \frac{n_1}{n} = \frac{3}{40} = 0,075$ ;  $\omega_2 = \frac{n_2}{n} = \frac{4}{40} = 0,1$ ;  $\omega_3 = \frac{n_3}{n} = \frac{6}{40} = 0,15$ ;  
 $\omega_4 = \frac{n_4}{n} = \frac{9}{40} = 0,225$ ;  $\omega_5 = \frac{n_5}{n} = \frac{8}{40} = 0,2$ ;  $\omega_6 = \frac{n_6}{n} = \frac{6}{40} = 0,15$ ;  
 $\omega_7 = \frac{n_7}{n} = \frac{2}{40} = 0,05$ ;  $\omega_8 = \frac{n_8}{n} = \frac{2}{40} = 0,05$ .

Табл. 2.3. Дискретний статистичний розподіл

$x_i$	0,81	0,82	0,85	0,87	0,90	0,94	0,97	0,99
$n_i$	3	4	6	9	8	6	2	2
$\omega_i$	0,075	0,1	0,15	0,225	0,2	0,15	0,05	0,05

в) За обсягом вибірки  $n = 40$  та заданому кількості частинних інтервалів в інтервальному статистичному розподілі  $k = 6$ , обчислюємо крок інтервалу:

$$h = \frac{0,99 - 0,81}{6} = 0,03.$$

Тобі інтервальний варіаційний ряд запишеться у вигляді:

$$(0,81;0,84), (0,84;0,87), (0,87;0,90), (0,90;0,93), (0,93;0,96), (0,96;0,99).$$

Підсумуємо частоти варіант, які потрапили в кожний із частинних інтервалів, при цьому частоти варіант, які збіглися з межами інтервалів (0,87 і 0,90), поділимо порівну між суміжними інтервалами. Тоді інтервальний статистичний розподіл вибірки набере вигляду:

Табл. 2.4. Інтервальний варіаційний ряд

інтервали	(0,81;0,84)	(0,84;0,87)	(0,87;0,90)	(0,90;0,93)	(0,93;0,96)	(0,96;0,99)
$n_i^*$	7	10	9	4	6	4

Як ми це зробили:

- на інтервальному ряду (0,81; 0,84) частоти варіант 3+4=7;
- на інтервальному ряду (0,84; 0,87) частоти варіант 6+4=10: (взяли з ряду 0,87 частину (4) статистичного розподілу, залишок - 5);
- на інтервальному ряду (0,87; 0,90) частоти варіант 5+4=9: (взяли з ряду 0,90 частину (4) статистичного розподілу, залишок - 4);
- на інтервальному ряду (0,90; 0,93) частоти варіант 4+0=4;
- на інтервальному ряду (0,93; 0,96) частоти варіант 6+0=6;
- на інтервальному ряду (0,96; 0,99) частоти варіант 2+2=4.

Щільності відносних частот  $\frac{n_i^*}{nh} = \frac{\omega_i^*}{h}$  для інтервального розподілу при

$n = 40$ ,  $h = 0,03$ :

$$\frac{\omega_1^*}{h} = \frac{n_1^*}{nh} = \frac{7}{40 \cdot 0,03} = \frac{7}{1,2} = 5,83; \quad \frac{\omega_2^*}{h} = \frac{n_2^*}{nh} = \frac{10}{40 \cdot 0,03} = \frac{10}{1,2} = 8,33; \quad \frac{\omega_3^*}{h} = \frac{9}{1,2} = 7,5;$$

$$\frac{\omega_4^*}{h} = \frac{4}{1,2} = 3,33; \quad \frac{\omega_5^*}{h} = \frac{6}{1,2} = 5; \quad \frac{\omega_6^*}{h} = \frac{4}{1,2} = 3,33.$$

Табл. 2.5. Інтервальний статистичний розподіл

інтервали	(0,81;0,84)	(0,84;0,87)	(0,87;0,90)	(0,90;0,93)	(0,93;0,96)	(0,96;0,99)
$n_i^*$	7	10	9	4	6	4
$\frac{\omega_i^*}{h}$	5,83	8,33	7,5	3,33	5	3,33

### 2.3. Графічні зображення статистичних розподілів

*Комулята* – графічне зображення варіаційного ряду з накопиченими частотами. При побудові комуляти за даними дискретного ряду по осі абсцис

відкладають значення ознаки (варіанти) ( $x_i$ ), а по осі ординат – накопичені частоти ( $N_m$ ). На перетині значень ознаки (варіант) і відповідним їм накопичених частот (частостей) будують точки, які також з'єднуються відрізками або кривою. Наприклад, підрахуємо накопичені частоти  $N_i$ , де  $i = \overline{1,8}$ , дані занесемо до таблиці:

$x_i$	0,81	0,82	0,85	0,87	0,90	0,94	0,97	0,99
$n_i$	3	4	6	9	8	6	2	2
$N_i$	3	7	13	22	30	36	38	40

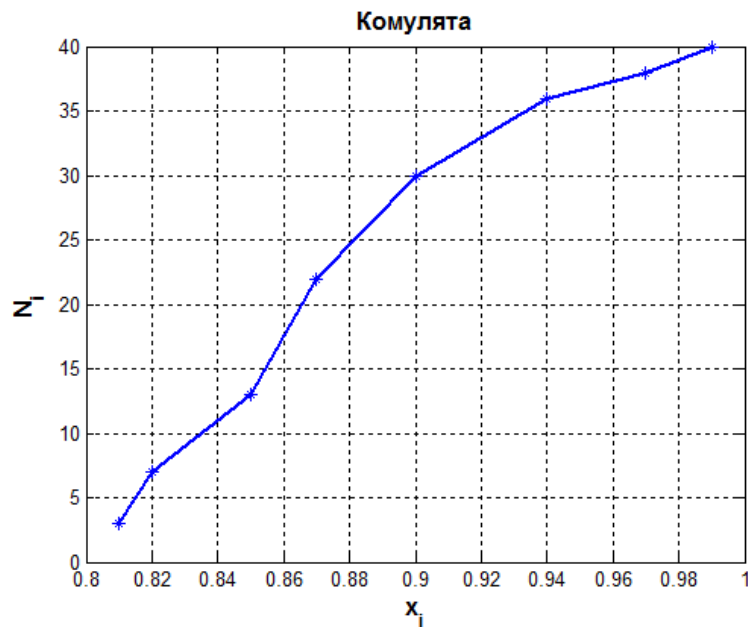


Рис. 2.2. Кумулята дискретного ряду розподілу

*Огіва* – будується аналогічно до кумуляти з тією тільки різницею, що на осі абсцис відкладаються точки, що відповідають накопиченим частотам ( $N_m$ ), а по осі ординат – значення ознаки ( $x_i$ ).

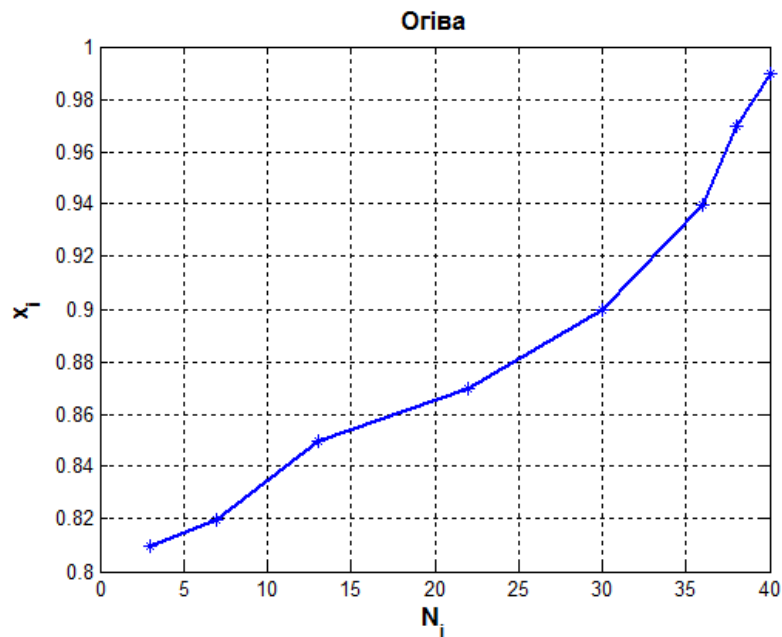


Рис. 2.2. Огіва дискретного ряду розподілу

*Крива Лоренца* –це спосіб графічного зображення рівня концентрації явища. Для її побудови на осі координат наносять процентну масштабну шкалу від 0 до 100. На осі абсцис відкладають накопичені частоти, а на осі ординат – накопичені відносні обсяги варіючої ознаки, вираженої в процентах.

Наприклад, проаналізуємо розподіл доходів у суспільстві за допомогою кривої Лоренца:

Група населення	Початково задана частка групи у %		Накопичені частки групи у %	
	У чисельності населення	У сукупності доході	У чисельності населення	У сукупності доході
Перша	10	2,5	10	2,5
Друга	20	7,5	30	10
Третя	20	10	50	20
Четверта	20	15	70	35
П'ята	20	20	90	55
Шоста	10	45	100	100
Разом	100	100	100	100

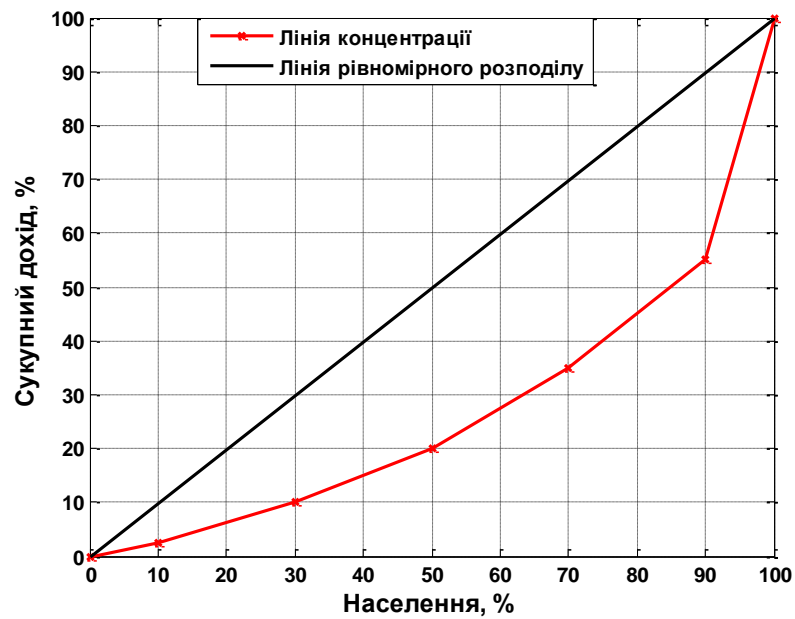


Рис. 2.3. Крива Лоренца

Полігоном частот (або відносних частот) називається ламана лінія, відрізки якої сполучають на площині точки з координатами  $(x_i, n_i)$  або  $(x_i, \omega_i)$ . Одержана ламана лінія є полігоном частот.

$x_i$	0,81	0,82	0,85	0,87	0,90	0,94	0,97	0,99
$n_i$	3	4	6	9	8	6	2	2
$\omega_i$	0,075	0,1	0,15	0,225	0,2	0,15	0,05	0,05

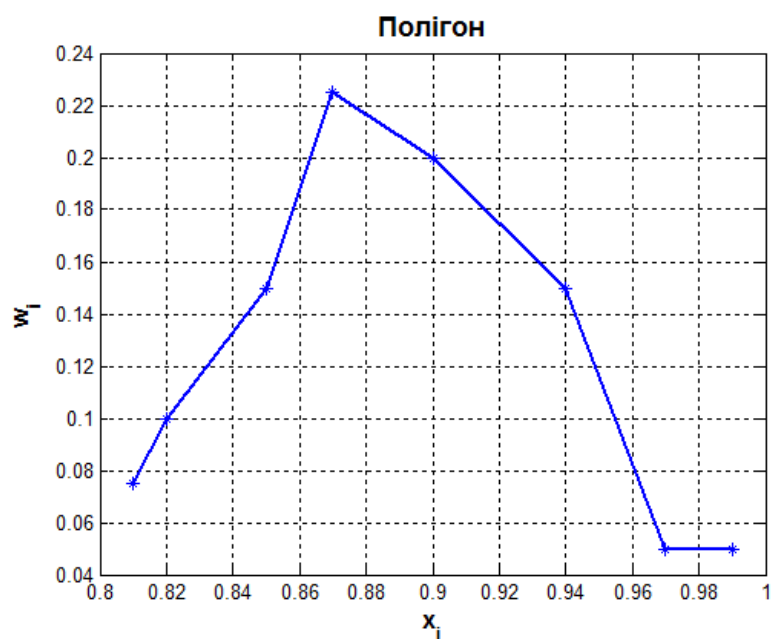


Рис. 2.4. Полігон відносних частот дискретного ряду

Гістограма застосовується для графічного зображення інтервальних статистичних розподілів. При її побудові по осі абсцис відкладаються значення досліджуваної ознаки (межі інтервалів), і на цих відрізках як на основах будуються прямокутники з висотами що дорівнюють щільностям частот (відносних частот) на відповідних інтервалах.

інтервали	(0,81;0,84)	(0,84;0,87)	(0,87;0,90)	(0,90;0,93)	(0,93;0,96)	(0,96;0,99)
$n_i^*$	7	10	9	4	6	4
$\frac{\omega_i^*}{h}$	5,83	8,33	7,5	3,33	5	3,33

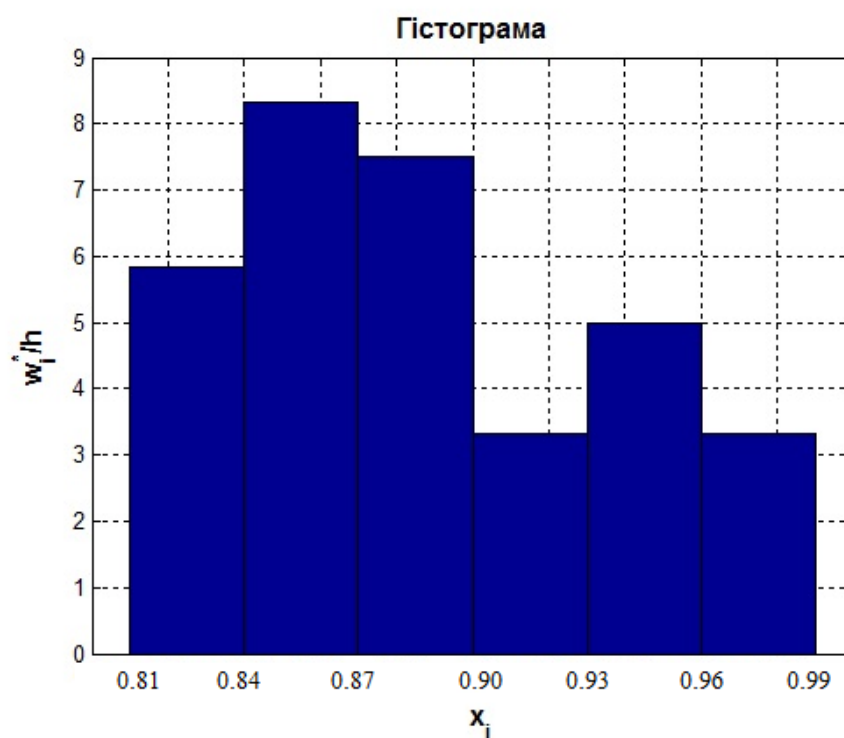


Рис. 2.5. Гістограма щільності відносних частот інтервального ряду

## 2.4. Емпірична функція розподілу

Емпіричною функцією статистичного розподілу називається функція  $F^*(x)$  яка для кожного значення  $x$  дорівнює відносній частоті події  $X < x$ , тобто:

$$F^*(x) = W(X < x) = \omega_x^{\text{нак}} = \frac{n_x^{\text{нак}}}{n},$$



де  $n_x^{нак}$  — накопичена частота всіх варіант  $x_i$ , менших за  $x$ ,  $n$  - обсяг вибірки.

Для дискретного статистичного розподілу емпірична функція  $F^*(x)$  набуває таких значень:

при  $x \leq x_1$   $F^*(x) = 0$  оскільки в цьому випадку немає варіант, менших за  $x$ ;

при  $x_1 < x \leq x_2$   $F^*(x) = \frac{n_1}{n}$ ;

при  $x_2 < x \leq x_3$   $F^*(x) = \frac{n_1 + n_2}{n}$ ;

.....

при  $x_{m-1} < x \leq x_m$   $F^*(x) = \frac{n_1 + n_2 + \dots + n_{m-1}}{n}$ ;

при  $x > x_m$   $F^*(x) = \frac{n_1 + n_2 + \dots + n_m}{n} = 1$ ;

Графік функції  $F^*(x)$  для дискретного ряду розподілу будується аналогічно графіку функції розподілу випадкової величини (розривний ступінчатий).

Для інтервального статистичного розподілу емпірична функція  $F^*(x)$  набуває таких значень:

при  $x = x_1$   $F^*(x) = 0$  оскільки варіанти, менші за  $x_1$  у вибірці інтервального ряду відсутні, де  $x_1$  - початок першого інтервалу;

при  $x = x_2$   $F^*(x) = \frac{n_1^*}{n}$ , оскільки варіанти, менші за  $x_2$ , містяться у першому інтервалі, і їх кількість  $n_1^*$ , де  $x_2$  - початок другого інтервалу;

при  $x = x_3$   $F^*(x) = \frac{n_1^* + n_2^*}{n}$  оскільки варіанти, менші за  $x_3$ , містяться у

першому та другому інтервалах, і їх кількість  $n_1^* + n_2^*$ , де  $x_3$  - початок третього інтервалу;

.....

при  $x = x_{k-1}$   $F^*(x) = \frac{n_1^* + n_2^* + \dots + n_{k-1}^*}{n}$ , де  $x_{k-1}$  - початок  $k$ -того інтервалу;

при  $x = x_k$   $F^*(x) = \frac{n_1^* + n_2^* + \dots + n_k^*}{n} = 1$ , де  $x_k$  - початок  $k+1$ -того інтервалу

тобто кінець  $k$ -того інтервалу.

Знайдені точки  $(x_i, F^*(x_i))$ ,  $i = \overline{1, k}$  наносять на графік і сполучають прямолінійними відрізками, тобто графіком функції буде комулята що зображується пунктиром.

**Приклад 2.3.** Побудувати емпіричну функцію розподілу  $F^*(x)$  за дискретним (табл. 2.4) та інтервальним (табл. 2.5) статистичними розподілами вибірки з прикладу 2.2.

*Розв'язання.* Для обчислення значень емпіричної функції скористаємось дискретним статистичним розподілом із відносними частотами  $\omega_i$  (табл. 2.4), який знайдено у прикладі 2.2.

$x_i$	0,81	0,82	0,85	0,87	0,90	0,94	0,97	0,99
$n_i$	3	4	6	9	8	6	2	2
$\omega_i$	0,075	0,1	0,15	0,225	0,2	0,15	0,05	0,05

Значення функції  $F^*(x)$  обчислюються простим нагромадженням відносних частот:

$$F^*(x) = \begin{cases} 0 & \text{при } x \leq 0,81; \\ 0 + 0,075 = 0,075 & \text{при } 0,81 < x \leq 0,82; \\ 0,075 + 0,1 = 0,175 & \text{при } 0,82 < x \leq 0,85; \\ 0,175 + 0,15 = 0,325 & \text{при } 0,85 < x \leq 0,87; \\ 0,325 + 0,225 = 0,55 & \text{при } 0,87 < x \leq 0,90; \\ 0,55 + 0,2 = 0,75 & \text{при } 0,90 < x \leq 0,94; \\ 0,75 + 0,15 = 0,9 & \text{при } 0,94 < x \leq 0,97; \\ 0,9 + 0,05 = 0,95 & \text{при } 0,97 < x \leq 0,99; \\ 0,95 + 0,05 = 1 & \text{при } x > 0,99. \end{cases}$$

Графік функції  $F^*(x)$  подано на рис. 2.6 (східчаста розривна лінія).

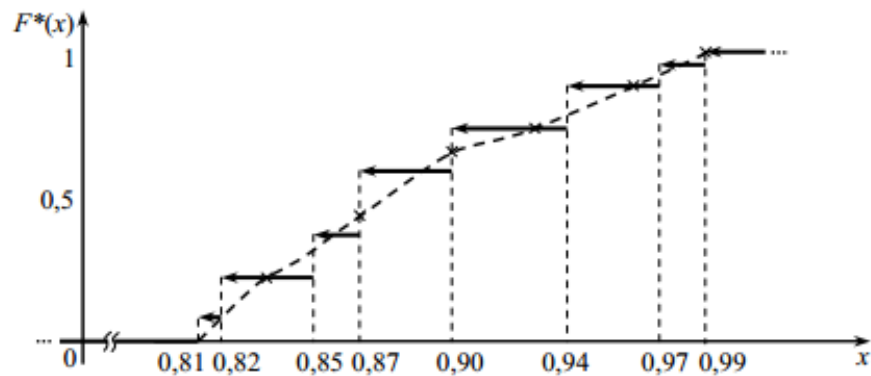


Рис. 2.6. Емпіричні функції розподілів

Для побудови  $F^*(x)$  за інтервальним статистичним розподілом (табл. 2.5) обчислюємо значення функції на межах інтервалів:

інтервали	(0,81;0,84)	(0,84;0,87)	(0,87;0,90)	(0,90;0,93)	(0,93;0,96)	(0,96;0,99)
$n_i^*$	7	10	9	4	6	4
$\frac{\omega_i^*}{h}$	5,83	8,33	7,5	3,33	5	3,33

$x = 0,81$ ,  $F^*(x) = 0$ , оскільки варіант, менших за 0,81, у вибірці немає;

$x = 0,84$ ,  $F^*(x) = \frac{7}{40} = 0,175$ , оскільки варіант, менших за 0,84,

містяться в першому інтервалі, їх кількість 7;

$$x = 0,87, \quad F^*(x) = \frac{1}{40}(7+10) = 0,425;$$

$$x = 0,90, \quad F^*(x) = \frac{1}{40}(7+10+9) = 0,65;$$

$$x = 0,93, \quad F^*(x) = \frac{1}{40}(7+10+9+4) = 0,75;$$

$$x = 0,96, \quad F^*(x) = \frac{1}{40}(7+10+9+4+6) = 0,9;$$

$$x = 0,99, \quad F^*(x) = \frac{1}{40}(7+10+9+4+6+4) = 1.$$

Нанесемо знайдені точки на графік рис. 2.6 (позначені \*) і сполучимо їх відрізками прямої. Дістанемо графік функції  $F^*(x)$ , зображений на рис. 2.6 пунктирною лінією.