

Оглавление

Занятие 1. Выборочные характеристики	1
Контрольные вопросы и задания	3
Аудиторные задачи	4
Домашнее задание	11
Занятие 2. Свойства оценок	25
Контрольные вопросы и задания	27
Аудиторные задачи	28
Домашнее задание	35
Занятие 3. Метод моментов построения оценок	40
Контрольные вопросы и задания	41
Аудиторные задачи	42
Домашнее задание	43

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.

1.11

Задание. Пусть задана выборка X_1, \dots, X_n из показательного распределения с параметром α .

- Докажите, что случайные величины $X_{(1)}, X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(n-1)}$ являются независимыми;
- найдите распределение разности $X_{(k+1)} - X_{(k)}$ соседних порядковых статистик.

Решение. $\vec{\xi} = (\xi_1, \dots, \xi_n)$ — случайный вектор с плотностью распределения $f_{\vec{\xi}}(\vec{x})$.

Линейное преобразование этого вектора $\vec{\eta} = A\vec{\xi}$, где A — некоторая n -мерная матрица.

$$f_{A\vec{\xi}}(\vec{y}) = \frac{1}{|\det A|} \cdot f_{\vec{\xi}}(A^{-1}\vec{y}).$$

Составим вектор из величин $X_{(1)}, X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(n-1)}$. Его плотность должна распадаться на произведение плотностей компонент.

Из задачи 1.10

$$f_{(X_{(1)}, \dots, X_{(n)})}(y_1, \dots, y_n) = n! f(y_1) \cdot \dots \cdot f(y_n) \cdot \mathbb{1}\{y_1 < y_2 < \dots < y_n\}.$$

Подставим плотность показательного распределения

$$\begin{aligned} n! f(y_1) \cdot \dots \cdot f(y_n) \cdot \mathbb{1}\{y_1 < y_2 < \dots < y_n\} = \\ = n! \alpha e^{-\alpha y_1} \cdot \mathbb{1}\{y_1 > 0\} \cdot \dots \cdot \alpha e^{-\alpha y_n} \cdot \mathbb{1}\{y_n > 0\} \cdot \mathbb{1}\{y_1 < y_2 < \dots < y_n\}. \end{aligned}$$

Перемножим

$$\begin{aligned} n! \alpha e^{-\alpha y_1} \cdot \mathbb{1}\{y_1 > 0\} \cdot \dots \cdot \alpha e^{-\alpha y_n} \cdot \mathbb{1}\{y_n > 0\} \cdot \mathbb{1}\{y_1 < y_2 < \dots < y_n\} = \\ = n! \alpha^n e^{-\alpha(y_1 + \dots + y_n)} \cdot \mathbb{1}\{0 < y_1 < y_2 < \dots < y_n\}. \end{aligned}$$

Нужно найти линейное преобразование

$$\begin{bmatrix} X_{(1)} \\ X_{(2)} - X_{(1)} \\ X_{(3)} - X_{(2)} \\ \dots \\ X_{(n)} - X_{(n-1)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ \dots & & & & & & & \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} X_{(1)} \\ X_{(2)} \\ X_{(3)} \\ \dots \\ X_{(n)} \end{bmatrix},$$

где

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ \dots & & & & & & & \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix} = A.$$

Определитель $\det A = 1$.

Ищем обратную матрицу

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & 0 & \dots & 0 \\ \dots & & & & & \\ 1 & 1 & 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} X_{(1)} \\ X_{(2)} - X_{(1)} \\ X_{(3)} - X_{(2)} \\ \dots \\ X_{(n)} - X_{(n-1)} \end{bmatrix} = \begin{bmatrix} X_{(1)} \\ X_{(2)} \\ X_{(3)} \\ \dots \\ X_{(n)} \end{bmatrix}.$$

Тогда имеем выражение

$$A^{-1}\vec{y} = \begin{bmatrix} y_1 \\ y_1 + y_2 \\ \dots \\ \sum_{i=1}^n y_i \end{bmatrix}.$$

Определим искомый вектор через

$$\vec{\eta} = (X_{(1)}, X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(n-1)}).$$

Тогда

$$\begin{aligned} & f_{\vec{\eta}}(y_1, \dots, y_n) = \\ & = n! \alpha^n e^{-\alpha(ny_1 + (n-1)y_2 + \dots + y_n)} \cdot \mathbb{1}\{0 < y_1 < y_1 + y_2 < \dots < y_1 + y_2 + \dots + y_n\}. \end{aligned}$$

Разобьём на n множителей

$$\begin{aligned} & n! \alpha^n e^{-\alpha(ny_1 + (n-1)y_2 + \dots + y_n)} \cdot \mathbb{1}\{0 < y_1 < y_1 + y_2 < \dots < y_1 + y_2 + \dots + y_n\} = \\ & = [n\alpha e^{-\alpha ny_1} \cdot \mathbb{1}\{0 < y_1\}] \cdot [(n-1)\alpha e^{-\alpha(n-1)y_2} \cdot \mathbb{1}\{y_2 > 0\}] \cdot \dots \times \\ & \quad \times [\alpha e^{-\alpha y_n} \cdot \mathbb{1}\{y_n > 0\}]. \end{aligned}$$

Имеем произведение плотностей компонент, значит, элементы вектора независимы и показательно распределены с параметром $\alpha(n-k)$, то есть $X_{(k+1)} - X_{(k)} \sim \Pi(\alpha(n-k))$. Считаем, что $X_{(0)} = 0$.

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

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

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

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

Выборочная дисперсия

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

1.19

Задание. По выборке объёма n из распределения Бернулли с параметром p постройте эмпирическую функцию распределения $F_n(y)$.

Решение. Случайная величина имеет распределение Бернулли, если она принимает всего два значения: 1 и 0 с вероятностями p и $(1 - p)$ соответственно. Таким образом: $P(x = 1) = p$, $P(x = 0) = 1 - p$.

Эмпирической функцией распределения, построенной по выборке

$$x_1, \dots, x_n,$$

называется функция

$$F_n(y) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}(x_k \leq y).$$

Пусть есть набор из n чисел (нулей и единиц) — выборка из распределения Бернулли.

Для удобства выстроим числа в порядке их возрастания:

$$0, 0, \dots, 0, 1, 1, \dots, 1.$$

Видим, что слева от нуля эмпирическая функция распределения будет равна нулю.

В точке 0 произойдёт скачок на

$$\frac{n-k}{n},$$

где k — количество единиц, а $(n-k)$ — количество нулей в выборке.

В точке 1 будет скачок на

$$1 - \frac{n-k}{n} = \frac{n-n+k}{n} = \frac{k}{n},$$

а значение самой функции будет равно единице.

$$F_n(y) = \begin{cases} 0, & x < 0, \\ \frac{n-k}{n}, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

Эмпирическая функция распределения будет выглядеть так, как показано на рис. 3.



Рис. 3: Эмпирическая функция распределения

1.20

Задание. Пусть X_1, \dots, X_n — выборка из распределения F . Докажите, что для произвольных $y \in \mathbb{R}$ и $k \in \{0, 1, \dots, n\}$ справедливо равенство

$$P\left(F_n(y) = \frac{k}{n}\right) = C_n^k F^k(y) (1 - F(y))^{n-k}.$$

Решение. Эмпирической функцией распределения, построенной по выборке X_1, \dots, X_n , называется функция

$$F_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq y).$$

Посмотрим, что значит событие в указанной вероятности

$$F_n(y) = \frac{k}{n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq y).$$

Сократим константы в знаменателях

$$k = \sum_{i=1}^n \mathbb{1}(X_i \leq y).$$

Это означает, что есть ровно k элементов выборки, не превышающих y . Следовательно, это биномиальное распределение с параметром $F(y)$, то есть

$$P\left(F_n(y) = \frac{k}{n}\right) = C_n^k F^k(y) (1 - F(y))^{n-k}.$$

1.21

Задание. Для выборки X_1, \dots, X_n из равномерного распределения на отрезке $[0, \theta]$ найдите плотность, математическое ожидание и дисперсию:

- а) максимального члена вариационного ряда $X_{(n)}$;
- б) минимального члена вариационного ряда $X_{(1)}$;
- в) совместную плотность распределения и ковариацию $X_{(n)}$ и $X_{(1)}$.

Решение.

$$\begin{aligned} \text{а) Найдем функцию распределения } n\text{-й порядковой статистики } F_{X_{(n)}}(y) &= \\ &= P(X_{(n)} \leq y) = P(n \text{ элементов выборки не превышают } y) = \\ &= C_n^n [F_{X_1}(y)]^n \cdot [1 - F_{X_1}(y)]^{n-n} = [F_{X_1}(y)]^n = [P(X_1 \leq y)]^n = \left(\frac{y}{\theta}\right)^n, \end{aligned}$$

при этом $y \in [0, \theta]$.

Продифференцируем

$$f_{X_{(n)}}(y) = \frac{\partial F_{X_{(n)}}(y)}{\partial y} = \frac{\partial}{\partial y} \left(\frac{y}{\theta}\right)^n = \frac{1}{\theta^n} \cdot \frac{dy^n}{dy} = \frac{ny^{n-1}}{\theta^n} \cdot \mathbb{1}\{y \in [0, \theta]\}.$$

По определению математического ожидания

$$MX_{(n)} = \int_0^\theta y dF^n(y) = \int_0^\theta y d\left(\frac{y}{\theta}\right)^n = \frac{1}{\theta^n} \int_0^\theta y dy^n = \frac{n}{\theta^n} \int_0^\theta y y^{n-1} dy.$$

Сложим степени сомножителей

$$\frac{n}{\theta^n} \int_0^\theta y y^{n-1} dy = \frac{n}{\theta^n} \int_0^\theta y^n dy = \frac{n}{\theta^n} \cdot \frac{y^{n+1}}{n+1} \Big|_0^\theta = \frac{\theta^{n+1} n}{\theta^n (n+1)} = \frac{\theta n}{n+1}.$$

Найдём второй момент

$$MX_{(n)}^2 = \int_0^\theta y^2 dF^n(y) = \frac{n}{\theta^n} \int_0^\theta y^2 y^{n-1} dy = \frac{n}{\theta^n} \int_0^\theta y^{n+1} dy.$$

Возьмём интеграл

$$\frac{n}{\theta^n} \int_0^\theta y^{n+1} dy = \frac{ny^{n+2}}{\theta^n (n+2)} \Big|_0^\theta = \frac{n\theta^{n+2}}{\theta^n (n+2)} = \frac{n\theta^2}{n+2}.$$

По свойствам дисперсии

$$DX_{(n)} = MX_{(n)}^2 - (MX_{(n)})^2 = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}.$$

Приведём к общему знаменателю

$$\begin{aligned} \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} &= \frac{n\theta^2 (n^2 + 2n + 1) - n^2\theta^2 (n+2)}{(n+2)(n+1)^2} = \\ &= \frac{n^3\theta^2 + 2n^2\theta^2 + n\theta^2 - n^3\theta^2 - 2n^2\theta^2}{(n+2)(n+1)^2} = \frac{n\theta^2}{(n+2)(n+1)^2}; \end{aligned}$$

b) найдём функцию распределения первой порядковой статистики

$$\begin{aligned} F_{X_{(1)}}(y) &= P(X_{(1)} \leq y) = P(\min(X_1, X_2, \dots, X_n) \leq y) = \\ &= P(\text{хотя бы 1 элемент выборки не превышает } y) = \end{aligned}$$

$$\begin{aligned} \sum_{k=1}^n P(\text{exactly } k \text{ elements of the sample does not exceed } y) &= \\ &= \sum_{k=1}^n C_n^k F^k(y) [1 - F(y)]^{n-k} = \\ &= \sum_{k=0}^n C_n^k F^k(y) [1 - F(y)]^{n-k} - C_n^0 F^0(y) [1 - F(y)]^{n-0}. \end{aligned}$$

Применяем формулу для бинома Ньютона

$$\begin{aligned} \sum_{k=0}^n C_n^k F^k(y) [1 - F(y)]^{n-k} - C_n^0 F^0(y) [1 - F(y)]^{n-0} = \\ = [F(y) + 1 - F(y)]^n - [1 - F(y)]^n = 1 - \left(1 - \frac{y}{\theta}\right)^n. \end{aligned}$$

Продифференцируем

$$f_{X_{(1)}}(y) = \frac{\partial F_{X_{(1)}}(y)}{\partial y} = \frac{\partial}{\partial y} \left(1 - \frac{(\theta - y)^n}{\theta^n}\right) = -\frac{1}{\theta^n} \cdot \frac{\partial}{\partial y} (\theta - y)^n.$$

Берём производную сложной функции

$$-\frac{1}{\theta^n} \cdot \frac{\partial}{\partial y} (\theta - y)^n - \frac{1}{\theta^n} \cdot n (\theta - y)^{n-1} (-1) = \frac{n (\theta - y)^{n-1}}{\theta^n}, \quad y \in [0, \theta].$$

Найдём математическое ожидание по определению

$$MX_{(1)} = \int_0^\theta y dF^n(y) = \int_0^\theta y d\left(1 - \frac{y}{\theta}\right)^n.$$

Сделаем замену

$$1 - \frac{y}{\theta} = z,$$

откуда $y = \theta(1 - z)$, при этом интегрирование происходит в пределах от одного до нуля

$$\int_0^\theta y d\left(1 - \frac{y}{\theta}\right)^n = \int_1^0 \theta(1 - z) dz^n = \theta n \int_1^0 (z - 1) z^{n-1} dz.$$

Разбиваем на 2 интеграла

$$\theta n \int_1^0 (z - 1) z^{n-1} dz = -\theta n \int_0^1 z^n dz + \theta n \int_0^1 z^{n-1} dz = -\theta n \cdot \frac{z^{n+1}}{n+1} \Big|_0^1 + \theta n \cdot \frac{z^n}{n} \Big|_0^1.$$

Подставляем пределы интегрирования

$$-\theta n \cdot \frac{z^{n+1}}{n+1} \Big|_0^1 + \theta n \cdot \frac{z^n}{n} \Big|_0^1 = -\theta n \cdot \frac{1}{n+1} + \theta n \cdot \frac{1}{n} = -\theta n \left(\frac{1}{n+1} - \frac{1}{n} \right).$$

Приводим к общему знаменателю

$$-\theta n \left(\frac{1}{n+1} - \frac{1}{n} \right) = -\theta n \cdot \frac{n - n - 1}{n(n+1)} = \frac{\theta n}{n(n+1)} = \frac{\theta}{(n+1)}.$$

Найдём второй момент

$$MX_{(1)}^2 = \int_0^\theta y^2 d\left(1 - \frac{y}{\theta}\right)^n.$$

Применяем такую же замену, как при поиске первого момента

$$\int_0^\theta y^2 d\left(1 - \frac{y}{\theta}\right)^n = \int_1^0 \theta^2 (1-z)^2 dz^n = \int_0^1 \theta^2 n (1-z)^2 z^{n-1} dz.$$

Выносим константу за знак интеграла и возводим скобку в квадрат

$$\begin{aligned} \int_0^1 \theta^2 n (1-z)^2 z^{n-1} dz &= \theta^2 n \int_0^1 (1-2z+z^2) z^{n-1} dz = \\ &= \theta^2 n \int_0^1 z^{n-1} dz - 2\theta^2 n \int_0^1 z^n dz + \theta^2 n \int_0^1 z^{n+1} dz = \\ &= \theta^2 n \cdot \frac{z^n}{n} \Big|_0^1 - 2\theta^2 n \cdot \frac{z^{n+1}}{n+1} \Big|_0^1 + \theta^2 n \cdot \frac{z^{n+2}}{n+2} \Big|_0^1 = \theta^2 - \frac{2\theta^2 n}{n+1} + \frac{\theta^2 n}{n+2} = \\ &= \theta^2 \left(1 - \frac{2n}{n+1} + \frac{n}{n+2}\right) = \\ &= \theta^2 \cdot \frac{(n+1)(n+2) - 2n(n+2) + n(n+1)}{(n+1)(n+2)} = \\ &= \frac{\theta^2 (n^2 + 3n + 2 - 2n^2 - 4n + n^2 + n)}{(n+1)(n+2)} = \frac{2\theta^2}{(n+1)(n+2)}. \end{aligned}$$

По свойствам дисперсии

$$DX_{(1)} = MX_{(1)}^2 - [MX_{(1)}]^2 = \frac{2\theta^2}{(n+1)(n+2)} - \frac{\theta^2}{(n+1)^2}.$$

Приводим к общему знаменателю

$$\begin{aligned} \frac{2\theta^2}{(n+1)(n+2)} - \frac{\theta^2}{(n+1)^2} &= \frac{2\theta^2(n+1) - \theta^2(n+2)}{(n+1)^2(n+2)} = \\ &= \frac{2\theta^2 n + 2\theta^2 - \theta^2 n - 2\theta^2}{(n+1)^2(n+2)} = \frac{\theta^2 n}{(n+1)^2(n+2)}; \end{aligned}$$

с) найдём функцию распределения вектора по определению

$$F_{(X_{(1)}, X_{(n)})}(y_1, y_n) = P(X_{(1)} \leq y_1, X_{(n)} \leq y_n).$$

Воспользовавшись формулой $P(A \cap B) = P(B) - P(\bar{A} \cap B)$, получим

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

Если максимум меньше какого-то значения, то все элементы меньше него

$$\begin{aligned} P(X_{(1)} > y_1, X_{(n)} \leq y_n) &= \\ &= P(X_1 \leq y_n, \dots, X_n \leq y_n) - \\ &= P(X_1 \in (y_1, y_n], X_2 \in (y_1, y_n], \dots, X_n \in (y_1, y_n]) = \\ &= [F(y_n)]^n - [F(y_1)]^n \end{aligned}$$

при $y_1 < y_n$.

Продифференцируем

$$f_{(X_{(1)}, X_{(n)})}(y_1, y_n) = \begin{cases} 0, & y_1 \geq y_n, \\ n(n-1) \cdot [F(y_n) - F(y_1)]^{n-2} \times \\ \times p_{X_1}(y_1) p_{X_n}(y_n) = \\ = n(n-1) \cdot \left[\frac{y_n}{\theta} - \frac{y_1}{\theta} \right]^{n-2} \cdot \frac{1}{\theta^2} = \\ = n(n-1) \cdot \frac{(y_n - y_1)^{n-2}}{\theta^{n-2}} \cdot \frac{1}{\theta^2} = \frac{n(n-1)(y_n - y_1)^{n-2}}{\theta^n}. \end{cases}$$

Найдём математическое ожидание вектора

$$M(X_{(1)}, X_{(n)}) = \frac{n(n-1)}{\theta^n} \cdot \int_0^\theta \int_{y_n}^\theta y_1 y_n (y_n - y_1)^{n-2} dy_1 dy_n.$$

Заменяем разность величин величиной x . Якобиан преобразования равен

$$\frac{\partial(x, y_n)}{\partial(x, y_1)} = \left| \begin{bmatrix} \frac{\partial x}{\partial y_1} & \frac{\partial y_n}{\partial y_1} \\ \frac{\partial x}{\partial y_n} & \frac{\partial y_1}{\partial y_1} \end{bmatrix} \right| = \left| \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \right| = 1.$$

Получаем

$$\begin{aligned}
& \frac{n(n-1)}{\theta^n} \cdot \int_0^\theta \int_{y_n}^\theta y_1 y_n (y_n - y_1)^{n-2} dy_1 dy_n = \\
& = \frac{n(n-1)}{\theta^n} \cdot \int_0^\theta \int_0^{\theta-y_n} (x + y_n) y_n x^{n-2} dx dy_n = \\
& = \frac{n(n-1)}{\theta^n} \cdot \int_0^\theta y_n dy_n \cdot \int_0^{\theta-y_n} (x^{n-1} + y_n x^{n-2}) dx = \\
& = \frac{n(n-1)}{\theta^n} \cdot \int_0^\theta y_n dy_n \cdot \left(\frac{x^n}{n} + y_n \cdot \frac{x^{n-1}}{n-1} \right) \Big|_0^{\theta-y_n} = \\
& = \frac{1}{\theta^n} \cdot \int_0^\theta \left[(n-1)(\theta - y_n) + y_n n (\theta - y_n)^{n-1} \right] \cdot y_n dy_n = \\
& = \frac{1}{\theta^n} \cdot \int_0^\theta (\theta - y_n)^{n-1} \cdot [(n-1)(\theta - y_n) + y_n n] \cdot y_n dy_n = \\
& = \frac{1}{\theta^n} \cdot \int_0^\theta (\theta - y_n)^{n-1} [n\theta - (\theta - y_n)] y_n dy_n.
\end{aligned}$$

Заменяем первую скобку в интеграле на t . Получаем

$$\frac{1}{\theta^n} \cdot \int_0^\theta (\theta - y_n)^{n-1} [n\theta - (\theta - y_n)] y_n dy_n = \frac{1}{\theta^n} \cdot \int_0^\theta t^{n-1} (n\theta - t) (\theta - t) dt.$$

Перемножаем скобки

$$\begin{aligned}
& \frac{1}{\theta^n} \cdot \int_0^\theta t^{n-1} (n\theta - t) (\theta - t) dt = \frac{1}{\theta^n} \cdot \int_0^\theta t^{n-1} (n\theta^2 - (n+1)\theta t + t^2) dt = \\
& = \frac{1}{\theta^n} \cdot \int_0^\theta (n\theta^2 t^{n-1} - (n+1)\theta t^n + t^{n+1}) dt = \\
& = \frac{1}{\theta^n} \cdot \left[\theta^2 t^n - \theta t^{n+1} + \frac{t^{n+2}}{n+2} \right] \Big|_0^\theta = \frac{1}{\theta^n} \cdot \left[\theta^{n+2} - \theta^{n+2} + \frac{\theta^{n+2}}{n+2} \right] = \frac{\theta^2}{n+2}.
\end{aligned}$$

По определению ковариации

$$cov(X_{(1)}, X_{(n)}) = M(X_{(1)}X_{(n)}) - MX_{(1)}MX_{(n)} = \frac{\theta^2}{n+2} - \frac{\theta}{n+1} \cdot \frac{\theta n}{n+1}.$$

Выносим общий множитель за скобки

$$\frac{\theta^2}{n+2} - \frac{\theta}{n+1} \cdot \frac{\theta n}{n+1} = \theta^2 \left(\frac{1}{n+2} - \frac{n}{(n+1)^2} \right).$$

Приводим к общему знаменателю дроби в скобках

$$\theta^2 \left(\frac{1}{n+2} - \frac{n}{(n+1)^2} \right) = \theta^2 \cdot \frac{n^2 + 2n + 1 - n^2 - 2n}{(n+2)(n+1)^2} = \frac{\theta^2}{(n+2)(n+1)^2}.$$

1.22

Задание. Пусть задана выборка X_1, \dots, X_n из показательного распределения с параметром α . Докажите, что

$$MX_{(k)} = \frac{1}{\alpha} \left(\frac{1}{n-k+1} + \dots + \frac{1}{n} \right),$$

не находя распределения порядковой статистики $X_{(k)}$.

Решение. Преобразуем левую часть того, что нужно доказать

$$MX_{(k)} = MX_{(1)} + \sum_{i=1}^{k-1} M(X_{(i+1)} - X_{(i)}).$$

Элементы выборки имеют плотность распределения

$$p(y) = \alpha e^{-\alpha y} \cdot \mathbb{1}\{y \geq 0\}.$$

Интегрируя плотность, получим функцию распределения

$$F(y) = \int_{-\infty}^y p(x) dx = \int_{-\infty}^y \alpha e^{-\alpha x} \cdot \mathbb{1}\{x \geq 0\} dx = \int_0^y \alpha e^{-\alpha x} dx = -e^{-\alpha x} \Big|_0^y.$$

Подставляем пределы интегрирования

$$-e^{-\alpha x} \Big|_0^y = 1 - e^{-\alpha y}, \quad y \geq 0.$$

Находим функцию распределения минимальной порядковой статистики

$$F_{X_{(1)}}(y) = P(X_{(1)} \leq y) = 1 - P(X_{(1)} > y).$$

Если минимальный элемент выборки больше какого-то значения, то все элементы выборки больше него

$$1 - P(X_{(1)} > y) = 1 - P(X_1 > y, X_2 > y, \dots, X_n > y).$$

Из независимости случайных величин следует, что

$$1 - P(X_1 > y, X_2 > y, \dots, X_n > y) = 1 - P(X_1 > y) \cdot P(X_2 > y) \cdot \dots \cdot P(X_n > y).$$

Переходим к противоположным событиям и учитываем то, что случайные величины одинаково распределены

$$1 - P(X_1 > y) \cdot P(X_2 > y) \cdot \dots \cdot P(X_n > y) = 1 - [1 - F(y)]^n.$$

Находим плотность минимальной порядковой статистики

$$f_{X_{(1)}}(y) = \frac{\partial F_{X_{(1)}}(y)}{\partial y} = \frac{\partial \{1 - [1 - F(y)]^n\}}{\partial y} = -n[1 - F(y)]^{n-1} \cdot (-1) f(y).$$

Упрощаем и учитываем, что $f(y)$ — это плотность распределения элементов выборки

$$-n[1 - F(y)]^{n-1} \cdot (-1) f(y) = n[1 - F(y)]^{n-1} p(y).$$

Подставляем функцию и плотность распределения

$$n[1 - F(y)]^{n-1} p(y) = n[1 - (1 - e^{-\alpha y})]^n \alpha e^{-\alpha y} \cdot \mathbb{1}\{y \geq 0\}.$$

Упрощаем выражение в скобках

$$n[1 - (1 - e^{-\alpha y})]^n \alpha e^{-\alpha y} \cdot \mathbb{1}\{y \geq 0\} = n\alpha e^{-\alpha(n-1)y} e^{-\alpha y} \cdot \mathbb{1}\{y \geq 0\}.$$

Суммируем показатели экспонент

$$n\alpha e^{-\alpha(n-1)y} e^{-\alpha y} \cdot \mathbb{1}\{y \geq 0\} = n\alpha e^{-\alpha n y} \cdot \mathbb{1}\{y \geq 0\}.$$

Отсюда следует, что $X_{(1)} \sim \text{Exp}(n\alpha)$.

Из задачи 1.11 б) $X_{(i+1)} - X_{(i)} \sim \text{Exp}(\alpha(n-i))$.

Тогда такая разность имеет математическое ожидание

$$M(X_{(i+1)} - X_{(i)}) = \frac{1}{\alpha(n-i)}.$$

Подставляем в полученное в начале выражение

$$MX_{(1)} + \sum_{i=1}^{k-1} M(X_{(i+1)} - X_{(i)}) = \frac{1}{n\alpha} + \sum_{i=1}^{k-1} \frac{1}{\alpha(n-i)}.$$

Записываем сумму в явном виде

$$\frac{1}{n\alpha} + \sum_{i=1}^{k-1} \frac{1}{\alpha(n-i)} = \frac{1}{\alpha} \cdot \left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \dots + \frac{1}{n-k+1} \right).$$

1.23

Задание. Для выборки X_1, \dots, X_n из равномерного распределения на отрезке $[0, \theta]$ найдите:

- а) ковариацию $X_{(n)}$ и $X_{(1)}$;

б) совместную плотность распределения $X_{(k)}$ и $X_{(j)}$, $1 \leq k \leq j \leq n$.

Решение.

а) Данный пункт решён в задаче 1.21 с);

б) по определению функции распределения

$$F_{(X_{(k)}, X_{(j)})}(y, z) = P(X_{(k)} \leq y, X_{(j)} \leq z).$$

Воспользовавшись формулой $P(A \cap B) = P(B) - P(\bar{A} \cap B)$, получим $P(X_{(k)} \leq y, X_{(j)} \leq z) = P(X_{(j)} \leq z) - P(X_{(k)} > y, X_{(j)} \leq z)$. Аналогично задаче 1.10 разбиваем на 2 решения согласно коэффициентам

$$\begin{aligned} & P(X_{(j)} \leq z) - P(X_{(k)} > y, X_{(j)} \leq z) = \\ & = \begin{cases} P(X_{(j)} \leq z), & y \geq z, \\ P(X_{(j)} \leq z) - P(X_{(k)} > y, X_{(j)} \leq z), & y < z. \end{cases} \end{aligned}$$

Найдём первую вероятность $P(X_{(j)} \leq z) = P(\text{хотя бы } j \text{ элементов выборки не превышает } z) =$

$$= \sum_{i=j}^n C_n^i F^i(z) [1 - F(z)]^{n-i} = \sum_{i=j}^n C_n^i \left(\frac{z}{\theta}\right)^i \left(1 - \frac{z}{\theta}\right)^{n-i}.$$

Приводим к общему знаменателю

$$\sum_{i=j}^n C_n^i \left(\frac{z}{\theta}\right)^i \left(1 - \frac{z}{\theta}\right)^{n-i} = \sum_{i=j}^n C_n^i \frac{z^i}{\theta^i} \cdot \frac{(\theta - z)^{n-i}}{\theta^{n-i}} = \sum_{i=j}^n C_n^i \frac{z^i (\theta - z)^{n-i}}{\theta^n}.$$

Найдём вторую вероятность $P(X_{(j)} \leq z, X_{(k)} > y) = P(\text{хотя бы } (j - k + 1)$

элементов выборки находится в интервале $(y, z]) =$

$$\begin{aligned} & = \sum_{i=j-k+1}^n C_n^i [F(z) - F(y)]^i \cdot \{1 - [F(z) - F(y)]\}^{n-i} = \\ & = \sum_{i=j-k+1}^n C_n^i \cdot \left(\frac{z}{\theta} - \frac{y}{\theta}\right)^i \cdot \left[1 - \left(\frac{z}{\theta} - \frac{y}{\theta}\right)\right]^{n-i} = \\ & = \sum_{i=j-k+1}^n C_n^i \cdot \frac{(z - y)^i}{\theta^i} \cdot \left(1 - \frac{z - y}{\theta}\right)^{n-i} = \\ & = \sum_{i=j-k+1}^n C_n^i \cdot \frac{(z - y)^i}{\theta^i} \cdot \frac{(\theta - z + y)^{n-i}}{\theta^{n-i}} = \\ & = \sum_{i=j-k+1}^n \frac{C_n^i}{\theta^n} (z - y)^i (\theta - z + y)^{n-i}. \end{aligned}$$

Продифференцируем. В первом случае будет 0. Найдём производную второй вероятности

$$\begin{aligned}
& -\frac{\partial^2 F}{\partial y \partial z} = \\
& = \frac{\partial}{\partial y} \left\{ \sum_{i=j-k+1}^n \frac{C_n^i}{\theta^n} \times \right. \\
& \times \left[i(z-y)^{i-1} (\theta-z+y)^{n-i} - (z-y)(n-i)(\theta-z+y)^{n-i-1} \right] \Big\} = \\
& = \sum_{i=j-k+1}^n \frac{C_n^i}{\theta^n} \times \\
& \times [-i(i-1)(z-y)^{i-2} (\theta-z+y)^{n-i} + \\
& + 2i(z-y)^{i-1} (n-i) \cdot (\theta-z+y)^{n-i-1} - \\
& - (z-y)^i (n-i)(n-i-1) \cdot (\theta-z+y)^{n-i-2}] = \\
& = \sum_{i=j-k+1}^n \frac{C_n^i (z-y)^{i-2} \cdot (\theta-z+y)^{n-i-2}}{\theta^n} \times \\
& \times [2i(z-y)(n-i)(\theta-z+y) - i(i-1)(\theta-z+y) - \\
& - (n-i)(n-i-1)(z-y)^2].
\end{aligned}$$

Тогда во втором случае плотность равна полученному выражению.

Занятие 2. Свойства оценок

Контрольные вопросы и задания

Что называют оценкой неизвестного параметра?

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

Приведите 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?

$M_{\theta} (\hat{\theta} - \theta)$ — среднеквадратическое отклонение.

Сформулируйте утверждение про поведение выборочных моментов.

Выборочный начальный момент M_k k -го порядка стремится к начальному моменту ν_k случайной величины X , то есть

$$\lim_{n \rightarrow \infty} P(|M_k - \nu_k| \geq \varepsilon) = 0,$$

для любого сколь угодно малого $\varepsilon > 0$, если моменты ν_{2k} и ν_k случайной величины X существуют и конечны.

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

В качестве оценки для математического ожидания естественно предложить среднеарифметическое наблюдаемых значений

$$\tilde{m} = \frac{\sum_{i=1}^n X_i}{n}.$$

Какая статистика является несмещённой оценкой для дисперсии распределения выборки?

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

— несмещённая оценка для $\sigma^2 = Dx_1$.

Аудиторные задачи

2.4

Задание. Для выборки равномерного распределения на отрезке $[\theta, 1]$ проверьте состоятельность и несмещённость оценки $X_{(1)}$ параметра θ .

Решение. θ — минимальное наблюдение. Проверяем, выполняется ли $X_{(1)} \xrightarrow{P} \theta, n \rightarrow \infty$.

По определению сходимости по вероятности

$$\forall \varepsilon > 0 \ P(|X_{(1)} - \theta| > \varepsilon) \rightarrow 0, n \rightarrow \infty.$$

Раскроем модуль

$$P\{X_{(1)} > \varepsilon + \theta\} = P(X_1 > \varepsilon + \theta, \dots, X_n > \varepsilon + \theta) = [P(X_1 > \varepsilon + \theta)]^n.$$

Подставим значение вероятности из геометрического эксперимента

$$[P(X_1 > \varepsilon + \theta)]^n = \left(\frac{1 - \theta - \varepsilon}{1 - \theta}\right)^n = \left(1 - \frac{\varepsilon}{1 - \theta}\right)^n \rightarrow 0, n \rightarrow \infty.$$

Число в скобках строго меньше единицы, так как $0 \leq \theta \leq 1$.

Отсюда следует, что оценка состоятельная.

Проверяем несмещённость оценки. Проверяем, выполняется ли

$$MX_{(1)} = \theta.$$

Нужно найти плотность

$$MX_{(1)} = \int_{\mathbb{R}} f_{X_{(1)}}(y) y dy.$$

Начинаем с функции распределения $F_{X_{(1)}}(y) = P(X_{(1)} \leq y)$. Переходим к противоположному событию

$$P(X_{(1)} \leq y) = 1 - P(X_{(1)} > y) = 1 - [P(X_1 > y)]^n.$$

Переходим к противоположному событию $1 - [P(X_1 > y)]^n = 1 - [1 - F(y)]^n$.

Продифференцируем

$$\frac{dF_{X_{(1)}}(y)}{dy} = n[1 - F(y)]^{n-1} f(y).$$

На отрезке $[\theta, 1]$ имеет равномерное распределение

$$n[1 - F(y)]^{n-1} f(y) = n \left[1 - \frac{y - \theta}{1 - \theta} \right]^{n-1} \cdot \mathbb{1}\{y \in [\theta, 1]\} \cdot \frac{1}{1 - \theta}.$$

Приведём к общему знаменателю

$$n \left[1 - \frac{y - \theta}{1 - \theta} \right]^{n-1} \cdot \mathbb{1}\{y \in [\theta, 1]\} \cdot \frac{1}{1 - \theta} = \frac{n}{(1 - \theta)^2} \cdot (1 - y)^{n-1} \cdot \mathbb{1}\{y \in [\theta, 1]\}.$$

Нашли плотность $X_{(1)}$ и теперь можем вычислить интеграл

$$MX_{(1)} = \int_{\theta}^1 y \cdot \frac{n}{(1 - \theta)^2} \cdot (1 - y)^{n-1} dy.$$

Замена:

$$1 - y = z, dy = -dz, y = 1 - z, y = 1 \Rightarrow z = 0, y = \theta \Rightarrow z = 1 - \theta.$$

Подставляя замену, получаем

$$\int_{\theta}^1 y \cdot \frac{n}{(1 - \theta)^2} \cdot (1 - y)^{n-1} dy = n \cdot \frac{1}{(1 - \theta)^n} \int_0^{1-\theta} (1 - z) z^{n-1} dz.$$

Вычислим интеграл

$$\begin{aligned} n \cdot \frac{1}{(1 - \theta)^n} \int_0^{1-\theta} (1 - z) z^{n-1} dz &= \frac{n}{(1 - \theta)^n} \left[\frac{(1 - \theta)^n}{n} - \frac{(1 - \theta)^{n+1}}{n + 1} \right] = \\ &= n \left(\frac{1}{n} - \frac{1 - \theta}{n + 1} \right) = 1 - \frac{n}{n + 1} (1 - \theta). \end{aligned}$$

Раскроем скобки

$$1 - \frac{n}{n + 1} (1 - \theta) = 1 - \frac{n}{n + 1} - \theta \cdot \frac{n}{n + 1} \neq \theta.$$

Отсюда следует, что оценка смещённая, но асимптотически несмещённая, потому что

$$1 - \frac{n}{n + 1} \rightarrow 0, n \rightarrow \infty$$

и

$$\frac{n}{n + 1} \rightarrow 1, n \rightarrow \infty.$$

2.5

Задание. Пусть X_1, \dots, X_n — выборка из распределения Пуассона с параметром $\lambda > 0$. Выясните, является ли статистика

$$\frac{1}{n} \sum_{i=1}^n X_i^2 :$$

- а) несмещённой оценкой для λ^2 ;
- б) состоятельной оценкой для λ^2 .

Решение.

- а) Нужно проверить, выполняется ли

$$M \frac{1}{n} \sum_{i=1}^n X_i^2 = \lambda^2.$$

Преобразуем левую часть

$$M \frac{1}{n} \sum_{i=1}^n X_i^2 = M X_1^2 = D X_1 + (M X_1)^2 = \lambda + \lambda^2 \neq \lambda^2.$$

Значит, оценка смещённая;

- б) проверяем, имеет ли место

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} \lambda^2, n \rightarrow \infty.$$

По закону больших чисел

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \rightarrow M X_1^2 = \lambda^2 + \lambda \neq \lambda^2,$$

значит, оценка не состоятельная.

2.6

Задание. Пусть X_1, \dots, X_n — выборка из показательного распределения с параметром $\alpha > 0$. Докажите, что статистика $1/\bar{X}$ является состоятельной оценкой для α .

Решение. Нужно показать, что

$$\frac{1}{\bar{X}} \xrightarrow{P} \alpha, n \rightarrow \infty.$$

Выборочное среднее

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

По закону больших чисел

$$\bar{X} \xrightarrow{P} MX_1 = \frac{1}{\alpha}.$$

Отсюда следует, что

$$\frac{1}{\bar{X}} \xrightarrow{P} \frac{1}{MX_1} = \alpha.$$

2.7

Задание. Пусть X_1, \dots, X_n — выборка из нормального распределения $N(a, \sigma^2)$. Докажите, что статистика

$$S_n = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2$$

является несмещённой и состоятельной оценкой для σ^1 .

Решение. Нужно проверить условие $MS_n = \sigma^2$.

Разность двух соседних элементов выборки имеет распределение

$$X_{i+1} - X_i \sim N(0, 2\sigma^2).$$

Найдём математическое ожидание статистики

$$MS_n = M \frac{1}{2(n-1)} \cdot \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2 = \frac{1}{2(n-1)} \sum_{i=1}^n M(X_{i+1} - X_i)^2.$$

Случайные величины одинаково распределены

$$\frac{1}{2(n-1)} \sum_{i=1}^n M(X_{i+1} - X_i)^2 = \frac{n}{2(n-1)} \cdot M(X_2 - X_1)^2.$$

В данном случае второй момент равен дисперсии

$$\frac{n}{2(n-1)} \cdot M(X_2 - X_1)^2 = \frac{1}{2} \cdot 2\sigma^2 = \sigma^2.$$

Отсюда следует, что оценка несмещённая.

Проверим состоятельность, то есть $S_n \xrightarrow{P} \sigma^2, n \rightarrow \infty$.

Разобьём S_n на две суммы

$$S_n = \frac{1}{2(n-1)} \cdot \left[\sum_{\text{even } i} (X_{i+1} - X_i)^2 + \sum_{\text{odd } i} (X_{i+1} - X_i)^2 \right].$$

В каждой из сумм слагаемые независимы

$$\begin{aligned} & \frac{1}{2(n-1)} \cdot \left[\sum_{\text{even } i} (X_{i+1} - X_i)^2 + \sum_{\text{odd } i} (X_{i+1} - X_i)^2 \right] = \\ & = \frac{1}{2(n-1)} \cdot \left[\frac{m}{m} \sum_{\text{even } i} (X_{i+1} - X_i)^2 + \frac{n-1-m}{n-1-m} \sum_{\text{odd } i} (X_{i+1} - X_i)^2 \right]. \end{aligned}$$

По закону больших чисел

$$\begin{aligned} & \frac{1}{2(n-1)} \cdot \left[\frac{m}{m} \sum_{\text{even } i} (X_{i+1} - X_i)^2 + \frac{n-1-m}{n-1-m} \sum_{\text{odd } i} (X_{i+1} - X_i)^2 \right] \xrightarrow{P} \\ & \xrightarrow{P} \frac{1}{2} \cdot M(X_2 - X_1)^2 = \frac{1}{2} \cdot D(X_2 - X_1) = \sigma^2, n \rightarrow \infty. \end{aligned}$$

Отсюда следует, что оценка состоятельная.

2.8

Задание. Пусть X_1, \dots, X_n — выборка из показательного распределения с параметром $\alpha > 1$. Для какого параметра $\theta = \theta(\alpha)$ статистика

$$\hat{\theta}_n = e^{\bar{X}}$$

является состоятельной оценкой? Является ли $\hat{\theta}_n$ сильно состоятельной оценкой того же параметра? Является ли $\hat{\theta}_n$ несмещённой оценкой того же параметра? Асимптотически несмещённой?

Решение.

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

По закону больших чисел

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} MX_1, n \rightarrow \infty.$$

Случайные величины в выборке имеют показательное распределение

$$MX_1 = \frac{1}{\alpha},$$

значит,

$$\bar{X} \xrightarrow{P} \frac{1}{\alpha}, n \rightarrow \infty.$$

Применяем непрерывную функцию e^x . Получаем $e^{\bar{X}} \xrightarrow{P} e^{\frac{1}{\alpha}}, n \rightarrow \infty$.

Проверяем, является ли оценка $e^{\bar{X}}$ несмещённой к параметру $e^{\frac{1}{\alpha}}$, то есть выполняется ли $Me^{\bar{X}} = e^{\frac{1}{\alpha}}$.

Вычисляем $Me^{\bar{X}} = Me^{\frac{1}{n} \sum_{i=1}^n X_i}$. Случайные величины независимы и одинаково распределены, поэтому $Me^{\frac{1}{n} \sum_{i=1}^n X_i} = \left(Me^{\frac{X_1}{n}} \right)^n$. По определению характеристической функции $\varphi_{X_1} = Me^{itX_1}$ получаем

$$\left(Me^{\frac{X_1}{n}} \right)^n = \left[\varphi_{X_1} \left(\frac{1}{in} \right) \right]^n.$$

Характеристическая функция показательного распределения

$$\varphi_{X_1}(t) = Me^{itX_1} = \frac{\alpha}{\alpha - it}.$$

Подставляем

$$\left[\varphi_{X_1} \left(\frac{1}{in} \right) \right]^n = \left(\frac{\alpha}{\alpha - \frac{1}{n}} \right)^n.$$

Прибавим и отнимем в числителе $1/n$ и поделим числитель на знаменатель

$$\left(\frac{\alpha}{\alpha - \frac{1}{n}} \right)^n = \left(\frac{\alpha + \frac{1}{n} - \frac{1}{n}}{\alpha - \frac{1}{n}} \right)^n = \left[1 + \frac{1}{n(\alpha - \frac{1}{n})} \right]^n = e^{\frac{1}{\alpha}}.$$

Значит, оценка смещённая, но несмещённая асимптотически.

2.9

Задание. Пусть X_1, \dots, X_n — выборка из геометрического распределения с параметром p . Найдите состоятельные оценки для параметров

$$p, p^2, \ln p, p \sin(1-p), pe^{\frac{q^2}{2}}.$$

Решение. Всё это — непрерывные функции от p . Применение непрерывной функции не нарушает сходимости по вероятности.

Если параметр каким-то образом связан со средним, но нужно пробовать выборочные моменты.

Сформируем выборочное среднее и применим закон больших чисел

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} MX_1.$$

Для геометрического распределения

$$MX_1 = \frac{1-p}{p}.$$

Прибавим единицу слева и справа

$$1 + \bar{X} \xrightarrow{P} 1 + \frac{1-p}{p} = 1 + \frac{1}{p} - 1 = \frac{1}{p}.$$

Функция

$$f(x) = \frac{1}{x}$$

— это непрерывная функция, значит, можем применить эту функцию слева и справа, и сходимость сохранится

$$\frac{1}{1 + \overline{X}} \xrightarrow{P} p.$$

Состоятельной оценкой для параметра p будет

$$\hat{p} = \frac{1}{1 + \overline{X}}.$$

Состоятельной оценкой для p^2 будет $\hat{p}^2 = \hat{p}^2 \xrightarrow{P} p^2$, потому что $f(x) = x^2$ — это непрерывная функция.

Логарифм — это непрерывная функция

$$\ln \hat{p} = \ln \hat{p} = \ln \frac{1}{1 + \overline{X}} \xrightarrow{P} \ln p,$$

поскольку

$$\frac{1}{1 + \overline{X}} \xrightarrow{P} p.$$

2.12

Задание. Пусть X_1, \dots, X_n — выборка из распределения Пуассона с параметром $\lambda > 0$. Докажите, что не существует несмещённой оценки для параметра $1/\lambda$.

Решение. Будем действовать от противного.

Допустим, что такая оценка существует, то есть существует $\hat{\theta}$ такое, что

$$M\hat{\theta} = \frac{1}{\lambda}.$$

Это условие несмещённости.

$\hat{\theta}$ — функция от выборки, то есть $\hat{\theta} = f(X_1, \dots, X_n)$.

Распределение Пуассона дискретное

$$\begin{aligned} M\hat{\theta} &= Mf(X_1, \dots, X_n) = \\ &= \sum_{k_1=0}^{\infty} \dots \sum_{k_n=0}^{\infty} f(X_1, \dots, X_n) P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n). \end{aligned}$$

Воспользуемся независимостью

$$\begin{aligned} &\sum_{k_1=0}^{\infty} \dots \sum_{k_n=0}^{\infty} f(X_1, \dots, X_n) \cdot P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) = \\ &= \sum_{k_1=0}^{\infty} \dots \sum_{k_n=0}^{\infty} f(X_1, \dots, X_n) \cdot P(X_1 = k_1) \cdot \dots \cdot P(X_n = k_n). \end{aligned}$$

Подставим в явном виде вероятности

$$\begin{aligned} \sum_{k_1=0}^{\infty} \dots \sum_{k_n=0}^{\infty} f(X_1, \dots, X_n) \cdot P(X_1 = k_1) \cdot \dots \cdot P(X_n = k_n) = \\ = \sum_{k_1=0}^{\infty} \dots \sum_{k_n=0}^{\infty} f(X_1, \dots, X_n) \cdot \frac{\lambda^{\sum_{i=1}^n k_i}}{\prod_{i=1}^n k_i!} \cdot e^{-\lambda n}. \end{aligned}$$

Обозначим

$$\sum k_i = k$$

и получим

$$e^{-\lambda n} \cdot \sum_{k=0}^{\infty} \lambda^k \cdot \sum_{k_i: k_1+k_2+\dots+k_n=k} \frac{f(k_1, \dots, k_n)}{k_1! \dots k_n!} = e^{-\lambda n} \sum_{k=0}^{\infty} \lambda^k c_k,$$

где c_k — число.

Допустим, что эта величина равна $1/\lambda$.

Посмотрим, возможно ли это

$$e^{-\lambda n} \sum_{k=0}^{\infty} \lambda^k c_k = \frac{1}{\lambda}.$$

Запишем так, чтобы с одной стороны было $e^{\lambda n}$, а всё остальное перенесём

$$e^{\lambda n} = \sum_{k=0}^{\infty} \lambda^{k+1} c_k.$$

Для экспоненты существует одно развитие в ряд. Из этого следует, что последнее равенство невозможно (развитие в ряд начинается со степени λ , равной единице).

Домашнее задание

2.17

Задание. Пусть X_1, \dots, X_n — выборка из равномерного распределения на отрезке $[0, \theta]$. Проверьте несмещённость, состоятельность и найдите среднеквадратическое отклонение следующих оценок параметра θ :

а) $X_{(1)} + X_{(n)}$;

б) $(n+1) X_{(1)}$.

Решение. θ — максимальное наблюдение.

а) Проверим, выполняется ли $X_{(1)} + X_{(n)} \xrightarrow{P} \theta, n \rightarrow \infty$.

По определению сходимости по вероятности $\forall \varepsilon > 0 P(X_{(1)} > \varepsilon) = P(X_1 > \varepsilon, X_2 > \varepsilon, \dots, X_n > \varepsilon) = [P(X_1 > \varepsilon)]^n$. Подставим значение вероятности из геометрического эксперимента

$$[P(X_1 > \varepsilon)]^n = \left(\frac{\theta - \varepsilon}{\theta}\right)^n \rightarrow 0, n \rightarrow \infty.$$

Значит, $X_{(1)} \xrightarrow{P} 0, n \rightarrow \infty$.

Остаётся проверить, выполняется ли $X_{(n)} \xrightarrow{P} \theta, n \rightarrow \infty$.

По определению сходимости по вероятности

$$P\{|X_{(n)} - \theta| > \varepsilon\} \rightarrow 0, n \rightarrow \infty.$$

Раскроем модуль

$$\begin{aligned} P\{\theta - X_{(n)} > \varepsilon\} &= P\{X_{(n)} - \theta < -\varepsilon\} = P\{X_{(n)} < \theta - \varepsilon\} = \\ &= P(X_1 < \theta - \varepsilon, X_2 < \theta - \varepsilon, \dots, X_n < \theta - \varepsilon) = [P(X_1 < \theta - \varepsilon)]^n. \end{aligned}$$

Подставим значение вероятности из геометрического эксперимента

$$[P(X_1 < \theta - \varepsilon)]^n = \left(\frac{\theta - \varepsilon}{\theta}\right)^n = \left(1 - \frac{\varepsilon}{\theta}\right)^n \rightarrow 0, n \rightarrow \infty.$$

Число в скобках строго меньше единицы, так как $0 \leq \theta \leq 1, \varepsilon > 0$.

Отсюда следует, что оценка состоятельная.

Проверим несмещённость оценки. Проверим, выполняется ли

$$M(X_{(1)} + X_{(n)}) = \theta.$$

Из задачи 1.21

$$MX_{(n)} = \frac{\theta n}{n+1}$$

и

$$MX_{(1)} = \frac{\theta}{n+1}.$$

Из свойства линейности математического ожидания

$$M(X_{(1)} + X_{(n)}) = MX_{(1)} + MX_{(n)} = \frac{\theta n}{n+1} + \frac{\theta}{n+1} = \frac{\theta(n+1)}{n+1} = \theta.$$

Отсюда следует, что оценка несмещённая.

Формула среднеквадратического отклонения имеет вид $\sigma = \sqrt{D\xi}$.

Найдём дисперсию оценки

$$\begin{aligned}
D(X_{(1)} + X_{(n)}) &= M(X_{(1)} + X_{(n)})^2 - [M(X_{(1)} + X_{(n)})]^2 = \\
&= M(X_{(1)}^2 + X_{(n)}^2 + 2X_{(1)}X_{(n)}) - (MX_{(1)} + MX_{(n)})^2 = \\
&= MX_{(1)}^2 + MX_{(n)}^2 + 2M(X_{(1)}X_{(n)}) - (MX_{(1)})^2 - 2MX_{(1)}MX_{(n)} - \\
&\quad - (MX_{(n)})^2 = DX_{(1)} + DX_{(n)} + 2cov(X_{(1)}, X_{(n)}).
\end{aligned}$$

Возьмём необходимые значения из задачи 1.21, а именно

$$DX_{(1)} = \frac{\theta^2 n}{(n+1)^2 (n+2)} = DX_{(n)}, \quad cov(X_{(1)}, X_{(n)}) = \frac{\theta^2}{(n+1)^2 (n+2)}.$$

Подставляя в найденное выражение, получаем

$$\begin{aligned}
DX_{(1)} + DX_{(n)} + 2cov(X_{(1)}, X_{(n)}) &= \\
&= \frac{\theta^2 n}{(n+1)^2 (n+2)} + \frac{\theta^2 n}{(n+1)^2 (n+2)} + \frac{2\theta^2}{(n+1)^2 (n+2)} = \\
&= \frac{2\theta^2 n + 2\theta^2}{(n+1)^2 (n+2)} = \frac{2\theta^2 (n+1)}{(n+1)^2 (n+2)} = \frac{2\theta^2}{(n+1)(n+2)}.
\end{aligned}$$

Извлекая корень, получаем

$$\sigma = \sqrt{\frac{2\theta^2}{(n+1)(n+2)}} = \theta \sqrt{\frac{2}{(n+1)(n+2)}};$$

b) проверим, выполняется ли $(n+1)X_{(1)} \xrightarrow{P} \theta, n \rightarrow \infty$.

По определению сходимости по вероятности

$$\forall \varepsilon > 0 \quad P(|(n+1)X_{(1)} - \theta| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty.$$

Перейдём к противоположному событию

$$P(|(n+1)X_{(1)} - \theta| > \varepsilon) = 1 - P\{|(n+1)X_{(1)} - \theta| \leq \varepsilon\}.$$

Раскроем модуль

$$\begin{aligned}
1 - P\{|(n+1)X_{(1)} - \theta| \leq \varepsilon\} &= \\
&= 1 - P\{-\varepsilon \leq (n+1)X_{(1)} - \theta \leq \varepsilon\} = \\
&= 1 - P\left\{\frac{-\varepsilon + \theta}{n+1} \leq X_{(1)} \leq \frac{\varepsilon + \theta}{n+1}\right\} = \\
&= 1 + P\left\{X_{(1)} \leq \frac{\theta - \varepsilon}{n+1}\right\} - P\left\{X_{(1)} < \frac{\varepsilon + \theta}{n+1}\right\} = \\
&= 1 + 1 - P\left(X_1 > \frac{\theta - \varepsilon}{n+1}, \dots, X_n > \frac{\theta - \varepsilon}{n+1}\right) - 1 + P\left(X_{(1)} > \frac{\varepsilon + \theta}{n+1}\right) = \\
&= 1 - \left[P\left(X_1 > \frac{\theta - \varepsilon}{n+1}\right)\right]^n + \left[P\left(X_1 > \frac{\theta + \varepsilon}{n+1}\right)\right]^n.
\end{aligned}$$

Подставим значения вероятностей из геометрического эксперимента

$$\begin{aligned} 1 - \left[P\left(X_1 > \frac{\theta - \varepsilon}{n+1}\right) \right]^n + \left[P\left(X_1 > \frac{\theta + \varepsilon}{n+1}\right) \right]^n &= \\ &= 1 - \left(\frac{\theta - \frac{\theta - \varepsilon}{n+1}}{\theta} \right)^2 + \left(\frac{\theta - \frac{\varepsilon + \theta}{n+1}}{\theta} \right)^n = \\ &= 1 - \left(1 - \frac{\theta - \varepsilon}{(n+1)\theta} \right)^n + \left(1 - \frac{\varepsilon + \theta}{(n+1)\theta} \right)^n \not\rightarrow 0, n \rightarrow \infty. \end{aligned}$$

Отсюда следует, что оценка несостоятельная.

Проверим несмещённость оценки. Проверим, выполняется ли

$$M(n+1)X_{(1)} = \theta.$$

Выносим константу из-под знака математического ожидания

$$(n+1)MX_{(1)} = (n+1) \cdot \frac{\theta}{n+1} = \theta.$$

Отсюда следует, что оценка несмещённая.

Найдём дисперсию оценки

$$D(n+1)X_{(1)} = (n+1)^2 DX_{(1)} = (n+1)^2 \cdot \frac{\theta^2 n}{(n+1)^2 (n+2)} = \frac{\theta^2 n}{n+2},$$

откуда среднееквадратическое отклонение равно

$$\sigma = \sqrt{D(n+1)X_{(1)}} = \sqrt{\frac{\theta^2 n}{n+2}} = \theta \sqrt{\frac{n}{n+2}}.$$

2.18

Задание. Пусть X_1, \dots, X_n — выборка из показательного распределения с параметром $1/\sqrt{\alpha}$. Выясните, является ли статистика $\hat{\alpha}_n = (\bar{X})^2$ несмещённой оценкой параметра α . Является ли эта оценка состоятельной?

Решение. Нужно проверить, выполняется ли $M(\bar{X})^2 = \alpha$.

Запишем, что означает выборочное среднее

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

Квадрат суммы запишем в виде двух сумм. Получим

$$\frac{1}{n^2} \cdot M\left(\sum_{i=1}^n X_i^2 + 2 \sum_{k < i} X_k X_i\right) = \frac{1}{n^2} \cdot M \sum_{i=1}^n X_i^2 + \frac{2}{n^2} M \sum_{k < i} X_k X_i.$$

Случайные величины независимы и одинаково распределены, поэтому

$$\frac{1}{n^2} \cdot M \sum_{i=1}^n X_i + \frac{2}{n^2} M \sum_{k < i} X_k X_i = \frac{1}{n^2} \sum_{i=1}^n M X_i^2 + \frac{2}{n^2} \sum_{k < i} M X_k M X_i.$$

Для показательного распределения

$$M X_i = \frac{1}{\lambda} = \frac{1}{\frac{1}{\sqrt{\alpha}}} = \sqrt{\alpha}, \quad M X_i^2 = \frac{2}{\lambda^2} = \frac{2}{\frac{1}{\alpha}} = 2\alpha.$$

Подставляем

$$\begin{aligned} \frac{1}{n^2} \sum_{i=1}^n M X_i^2 + \frac{2}{n^2} \sum_{k < i} M X_k M X_i &= \frac{1}{n^2} \cdot n \cdot 2\alpha + \frac{1}{n^2} \cdot (n-1)n (M X_1)^2 = \\ &= \frac{2\alpha}{n} + \frac{n-1}{n} \cdot \alpha = \frac{2\alpha}{n} + \alpha - \frac{\alpha}{n} = \frac{\alpha}{n} + \alpha \rightarrow \alpha, \quad n \rightarrow \infty, \end{aligned}$$

значит, оценка смещённая, но несмещённая асимптотически.

Проверим, имеет ли место $(\bar{X})^2 \xrightarrow{P} \alpha, n \rightarrow \infty$.

По закону больших чисел

$$(\bar{X})^2 = \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \rightarrow (M X_1)^2 = (\sqrt{\alpha})^2 = \alpha,$$

значит, оценка состоятельная.

2.21

Задание. Пусть X_1, \dots, X_n — выборка из биномиального распределения с параметрами 2 и p . Для какого параметра $\theta = \theta(p)$ статистика $\hat{\theta}_n = e^{\bar{X}}$ будет состоятельной? Является ли $\hat{\theta}_n$ сильно состоятельной оценкой того же параметра? Является ли $\hat{\theta}_n$ несмещённой оценкой того же параметра? Найдите среднеквадратическое отклонение этой оценки.

Решение.

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

По закону больших чисел

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} M X_1, \quad n \rightarrow \infty.$$

Случайные величины в выборке имеют биномиальное распределение с математическим ожиданием $M X_1 = np = 2p$, значит, $\bar{X} \xrightarrow{P} 2p, n \rightarrow \infty$.

Применим непрерывную функцию e^x . Получим $e^{\bar{X}} \xrightarrow{P} e^{2p}, n \rightarrow \infty$.

Проверим, является ли оценка несмещённой к параметру e^{2p} , то есть выполняется ли $M e^{\bar{X}} = e^{2p}$.

Вычисляем $Me^{\bar{X}} = Me^{\frac{1}{n} \sum_{i=1}^n X_i}$. Случайные величины независимы и одинаково распределены, поэтому $Me^{\frac{1}{n} \sum_{i=1}^n X_i} = \left(Me^{\frac{X_1}{n}}\right)^n$. По определению характеристической функции $\varphi_{X_1}(t) = Me^{itX_1}$, получим

$$\left(Me^{\frac{X_1}{n}}\right)^n = \left[\varphi_{X_1}\left(\frac{1}{in}\right)\right]^n.$$

Характеристическая функция биномиального распределения

$$\varphi_{X_1}(t) = Me^{itX_1} = [(e^{it} - 1)p + 1]^n.$$

Подставим и получим

$$\left[\varphi_{X_1}\left(\frac{1}{in}\right)\right]^n = \left[(e^{\frac{i}{in}} - 1)p + 1\right]^n = \left[(e^{\frac{1}{n}} - 1)p + 1\right]^n \neq e^{2p}.$$

Значит, оценка смещённая.

Проверим сильную состоятельность. По усиленному закону больших чисел

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{a.s.} MX_1 = 2p, n \rightarrow \infty.$$

Отсюда следует, что $e^{\bar{X}} \xrightarrow{a.s.} e^{2p}, n \rightarrow \infty$. Значит, оценка сильно состоятельная.

$$\text{Найдём дисперсию оценки } De^{\bar{X}} = Me^{2\bar{X}} - \left(Me^{\bar{X}}\right)^2.$$

$$\text{Нашли, что } Me^{\bar{X}} = \left[(e^{\frac{1}{n}} - 1)p + 1\right]^n.$$

Вычисляем $Me^{2\bar{X}} = Me^{\frac{2}{n} \sum_{i=1}^n X_i}$. Из независимости и одинаковой распределённости случайных величин следует, что $Me^{\frac{2}{n} \sum_{i=1}^n X_i} = \left(Me^{\frac{2X_1}{n}}\right)^n$. По определению характеристической функции $\varphi_{X_1}(t) = Me^{itX_1}$, получаем

$$\left(Me^{\frac{2X_1}{n}}\right)^n = \left[\varphi_{X_1}\left(\frac{2}{in}\right)\right]^n.$$

Подставляем характеристическую функцию биномиального распределения

$$\left[\varphi_{X_1}\left(\frac{2}{in}\right)\right]^n = \left[(e^{\frac{2i}{in}} - 1)p + 1\right]^n = \left[(e^{\frac{2}{n}} - 1)p + 1\right]^n.$$

Отсюда находим дисперсию оценки

$$De^{\bar{X}} = \left[(e^{\frac{2}{n}} - 1)p + 1\right]^n - \left[(e^{\frac{1}{n}} - 1)p + 1\right]^{2n}.$$

Извлекая корень, получим $\sigma = \sqrt{\left[(e^{\frac{2}{n}} - 1)p + 1\right]^n - \left[(e^{\frac{1}{n}} - 1)p + 1\right]^{2n}}$.

Занятие 3. Метод моментов построения оценок

Контрольные вопросы и задания

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?

$M_{\theta} (\hat{\theta} - \theta)$ — среднеквадратическое отклонение.

Сформулируйте утверждение про поведение выборочных моментов.

Выборочный начальный момент M_k k -го порядка стремится к начальному моменту ν_k случайной величины X , то есть

$$\lim_{n \rightarrow \infty} P(|M_k - \nu_k| \geq \varepsilon) = 0,$$

для любого сколь угодно малого $\varepsilon > 0$, если моменты ν_{2k} и ν_k случайной величины X существуют и конечны.

Сформулируйте основную идею метода моментов построения оценки неизвестного параметра.

x_1, \dots, x_n — выборка из распределения F_θ и $\theta : M_\theta f(x_1) = g(\theta)$.

Вычисляем математическое ожидание, считая, что x_1 имеет распределение с параметром θ , иными словами,

$$M_\theta f(x) = \int_{\mathbb{R}} x dF_\theta(x), \quad f, g \in C(\mathbb{R}), \quad g$$

— строго монотонная. Тогда в силу усиленного закона больших чисел

$$\theta \approx g^{-1} \left(\frac{1}{n} \sum_{k=1}^n f(x_k) \right) \stackrel{a.s.}{\rightarrow} \theta, \quad n \rightarrow \infty.$$

Если существуют непрерывные f и g , g — обратима и $g(\theta) = M_\theta f(x)$, то в качестве оценки можно выбрать

$$\hat{\theta} = g^{-1} \left(\int_{\mathbb{R}} f dF_n \right).$$

Аудиторные задачи

3.3

Задание. Пользуясь методом моментов, оцените параметр θ равномерного распределения на отрезке:

- a) $[0, \theta]$;
- b) $[\theta - 1, \theta + 1]$;
- c) $[0, 2\theta]$;
- d) $[-\theta, \theta]$.

Решение.

- a) $X_i \sim U([0, \theta])$. Записываем теоретический момент. Для равномерного распределения — это середина отрезка

$$MX_1 = \frac{\theta}{2}.$$

Должны приравнять

$$\frac{\theta^*}{2} = \bar{X},$$

откуда $\theta^* = 2\bar{X}$;

b) случайные величины имеют распределение $X_i \sim U([\theta - 1, \theta + 1])$.

Вычисляем теоретический момент $MX_1 = \theta$. Должны записать, что $\theta^* = \bar{X}$;

с) случайные величины имеют распределение $X_i \sim U([0, 2\theta])$.

Вычисляем теоретический момент $MX_1 = \theta$, откуда $\theta^* = \bar{X}$;

d) случайные величины имеют распределение $X_i \sim U([- \theta, \theta])$.

Вычисляем теоретический момент $MX_1 = 0$ — не подходит, потому что не является функцией от θ . Можем вычислить второй момент, который в данном случае совпадает с дисперсией $MX_1^2 = DX_1$. Для равномерного распределения

$$DX_1 = \frac{4\theta^2}{12} = \frac{\theta^2}{3}.$$

Должны записать уравнение

$$\frac{(\theta^2)^*}{3} = \overline{X^2},$$

откуда $(\theta^2)^* = 3\overline{X^2}$. Извлекая корень, получаем $\theta^* = \sqrt{3\overline{X^2}}$.

3.4

Задание. Пусть X_1, \dots, X_n — выборка из распределения Пуассона с параметром λ . Пользуясь методом моментов, постройте оценку параметра λ и убедитесь, что эта оценка является несмещённой и состоятельной.

Решение. Первый теоретический момент $MX_1 = \lambda$. Должны приравнять

$$\lambda^* = \frac{1}{n} \sum_{i=1}^n X_i.$$

По закону больших чисел $\lambda^* \xrightarrow{P} \lambda$. Отсюда следует состоятельность. $M\lambda^* = MX_1 = \lambda$. Отсюда следует несмещённость.

3.5

Задание. Пользуясь методом моментов с пробной функцией $g(y) = y$, оцените параметр сдвига $\beta \in \mathbb{R}$ показательного распределения с плотностью

$$f_\beta(y) = \begin{cases} e^{\beta-y}, & t \geq \beta, \\ 0, & y < \beta. \end{cases}$$

Решение. $Mg(X_1) = \overline{g(X)}$ — метод моментов.

Ищем теоретический момент, зная плотность распределения

$$MX_1 = \int_{\beta}^{\infty} ye^{\beta-y} dy.$$

Делаем замену $y - \beta = z$ и получаем

$$\int_{\beta}^{\infty} ye^{\beta-y} dy = \int_0^{\infty} (\beta + z) e^{-z} dz = \beta + 1.$$

Отсюда должны решить уравнение $\beta^* + 1 = \bar{X}$.

Отсюда $\beta^* = \bar{X} - 1$. Эта оценка несмещённая и состоятельная.

Домашнее задание