Распределения и статистические оценки

1. Найти математическое ожидание выборочных начальных моментов.

2. Найти дисперсию выборочного среднего.

3. Найти математическое ожидание выборочной дисперсии.

4. Выборка $X_1, \dots, X_n$ соответствует распределению F(x). Найти распределение экстремальной порядковой статистики $X_{(n)}$.

5. Выборка $X_1, \dots, X_n$ соответствует распределению F(x). Найти распределение экстремальной порядковой статистики $X_{(1)}$.

6. Выборка $X_1, \dots, X_n$ соответствует распределению R(0,1). Найти математическое ожидание экстремальных порядковых статистик $X_{(1)}$ и $X_{(n)}$.

7. Выборка $X_1, \dots, X_n$ соответствует распределению F(x), а $\widehat{F}_n(x)$ - эмпирическая функция распределения, построенная по выборке $X_1, \dots, X_n$. Используя неравенство Чебышёва, для любого t>0 укажите верхнюю границу для следующей вероятности $P(\sqrt{n}\left|\widehat{F}_n(x) - F(x)\right| > t)$.

8. Выборка $X_1, \dots, X_n$ соответствует распределению F(x), а $\widehat{F}_n(x)$ - эмпирическая функция распределения, построенная по выборке $X_1, \dots, X_n$ достаточно большого объёма. Используя теорему Муавра-Лапласа, вычислите следующую вероятность $P\left(\sqrt{n}\left|\widehat{F}_n(x) - F(x)\right| > t\right)$ при любом t>0.

Домашнее задание

1. Подготовьте данные для вычисления выборочного коэффициента корреляции между экзаменационными оценками студентов вашей группы по двум (на ваш выбор) предметам.

2. Выборка $X_1, \dots, X_n$ соответствует экспоненциальному распределению E(1) и имеет достаточно большой объём, а $\widehat{F}_n(x)$ - эмпирическая функция распределения, построенная по этой выборке. Вычислить $P(\left|\widehat{F}_n(1) - F(1)\right| \le 1/\sqrt{n})$.

3. Выборка $X_1, \dots, X_n$ соответствует распределению $N(m, \theta^2)$ (параметр m известен, а θ^2 - неизвестен). Рассмотрим следующую статистику $T(X_1, \dots, X_n) = \sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_1^n |X_i - m|$. Найдите математическое ожидание СВ $T(X_1, \dots, X_n)$. Докажите, что при $n \to \infty$ последовательность $T(X_1, \dots, X_n)$ сходится почти наверное к θ.

4. Найти дисперсию к-го начального выборочного момента.

10.01 Семинар 1

по всем кого не было тоже скучаем :-

$Z_n = (X_1, X_2, \dots, X_n)$ – выборка

$z_n = (x_1, x_2, \dots, x_n)$ – реализ. выборки

работаем только с однородными одинаково распред. выборками

$X_{(1)} \le \dots \le X_{(n)}$ – вариац. ряд

$X_{(1)}, X_{(n)}$ – экстремальный элемент

Теоретич. момент	Выборочн. момент
$\mu_r = E[X^r]$	$\hat{\mu}_r = \frac{1}{n}\sum_{i=1}^{n}(X_i)^r$
EX	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
$D_r = E[(X-EX)^r]$	$\hat{D}_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^r$
$DX = E[(X-EX)^2]$	$\hat{d}_x = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$
$K_{xy} = E[(X-EX)(Y-EY)]$	$\hat{K}_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$
$\rho_{xy} = \frac{K_{xy}}{\sqrt{DX \cdot DY}}$	$\hat{\rho}_{xy} = \frac{\hat{K}_{xy}}{\sqrt{\hat{d}_x \cdot \hat{d}_y}}$

1. Найти математическое ожидание выборочных начальных моментов.

$$E[\hat{\mu}_k] = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i)^k\right] = \frac{1}{n}\sum_{i=1}^{n}E[(X_i)^k] = \cancel{\frac{1}{n}}\cdot\cancel{n}\cdot E[(X_1)^k] = E[(X_1)^k] = \mu_k$$

2. Найти дисперсию выборочного среднего.

$$D[\overline{X}] = D\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}D\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\cdot\sum_{i=1}^{n} D[X_i] = \frac{1}{n^2}\cdot n\cdot D[X_1] = \frac{1}{n}D[X_1]$$

$D[aX] = a^2 DX$

$D[X \pm Y] = DX + DY \pm \underbrace{2\,cov(x,y)}_{0}$

т.к. независим

$$\boxed{D[\overline{X}] = \frac{DX}{n}}$$

$E\overline{X} = EX$

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}E\left[\sum X_i\right] = \frac{1}{n}\sum EX_i = \frac{1}{n}\cdot nEX_1 = EX_1$$

3. Найти математическое ожидание выборочной дисперсии.

если смещенное

3. Найти математическое ожидание выборочной дисперсии.

$$\hat{d}_x = \frac{1}{n}\sum_{i=1}^{n}(x_k - \overline{x})^2$$

$$E\left[\frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2\right] = \frac{1}{n}\sum_{k=1}^{n}E[(x_k - \overline{x})^2] = (1)$$

$$\hat{d}_x = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2 = \frac{1}{n}\sum_{k=1}^{n}((x_k - m_x) - (\overline{x} - m_x))^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - m_x)^2 - 2(\overline{x} - m_x)\cdot\frac{1}{n}\sum_{k=1}^{n}(x_k - m_x) + \frac{1}{n}\sum_{k=1}^{n}(\overline{x} - m_x)^2$$

$$\frac{1}{n}\sum_{k=1}^{n}(x_k - m_x) = \frac{1}{n}\sum_{k=1}^{n}x_k - \underbrace{\frac{1}{n}\sum_{k=1}^{n}m_x}_{\frac{1}{n}\cdot n\cdot m_x} = \overline{x} - m_x$$

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

$$(1) = \frac{1}{n}E\left[\sum_{k=1}^{n}(x_k - m_x)^2\right] - E[(\overline{x} - m_x)^2] = \frac{1}{n}\sum_{k=1}^{n}\underbrace{E[(x_k - m_x)^2]}_{D_{x_k}} - \underbrace{E[(\overline{x} - m_x)^2]}_{D\overline{x}} =$$

$$= \frac{1}{n}\cdot n\cdot D_{x_k} - D\overline{x} = Dx - D\overline{x} = Dx - \frac{Dx}{n} = Dx\left(1 - \frac{1}{n}\right)$$

спасибо Арсению, что записал, пока я писала на доске

$\frac{n-1}{n}$ если несмещенная нужна

4. Выборка $X_1, \dots, X_n$ соответствует распределению F(x). Найти распределение экстремальной порядковой статистики $X_{(n)}$.

$$F_{X_{(n)}} = P(X_{(n)} \le x) = P(X \le x \;\dots\; X_n \le x) = \prod_{i=1}^{n} P(X_i \le x) = F_X^n(x)$$

5. Выборка $X_1, \dots, X_n$ соответствует распределению F(x). Найти распределение экстремальной порядковой статистики $X_{(1)}$.

$$F_{X_{(1)}}(x) = P(X_{(1)} \le x) = 1 - P(X_{(1)} \ge x) = 1 - P(X_1 \ge x, \dots, X_n \ge x) = 1 - \prod_{i=1}^{n}P(X_i \ge x) = 1 - (1 - F_X(x))^n$$

$$\begin{cases} f_{X_{(n)}}(x) = F'_{X_{(n)}}(x) = n f_X^{n-1}(x) \\ f_{X_{(1)}}(x) = n(1 - F_X(x))^{n-1} f_X(x) \end{cases}$$

6. Выборка $X_1, \dots, X_n$ соответствует распределению R(0,1). Найти математическое ожидание экстремальных порядковых статистик $X_{(1)}$ и $X_{(n)}$.

$$F_X(x) = \begin{cases} 1, & x>1 \\ x, & x \in [0,1] \\ 0, & x<1 \end{cases}$$

подставили в формулу $\frac{x-a}{b-a}$ сразу числа из условия

$$F_{X_{(n)}}(x) = \begin{cases} 1, & x>1 \\ x^n, & x \in [0,1] \\ 0, & x<0 \end{cases}$$

$$E X_{(n)} = \int_{-\infty}^{\infty} x f_{X_{(n)}}(x)\,dx = \int_0^1 n x^n\,dx = n \cdot \frac{x^{n+1}}{n+1}\Big|_0^1 = \frac{n}{n+1}$$

$$E[X_{(1)}] = \frac{1}{n+1}$$

13.01 Домашняя работа 1

1. Подготовьте данные для вычисления выборочного коэффициента корреляции между экзаменационными оценками студентов вашей группы по двум (на ваш выбор) предметам.

X – экзаменационные оценки за Теорвер

Y – экзаменационные оценки за Матстат

$X = \{10, 9, 8, 10, 7, 6, 10\}$

$Y = \{8, 10, 7, 9, 10, 6, 8\}$

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{7}(10+9+8+10+7+6+10) = 8\frac{4}{7} \approx 8{,}57$

$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{7}(8+10+7+9+10+6+8) = 8\frac{2}{7} \approx 8{,}29$

Теоретич. момент	Выборочн. момент
$\mu_r = E[X^r]$	$\hat{\mu}_r = \frac{1}{n}\sum_{i=1}^{n}(X_i)^r$
EX	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
$D_r = E[(X-EX)^r]$	$\hat{D}_r = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{X})^r$
$DX = E[(X-EX)^2]$	$\hat{D}x = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$
$K_{xy} = E[(X-EX)(Y-EY)]$	$\hat{K}_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{X})(y_i-\overline{Y})$
$\rho_{xy} = \frac{K_{xy}}{\sqrt{DX \cdot DY}}$	$\hat{\rho}_{xy} = \frac{\hat{K}_{xy}}{\sqrt{\hat{D}x \cdot \hat{D}y}}$

2. Выборка $X_1, \dots, X_n$ соответствует экспоненциальному распределению E(1) и имеет достаточно большой объём, а $\widehat{F}_n(x)$ - эмпирическая функция распределения, построенная по этой выборке. Вычислить $P(|\widehat{F}_n(1) - F(1)| \leq 1/\sqrt{n})$.

$F(x) = 1 - e^{-1} \approx 0{,}6321$ т.к. $X_1, \dots, X_n$ соответствует эксп. экспоненциальному распределению

по ЦПТ при достаточно большом n $\hat{F}_n(x)$ отклоняется от $F(x)$: $F_n(1) - F(1) \sim N\left(0, \frac{F(1)(1-F(1))}{n}\right)$

$\sigma = \sqrt{\frac{F(1)(1-F(1))}{n}}$

$P\left(|\hat{F}_n(1) - F(1)| \leq \frac{1}{\sqrt{n}}\right) = P\left(-\frac{1}{\sqrt{n}} \leq \hat{F}_n(1) - F(1) \leq \frac{1}{\sqrt{n}}\right)$

$z = \frac{\hat{F}_n(1) - F(1)}{\sigma}$ $\quad z \sim N(0,1)$

$1 = \frac{1}{\sqrt{n} \cdot \sigma}$

$P(-a \leq z \leq a) = 2\Phi_0(a) = 2\Phi_0\left(\sqrt{\frac{n}{0{,}2325}}\right) - 1$

для $\sigma = \sqrt{\frac{F(1)(1-F(1))}{n}}$

$a = \frac{1}{\sqrt{n}\sqrt{\frac{F(1)(1-F(1))}{n}}} = \sqrt{\frac{n}{F(1)(1-F(1))}} = \sqrt{\frac{n}{0{,}2325}}$

$F(1) = 1 - e^{-1} \approx 0{,}6321$

$1 - F(1) \approx 0{,}3676$

3. Выборка $X_1, \dots, X_n$ соответствует распределению $N(m, \theta^2)$ (параметр m известен, а θ^2 - неизвестен). Рассмотрим следующую статистику $T(X_1, \dots, X_n) = \sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_1^n |X_i - m|$. Найдите математическое ожидание СВ $T(X_1, \dots, X_n)$. Докажите, что при $n \to \infty$ последовательность $T(X_1, \dots, X_n)$ сходится почти наверное к θ.

$T(X_1, \dots, X_n) = \sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_{i=1}^{n}|X_i - m| \quad X_i \sim N(m, \theta^2)$

$z \sim N(0, \sigma^2)$

$E[|z|] = \sigma\sqrt{\frac{2}{\pi}}$ обобщенная формула для всех нормальных распределений $N(\mu, \sigma^2)$, где $\mu = 0$

(не классическое мат. ожидание E, т.к. $|z|$ нелинейная)

при $z = X_i - m \sim N(0, \theta^2)$, т.е. $\sigma = \theta$

$E[|X_i - m|] = \theta\sqrt{\frac{2}{\pi}}$

$E[T(X_1, \dots X_n)] = \sqrt{\frac{\pi}{2}} \cdot \frac{1}{n}\sum_{i=1}^{n} E[|X_i - m|] = \sqrt{\frac{\pi}{2}} \cdot \theta\sqrt{\frac{2}{\pi}} = \theta$

$\frac{1}{n}\sum_{i=1}^{n}|X_i - m| \xrightarrow{\text{п.н.}} E[|X_1 - m|] = \theta\sqrt{\frac{2}{\pi}}$ по ЗБЧ т.к. $|X_i - m|$ независима и имеет конечный второй момент

$$T(X_1, \dots, X_n) = \sqrt{\frac{\pi}{2}} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} |X_i - m| \xrightarrow{\text{п.н.}} \sqrt{\frac{\pi}{2}} \cdot \Theta \sqrt{\frac{2}{\pi}} = \Theta$$, чтд

4. Найти дисперсию к-го начального выборочного момента.

$$\mathcal{D}\left(\frac{1}{n} \sum_{i=1}^{n} (X_i)^k\right) = \frac{1}{n^2} \mathcal{D}\left(\sum_{i=1}^{n} (X_i)^k\right) =$$

$$= \frac{1}{n^2} \cdot n \cdot \mathcal{D}(X^k) = \frac{1}{n} \mathcal{D}(X^k)$$

Теоретич момент	Выборочн. момент
$\mu_r = E[X^r]$	$\hat{\mu}_r = \frac{1}{n} \sum_{i=1}^{n} (X_i)^r$
EX	$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
$D_r = E[(X - EX)^r]$	$\hat{D}_r = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^r$
$\mathcal{D}X = E[(X - EX)^2]$	$\hat{\mathcal{D}}x = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$
$K_{xy} = E[(x - EX)(y - EY)]$	$\hat{K}_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{y})$
$\rho_{xy} = \frac{K_{xy}}{\sqrt{\mathcal{D}x \cdot \mathcal{D}y}}$	$\hat{\rho}_{xy} = \frac{\hat{K}_{xy}}{\sqrt{\hat{\mathcal{D}}x \cdot \hat{\mathcal{D}}y}}$