Оглавление

Занятие 1. Выборочные характеристики

Контрольные вопросы и задания

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

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

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

Вариационный ряд и его члены являются порядковыми статистиками.

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

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

Какими свойствами обладает эмпирическая функция распределения?

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

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

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

— выборочное среднее.

Выборочная дисперсия

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

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

Выборочный момент порядка k — это случайная величина

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

Аудиторные задачи

1.4

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

a) $\overline{X}$;

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 $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Вычислите математическое ожидание и дисперсию статистики

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

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

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

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

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

Из независимости X_i получаем

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

Так как X_i одинаково распределены, то все дисперсии одинаковы

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

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

$\overline{X}$ не обязательно буде принимать целые значения.

1.6

Задание. Вычислите математическое ожидание статистик:

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

b) $S_0^2 = 1/(n-1)\cdot \sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2$.

Решение.

a) Распишем каждую из величин

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

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

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

Подставляем C_n^2 и группируем

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

Вынесем общий множитель за скобки

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

Эта оценка смещена ассимптотически;

b) выразим S_0 через S. Раскроем квадрат

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

Имеем сумму n одинаковых слагаемых

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

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 $X_1, \ldots, X_n$:

a) распределение k-ой порядковой статистики $X_{(k)}$;

b) вероятность $P\left(X_{(k)} < y,\, X_{(k+1)} \geq y\right)$.

Решение.

a) Сделали упорядочивание случайных величин

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

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

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

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

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

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

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

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

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).

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

Выборочное среднее

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

Выборочная дисперсия

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

Она является несмещённой.

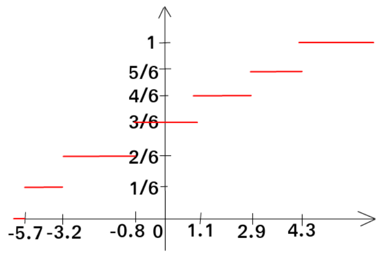

Рис. 1: Эмпирическая функция распределения

1.9

Задание. Вычислите вероятность $P\left(F_n\left(y\right) < F_n\left(z\right)\right)$.
Решение.

a) $y \geq z$. Событие невозможное, потому что $F_n\left(y\right) \geq F_n\left(z\right)$;

b) рассмотрим случай, когда $y < z$.

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

1.10

Задание. Пусть $X1, \ldots, X_n$ — выборка из распределения F с плотностью f. Найдите совместную плотность распределения всех порядковых статистик, то есть плотность распределения случайного вектора $\left(X_{(1)}, \ldots, X_{(n)}\right)$.

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

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

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

$$P\left(X_{(1)} \leq y_2\right) - P\left(X_{(1)} > y_1,\, X_{(2)} \leq y_2\right) =$$
$$= P\left(X_1 \leq y_2,\, X_2 \leq y_2\right) - P\left(X_1 \in \left(y_1, y_2\right],\, X_2 \in \left(y_1, y_2\right]\right).$$

Случайные величины X_1, X_2 — независимые и одинаково распределённые

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

Продифференцируем

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

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

$\left(X_{(1)},\ldots,X_{(n)}\right)\in\Delta$.

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

$$P\left\{\left(X_{(1)},\ldots,X_{(n)}\right)\in\Gamma\right\}=\int\limits_{\Gamma}f_{\left(X_{(1)},\ldots,X_{(n)}\right)}\left(z_1,\ldots,z_n\right)dz_1\ldots dz_n.$$

С другой стороны,

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

Учтём все перестановки

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

Подставим найденное выражение для вероятности

$$n!P\left\{\left(X_1,\ldots,X_n\right)\in\Gamma\right\}n!\cdot\int\limits_{\Gamma}f\left(z_1\right)\cdot\ldots\cdot f\left(z_n\right)dz_1\ldots dz_n.$$

Сравниваем полученные выражения

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

— плотность вектора упорядоченных статистик.

1.11

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

a) Докажите, что случайные величины $X_{(1)},X_{(2)}-X_{(1)},\ldots,X_{(n)}-X_{(n-1)}$ являются независимыми;

b) найдите распределение разности $X_{(k+1)}-X_{(k)}$ соседних порядковых статистик.

Решение. $\vec{\xi} = (\xi_1, \ldots, \xi_n)$ — 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.

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

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

Из задачи 1.10

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

Подставим плотность показательного распределения

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

Перемножим

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

Нужно найти линейное преобразование

$$\begin{bmatrix} X_{(1)} \\ X_{(2)} - X_{(1)} \\ X_{(3)} - X_{(2)} \\ \ldots \\ X_{(n)} - X_{(n-1)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 9 & \ldots & 0 & 0 & 0 \\ -1 & 1 & 0 & 9 & \ldots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \ldots & 0 & 0 & 0 \\ \ldots \\ 0 & 0 & 0 & 0 & \ldots & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} X_{(1))} \\ X_{(2)} \\ X_{(3)} \\ \ldots \\ X_{(n)} \end{bmatrix},$$

где

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

Определитель $detA = 1$.

Ищем обратную матрицу

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

Тогда имеем выражение

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

Определим искомый вектор через

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

Тогда

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

Разобъём на n множителей

$$n!\alpha^n e^{-\alpha(ny_1 + (n-1)y_2 + \dots + y_n)} \cdot \mathbb{1}\left\{0 < y_1 < y_1 + y_2 < \dots < y_1 + y_2 + \dots + y_n\right\} =$$
$$= \left[n\alpha e^{-\alpha n y_1} \cdot \mathbb{1}\left\{0 < y_1\right\}\right] \cdot \left[(n-1)\,\alpha e^{-\alpha(n-1)y_2} \cdot \mathbb{1}\left\{y_2 > 0\right\}\right] \cdot \ldots \times$$
$$\times \left[\alpha e^{-\alpha y_n} \cdot \mathbb{1}\left\{y_n > 0\right\}\right].$$

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

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

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

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

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

Из независимости X_i получаем

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

Так как X_i одинаково распределены, то все дисперсии одинаковы

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

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

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

находим

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

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

1.16

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

$$Y_i = F(X_i).$$

Решение. По определению

$$F_{\eta_1,\ldots,\eta_n}(X_1, \ldots, X_n) = P(\eta_1 \le X_1, \ldots, \eta_n \le X_n).$$

Воспользуемся независимостью

$$P(\eta_1 \le X_1, \ldots, \eta_n \le X_n) = P(\eta_1 \le X_1) \cdot \ldots \cdot P(\eta_n \le X_n).$$

Функция распределения i-й компоненты вектора равна

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

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

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

1.17

Задание. Пусть $X_1, \ldots, X_n$ — 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

$$X_{(1)} \le X_{(2)} \le \ldots \le X_{(k)} \le \ldots \le X_{(n)}.$$

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

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

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

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

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

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

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

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

Подставим полученное выражение в функцию распределения

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

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)

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

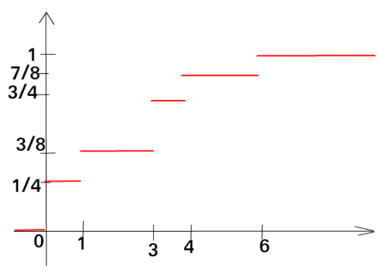

Рис. 2: Эмпирическая функция распределения

Выборочное среднее

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

Выборочная дисперсия

$$\hat{\sigma^2} = \frac{1}{7}\sum_{i=1}^{8}(X_i - 2.5)^2 =$$

$$= \frac{1}{7}\left[2(0-2.5)^2 + (1-2.5)^2 + 3(3-2.5)^2 + (4-2.5)^2 + (6-2.5)^2\right] =$$

$$= \frac{1}{7}(12.5 + 2.25 + 0.75 + 2.25 + 12.25) = \frac{30}{7} \approx 4.29.$$

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

$$x_1, \ldots, x_n,$$

называется функция

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

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

Для удобства выстроим числа в порядке их возрастания:

$$0, 0, \ldots, 0, 1, 1, \ldots, 1.$$

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.

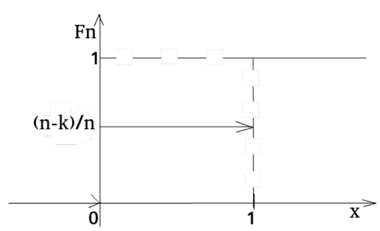

Рис. 3: Эмпирическая функция распределения

1.20

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

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

Посмотрим, что значит событие в указанной вероятности

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

Сократим константы в знаменателях

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

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

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

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, ассимптотически несмещённой, состоятельной, сильно состоятельной, оптимальной.

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

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

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

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

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 $\forall\theta \in \Theta$, $M_\theta\left(\hat{\theta} - \theta\right)^{\leq} M_\theta\left(\tilde{\theta} - \theta\right)^2$.

Что называется среднеквадратическим отклонением оценки?

$M_\theta\left(\hat{\theta} - \theta\right)$ — среднеквадратическое оклонение.

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?

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

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

Аудиторные задачи

2.4

Задание. Для выборки равномерного распределения на отрезке $[0,1]$ проверьте состоятельность и несмещённость оценки $X_{(1)}$ параметра θ.

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

По определению сходимости по вероятности

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

Раскроем модуль

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

Подставим значение вероятности из геометрического эксперимента

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

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

Отсюда следует, что оценка состоятельная.

Проверяем несмещённость оценки. Проверяем, выполняется ли

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

Нужно найти плотность

$$MX_{(1)}=\int\limits_{\mathbb{R}} f_{X_{(1)}}(y)\,ydy.$$

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

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

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

Продифференцируем

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

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

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

Приведём к общему знаменателю

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

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

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

Замена:

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

Подставляя замену, получаем

$$\int\limits_{\theta}^{1} y \cdot \frac{n}{(1-\theta)^2} \cdot (1-y)^{n-1}\, dy = n \cdot \frac{1}{(1-\theta)^n} \int\limits_{0}^{1-\theta} (1-z)\, z^{n-1} dz.$$

Вычислим интеграл

$$n \cdot \frac{1}{(1-\theta)^n} \int\limits_{0}^{1-\theta} (1-z)\, z^{n-1} dz \frac{n}{(1-\theta)^n}\left[\frac{(1-\theta)^n}{n} - \frac{(1-\theta)^{n+1}}{n+1}\right] =$$
$$= n\left(\frac{1}{n} - \frac{1-\theta}{n+1}\right) = 1 - \frac{n}{n+1}(1-\theta).$$

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 ассимптотически несмещённая, потому что

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

и

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

2.5

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

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

a) несмещённой оценкой для λ^2;

b) состоятельной оценкой для λ^2.

Решение.

a) Нужно проверить, выполняется ли

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

Преобразуем левую часть

$$M\frac{1}{n}\sum_{i=1}^{n} X_i^2 = MX_1^2 = DX_1 + (MX_1)^2 = \lambda + \lambda^2 \neq \lambda^2.$$

Значит, оценка смещённая;

b) проверяем, имеет ли место

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

По закону больших чисел

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \to MX_1^2 = \lambda^2 + \lambda \neq \lambda^2,$$

значит, оценка не состоятельная.

2.6

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

Решение. Нужно показать, что

$$\frac{1}{\overline{X}} \xrightarrow{P} \alpha,\ n \to \infty.$$

Выборочное среднее

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

По закону больших чисел

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

Отсюда следует, что

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

2.7

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

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

является несмещённой и состоятельной оценкой для σ^1.

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

Разность двух соседних элементов выборки имеет распределение

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

Найдём математическое ожидание статистики

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

Случайные величины одинаково распределены

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

В данном случае второй момент равен дисперсии

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

Отсюда следует, что оценка несмещённая.

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

Разобъём S_n на две суммы

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

В каждой из сумм слагаемые независимы

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

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

По закону больших чисел

$$\frac{1}{2(n-1)} \cdot \left[\frac{m}{m} \sum_{even\, i} (X_{i+1} - X_i)^2 + \frac{n-1-m}{n-1-m} \sum_{odd\, i} (X_{i+1} - X_i)^2\right] \xrightarrow{P}$$

$$\xrightarrow{P} \frac{1}{2} \cdot M (X_2 - X_1)^2 = \frac{1}{2} \cdot D (X_2 - X_1) = \sigma^2,\ n \to \infty.$$

Отсюда следует, что оценка состоятельная.

2.8

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

$$\hat{\theta_n} = e^{\overline{X}}$$

является состоятельной оценкой? Является ли $\hat{\theta_n}$ сильно состоятельной оценкой того же параметра? Является ли $\hat{\theta_n}$ несмещённой оценкой того же параметра? Ассимптотически немещённой?

Решение.

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

По закону больших чисел

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

СЛучайные величины в выборке имеют показательное распределение

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

значит,

$$\overline{X} \xrightarrow{P} \frac{1}{\alpha},\ n \to \infty.$$

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

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

Вычисляем $Me^{\overline{X}} = Me^{\frac{1}{n}\sum\limits_{i=1}^{n} X_i}$. Случайные величины независимы и одинаково распределены, поэтому $Me^{\frac{1}{n}\sum\limits_{i=1}^{n} X_i} = \left(Me^{\frac{X_1}{n}}\right)^n$. 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

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

Значит, оценка смещённая, но несмещённая ассимптотически.

Домашнее задание