Оглавление

Занятие 1. Выборочные характеристики **1**
Контрольные вопросы и задания 3
Аудиторные задачи 4
Домашнее задание 7

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

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

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 $F_{X_{(k)}}(y) = P\left(X_{(k)} \leq y\right) = P\{$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 — $X_i \leq y$.

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

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

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

Домашнее задание