

Оглавление

Занятие 1. Выборочные характеристики	1
Контрольные вопросы и задания	3
Аудиторные задачи	4
Домашнее задание	9
Занятие 2. Свойства оценок	12
Контрольные вопросы и задания	13
Аудиторные задачи	14
Домашнее задание	14

Занятие 1. Выборочные характеристики

Контрольные вопросы и задания

Приведите определение выборки, вариационного ряда, статистики, порядковой статистики, эмпирической функции распределения.

x_1, \dots, x_n — наблюдаемые значения — независимые одинаково распределённые случайные величины с неизвестной функцией распределения $F(x)$.

Такой набор случайных величин называется выборкой из распределения F .

Вариационный ряд — последовательность $x_{(1)}, \dots, x_{(n)}$, полученная в результате расположения в порядке неубывания исходной последовательности независимых одинаково распределённых случайных величин x_1, \dots, x_n .

Статистикой называют функцию S от выборки $X = (x_1, x_2, \dots, x_n)$ такую, что $S(X) = S(x_1, x_2, \dots, x_n)$.

Вариационный ряд и его члены являются порядковыми статистиками.

Эмпирической (выборочной) функцией распределения, построенной по выборке x_1, \dots, x_n называется функция

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{x_k \leq x}, \quad x \in \mathbb{R}.$$

Какими свойствами обладает эмпирическая функция распределения?

Есть множество полной вероятности, на котором эмпирическая функция распределения аппроксимирует функцию распределения, то есть почти наверное $F_n \Rightarrow F, n \rightarrow \infty$.

Запишите выражения для выборочного среднего, выборочной дисперсии, выборочных моментов.

$$\frac{1}{n} \sum_{k=1}^n x_k$$

— выборочное среднее.

Выборочная дисперсия

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2.$$

Выборочные моменты в математической статистике — это оценка теоретических моментов распределения на основе выборки.

Выборочный момент порядка k — это случайная величина

$$a_n(k) = \frac{1}{n} \sum_{i=1}^n x_i^k.$$

Аудиторные задачи

1.4

Задание. Пусть X_1, \dots, X_n — выборка из равномерного распределения на отрезке $[0, \theta]$ с неизвестным параметром θ . Какие из приведённых ниже функций являются статистиками?

- a) \bar{X} ;
- b) $5X_{(n)}$;
- c) $\theta/2$;
- d) X_1/θ ;
- e) $X_{(1)} + X_1 + X_n$.

Решение.

- a) Да;
- b) да;
- c) нет, так как не функция от выборки;
- d) функция не только от выборки (зависит от неизвестного параметра). Отсюда следует, что это не статистика;
- e) да.

1.5

Задание. Пусть X_1, \dots, X_n — выборка из распределения Пуассона с параметром λ . Вычислите математическое ожидание и дисперсию статистики

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Выясните, имеет ли статистика \bar{X} распределение Пуассона.

Решение. Все X_i одинаково распределены. Отсюда следует, что все математические ожидания одинаковы

$$M\bar{X} = \frac{1}{n} \sum_{i=1}^n MX_i = \frac{1}{n} \cdot nMX_1 = MX_1 = \lambda.$$

Для всякой выборки справедливо $M\bar{X} = MX_1$.

Из независимости X_i получаем

$$D\bar{X} = \frac{1}{n^2} \sum_{i=1}^n DX_i.$$

Так как X_i одинаково распределены, то все дисперсии одинаковы

$$D\bar{X} = \frac{DX_1}{n} = \frac{\lambda}{n}.$$

Математическое ожидание и дисперсия для распределения Пуассона совпадают. Отсюда следует, что эта случайная величина не имеет распределения Пуассона.

\bar{X} не обязательно будет принимать целые значения.

1.6

Задание. Вычислите математическое ожидание статистик:

a) $S^2 = \overline{X^2} - (\bar{X})^2$;

b) $S_0^2 = 1/(n-1) \cdot \sum_{i=1}^n (X_i - \bar{X})^2$.

Решение.

a) Распишем каждую из величин

$$S^2 = \overline{X^2} - (\bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2.$$

Распишем квадрат

$$\frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n^2} \left(\sum_{i=1}^n X_i^2 + 2 \sum_{i,j=1, i < j}^n X_i X_j \right).$$

Берём слева и справа математическое ожидание. Из того, что случайные величины в выборке одинаково распределены

$$MS^2 = MX_1^2 - \frac{1}{n^2} [nMX_1 + 2C_n^2 (MX_1)^2].$$

Подставляем C_n^2 и группируем

$$MX_1^2 - \frac{1}{n^2} [nMX_1 + 2C_n^2 (MX_1)^2] = \frac{n-1}{n} \cdot MX_1^2 - \frac{n-1}{n} (MX_1)^2.$$

Вынесем общий множитель за скобки

$$\frac{n-1}{n} \cdot MX_1^2 - \frac{n-1}{n} (MX_1)^2 = \frac{n-1}{n} [MX_1^2 - (MX_1)^2] = \frac{n-1}{n} \cdot SX_1.$$

Эта оценка смещена асимптотически;

b) выразим S_0 через S . Раскроем квадрат

$$S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2).$$

Имеем сумму n одинаковых слагаемых

$$S_0^2 = \frac{1}{n-1} (n\bar{X}^2 - 2(\bar{X})^2 n + n\bar{X}^2) = \frac{n}{n-1} [\bar{X}^2 - (\bar{X})^2] = \frac{n-1}{n} \cdot S^2.$$

Отсюда следует, что

$$MS_0^2 = \frac{n}{n-1} \cdot \frac{n-1}{n} \cdot DX_1 = DX_1.$$

1.7

Задание. Найдите в терминах функции распределения F выборки X_1, \dots, X_n :

- a) распределение k -ой порядковой статистики $X_{(k)}$;
- b) вероятность $P(X_{(k)} < y, X_{(k+1)} \geq y)$.

Решение.

- a) Сделали упорядочивание случайных величин

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(k)} \leq \dots \leq X_{(n)}.$$

По определению $F_{X_{(k)}}(y) = P(X_{(k)} \leq y) = P\{\text{хотя бы } k \text{ элементов выборки не превышает } y\} =$

$$= \sum_{i=k}^n P(A_i),$$

где $A_i = \{\text{ровно } i \text{ элементов выборки не превышают } y\}$.

Есть n испытаний, успех — $X_i \leq y$.

Вероятность успеха — это $F(y)$, вероятность неудачи — это $[1 - F(y)]$.

Это биномиальное распределение

$$F_{X_{(k)}} = \sum_{i=k}^n C_n^i F^i(y) [1 - F(y)]^{n-i};$$

б) согласно с предыдущим пунктом $P(X_{(k)} < y, X_{(k+1)} \geq y) = P\{\text{ровно } i \text{ элементов выборки не превышает } y\} = C_n^k F^k(y) [1 - F(y)]^{n-k}$.

1.8

Задание. Пусть $(-0.8; 2.9; 4.5; -5.7; 1.1; -3.2)$ — наблюдаемые значения выборки. Составьте вариационный ряд, постройте эмпирическую функцию распределения $F_6(x)$ и её график. Вычислите выборочное среднее и выборочную дисперсию.

Решение. Вариационный ряд: $(-5.7; -3.2; -0.8; 1.1; 2.9; 4.3)$.

Эмпирическая функция распределения (рис. 1).

$$F_6(y) = \frac{1}{6} \sum_{i=1}^6 \mathbb{1}\{x_i \leq y\} = \begin{cases} 0, & x < -5.7; \\ \frac{1}{6}, & -5.7 \leq x < -3.2, \\ \frac{2}{6}, & -3.2 \leq x < -0.8, \\ \frac{3}{6}, & -0.8 \leq x < 1.1, \\ \frac{4}{6}, & 1.1 \leq x < 2.9, \\ \frac{5}{6}, & 2.9 \leq x < 4.3, \\ 1, & x \geq 4.3. \end{cases}$$

Выборочное среднее

$$\bar{X} = \frac{1}{6} (-5.7 - 3.2 - 0.8 + 1.1 + 2.9 + 4.3) = \frac{1}{6} (-9.7 + 8.3) = -\frac{1}{6} \cdot 1.4 = -0.23.$$

Выборочная дисперсия

$$\hat{\sigma}^2 = \frac{1}{5} \sum_{i=1}^6 (X_i + 0.23)^2.$$

Она является несмещённой.

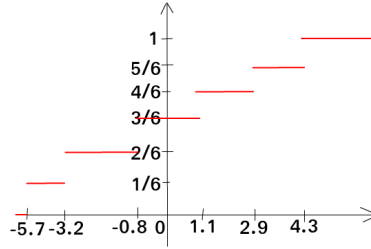


Рис. 1: Эмпирическая функция распределения

1.9

Задание. Вычислите вероятность $P(F_n(y) < F_n(z))$.

Решение.

а) $y \geq z$. Событие невозможное, потому что $F_n(y) \geq F_n(z)$;

б) рассмотрим случай, когда $y < z$.

Тогда искомая вероятность равна $P(F_n(y) < F_n(z)) = P\{\text{в } (y, z) \text{ попал хотя бы 1 элемент выборки}\} = 1 - P\{\text{в } (y, z) \text{ ни один элемент выборки не попал}\}$. Случайные величины одинаково распределены, поэтому $1 - P\{\text{в } (y, z) \text{ ни один элемент выборки не попал}\} =$
 $= [1 - P\{x_i \notin (y, z)\}]^n = 1 - [1 - P\{x_1 \in (y, z)\}]^n =$
 $= 1 - [1 - P(x_1 < z) + P(x_1 < y)]^n = 1 - [1 - F(z) + F(y)]^n.$

1.10

Задание. Пусть X_1, \dots, X_n — выборка из распределения F с плотностью f . Найдите совместную плотность распределения всех порядковых статистик, то есть плотность распределения случайного вектора $(X_{(1)}, \dots, X_{(n)})$.

Решение. $F_{(X_{(1)}, X_{(2)})}(y_1, y_2) = P(X_{(1)} \leq y_1, X_{(2)} \leq y_2)$. Воспользуемся формулой $P(A \cap B) = P(B) - P(\bar{A} \cap B)$. Получим

$$P(X_{(1)} \leq y_1, X_{(2)} \leq y_2) = P(X_{(1)} \leq y_2) - P(X_{(1)} > y_1, X_{(2)} \leq y_2).$$

Среди $X_{(1)}$ и $X_{(2)}$ случайная величина $X_{(2)}$ является максимальной.

$$\begin{aligned} & P(X_{(1)} \leq y_2) - P(X_{(1)} > y_1, X_{(2)} \leq y_2) = \\ & = P(X_1 \leq y_2, X_2 \leq y_2) - P(X_1 \in (y_1, y_2], X_2 \in (y_1, y_2]). \end{aligned}$$

Случайные величины X_1, X_2 — независимые и одинаково распределённые

$$\begin{aligned} & P(X_1 \leq y_2, X_2 \leq y_2) - P(X_1 \in (y_1, y_2], X_2 \in (y_1, y_2]) = \\ & = \begin{cases} [F(y_2)]^2, & y_1 \geq y_2, \\ [F(y_2)]^2 - [F(y_2) - F(y_1)]^2, & y_1 < y_2. \end{cases} \end{aligned}$$

Продифференцируем

$$f_{(X_{(1)}, X_{(2)})}(y_1, y_2) = \begin{cases} 0, & y_1 \geq y_2, \\ 2f(y_1)f(y_2), & y_1 < y_2. \end{cases}$$

Рассматриваем множество всех векторов, которые имеют упорядоченные координаты $\Delta = \{\vec{x} \in \mathbb{R}^n : z_1 < z_2 < \dots < z_n\}$, $\Gamma \subseteq \Delta$ — произвольное подмножество.

$$(X_{(1)}, \dots, X_{(n)}) \in \Delta.$$

Чтобы найти вероятность того, что данный вектор принадлежит Γ , должны проинтегрировать плотность этого вектора по этому множеству

$$P\{(X_{(1)}, \dots, X_{(n)}) \in \Gamma\} = \int_{\Gamma} f_{(X_{(1)}, \dots, X_{(n)})}(z_1, \dots, z_n) dz_1 \dots dz_n.$$

С другой стороны,

$$P\{(X_{(1)}, \dots, X_{(n)}) \in \Gamma\} = \sum_{\sigma \in S_n} P\{(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in \Gamma\}.$$

Учтём все перестановки

$$\sum_{\sigma \in S_n} P\{(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in \Gamma\} = n! P\{(X_1, \dots, X_n) \in \Gamma\}.$$

Подставим найденное выражение для вероятности

$$n! P\{(X_1, \dots, X_n) \in \Gamma\} = n! \cdot \int_{\Gamma} f(z_1) \cdot \dots \cdot f(z_n) dz_1 \dots dz_n.$$

Сравниваем полученные выражения

$$f_{(X_{(1)}, \dots, X_{(n)})}(z_1, \dots, z_n) = n! f(z_1) \cdot \dots \cdot f(z_n) \cdot \mathbb{1}\{z_1 < z_2 < \dots < z_n\}$$

— плотность вектора упорядоченных статистик.

Домашнее задание

1.15

Задание. Пусть X_1, \dots, X_n — выборка из равномерного на отрезке $[a, b]$ распределения. Вычислите математическое ожидание и дисперсию статистики

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Выясните, имеет ли статистика \bar{X} равномерное распределение; нормальное распределение.

Решение. Все X_i одинаково распределены. Отсюда следует, что все математические ожидания одинаковы

$$M\bar{X} = \frac{1}{n} \sum_{i=1}^n MX_i = \frac{1}{n} \cdot nMX_1 = MX_1 = \frac{a+b}{2}.$$

Из независимости X_i получаем

$$D\bar{X} = \frac{1}{n^2} \sum_{i=1}^n DX_i.$$

Так как X_i одинаково распределены, то все дисперсии одинаковы

$$D\bar{X} = \frac{DX_1}{n} = \frac{(b-a)^2}{12n}.$$

Чтобы выяснить, распределена ли статистика \bar{X} по нормальному или равномерному распределению, найдём её характеристическую функцию $\varphi_{\bar{X}}(t)$. Учитывая независимость элементов выборки и то, что

$$\varphi_{X_1}(t) = \dots = \varphi_{X_n}(t) = \frac{e^{itb} - e^{ita}}{it(b-a)},$$

находим

$$\varphi_{\bar{X}}(t) = \varphi_{X_1}\left(\frac{t}{n}\right) \cdot \dots \cdot \varphi_{X_n}\left(\frac{t}{n}\right) = \left[\frac{(e^{itb} - e^{ita})}{it(b-a)} \right]^n.$$

Отсюда следует, что \bar{X} не имеет указанных распределений.

1.16

Задание. Пусть X_1, \dots, X_n — выборка из некоторого распределения вероятностей, функция распределения которого F является непрерывной и строго возрастающей. Найдите распределение выборки Y_1, \dots, Y_n , где

$$Y_i = F(X_i).$$

Решение. По определению

$$F_{\eta_1, \dots, \eta_n}(X_1, \dots, X_n) = P(\eta_1 \leq X_1, \dots, \eta_n \leq X_n).$$

Воспользуемся независимостью

$$P(\eta_1 \leq X_1, \dots, \eta_n \leq X_n) = P(\eta_1 \leq X_1) \cdot \dots \cdot P(\eta_n \leq X_n).$$

Функция распределения i -й компоненты вектора равна

$$F_{\eta_i}(x) = P(F_{\xi_i}(X_i) \leq x) = \begin{cases} 0, & x \leq 0, \\ 1, & x > 1. \end{cases}$$

Рассмотрим $[0, 1]$.

Поскольку F — непрерывная и строго возрастающая, то существует $F^{-1}(x)$. Обозначим через z точку $F^{-1}(x)$ такую, что $F(z) = x$. Событие $\{\eta = F(\xi) < x\}$ происходит тогда и только тогда, когда происходит событие $\{\xi < z\}$.

Получаем на отрезке $[0, 1]$ равномерное распределение

$$F_{\eta}(x) = F_{\xi}(z) = F_{\xi}(F_{\xi}^{-1}(x)) = \begin{cases} 0, & x \leq 0, \\ x, & x \in (0, 1], \\ 1, & x > 1. \end{cases}$$

1.17

Задание. Пусть X_1, \dots, X_n — выборка из дискретного распределения с вероятностями $P(X_1 = m) = p_m$, где

$$\sum_{m=0}^N p_m = 1.$$

Найдите распределение k -й порядковой статистики $X_{(k)}$.

Решение. Сделали упорядочивание случайных величин

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(k)} \leq \dots \leq X_{(n)}.$$

По определению $F_{X_{(k)}}(y) = P(X_{(k)} \leq y) = P\{\text{хотя бы } k \text{ элементов выборки не превышает } y\} =$

$$= \sum_{i=k}^n P(A_i),$$

где $A_i = \{\text{ровно } i \text{ элементов выборки не превышают } y\}$.

Есть n испытаний, успех — $X_i \leq y$.

Вероятность успеха — это $F(y)$, вероятность неудачи — это $[1 - F(y)]$. Это биномиальное распределение

$$F_{X_{(k)}} = \sum_{i=k}^n C_n^i F^i(y) [1 - F(y)]^{n-i}.$$

Представим $F_{X_i}(y) = F(y)$ через m . Запишем по определению

$$F_{X_1}(y) = P(X_1 \leq y) = \sum_{m=1}^n P(X_1 = m) = \sum_{m=1}^n p_m.$$

Подставим полученное выражение в функцию распределения

$$F_{X_{(k)}} = \sum_{i=k}^n C_n^i \sum_{m=1}^n p_m \left(1 - \sum_{m=1}^n p_m\right)^{n-i}.$$

1.18

Задание. Пусть $(3, 0, 4, 3, 6, 0, 3, 1)$ — наблюдаемые значения выборки. Составьте вариационный ряд, постройте эмпирическую функцию распределения $F_8(x)$ и её график. Вычислите выборочное среднее и выборочную дисперсию.

Решение. Вариационный ряд: $(0, 0, 1, 3, 3, 3, 4, 6)$.

Эмпирическая функция распределение (рис. 2)

$$F_8(y) = \frac{1}{8} \sum_{i=1}^8 \mathbb{1}\{x_i \leq y\} = \begin{cases} 0, & x < 0, \\ \frac{2}{8} = \frac{1}{4}, & 0 \leq x < 1, \\ \frac{3}{8}, & 1 \leq x < 3, \\ \frac{6}{8} = \frac{3}{4}, & 3 \leq x < 4, \\ \frac{7}{8}, & 4 \leq x < 6, \\ 1, & x \geq 6. \end{cases}$$

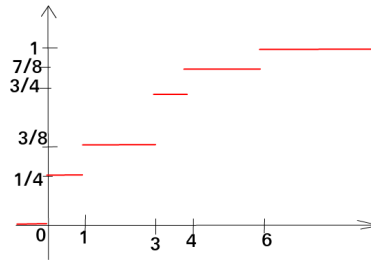


Рис. 2: Эмпирическая функция распределения

Выборочное среднее

$$\bar{X} = \frac{1}{8} (0 + 0 + 1 + 3 + 3 + 3 + 4 + 6) = \frac{1}{8} \cdot 20 = \frac{10}{4} = 2.5.$$

Выборочная дисперсия

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{7} \sum_{i=1}^8 (X_i - 2.5)^2 = \\ &= \frac{1}{7} \left[2(0 - 2.5)^2 + (1 - 2.5)^2 + 3(3 - 2.5)^2 + (4 - 2.5)^2 + (6 - 2.5)^2 \right] = \\ &= \frac{1}{7} (12.5 + 2.25 + 0.75 + 2.25 + 12.25) = \frac{30}{7} \approx 4.29. \end{aligned}$$

Занятие 2. Свойства оценок

Контрольные вопросы и задания

Что называют оценкой неизвестного параметра?

Статистику, значение которой заменяет неизвестный параметр, называют оценкой этого параметра.

Преведиты определение оценки: несмещённой, асимптотически несмещённой, состоятельной, сильно состоятельной, оптимальной.

Оценка $\hat{\theta}$ несмещённая, если $\forall \theta \in \Theta : M_{\theta} \hat{\theta} = \theta$.

Асимптотически несмещённая оценка — такая оценка, математическое ожидание которой совпадает с оцениваемым параметром при $n \rightarrow \infty$.

Оценка $\hat{\theta}$ называется состоятельной, если стремится к истинному значению θ по вероятности $\hat{\theta} \xrightarrow{P} \theta, n \rightarrow \infty$.

Оценка $\hat{\theta}$ называется сильно состоятельной, если стремится к истинному значению θ почти наверное $\hat{\theta} \xrightarrow{a.s.} \theta, n \rightarrow \infty$.

Несмещённая оценка $\hat{\theta} \in K$ называется оптимальной в классе квадратично интегрируемых оценок K , если для всякой другой несмещённой оценки $\tilde{\theta} \in \Theta \forall \theta \in \Theta : D_{\theta} \hat{\theta} \leq D_{\theta} \tilde{\theta}$ или же $\forall \theta \in \Theta, M_{\theta} (\hat{\theta} - \theta)^2 \leq M_{\theta} (\tilde{\theta} - \theta)^2$.

Что называется среднеквадратическим отклонением оценки?

Сформулируйте утверждение про поведение выборочных моментов.

Какая оценка является несмещённой и содержательной для математического ожидания распределения выборки?

Какая статистика является несмещённой оценкой для дисперсии распределения выборки?

Аудиторные задачи

Домашнее задание