Ime in priimek: ______________________ Vpisna št: | | | | | | | | |

FAKULTETA ZA MATEMATIKO IN FIZIKO

Oddelek za matematiko

Statistika

Pisni izpit

30. avgust 2019

Navodila

Pazljivo preberite besedilo naloge, preden se lotite reševanja. Za pozitiven rezultat morate zbrati vsaj 45 točk od 100 možnih. Veliko uspeha!

Naloga	a.	b.	c.	d.	
1.			•	•	
2.				•	
3.			•	•	
4.			•	•	
Skupaj	•	•	•	•	

1. (25) Hazarder Marko obišče 100 igralnih avtomatov. Pred vsakim vrže pošten kovanec. Če pade cifra, na tem avtomatu igra, sicer pa ne. Na vsakem avtomatu ima, če igra, pričakovano izgubo 1 evro, standardni odklon pa je 3 evre. Privzamemo, da so posamezne igre in meti kovanca med seboj neodvisni.

a. (15) Izračunajte pričakovano vrednost in varianco Markovega izkupička po obisku vseh igralnih avtomatov.

Rešitev: Če z S označimo Markov izkupiček, potem ko obrede vse avtomate, lahko pišemo:

$$S = I_1 X_1 + I_2 X_2 + \cdots + I_{100} X_{100} \,,$$

kjer je $I_k = 1$, če je Marko igral na k-tem avtomatu, sicer pa je $I_k = 0$; X_k je Markov izkupiček na k-tem avtomatu (namišljen, če Marko tam ni igral). Računamo:

$$\begin{aligned}
E(I_k X_k) &= E(I_k) E(X_k) = -\frac{1}{2} \,, \\
E(X_k^2) &= \mathrm{var}(X_k) + \big(E(X_k)\big)^2 = 10 \,, \\
E\big((I_k X_k)^2\big) &= E(I_k) E(X_k^2) = 5 \,, \\
\mathrm{var}(I_k X_k) &= E\big((I_k X_k)^2\big) - \big(E(I_k X_k)\big)^2 = \frac{19}{4} \,.
\end{aligned}$$

ali alternativno:

$$E(I_k X_k | X_k = 0) = 0 \,, \qquad E(I_k X_k | X_k = 1) = -1 \,,$$
$$\mathrm{var}(E(I_k X_k | X_k)) = \frac{1}{4} \,,$$
$$\mathrm{var}(I_k X_k | X_k = 0) = 0 \,, \qquad \mathrm{var}(I_k X_k | X_k = 1) = 9 \,,$$
$$E(\mathrm{var}(I_k X_k | X_k)) = \frac{9}{2} \,,$$
$$\mathrm{var}(I_k X_k | X_k = 1) = \mathrm{var}(E(I_k X_k | X_k)) + E(\mathrm{var}(I_k X_k | X_k)) = \frac{19}{4} \,.$$

Torej je $E(S) = -100/2 = -50$ in $\mathrm{var}(S) = 100 \cdot 19/4 = 475$.

b. (10) Približno izračunajte verjetnost, da bo imel Marko na koncu dobiček.

Rešitev: Po centralnem limitnem izreku je:

$$P(S > 0) \approx 1 - \Phi\left(\frac{50}{\sqrt{475}}\right) \doteq 0{,}011 \,.$$

2. (25) V populaciji znane velikosti N vsaki enoti pripada vrednost statistične spremenljivke X. V populaciji so enote tipa A in tipa B, in sicer je prvih N_A, drugih pa N_B, pri čemer sta N_A in N_B neznana, vemo le, da je $N_B < n$ za neki dani n. Iz populacije izberemo enostavni slučajni vzorec velikosti n. Za enote, izbrane v vzorec, lahko določimo, ali so tipa A ali B, in zabeležimo vrednost statistične spremenljivke. Iz zgornjih predpostavk sledi, da bo v vzorcu vsaj ena enota tipa A. Oceniti želimo povprečje vrednosti statistične spremenljivke X **samo** za enote tipa A. Označimo to povprečje z μ_A.

Označite z M_A število enot tipa A v vzorcu, z $\hat{\mu}_A$ vzorčno povprečje za enote tipa A in s $\hat{\sigma}_A^2$ vzorčno varianco za enote tipa A.

a. (5) Predlagajte nepristransko cenilko parametra μ_A. Utemeljite, da je nepristranska.

Rešitev: Za cenilko predlagamo povprečje vrednosti statistične spremenljivke za vzorčne enote tipa A. Označimo z M_A število enot tipa A v vzorcu. Pogojno na $\{M_A = m\}$ je m enot tipa A, zajetih v vzorec, enostavni slučajni vzorec iz populacije enot tipa A, zato je $E(\hat{\mu}_A | M_A = m) = \mu_A$. Posledično je $E(\hat{\mu}_A) = \mu_A$.

b. (10) Utemeljite, da je

$$\mathrm{var}(\hat{\mu}_A | M_A = m) = \frac{\sigma_A^2}{m} \cdot \frac{N_A - m}{N_A - 1}$$

kjer je σ_A^2 populacijska varianca za enote tipa A.

Rešitev: Označimo z M_A število enot tipa A v vzorcu. Spet uporabimo, da m enot tipa A, zajetih v vzorec, pogojno na $\{M_A = m\}$ tvori enostavni slučajni vzorec iz populacije enot tipa A. Sledi, da je

$$\mathrm{var}(\hat{\mu}_A | M_A = m) = \frac{\sigma_A^2}{m} \cdot \frac{N_A - m}{N_A - 1}\,.$$

c. (10) Naj bo $\hat{\sigma}_A^2$ vzorčna varianca. Če so $x_1^A, \ldots, x_m^A$ vzorčne vrednosti za enote tipa A, je

$$\hat{\sigma}_A^2 = \frac{1}{m} \sum_{k=1}^{m} \left(x_k^A - \bar{x}^A\right)^2\,.$$

Izračunajte

$$E(\hat{\sigma}_A^2 | M_A = m)\,.$$

Rešitev: Ponovno uporabimo, da enote tipa A, zajete v vzorec, pogojno na $\{M_A = m\}$ tvorijo enostavni slučajni vzorec velikosti m enot tipa A. Vemo, da je v tem primeru

$$E\left(\hat{\sigma}_A^2 | M_A = m\right) = \sigma_A^2 \cdot \frac{(m-1)N_A}{m(N_A - 1)}\,.$$

3. (25) Naj bo $\lambda, q > 0$, $X \sim \mathrm{Pois}(\lambda)$ in pogojno na $\{X = x\}$ naj bo $Y \sim \mathrm{Pois}(qx)$. Podatki $(x_1, y_1), \ldots, (x_n, y_n)$ so nastali kot neodvisne realizacije para (X, Y).

a. (15) Poiščite cenilki za λ in q po metodi največjega verjetja.

Rešitev: Najprej izračunamo

$$P(X = x, Y = y) = P(X = x)\, P(Y = y|X = x) = \frac{\lambda^x (qx)^y e^{-\lambda - qx}}{x!} y!\,.$$

Funkcija verjetja za vse podatke je torej enaka

$$\begin{aligned} &L(\lambda, q|x_1, y_1, \ldots, x_n, y_n) \\ &\quad = \frac{\lambda^{x_1+x_2+\cdots+x_n} (qx_1)^{y_1} (qx_2)^{y_2} \cdots (qx_n)^{y_n} e^{-n\lambda - q(x_1+x_2+\cdots+x_n)}}{x_1!\, x_2! \cdots x_n!\, y_1!\, y_2! \cdots y_n!}\,, \end{aligned}$$

njen logaritem pa je

$$\begin{aligned} \ell(\lambda, q|x_1, y_1, \ldots, x_n, y_n) = \sum_{k=1}^n x_k \log\lambda - n\lambda - q\sum_{k=1}^n x_k + \sum_{k=1}^n y_k(\log q + \log x_k) \\ - \sum_{k=1}^n \log(x_k!) - \sum_{k=1}^n \log(y_k!)\,. \end{aligned}$$

Odvajamo:

$$\frac{\partial\ell}{\partial\lambda} = \frac{1}{\lambda}\sum_{k=1}^n x_k - n\,, \qquad \frac{\partial\ell}{\partial q} = \frac{1}{q}\sum_{k=1}^n y_k - \sum_{k=1}^n x_k\,,$$

izenačimo z nič in dobimo cenilki

$$\hat\lambda = \frac{1}{n}\sum_{k=1}^n x_k\,, \qquad \hat q = \frac{\sum_{k=1}^n y_k}{\sum_{k=1}^n x_k}\,.$$

b. (10) Pod predpostavko, da je vzorec velik, aproksimirajte standardni napaki cenilk iz prejšnje točke.

Rešitev: Za eno samo opažanje, dokler je še neznano, pride

$$\begin{aligned} \frac{\partial^2\ell}{\partial\lambda^2}(\lambda, q|X, Y) &= -\frac{X}{\lambda^2}\,, \\ \frac{\partial^2\ell}{\partial\lambda\,\partial q}(\lambda, q|X, Y) &= 0\,, \\ \frac{\partial^2\ell}{\partial q^2}(\lambda, q|X, Y) &= -\frac{Y}{q^2}\,. \end{aligned}$$

Iz definicije porazdelitve dobimo $E(X) = \lambda$ *in* $E(Y|X) = qX$*, torej* $E(Y) = q\lambda$*. Fisherjeva matrika je torej enaka*

$$I(\lambda, q) = \begin{pmatrix} 1/\lambda & 0 \\ 0 & \lambda/q \end{pmatrix}\,,$$

torej aproksimativni standardni napaki znašata

$$\mathrm{se}(\lambda) \approx \sqrt{\frac{\lambda}{n}}\,, \qquad \mathrm{se}(q) \approx \sqrt{\frac{q}{n\lambda}}\,.$$

4. (25) Dana imamo dva kompleta regresijskih enačb:

$$Y_i = \alpha_1 + \beta x_i + \epsilon_i$$

za $i = 1, 2, \ldots, m$ in

$$Z_j = \alpha_2 + \beta w_j + \eta_j$$

za $j = 1, 2, \ldots, n$. Predpostavljamo, da vsak model zase ustreza standardnim predpostavkam regresijskega modela, torej, da za vse i, j velja $E(\epsilon_i) = E(\eta_j) = 0$, $\mathrm{var}(\epsilon_i) = \mathrm{var}(\eta_i) = \sigma^2$ in so tako ϵ_i kot tudi η_j med sabo nekorelirani. Predpostavljamo tudi, da so vsi ϵ_i in η_j nekorelirani.

a. (15) Predpostavite, da je

$$\sum_{i=1}^{m} x_i = 0 \qquad \text{in} \qquad \sum_{j=1}^{n} w_j = 0$$

ter

$$\sum_{i=1}^{m} x_i^2 = 1 \qquad \text{in} \qquad \sum_{j=1}^{n} w_j^2 = 1\,.$$

Podajte najboljšo nepristransko linearno cenilko parametra β in izračunajte njeno standardno napako.

Rešitev: Dana kompleta sestavimo v enoten regresijski model $\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, kjer je

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \\ Z_1 \\ \vdots \\ Z_n \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} 1 & 0 & x_1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_m \\ 0 & 1 & w_1 \\ \vdots & \vdots & \vdots \\ 0 & 1 & w_n \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{pmatrix}, \qquad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \\ \eta_1 \\ \vdots \\ \eta_n \end{pmatrix},$$

za katerega iz danih predpostavk sledi, da izpolnjuje vse standardne predpostavke. Izračunamo

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} m & 0 & 0 \\ 0 & n & 0 \\ 0 & 0 & 2 \end{pmatrix}, \qquad \left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{pmatrix} 1/m & 0 & 0 \\ 0 & 1/n & 0 \\ 0 & 0 & 1/2 \end{pmatrix}$$

in še

$$\mathbf{X}^T\boldsymbol{Y} = \begin{pmatrix} \sum_{i=1}^m Y_i \\ \sum_{j=1}^n Z_j \\ \sum_{i=1}^m x_i Y_i + \sum_{j=1}^n w_j Z_j \end{pmatrix}.$$

Sledi

$$\hat{\beta} = \frac{1}{2}\left(\sum_{i=1}^{m} x_i Y_i + \sum_{j=1}^{n} w_j Z_j\right).$$

Če pišemo še $\mathbf{C} = [c_{ij}]_{i,j} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}$, za standardno napako po formuli dobimo

$$\mathrm{se}(\hat{\beta}) = \sigma\sqrt{c_{33}} = \frac{\sigma}{\sqrt{2}}\,.$$

b. (10) Podajte še najboljšo nepristransko linearno cenilko karakteristike $\delta := \alpha_1 - \alpha_2$ in izračunajte njeno standardno napako.

Rešitev: Iz matrik $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$ *in* $\mathbf{X}^T\boldsymbol{Y}$ *dobimo, da sta najboljši nepristranski linearni cenilki parametrov* α_1 *in* α_2 *enaki* $\hat{\alpha}_1 = \bar{Y}$ *in* $\hat{\alpha}_2 = \bar{Z}$. *Iz izreka Gaussa in Markova sledi, da je najboljša nepristranska linearna cenilka karakteristike* δ *enaka* $\hat{\delta} = \bar{Y} - \bar{Z}$. *Neposredno ali pa iz matrike* $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$ *razberemo, da je* $\mathrm{var}(\hat{\alpha}_1) = \sigma^2/m$ *in* $\mathrm{var}(\hat{\alpha}_2) = \sigma^2/n$. *Ker sta* $\hat{\alpha}_1$ *in* $\hat{\alpha}_2$ *nekorelirani, je* $\mathrm{var}(\hat{\delta}) = \mathrm{var}(\hat{\alpha}_1) + \mathrm{var}(\hat{\alpha}_2)$, *torej*

$$\mathrm{se}(\hat{\delta}) = \sigma\sqrt{\frac{1}{m} + \frac{1}{n}}\,.$$