Круглов А.И. Б05-202 ST6 ~~1~~ ~~2~~ 3 4 5 ~~6~~ ~~7~~ 8

$\boxed{1}$ Найти достаточные статистики для параметрических семейств

a) $\mathcal{P} = \{\Gamma(\alpha, \beta): \alpha > 0, \beta > 0\}$. $p_\theta(x) = \frac{\alpha^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\alpha x}$, $p_\theta(x_1, \dots, x_n) = \frac{\alpha^{n\beta}}{\Gamma^n(\beta)} (\prod X_i)^{\beta - 1} e^{-\alpha \sum X_i}$,

по крит. Н.-Ф.: $(\prod_{i=1}^n X_i, \sum_{i=1}^n X_i)$ – дост. статистика (было на лекции)

b) $\mathcal{P} = \{U[a,b]: a < b\}$. $p_\theta(x) = \frac{1}{b-a} \mathbb{1}[a,b]$, $p_\theta(x_1, \dots, x_n) = \frac{1}{(b-a)^n} \mathbb{1}\{a \le X_{(1)} \le \dots \le X_{(n)} \le b\}$,

по крит. Н.-Ф.: $(X_{(1)}, X_{(n)})$ – дост. статистика

c) $\mathcal{P} = \{Pois(\theta): \theta > 0\}$. $p_\theta(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, $p_\theta(x_1, \dots, x_n) = \frac{\lambda^{\sum X_i}}{\prod(X_i!)} e^{-n\lambda}$, по крит. Н.-Ф.: $\sum_{i=1}^n X_i$ – дост. ст.

d) $\mathcal{P} = \{Geom(\theta): \theta \in (0,1)\}$. $p_\theta(x) = (1-p)^{x-1} p$, $p_\theta(x_1, \dots, x_n) = (1-p)^{\sum X_i - n} p^n$, по крит. Н.-Ф.: $\sum_{i=1}^n X_i$ – дост. ст.

Выписать правило обновления дост. статистик для последовательно поступающей выборки:

сумму надо сложить с новым эл-том ($\sum_{i=1}^n X_i \xrightarrow{X_{n+1}} \Sigma + X_{n+1}$), произведение – умножить на новый эл-т ($\prod_{i=1}^n X_i \xrightarrow{X_{n+1}} \Pi \cdot X_{n+1}$), для min\max взять новый min\max ($X_{(1)} \xrightarrow{X_{n+1}} \min(X_{(1)}, X_{n+1})$, $X_{(n)} \xrightarrow{X_{n+1}} \max(X_{(n)}, X_{n+1})$).

$\boxed{2}$ X – СВ с $p_\theta(x) = \frac{g(x)}{h(\theta)} e^{\theta^T u(x)}$, $\theta \in \Theta \subset \mathbb{R}^d$. Док-ть: $\operatorname{cov}_\theta(u_j(X), u_k(X)) = \frac{\partial^2 \ln h(\theta)}{\partial\theta_j \partial\theta_k}$

По виду p_θ ясно, что работаем с экспоненциальным семейством. Из условия нормировки $\int p_\theta(x) = 1$ следует, что $h(\theta) = \int g(x) e^{\theta^T u(x)} dx$. Тогда $\nabla_\theta \ln h(\theta) = \frac{\nabla_\theta \int g(x) e^{\theta^T u(x)} dx}{h(\theta)} = \int u(x) \cdot \frac{g(x)}{h(\theta)} e^{\theta^T u(x)} dx =$
$= \int u(x) \cdot p_\theta(x) dx = \mathbb{E} u(x)$; $\frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln h(\theta) = \mathbb{E}(u_j(x) u_k(x)) - \mathbb{E} u_j(x) \cdot \mathbb{E} u_k(x) = \operatorname{cov}_\theta(u_j(x), u_k(x))$, ч.т.д.

$\boxed{6}$ Найти асимп. толер. $\tau(\hat\theta)$ оценки Гальтона $\hat\theta = \operatorname{median}\{\frac{X_{(i)} + X_{(n-i+1)}}{2} : i = 1, \dots, \lfloor \frac{n+1}{2} \rfloor\}$

Упорядочим выборку по возрастанию: $X_{(1)}, \dots, X_{(n-k)}, X_{(n-k+1)}, \dots, X_{(n)}$, заменим k наибольших элементов на бесконечности: $X_{(1)}, \dots, X_{(n-k)}, +\infty, \dots, +\infty$.

$\hat\theta = \operatorname{median}\{\frac{X_{(1)} + X_{(n)}}{2}, \dots, \frac{X_{\lfloor \frac{n+1}{2} \rfloor} + X_{\lceil \frac{n+1}{2} \rceil}}{2}\} = \operatorname{median}\{\underbrace{+\infty, \dots, +\infty}_{k \text{ штук}}, \frac{X_{(k+1)} + X_{(n-k)}}{2}, \dots, \frac{X_{\lfloor \frac{n+1}{2} \rfloor} + X_{\lceil \frac{n+1}{2} \rceil}}{2}\}$

Чтобы median $\ne \infty$, нужно, чтобы конечных слагаемых оставалось хотя бы $k+1$ штук. Т.е.

$\lfloor \frac{n+1}{2} \rfloor - k \ge k+1 > k$, $\lfloor \frac{n+1}{2} \rfloor > 2k \Rightarrow k^* = \frac{n+1}{4}$. Тогда $\tau(\hat\theta) = \lim_{n\to\infty} \frac{k^*}{n} = \lim_{n\to\infty} \frac{n+1}{4n} = \frac{1}{4}$

$\boxed{7}$ Найти асимп. толер. $\tau(\hat\theta)$ медианы средних Уолша $W = \hat\theta = \operatorname{median}\{\frac{X_i + X_j}{2} : 1 \le i \le j \le n\}$

					∞		∞	∞
	j \ i	1	...	n-k	n-k+1	...	n
	1						
	⋮						
	n-k						
∞	n-k+1	∞	∞	∞	∞		
∞	⋮	∞	∞	∞	∞	∞	
∞	n	∞	∞	∞	∞	∞	∞

В таблице $1 \le i \le j \le n$ – левый нижний треугольник (с жирной границей). Пусть $X_{n-k+1}, \dots, X_n$ обратились в ∞ (т.е. последние k элементов). В интересующем нас треугольнике ячейки заполнены $\frac{X_i + X_j}{2}$, поэтому k нижних строк обратятся в ∞. В треугольнике $\frac{(1 + (n-k))(n-k)}{2}$ конечных и $\frac{(1+n)n}{2} -$
$- \frac{(1+n-k)(n-k)}{2} = \frac{k + 2kn - k^2}{2} = \frac{k(1+2n-k)}{2}$ бесконечных ячеек. Чтобы медиана по ячейкам тре-

$\boxed{7}$ (продолжение)

выборка осталась конечной, нужно гарантировать, чтобы $\frac{(1+n-k)(n-k)}{2} > \frac{k(1+2n-k)}{2}$. Решим:

$n-k+n^2-kn-kn+k^2 > k+2kn-k^2$, $n-2k+n^2-4kn+2k^2 > 0$, $k^2-k(1+2n)+\frac{n^2-n}{2} > 0$,

$k_{1,2} = \frac{1+2n \pm \sqrt{1+6n+2n^2}}{2} = n+\frac12 \pm \frac12\sqrt{1+6n+2n^2}$. Т.к. $k<n$, берем $\overset{+}{\ominus}$ в качестве k^*.

$\tau(\hat\theta) = \lim\limits_{n\to\infty}\frac{k^*}{n} = \lim\limits_{n\to\infty}\left(1+\cancel{\frac{1}{2n}}-\frac12\sqrt{\cancel{\frac{1}{n^2}}+\cancel{\frac{6}{n}}+2}\right) = 1-\frac{\sqrt2}{2} \approx 0{,}29$

$\boxed{8}$ $X_1,\dots,X_n \sim N(\theta,1)$ с шумом из $N(\theta,9)$, т.е. $p_\theta(x) = (1-\varepsilon)p_{\theta 1}(x) + \varepsilon p_{\theta 9}(x)$. ε: мед. эффективнее выборочного среднего —?

А.Д. медианы : $\frac{1}{4n p_\theta^2(\mu)} = \frac{1}{4n p_\theta^2(\theta)} = \frac{1}{4n((1-\varepsilon)p_{\theta 1}(\theta)+\varepsilon p_{\theta 9}(\theta))^2} =$

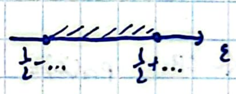
$= \frac{1}{4n\left((1-\varepsilon)\frac{1}{\sqrt{2\pi}}+\varepsilon\frac{1}{\sqrt{18\pi}}\right)^2} = \frac{2\pi}{4n((1-\varepsilon)+\varepsilon/3)^2} = \frac{\pi}{2n(1-\frac23\varepsilon)^2} = f$

А.Д. выборочного среднего: $\frac{\sigma^2}{n} = \frac{((1-\varepsilon)\cdot 1+\varepsilon\cdot 3)^2}{n} = \frac{(1+2\varepsilon)^2}{n} = g$

$\frac{g}{f} = \frac{(1+2\varepsilon)^2\cdot 2n(1-\frac23\varepsilon)^2}{n\cdot\pi} = \frac{2}{\pi}(1+2\varepsilon)^2(1-\frac23\varepsilon)^2 > 1 \longrightarrow (1+2\varepsilon)(1-\frac23\varepsilon) > \sqrt{\pi/2} \longrightarrow$

$-\frac43\varepsilon^2+\frac43\varepsilon+(1-\sqrt{\pi/2}) > 0 \longrightarrow \varepsilon^2-\varepsilon-\frac34(1-\sqrt{\pi/2}) < 0$, $\varepsilon_{1,2} = \frac{1\pm\sqrt{1+3(1-\sqrt{\pi/2})}}{2} =$

$= \frac{1\pm\sqrt{4-3\sqrt{\pi/2}}}{2} = \frac12 \pm \sqrt{1-\frac34\sqrt{\frac{\pi}{2}}}$, $\varepsilon_1 \approx 0.255$, $\varepsilon_2 \approx 0.745$



Ответ: при $\varepsilon \in (0.255,\ 0.745)$ (примерно)

$\boxed{5}$ $X_1,\dots,X_n \sim Laplace(\theta)$, т.е. $p_\theta(x) = \frac12 e^{-|x-\theta|}$. Сравнить в асимпт. подходе 4 оценки

Посчитаем асимптотические дисперсии всех оценок и сравним.

а) $\bar X$: $\sigma^2_{\bar X} = \frac1n\cdot\sigma^2_{Laplace} = \frac2n$; б) $\bar X_\alpha$:

в) $\hat\mu$:

$\boxed{3}$ $\mathcal{P} = \{\Gamma(\alpha,\beta) : \alpha>0, \beta>0\}$

естественная параметризация ✓

а) $p(x) = \frac{\alpha^\beta x^{\beta-1}}{\Gamma(\beta)}e^{-\alpha x} = \frac{g(x)}{h(\theta)}e^{a(\theta)^T u(x)}$, где $g(x) = x^{\beta-1}$, $h(\theta) = \frac{\Gamma(\beta)}{\alpha^\beta}$, $a(\theta) = \alpha$, $u(x) = x$

б) докажем, что $S(X) = (\sum_{i=1}^n X_i, \sum_{i=1}^n \ln X_i)$ — достаточная по крит. Н-Ф. $\prod_{i=1}^n p(x) = \frac{\alpha^{n\beta}}{\Gamma^n(\beta)}\prod_{i=1}^n x_i^{\beta-1}e^{-\alpha\sum X_i} =$

$= \frac{\alpha^{n\beta}}{\Gamma^n(\beta)}\exp\left(\ln\prod_{i=1}^n X_i^{\beta-1}\right)\exp\left(-\alpha\sum_{i=1}^n X_i\right) = \frac{\alpha^{n\beta}}{\Gamma^n(\beta)}\exp\left((\beta-1)\underbrace{\sum_{i=1}^n \ln X_i}-\alpha\underbrace{\sum_{i=1}^n X_i}\right) \Rightarrow S$ — дост. ст. по кр. Н-Ф.

в) $E\,S(X) = (E\sum X_i, E\sum\ln X_i) = (\sum EX_i, \sum E\ln X_i) = (n\cdot\frac{\beta}{\alpha}, n\cdot\ln\frac{\beta}{\alpha})$

$\operatorname{cov} S(x) = \begin{pmatrix} Var\sum X_i & cov(\sum X_i, \sum\ln X_i) \\ cov(\sum\ln X_i, \sum X_i) & var\sum\ln X_i \end{pmatrix} = \begin{pmatrix} n\frac{\beta}{\alpha^2} & n\beta\frac{\ln\alpha}{\alpha} \\ n\beta\frac{\ln\alpha}{\alpha} & n\ln\frac{\beta}{\alpha^2} \end{pmatrix}$

г) ОМП: $L_X(\theta) = \prod_{i=1}^n p_\theta(x) \overset{(б)}{=} \frac{\alpha^{n\beta}}{\Gamma^n(\beta)}\exp\left((\beta-1)\sum_{i=1}^n\ln X_i - \alpha\sum_{i=1}^n X_i\right)$, $\ell_X(\theta) = n\beta\ln\alpha - n\ln\Gamma(\beta) +$

$+(\beta-1)\sum_{i=1}^n \ln X_i - \alpha\sum_{i=1}^n X_i$; $\frac{\partial}{\partial\alpha}\ell_X(\theta) = \frac{n\beta}{\alpha} - \sum_{i=1}^n X_i = 0 \Rightarrow \beta = \alpha\bar X$, $\frac{\partial}{\partial\beta}\ell_X(\theta) = n\ln\alpha - \frac{n\Gamma'(\beta)}{\Gamma(\beta)} + \sum_{i=1}^n \ln X_i =$

$= 0 \Rightarrow \ln\alpha - \psi_0(\beta) + \overline{\ln X} = 0$. Тогда $\alpha = \exp(\psi_0(\alpha\bar X) - \overline{\ln X})$, $\beta = \bar X\exp(\psi_0(\beta) - \overline{\ln X})$

α, β — ? ...

4 Формулировки

а) а) Эксп. класс — сем-во распределений с $p_\theta(x) = \frac{g(x)}{h(\theta)} e^{a(\theta)^T u(x)}$. Если $a(\theta) = \theta$, пар-ия естественная

б) Достаточная статистика — статистика, хранящая всю информацию о выборке. Удобно хранить ее, а не всю выборку, и просто обновлять по ходу получения новых элементов выборки

в) матрица ковариаций $\mathrm{cov}(X) = E(XX^T) - EX\,EX^T$

г) А.Н.О. — оценка, чье распределение стремится к нормальному с ростом размера выборки

д) ОМП — оценка, оптимизирующая ф-цию правдоподобия $\prod_{i=1} p_\theta(x_i)$

б) Крит. Неймана-Фишера: S - дост. статистика $\Leftrightarrow$ $p(x) = h(S,\theta) \cdot g(x)$