Rešitve kolokvija iz verjetnosti in statistike z dne 22. 4. 2010

Matematika – UNI-BOL

1. Uporabimo lahko centralni limitni izrek. Iz:

$$E(X_i) = 2\,, \quad D(X_i) = 0{\dot{}}6\,, \quad E(S) = 800\,, \quad D(S) = 240\,,$$

kjer je $S = X_1 + X_2 + \cdots + X_{400}$, dobimo, da je približno $S \sim \mathrm{N}\big(800, \sqrt{240}\big)$. Če ustrezno verjetnost aproksimiramo tako, kot je napisana, dobimo:

$$P(S < 780) \approx \frac{1}{2} - \Phi\left(\frac{20}{\sqrt{240}}\right) \doteq 0{\dot{}}5 - \Phi(1{\dot{}}2910) \doteq 0{\dot{}}0984\,.$$

Lahko pa upoštevamo, da ima S zalogo vrednosti na celoštevilski mreži, in aproksimiramo:

$$P(S < 779{\dot{}}5) \approx \frac{1}{2} - \Phi\left(\frac{20{\dot{}}5}{\sqrt{240}}\right) \doteq 0{\dot{}}5 - \Phi(1{\dot{}}3233) \doteq 0{\dot{}}0929\,.$$

Točen rezultat: $0{\dot{}}0914338$.

2. Iz navzkrižne gostote vzorca:

$$\begin{aligned} p(x_1, \ldots x_n) &= \frac{a^{2an}}{\big(\Gamma(2a)\big)^n}\, x_1^{2a-1} x_2^{2a-1} \cdots x_n^{2a-1}\, e^{-ax_1}\, e^{-ax_2} \cdots e^{-ax_n} = \\ &= \frac{a^{2an}}{\big(\Gamma(2a)\big)^n\, x_1\, x_2 \cdots x_n}\, e^{a(2\ln x_1 + 2\ln x_2 + \cdots + 2\ln x_n - x_1 - x_2 - \cdots - x_n)} \end{aligned}$$

razberemo, da je minimalna zadostna statistika recimo:

$$2(\ln X_1 + \ln X_2 + \cdots + \ln X_n) - X_1 - X_2 - \cdots - X_n\,.$$

3. a) Slučajne spremenljivke $Z_i = (X_i - \mu)/\sigma$ so neodvisne in porazdeljene standardno normalno, torej neodvisno od μ in σ. Ker je:

$$U = \frac{(Z_1 - \bar{Z})(Z_2 - \bar{Z}) \cdots (Z_n - \bar{Z})}{\big[(Z_1 - \bar{Z})^2 + (Z_2 - \bar{Z})^2 + \cdots + (Z_n - \bar{Z})^2\big]^{n/2}}$$

kjer je $\bar{Z} = (Z_1 + \cdots + Z_n)/n$, je tudi porazdelitev te statistike neodvisna od μ in σ, se pravi, da je statistika U postranska.

b) Znano je, da je normalna družina eksponentna družina, katere naravni parametrični prostor ima neprazno notranjost. V takih modelih so postranske statistike neodvisne od minimalne zadostne statistike, ki pa je v našem primeru $\big(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\big)$. Torej je U neodvisna od $\sum_{i=1}^n X_i^2$.

4. a) Iz funkcije verjetja in njenega odvoda:

$$L(k,\lambda) = \frac{\lambda^k\, e^{-\lambda}}{k!}\,, \qquad \frac{\partial L}{\partial \lambda} = \frac{(k-\lambda)\lambda^{k-1}\, e^{-\lambda}}{k!}$$

dobimo cenilko $\hat{\lambda} = X$, cenilka za λ^2 pa je X^2. Le-ta je pristranska, saj je $E(X^2) = \lambda^2 + \lambda$.

b) Ker gre za eksponentno družino, katere naravni parametrični prostor ima neprazno notranjost, bo nepristranska cenilka imela najmanjšo možno disperzijo, brž ko bo funkcija minimalne zadostne statistike X. Iščemo torej tako funkcijo h, da bo $E_\lambda\big(h(X)\big) = \lambda^2$ za vse λ. To lahko naredimo z nastavkom:

$$\sum_{k=0}^{\infty} \frac{h(k)\, \lambda^k\, e^{-\lambda}}{k!} = \lambda^2$$

oziroma:

$$\sum_{k=0}^{\infty} \frac{h(k)\, \lambda^k}{k!} = \lambda^2\, e^{\lambda}\,,$$

nakar z razvojem:

$$\lambda^2\, e^{\lambda} = \sum_{n=0}^{\infty} \frac{\lambda^{n+2}}{n!} = \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} = \sum_{k=2}^{\infty} \frac{k(k-1)\lambda^k}{k!} = \sum_{k=0}^{\infty} \frac{k(k-1)\lambda^k}{k!}$$

in primerjavo koeficientov dobimo, da mora biti $h(k) = k(k-1)$, torej je iskana statistika $X(X-1)$. Le-to lahko tudi kar uganemo iz prvih dveh momentov Poissonove porazdelitve.

c) Velja:

$$\begin{aligned} q(X^2 - aX) &= E\big[(X^2 - aX - \lambda^2)^2\big] = \\ &= E(X^4) + a^2\, E(X^2) + \lambda^4 - 2a\, E(X^3) - 2\lambda^2\, E(X^2) + 2a\lambda^2\, E(X) = \\ &= 4\lambda^3 + (7 - 6a + a^2)\lambda^2 + (1 - 2a + a^2)\lambda\,. \end{aligned}$$

Smiselne izbire parametra a so tiste, pri katerih ne obstaja nobena druga vrednost, pri kateri bi bila srednja kvadratična napaka manjša za *vse* λ. To pa so tiste, pri katerih se koeficienta $c_1(a) := 1 - 2a + a^2$ in $c_2(a) = 7 - 6a + a^2$ ne moreta hkrati zmanjšati. Za $a < 1$ sta oba koeficienta strogo padajoča, za $a > 3$ pa oba strogo naraščajoča v a, torej ju lahko hkrati zmanjšamo in izbira v teh dveh območjih ni smiselna. Za $1 \le a \le 3$ pa je c_1 naraščajoč, c_2 pa padajoč v a, torej se ne moreta hkrati zmanjšati (poljubni dve različni izbiri iz $1 \le a \le 3$ sta neprimerljivi). Zato je iz intervala $[1,3]$ smiselno izbirati.