Rešitve kolokvija iz verjetnosti in statistike z dne 3. 6. 2010

Matematika – UNI-BOL

1. a) *Prvi način*: parameter s stabilno disperzijo iščemo kot funkcijo naravnega parametra. Tu gre namreč za enoparametrično eksponentno družino, kjer verjetnostno funkcijo zapišemo v obliki:

$$p(x;\lambda) = \frac{e^{-\lambda}}{x!}\, e^{x\ln\lambda}\,,$$

torej je naravni parameter $\theta = \ln\lambda$. Parameter s stabilno disperzijo iščemo v obliki $g(\theta)$, kjer je $g'(\theta) = \sigma(X) = e^{\theta/2}$, kar bo res za $g(\theta) = 2\,e^{\theta/2} = 2\sqrt{\lambda}$.

Drugi način: parameter s stabilno disperzijo iščemo kot funkcijo matematičnega upanja, v tem primeru $E(X) = \lambda$. Za dovolj gladko naraščajočo funkcijo h bo približno veljalo:

$$\sigma\big(h(X)\big) \approx h'\big(E(X)\big)\sigma(X) = h'(\lambda)\sqrt{\lambda}\,.$$

Desna stran bo enaka 1 za $h'(\lambda)\sqrt{\lambda} = 1$, to pa velja za $h(\lambda) = 2\sqrt{\lambda}$. Dobimo isti parameter kot pri prvem načinu (v eksponentnih družinah to vedno velja).

b) Interval zaupanja konstruiramo na podlagi aproksimacije z normalno porazdelitvijo, ta pa velja tudi za eno samo opažanje, brž ko je λ dovolj velik. Za velike λ se namreč Poissonova porazdelitev bliža normalni, to pa zato, ker je to tudi porazdelitev velikega števila vsote neodvisnih Poissonovih slučajnih spremenljivk s parametrom blizu 1, kjer je Poissonova porazdelitev še dovolj “lepa”. Glede na točko a), normalno aproksimacijo in dejstvo, da je cenilka po metodi največjega verjetja enaka $\hat{\lambda} = X$, je iskani aproksimativni interval zaupanja določen z neenačbo:

$$2\big|\sqrt{\lambda} - \sqrt{X}\big| < \Phi^{-1}\left(\frac{1-\alpha}{2}\right) \doteq 1{\cdot}96\,,$$

kar za naše opažanje (ustrezno zaokroženo) znese:

$$81{\cdot}3 < \lambda < 120{\cdot}6\,.$$

2. $\bar{X} = 19{\cdot}71$, $\quad S_{\mathrm{p}} \doteq 2{\cdot}925$, $\quad df = 9$, $\quad c = t_{0.975} \doteq 3{\cdot}25$, $\quad \Delta \doteq 3{\cdot}01$.
Interval zaupanja: $16{\cdot}70 < \mu < 22{\cdot}72$.

3. Glede na to, da sta ničelna in alternativna hipoteza obe enostavni, je test, ki temelji na razmerju verjetij, najmočnejši. Če z L_0 označimo verjetje pri ničelni, z L_1 pa pri alternativni hipotezi, za $x > 0$ velja:

$$\frac{L_0(x)}{L_1(x)} = \frac{\frac{1}{500}\,e^{-x/500}}{\frac{1}{100}\,e^{-x/100}} = \frac{1}{5}\,e^{4x/500}\,.$$

Če je torej X opažena življenjska doba žarnice, bomo ničelno hipotezo zavrnili, če bo testna statistika $\frac{1}{5}\,e^{4X/500}$ premajhna, to pa bo tedaj, ko bo življenjska doba X

premajhna.

Isti sklep lahko dobimo tudi iz dejstva, da gre za enoparametrično eksponentno družino z naravno zadostno statistiko $-X$ glede na parametrizacijo z λ, in dejstva, da je vrednost parametra pri alternativni hipotezi večja od tiste pri ničelni hipotezi. Ker je X pri ničelni hipotezi porazdeljena zvezno, randomizacija ni potrebna in lahko H_0 zavrnemo, brž ko je $X \le c$, kjer je $P_{H_0}(X \le c) = \alpha$. Iz:

$$P_{H_0}(X \le c) = \int_0^c \frac{1}{500}\, e^{-x/500}\, \mathrm{d}x = 1 - e^{-c/500}$$

dobimo $c = -500 \ln(0\text{˙}95) \doteq 25\text{˙}64$ (zaokroženo navzdol). Moč tega testa je:

$$P_{H_1}(X \le c) = \int_0^c \frac{1}{100}\, e^{-x/100}\, \mathrm{d}x = 1 - e^{-c/100} \doteq 0\text{˙}226\,.$$

4. Pri modelu, ki ga obravnava naloga, lahko funkcijo verjetja zapišemo v obliki:

$$L(x_1, \ldots x_n; \mu, \sigma) = \left(\sigma\sqrt{2\pi}\right)^{-n} \exp\left[-\frac{1}{\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2\right)\right]$$

in maksimum je pri celem modelu dosežen pri:

$$\mu = \hat{\mu}_1 := \frac{1}{n}\sum_{i=1}^n X_i\,, \quad \sigma^2 = \hat{\sigma}_1^2 := \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2.$$

Če se omejimo na ničelno hipotezo $\mu = \sigma^2$, pa je:

$$L(x_1, \ldots x_n; \sigma^2, \sigma) = \left(\sigma\sqrt{2\pi}\right)^{-n} \exp\left[-\frac{1}{\sigma^2}\sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i - \frac{n\sigma^2}{2}\right]$$

in iz:

$$\frac{\mathrm{d}}{\mathrm{d}\sigma} \ln L(x_1, \ldots x_n; \sigma^2, \sigma) = \frac{1}{\sigma^3}\sum_{i=1}^n x_i^2 - n\sigma - \frac{n}{\sigma}$$

dobimo, da je maksimum dosežen pri:

$$\sigma^2 = \hat{\sigma}_2^2 := \hat{\mu}_2 := \frac{1}{2}\left[\sqrt{1 + \frac{4}{n}\sum_{i=1}^n x_i^2} - 1\right].$$

Če z Λ označimo razmerje verjetij:

$$\Lambda = \frac{\sup_{\Theta_{H_0}} L(X_1, \ldots X_n; \mu, \sigma)}{\sup_{\Theta} L(X_1, \ldots X_n; \mu, \sigma)} = \frac{L(X_1, \ldots X_n; \hat{\mu}_2, \hat{\sigma}_2)}{L(X_1, \ldots X_n; \hat{\mu}_1, \hat{\sigma}_1)}\,,$$

ima $-2\ln(\Lambda)$ približno porazdelitev hi kvadrat z eno prostostno stopnjo. Za naše opažanje dobimo $-2\ln(\Lambda) \doteq 5\text{˙}61$, kritična vrednost pa je $\chi^2_{0.99} \doteq 6\text{˙}63$. Hipoteze torej ne moremo zavrniti.