Замечание: если для посл-ти с.в. вып-ся усл. вн Чеб, то (эта посл. удовл. з-ну ЗБЧ) говорят, что эта посл-ть удовл. з-ну б.ч. в форме Чебышева.

В нашем примере посл-ть $X_1, X_2, \dots$ не удовл. ЗБЧ в форме Чеб, но при этом, возможно, удовл. ЗБЧ.

5) Проверим, выполняется ли для посл. $X_1, X_2, \dots$ определение ЗБЧ

$$\forall \varepsilon > 0 \quad P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} m_i\right| > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0$$

29.02.24. Семинар 4.

Математическая статистика.

Предварительная обработка результатов эксперимента.

Одну из задач ("больших") мат. стат. можно сформ-ть так:

X – с.в., з-н расп. которой неизвестен. Требуется по результатам наблюдений (по имеющимся реализациям) с.в. X сделать выводы о ее з-не расп-я.

~~Прим. Допустим~~

Опр. Случайной выборкой (объема n) из генеральной сов-ти X наз. с.в.

$$\vec{X} = (X_1, X_2, \dots, X_n),$$

где

1) с.в. $X_1, X_2, \dots, X_n$ незав. в совокупности
2) $X_i \sim X$, $i = \overline{1;n}$

Опр. Выборкой объема n из ген. сов-ти X наз (любую) реализацию случайной выборки $\vec{X}$ объема n из этой ген. сов-ти.

$$\vec{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \quad \text{(обычный числ. в-р)}$$

$\boxed{N}$ Задана выборка $\vec{x} = (13.8,\ 11.9,\ 15.5,\ 12.5,\ 14.7,\ 13.4,\ 13.7,\ 14.7,\ 13.7,\ 13.4,\ 12.3,\ 13.0,\ 9.7,\ 16.2,\ 13.4)$

Для этой выборки:
1) построить вариационный ряд,
2) —"— статистический ряд,
3) —"— интерв. стат. ряд,
4) —"— гистограмму
5) найти выборочное среднее, выб. дисперсию и испр. выб. дисперсию.

Решение:

1) Вариационный ряд

Пусть $\vec{x} = (x_1, \dots, x_n)$ – выборка из ген. совок-ти X.
Расположим значения x_i, $i=\overline{1,n}$, в порядке неубывания:

$$x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$$

Вектор $(x_{(1)}, \dots, x_{(n)})$ наз. вариационным рядом выборки $\vec{x}$.

Вариационный ряд:

(9.7, 11.9, 12.3, 12.5, 13.0, 13.4, 13.4, 13.7, 13.7, 13.7, 13.8, 14.7, 14.7, 15.5, 16.2)

2) Статистический ряд:

Опр. Стат. рядом вектора $\vec{x} = (x_1, \dots, x_n)$ наз. таблица вида

z_1	z_2	...	z_m
n_1	n_2	...	n_m

Здесь z_i, $i=\overline{1,m}$, – все попарно различные значения компонент x_j, $j=\overline{1,n}$, в-ра $\vec{x}$;
n_i – количество компонент вектора $\vec{x}$, которые равны z_i.

9.7	11.9	12.3	12.5	13.0	13.4	13.7	13.8	14.7	15.5	16.2
1	1	1	1	1	2	3	1	2	1	1

3) Интервальный стат. ряд.

Если число возм. значений эл-в выборки велико ($m \gtrsim 50$), то эти значения часто группируют в интервалы.
Число интервалов определяют по формуле

$$k = [\log_2 n] + 2$$

Интервалы принимают равновеликими, поэтому длина каждого интервала

$$\Delta = \frac{x_{(n)} - x_{(1)}}{k} \to \text{размах выборки}$$

Тогда

$$J_i = [x_{(1)} + \Delta\cdot(i-1),\ x_{(1)} + i\cdot\Delta),\quad i=\overline{1,k-1}$$

$$J_k = [x_{(1)} + \Delta\cdot(k-1),\ x_{(n)}]$$

Опр. Интервальным стат. рядом выборки $\vec{x}$ наз. таблица вида

J_1	...	J_i	...	J_k
l_1	...	l_i	...	l_k

Здесь l_i, $i=\overline{1,k}$, – число компонент выборки $\vec{x}$, кот. попали в J_i

$x_{(1)} = 9.7 \quad x_{(n)} = 16.2.$

$k = [\log_2 n] + 2 = [\log_2 15] + 2 = 5$

$\Delta = \frac{x_{(n)} - x_{(1)}}{k} = \frac{16.2 - 9.7}{5} = \frac{6.5}{5} = 1.3$

[9.7, 11.0)	[11.0, 12.3)	[12.3, 13.6)	[13.6, 14.9)	[14.9, 16.2]
1	1	5	6	2

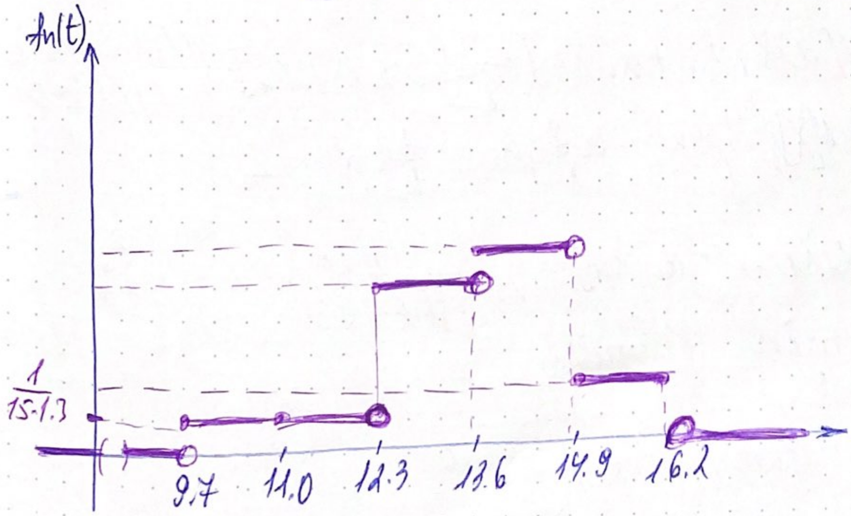
4) Гистограмма

По опр. <u>эмпирической плотностью</u>, отв. выборке $\vec{x}$ наз. функцию

$$f_n(t) = \begin{cases} \frac{\nu_i}{n\Delta}, & \text{если } t \in J_i \\ 0 & \text{иначе} \end{cases}$$

Опр. График эмпирической ф-ии пл-ти наз. <u>гистограммой</u>.

$$f_n(t) = \begin{cases} 0, & t \notin [9.7, 16.2] \\ \frac{1}{15 \cdot 1.3}, & t \in J_1 \\ \frac{1}{15 \cdot 1.3}, & t \in J_2 \\ \frac{5}{15 \cdot 1.3}, & t \in J_3 \\ \frac{6}{15 \cdot 1.3}, & t \in J_4 \\ \frac{2}{15 \cdot 1.3}, & t \in J_5 \end{cases}$$



5) выборочное среднее:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{15}(9.7+\dots+16.2) \approx 13.48$$

выборочная дисперсия:

$$\hat{\sigma}^2(\vec{x}) = \cancel{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \approx 2.27$$

исправленная выборочная дисперсия:

$$s^2(\vec{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{n}{n-1}\cdot\hat{\sigma}^2(\vec{x}) \approx 2.43$$

Замечание: выборка из пред. примера смоделирована из $N(13, 2.5)$

$$\begin{cases} \bar{x} = 13.48 \\ s^2(\vec{x}) = 2.43 \end{cases}$$

07.03.24. Семинар 5. (замена)

① $\vec{X} = (X_1, \dots, X_n)$ ген. выборка

Док-ть, что из линейных оценок величины $a = M[X]$ вида

$$\hat{a}(X) = \alpha_1 X_1 + \dots + \alpha_n X_n$$

лучшей является оценка $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$

Док-во:

- $M[\hat{a}(X)] = M[\alpha_1 X_1 + \dots + \alpha_n X_n] = \sum_{k=1}^{n}\alpha_k M[X_k] = \left\{ \begin{array}{l} M[X_i] = M[X_j] = \\ = M[X] = a \end{array} \right\} =$
$= M[X]\cdot\sum_{k=1}^{n}\alpha_k = M[X]\sum_{k=1}^{n}\alpha_k = a\sum_{k=1}^{n}\alpha_k \Rightarrow \sum_{k=1}^{n}\alpha_k = 1$

- $D[\hat{a}(X)] = D[\alpha_1 X_1 + \dots + \alpha_n X_n] = \left\{ \begin{array}{l} X_i \text{ – незав.,} \\ \text{расп. по} \\ \text{одному з.} \end{array} \right\} = \sigma^2 \sum_{k=1}^{n}\alpha_k^2$

$$\sigma^2 \sum_{k=1}^{n}\alpha_k^2 \xrightarrow[\sum_{k=1}^{n}\alpha_k = 1]{} \min \quad (\text{т.е. лучшая})$$

Необх. условие условного экстремума:

$$L = \sum_{k=1}^{n}\alpha_k^2 + \lambda\left(\sum_{k=1}^{n}\alpha_k - 1\right)$$