Практика 8382, 26.09.2020

Тема: Эмпирическое распределение

⊖: Эмпирическая функция распределения, $X_1,\dots,X_n$ – набор наблюдений

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i<x\}} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_{(i)}<x\}},\ x\in\mathbb{R}$$

$X_{(1)}\le\dots\le X_{(n)}$ – порядковые статистики.

⊖: Эмпирическое распределение – распределение, имеющее ф.р. F_n

⊖: $\alpha: \mathcal{C}\to\mathbb{R}$ – числовая характеристика $\Rightarrow$ $\alpha(F_n)$ – выборочная хар-ка.

1. Получены результаты наблюдений $(-1, 0, 1, -2, 0, 1, 0, 2, -2, 3)$

а) Построить вар. ряд и ранги

б) Найти эмпирическую ф.р., эмпирическое распределение

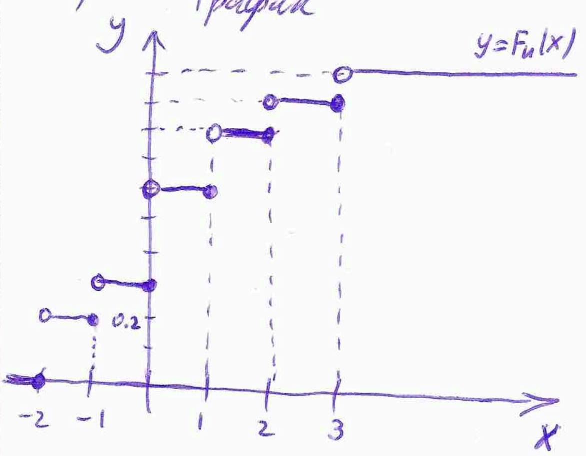
в) Найти выборочные аналоги мат. ожидания, дисперсии, медианы, квартилей.

Решение

а) Вар. ряд $(-2,-2,-1,0,0,0,1,1,2,3)$

Ранги $(3,\ 4\div6,\ 7\div8,\ 1\div2,\ 4\div6,\ 7\div8,\ 4\div6,\ 9,\ 1\div2,\ 10) \mapsto (3,4,7,1,5,8,6,9,2,10)$

б) График



$$F_n(x) = \begin{cases} 0, & x\le -2 \\ 0.2, & x\in(-2,-1] \\ 0.3, & x\in(-1,0] \\ 0.6, & x\in(0,1] \\ 0.8, & x\in(1,2] \\ 0.9, & x\in(2,3] \\ 1, & x>3 \end{cases}$$

Эмпирическое распределение – дискретное

зн s_n	−2	−1	0	1	2	3	Σ
вер	0.2	0.1	0.3	0.2	0.1	0.1	1

$$E_{F_n}X = \bar X = \frac{1}{10}\left(-2\cdot0.2+(-1)\cdot0.1+0\cdot0.3+1\cdot0.2+2\cdot0.1+3\cdot0.1\right) = 0.2$$

$$D_{F_n}X = S^2 = \frac{1}{10}\left((-2)^2\cdot0.2+(-1)^2\cdot0.1+0^2\cdot0.3+1^2\cdot0.2+2^2\cdot0.1+3^2\cdot0.1\right)-(\bar X)^2 = 2.36$$

$Med = z_{1/2} = X_{(5)} = 0$ $([X_{(5)}, X_{(6)}])$ – медиана

$Q1 = z_{1/4} = X_{(3)} = -1$ – квартиль $1/4$; $Q3 = z_{3/4} = X_{(8)} = 1$.

$\boxed{Д/З}$ 2. Получены результаты наблюдений $(1, -3, 0, 2, -1, 2, 0, 1, -2)$

Выполнить задание (а)–(в) из задачи 1 по полученным данным.

Тема: Достаточные статистики (параметрическое оценивание)

$X_1,\dots,X_n$ — выборка из распределения P_θ, $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$

p_θ — плотность распределения P_θ по отношению к доминирующей мере μ: $p_\theta = \frac{dP_\theta}{d\mu}$

а) μ — мера Лебега, если X_1 - абс. непрерывна

б) μ — считающая мера на множестве значений случайной величины X_1, если X_1 - дискретна

③ Если μ - мера Лебега $\Rightarrow p_\theta$ - плотность распределения;
если μ - считающая мера $\Rightarrow p_\theta(x) = P_\theta(X_1 = x)$, $x \in supp(X_1)$.

В силу независимости совместная плотность распределения

$$p_\theta(x_1 \dots x_n) = \prod_{i=1}^{n} p_\theta(x_i), \quad \theta \in \Theta.$$

Ⓞ: Функция правдоподобия

$$L(\vec{X};\theta) = p_\theta(x_1,\dots,x_n)$$

Ⓞ: Статистика $T: \mathbb{R}^n \to \mathbb{R}^k$ достаточная, если

$$P_\theta(\vec{X} \in A \mid T) = f(T) \text{ - не зависит от } \theta.$$

Теорема /Нейман - Фишер/

Статистика T - достаточная $\Leftrightarrow \exists\, g_\theta, h : L(\vec{X};\theta) = g_\theta(T(\vec{X}))h(\vec{X})$.

Ⓞ: Достаточная статистика T наз. минимальной, если для любой д.с. T^* существует функция g: $T = g(T^*)$

Некоторые примеры вычисления д.с. обсуждаются в курсе лекций.

3. Пусть $X_1,\dots,X_n$ - выборка из распределения с плотностью распределения p_θ. Найти минимальную достаточную статистику, если

а) $p_\theta(x) = \begin{cases} \theta e^{-\theta x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ (показательное)

б) $p_\theta(x) = \begin{cases} \frac{x^{p-1}e^{-x/b}}{\Gamma(p)b^p}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ $\theta = (p,b)$, $p > 0, b > 0$ (Гамма)

в) $p_\theta(x) = \begin{cases} a e^{-(x-b)\cdot a}, & x \geq b \\ 0, & x < 0 \end{cases}$ $\theta = (a,b)$ (показательное со сдвигом)

г) $p_\theta(x) = \begin{cases} \frac{1}{\theta}, & x \in [0,\theta] \\ 0, & x \notin [0,\theta] \end{cases}$ (равномерное)

д) $p_\theta(x) = \frac{1}{2b} e^{-\frac{|x-a|}{b}}$, $x \in \mathbb{R}$ (Лапласа)
$\theta = (a,b)$ $a \in \mathbb{R}, b > 0$

Решение (в) Перепишем плотность распределение в виде ③

$$p_\theta(x) = a\, e^{-(x-b)\cdot a}\, \mathbb{1}_{\{x \geq b\}}$$

Функция правдоподобия

$$L(\vec{X};\theta) = \prod_{i=1}^{n} a\, e^{-(X_i-b)\cdot a}\, \mathbb{1}_{\{X_i \geq b\}} = a^n e^{-a\sum_{i=1}^{n} X_i + ab} \underbrace{\prod_{i=1}^{n} \mathbb{1}_{\{X_i \geq b\}}}_{} =$$

равно 1 только если все $X_i \geq b \Leftarrow>$ $X_{(1)} \geq b$

$$= \underbrace{a^n e^{-a\boxed{\sum_{i=1}^{n} X_i} + ab}\, \mathbb{1}_{\{\boxed{X_{(1)}} \geq b\}}}_{=g_\theta(T(\vec{X}))}$$

$h(\vec{X}) \equiv 1$.

$\Leftarrow>$ $T(\vec{X}) = (\overline{X}, X_{(1)}) \sim \left(\sum_{i=1}^{n} X_i, X_{(1)}\right)$ — минимальная Д.С.

(г) Функция правдоподобия

$$L(\vec{X};\theta) = \prod_{i=1}^{n} \frac{1}{2b} e^{-\frac{|X_i-a|}{b}} = \frac{1}{2^n b^n} e^{-\frac{1}{b}\sum_{i=1}^{n}|X_i-a|} =$$

$$= \frac{1}{2^n b^n} e^{-\frac{1}{b}\sum_{i=1}^{n}|X_{(i)}-a|} = \underbrace{\frac{1}{2^n}}_{h(\vec{X})} \overbrace{\frac{1}{b^n} e^{-\frac{1}{b}\left(\sum_{i=1}^{K_a(\vec{X})}(a-X_{(i)}) + \sum_{i=K_a(\vec{X})+1}^{n}(X_{(i)}-a)\right)}}^{g_\theta(T)}$$

$K_a(\vec{X})$ - наибольшее значение i, при котором $X_{(i)} \leq a$

Очевидно $(X_{(1)},\dots,X_{(n)})$ — достаточная статистика. Для вычисления $g_\theta(T) = \frac{1}{b^n} e^{-\frac{1}{b}\left(a\cdot(2K_a(\vec{X})-n) + \sum_{i=1}^{K_a(\vec{X})}(-X_{(i)}) + \sum_{i=K_a(\vec{X})+1}^{n} X_{(i)}\right)}$ при всех значениях (a,b) необходимы значения $\sum_{i=1}^{t}(-X_{(i)}) + \sum_{i=t+1}^{n} X_{(i)}$ при всех $t = 1,2,\dots,n$, поскольку $K_a(\vec{X})$ принимает все значения от $1,\dots,n$ при подходящем выборе a, т.е

$$\left.\begin{array}{l} \sum_{i=1}^{n} X_{(i)} \\ -X_{(1)} + \sum_{i=2}^{n} X_{(i)} \\ \dots \\ -\sum_{i=1}^{n-1} X_{(i)} + X_{(n)} \end{array}\right] \longleftrightarrow (X_{(1)},\dots,X_{(n)}) = T(\vec{X})$$

Т.о. $T(\vec{X})$ — минимальная д.с.

Д/з. Выполнить задание (3.а); (3.б); (3.г).

4. Пусть $X_1, \dots, X_n$ – выборка из дискретного распределения с дискретной плотностью $q_\theta(x) = P_\theta(X_1 = x)$, $x \in E \subseteq \mathbb{Z}$. Найти минимальную достаточную статистику, если

а) $q_\theta(x) = p^x (1-p)^{1-x}$, $x \in \{0; 1\}$ (Бернулли), $\theta = p \in [0, 1]$

б) $q_\theta(x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$, $x \in \mathbb{N} \cup \{0\}$ (Пуассона), $\theta = \lambda > 0$

в) $q_\theta(x) = p(1-p)^x$, $x \in \mathbb{N} \cup \{0\}$ (Геометрическое), $\theta = p \in (0, 1)$

г) $q_\theta(x) = \dfrac{\Gamma(x+s)}{\Gamma(x+1)\Gamma(s)} p^s (1-p)^x$, $x \in \mathbb{N} \cup \{0\}$ (Отрицательное биномиальное) $\theta = (p, s)$, $p \in (0, 1)$; $s > 0$.

<u>Решение</u> (а) Функция правдоподобия

$$L(\vec{X}; \theta) = \prod_{i=1}^n p^{X_i} (1-p)^{1-X_i} = (1-p)^n p^{\sum_{i=1}^n X_i} (1-p)^{-\sum_{i=1}^n X_i} =$$

$$= \underbrace{(1-p)^n \left(\frac{p}{1-p}\right)^{\boxed{\sum_{i=1}^n X_i}}}_{g_\theta(T(\vec{X}))}, \quad X_i \in \{0; 1\}$$

$$T(\vec{X}) = \sum_{i=1}^n X_i \longleftrightarrow \overline{X}.$$

Д/З. Выполнить задание (4.б), (4.в), (4.г).