ВЕРОВАТНОЋА И СТАТИСТИКА Б - ТЕСТ 2
ЈУН 2015.

1. Фамилија расподела $F(t;\theta)$ зависи од непознатог параметра θ.

а) Дефинисати регуларност фамилије $F(t;\theta)$. (1 поен)

б) Навести пример фамилије која није регуларна (и образложити). (1 поен)

Решење:

а) Фамилија је регуларна ако важи:

– $\frac{\partial}{\partial\theta}\sum\limits_{k=0}^{+\infty} p(t_k;\theta) = \sum\limits_{k=0}^{+\infty} \frac{\partial}{\partial\theta} p(t_k;\theta) \Leftrightarrow \sum\limits_{k=0}^{+\infty} \frac{\partial}{\partial\theta} p(t_k;\theta) = 0$, где је $p(t_k;\theta) = P\{X = t_k\}$ (дискретан случај)

– $\frac{\partial}{\partial\theta}\int\limits_{-\infty}^{+\infty} f(t;\theta)dx = \int\limits_{-\infty}^{+\infty} \frac{\partial}{\partial\theta} f(t;\theta)dx \Leftrightarrow \int\limits_{-\infty}^{+\infty} \frac{\partial}{\partial\theta} f(t;\theta)dx = 0$, где је $f(t;\theta)$ густина расподеле (непрекидан случај)

б) Било која фамилија код које се параметар налази у границама области дефинисаности. На пример, $U[0,\theta]$:
$\int\limits_{-\infty}^{+\infty} \frac{\partial}{\partial\theta} f(t;\theta)dt = \int\limits_{0}^{\theta} \frac{\partial}{\partial\theta}\frac{1}{\theta}dt = \int\limits_{0}^{\theta}(-\frac{1}{\theta^2})dt = -\frac{1}{\theta} \neq 0$

2. Нека је статистика $\hat\theta = \hat\theta(X_1,\ldots,X_n)$ оцена непознатог параметра θ расподеле $G(t;\theta)$.

а) Под којим условом је оцена $\hat\theta$ асимптотски непристрасна? (1 поен)

б) Под којим условом је оцена $\hat\theta$ постојана у средњем реда 4.4? (1 поен)

Решење:

а) Оцена је асимптотски непристрасна ако важи: $E\widehat{\theta_n} \to \theta, n \to +\infty$

б) Оцена је постојана у средњем реда 4.4 ако конвергира у средњем реда 4.4 ка параметру θ, тј. ако важи: $E|\widehat{\theta_n} - \theta|^{4.4} \to 0, n \to +\infty$.

3. Фамилија раподела $F(t;\theta)$ зависи од непознатог параметра θ.

а) Ако је расподела $F(t;\theta)$ дискретна, написати функцију веродостојности. (1 поен)

б) Написати функцију веродостојности у случају униформне $\mathscr{U}[0,\theta]$ расподеле. (2 поена)

Решење:

а) $L(\theta) = \prod\limits_{k=1}^{n} p(t_k;\theta)$, где је $p(t_k;\theta) = P\{X = t_k\}, k = 1,...,n$

б) $L(\theta) = \prod\limits_{k=1}^{n} f(t_k;\theta) = \prod\limits_{k=1}^{n} \frac{1}{\theta} I\{t_k \le \theta\} = \frac{1}{\theta^n} I\{\theta \ge t_{(n)}\}$, где је $t_{(n)} = \max\{t_1,...,t_n\}$

4. У 200 бацања две коцкице добијен је збир 7 у 20 бацања, а збир 11 у 14 бацања (остали збирови су добијени 166 пута). Пирсоновим тестом са прагом значајности 0.05 испитати да ли су коцкице исправне. (3 поена)

Решење:

X_k	7	11	остали
M_k	20	14	166
p_k	$\frac{1}{6}$	$\frac{2}{36}$	$\frac{28}{36}$
np_k	33.33	11.11	155.56

$p_1 = P\{X = 7\} = \frac{6}{6}\cdot\frac{1}{6} = \frac{1}{6}$ (повољни исходи: 16,25,34,43,52,61)

$p_2 = P\{X = 11\} = \frac{2}{6}\cdot\frac{1}{6} = \frac{1}{18}$ (повољни исходи: 56 и 65)

$p_3 = 1 - p_1 - p_2 = \frac{28}{36}$

$\chi_0^2 = \sum\limits_{k=1}^{3} \frac{(M_k - np_k)^2}{np_k} = \frac{(20-33.33)^2}{33.33} + \frac{(14-11.11)^2}{11.11} + \frac{(166-155.56)^2}{155.56} = 6.7842$

$W = \{\chi_0^2 \ge c\}, c = \chi^2_{3-1,0.05} = 5.991$

$6.7842 > 5.991 \Rightarrow$ одбацује се H_0, тј. коцкице нису исправне

ВЕРОВАТНОЋА И СТАТИСТИКА Б - КОЛОКВИЈУМ 2
ЈУН 2015.

1. Обележје X има гама расподелу (θ - непознат параметар) са густином:

$$f(t) = \frac{t \cdot e^{-\frac{t}{\theta}}}{\theta^2},\ t \geq 0\ .$$

а) Записати функцију веродостојности (на основу узорка обима n). (2 поена)

б) Оценити непознати параметар θ методом максималне веродостојности. (4 поена)

в) Испитати ефикасност тако добијене оцене. (4 поена)

Напомена: Математичко очекивање случајне величине X је $E(X) = 2\theta$, а дисперзија је $D(X) = 2\theta^2$.

Решење:

а) $L(\theta) = \prod\limits_{k=1}^{n} f(t;\theta) = \frac{\prod\limits_{k=1}^{n} t_k}{\theta^{2n}} e^{-\frac{1}{\theta}\sum\limits_{k=1}^{n} t_k}$

б) $\ln L(\theta) = -\frac{1}{\theta}\sum\limits_{k=1}^{n} t_k + \ln\prod\limits_{k=1}^{n} t_k - 2n\ln\theta$

Оцена максималне веродостојности се добија из једначине: $\frac{\partial \ln L(\theta)}{\partial\theta} = 0$

$\frac{\partial \ln L(\theta)}{\partial\theta} = \frac{1}{\theta^2}\sum\limits_{k=1}^{n} t_k - \frac{2n}{\theta} = 0 \Rightarrow \theta = \frac{\sum\limits_{k=1}^{n} t_k}{2n} \Rightarrow \hat{\theta} = \frac{\overline{X_n}}{2}$

в) – **регуларност фамилије:**

$\int\limits_0^{+\infty} \frac{\partial}{\partial\theta}\frac{te^{-\frac{t}{\theta}}}{\theta^2}dt = \int\limits_0^{+\infty} \frac{te^{-\frac{t}{\theta}}\frac{x}{\theta^2}\theta^2 - 2\theta te^{-\frac{t}{\theta}}}{\theta^4}dt = \int\limits_0^{+\infty} \frac{t^2e^{-\frac{t}{\theta}} - 2\theta te^{-\frac{t}{\theta}}}{\theta^4}dt = \left| \begin{array}{c} \frac{t}{\theta} = x \\ \frac{dt}{\theta} = dx \end{array} \right|$

$= \frac{1}{\theta^3}\int\limits_0^{+\infty} \theta^2 x^2 e^{-x}dx - \frac{2}{\theta^2}\int\limits_0^{+\infty} \theta x e^{-x}dx = \frac{1}{\theta}\Gamma(3) - \frac{2}{\theta}\Gamma(2) = \frac{2}{\theta} - \frac{2}{\theta} = 0$

$\Rightarrow$ фамилија је регуларна

– **непристрасност:**

$E\hat{\theta} = E(\frac{1}{2}\overline{X_n}) = \frac{1}{2n}\sum\limits_{k=1}^{n} EX_k = \frac{1}{2n}n2\theta = \theta$

$\Rightarrow$ оцена је непристрасна

– **Рао-Крамерова граница:** $G = \frac{1}{nI(\theta)}$

$I(\theta) = E\left(\frac{\partial \ln f(t;\theta)}{\partial\theta}\right)^2$

$\frac{\partial \ln f(t;\theta)}{\partial\theta} = \frac{\partial(\ln t - \frac{t}{\theta} - 2\ln\theta)}{\partial\theta} = \frac{t}{\theta^2} - \frac{2}{\theta}$

$I(\theta) = E\left(\frac{X}{\theta^2} - \frac{2}{\theta}\right) = \frac{1}{\theta^4}EX^2 - \frac{4}{\theta^3}EX + \frac{4}{\theta^2} = \frac{1}{\theta^4}(2\theta^2 + 4\theta^2) - \frac{4}{\theta^3}2\theta + \frac{4}{\theta^2} = \frac{2}{\theta^2}$

$G = \frac{1}{n\frac{2}{\theta^2}} = \frac{\theta^2}{2n}$

– **Дисперзија оцене:**

$D\hat{\theta} = D(\frac{1}{2}\overline{X_n}) = \frac{1}{4n^2}\sum\limits_{k=1}^{n} D(X_k) = \frac{\theta^2}{2n}$

$\Rightarrow G = D\hat{\theta}$ па је оцена ефикасна.

2. Сабрано је 200 бројева из интервала $[0, 2\theta]$ и добијен је збир 65.044 (сабирци имају униформну $\mathscr{U}[0, 2\theta]$ расподелу). Одредити 90%-тни интервал поверења за непознати параметар θ. (2 поена)

Решење:

$n > 30 \Rightarrow$ може се применити Централна гранична теорема: $\frac{\sum\limits_{k=1}^{n} X_k - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}} \to X^* : \mathscr{N}(0,1), n \to +\infty$

$$P\left\{\left|\frac{\sum\limits_{k=1}^{n} X_k - E(\sum\limits_{k=1}^{n} X_k)}{\sqrt{D(\sum\limits_{k=1}^{n} X_k)}}\right| < z\right\} = \beta$$

$z = \Phi^{-1}(0.45) = 1.65, EX_k = \frac{2\theta}{2}, DX_k = \frac{4\theta^2}{12}$

$$-z < \frac{\sum\limits_{k=1}^{n} X_k - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}} < z$$

$$-1.65 < \frac{65.044 - 200\theta}{\sqrt{200\frac{\theta^2}{3}}} < 1.65$$

$$-1.65 < \frac{7.9662}{\theta} - 24.4948 < 1.65$$

$$I_\theta = (0.3047, 0.3487)$$

3. Обележје X има бета $\beta(\theta, 1)$ расподелу са густином:

$$f(t) = \theta \cdot t^{\theta-1},\ t \in (0,1)\ .$$

а) Са прагом значајности 0.1, одредити најбољу критичну област за тестирање $H_0(\theta = 1)$ против $H_1(\theta > 1)$ на основу узорка обима 10. (4 поена)

б) Да ли је тај тест униформно најмоћнији? Одговор образложити. (1 поен)

в) Одредити моћ теста ако је алтернативна хипотеза $H_1(\theta = 2)$ (праг значајности и обим узорка су исти као у делу а)). (3 поена)

Напомена 1: Корисно је одредити расподелу случајне величине $Y = -\ln X$.
Напомена 2: Хи-квадрат расподела је посебан случај гама расподеле.
Напомена 3: Ако је $U \in \chi^2_{20}$, онда је $F_U(24.886) = 0.79417$.

Решење:

а) Треба преформулисати хипотезу H_1 тако да буде проста како би могла да се примени Нојман-Пирсонова лема. Тестирамо хипотезу $H_0(\theta = 1)$ против алтернативе $H_1(\theta = \theta_1), \theta_1 > 1$. Сада се може применити Нојман-Пирсонова лема:

$L(\theta) = \prod\limits_{k=1}^{n} f(t;\theta) = \theta^n (\prod\limits_{k=1}^{n} t_k)^{\theta-1}$

$W = \left\{\frac{L(\theta_1)}{L(1)} \geq c\right\} = \left\{\frac{\theta_1^n (\prod\limits_{k=1}^{n} t_k)^{\theta_1-1}}{1} \geq c\right\} = \left\{\prod\limits_{k=1}^{n} t_k^{\theta_1-1} \geq c_1\right\} = \left\{ln(\prod\limits_{k=1}^{n} t_k) \geq c_3\right\} = \left\{\sum\limits_{k=1}^{n}(-\ln x_k) \leq K\right\}$

$X : \beta(\theta,1) \Rightarrow -\ln X : \varepsilon(\theta) \Rightarrow \sum\limits_{k=1}^{n}(-\ln X) : \gamma(n,\theta) \Rightarrow 2\theta \sum\limits_{k=1}^{n}(-\ln X) : \chi^2_{2n}$

При H_0: $2\sum\limits_{k=1}^{n}(-\ln X_k) : \chi^2_{20}$.

$K = \frac{1}{2}\chi^2_{20,0.90} = 6.2215$

Најбоља критична област је: $W = \left\{\sum\limits_{k=1}^{n}(-\ln x_k) \leq 6.2215\right\}$

б) Тест јесте униформно најмоћнији јер критична област не зависи од вредности параметра у алтернативној хипотези.

в) Моћ теста је: $\gamma = P_{H_1}\{\sum\limits_{k=1}^{n}(-\ln X_k) \leq 6.2215\}$

При хипотези H_1 важи: $2 \cdot 2\sum\limits_{k=1}^{n}(-\ln X_k) : \chi^2_{20}$.

$\gamma = P_{H_1}\{4\sum\limits_{k=1}^{n}(-\ln X_k) \leq 4 \cdot 6.2215\} = F_{\chi^2_{20}}(24.886) = 0.79417$