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

FAKULTETA ZA MATEMATIKO IN FIZIKO

Oddelek za matematiko

Statistika

Pisni izpit

9. junij 2020

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.	e.	Skupaj
1.				•	•	
2.						
3.				•	•	
4.			•	•	•	
Skupaj	•	•	•	•	•	

1. (25) Znani angleški statistik M. G. Kendall v svojem obsežnem članku *The Random Character of Stock Market Prices* pravi:

> ... dnevno zaporedje cen delnic zgleda "blodeče", kot da bi nek škrat vsak dan naključno izbral število iz velike škatle, ki ima povprečje 0 in nek standardni odklon, in to izbrano število prištel ceni delnice prejšnega dne.

Predpostavite, da je povprečje škatle res 0, standardni odklon pa 2. Razlika cene delnice na začetku leta in na koncu leta je tako enaka vsoti 365 naključno izbranih števil iz te škatle.

a. (5) Recimo, da je bila cena delnice na začetku leta enaka 150 (v ustreznih enotah). Kolikšna je približno verjetnost, da bo na koncu vredna 180 ali več?

Rešitev: Potrebno je izračunati verjetnost, da bo vsota 365 naključno izbranih števil iz škatle enaka 10 ali več. Z uporabo centralnega limitnega izreka računamo

$$\begin{aligned} P(S_{365} \geq 30) &= P\left(\frac{S_{365}}{\sqrt{365 \cdot 2}} \geq \frac{30}{2 \cdot \sqrt{365}}\right) \\ &\approx P(Z \geq 0,79) \\ &= 0,21\,. \end{aligned}$$

b. (10) Nekdo vam ponuja naslednjo stavo: na začetku leta je cena delnice 100. Če bo cena delnice na koncu leta 110 ali več, ti plačam 20 enot, če ne pa ti meni plačaš 5 enot. Kaj menite o tej stavi?

Namig: Kolikšen je vaš pričakovan dobiček?

Rešitev: Najprej izračunamo verjetnost, da bo vrednost delnice na koncu leta 110 *ali več.*

$$\begin{aligned} P(S_{365} \geq 10) &= P\left(\frac{S_{365}}{2 \cdot \sqrt{365}} \geq \frac{10}{2 \cdot \sqrt{365}}\right) \\ &\approx P(Z \geq 0,26) \\ &= 0,40\,. \end{aligned}$$

Torej bo naše izplačilo z verjetnostjo $0,4$ *enako* 20*, z verjetnostjo* $0,6$ *pa* -5*. Pričakovano izplačilo je* 5*. Stavo s pozitivnim pričakovanim izplačilom seveda vedno sprejmemo.*

c. (10) Recimo, da bo škrat vsakemu številu v škatli dodal število a. Koliko mora dodati, da bo z verjetnostjo 0,9 cena delnice na koncu leta višja kot na začetku.

Rešitev: Povprečje se z dodajanjem a poveča za a*. Računamo*

$$P(S_n > 0) = P\left(\frac{S_n - na}{2\sqrt{n}} > -\frac{a\sqrt{n}}{2}\right) \overset{\text{CLT}}{\approx} P\left(Z > -\frac{a\sqrt{n}}{2}\right)\,.$$

Iz tabel dobimo, da bi moralo biti

$$-\frac{a\sqrt{n}}{2} = -1,28\,.$$

Sledi $a \approx 0,13$.

2. (25) Dva statistika iz populacije velikosti N izbereta vsak svoj enostavni slučajni vzorec velikosti n. Najprej ga izbere prvi in enote vrne v populacijo, nato pa neodvisno od prvega še drugi. Predpostavljamo, da je $2n \leq N$. Naš vzorec je unija vzorcev obeh statistikov. Populacijsko povprečje ocenimo z vzorčnim povprečjem. Označite slučajno velikost vzorca z N_V. Vzorčno povprečje označite z $\bar{X}$. Populacijsko povprečje in populacijsko varianco označite z μ in σ^2.

a. (5) Izračunajte $E(\bar{X}|N_V)$.

Rešitev: Pri dani velikosti vzorca je združeni vzorec enostavni slučajni. Ker je vzorčno povprečje za enostavni slučajni vzorec nepristranska cenilka populacijskega povprečja, je

$$E(\bar{X}|N_V) = \mu\,.$$

b. (5) Izračunajte $\mathrm{var}(\bar{X}|N_V)$.

Rešitev: Iz istega razloga kot v prvem delu je

$$\mathrm{var}(\bar{X}|N_V) = \frac{\sigma^2}{N_V} \cdot \frac{N - N_V}{N - 1}\,.$$

c. (10) Označite

$$a = E\left(\frac{1}{n+Y}\right)\,,$$

kjer je $Y \sim \mathrm{HiperGeom}(n, N-n, N)$. Izračunajte $\mathrm{var}(\bar{X})$. V izrazu bo nastopala količina a.

Rešitev: Slučajna spremenljivka N_V ima enako porazdelitev kot $n + Y$, zato bo

$$E\left(\frac{1}{N_V}\right) = a\,.$$

Po formuli za razcep variance sledi

$$\mathrm{var}(\bar{X}) = \frac{\sigma^2}{(N-1)} E\left(\frac{N}{N_V} - 1\right)\,,$$

kar se poenostavi v

$$\mathrm{var}(\bar{X}) = \frac{\sigma^2}{(N-1)}(Na - 1)\,.$$

d. (5) Predlagajte nepristransko cenilko populacijske variance σ^2.

Rešitev: Po točki b. bo nepristanska cenilka

$$\hat{\sigma}^2 = \frac{N-1}{(N_V - 1)N} \sum_{i=1}^{N_V} \left(X_i - \bar{X}\right)^2\,,$$

kjer so $X_1, \ldots, X_{N_V}$ opazovane vrednosti v združenem vzorcu.

3. (25) Opazovane vrednosti naj bodo pari $(x_1, y_1), \ldots, (x_n, y_n)$. Predpostavljamo, da so pari nastali kot neodvisni slučajni pari $(X_1, Y_1), \ldots, (X_n, Y_n)$ z gostoto

$$f(x, y) = e^{-x} \cdot \frac{1}{\sigma\sqrt{2\pi x}}\, e^{-\frac{(y-\theta x)^2}{2\sigma^2 x}}$$

za $x > 0$ in $-\infty < y < \infty$ ter $\sigma^2 > 0$. Kot znano upoštevajte, da je spremenljivka

$$Z = \frac{Y - \theta X}{\sqrt{X}}$$

porazdeljena normalno $\mathrm{N}(0, \sigma^2)$ in neodvisna od X.

a. (5) Poiščite cenilki parametrov θ in σ^2 po metodi največjega verjetja.

Rešitev: Logaritemska funkcija verjetja je

$$\ell\left(\theta, \sigma | \mathbf{x}, \mathbf{y}\right) = \sum_{k=1}^{n} \left(-\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2}\sum_{k=1}^{n}\log x_k - \frac{(y_k - \theta x_k)^2}{2\sigma^2 x_k} \right).$$

Parcialno odvajamo in dobimo

$$\begin{aligned} \frac{\partial \ell}{\partial \theta} &= \sum_{k=1}^{n} \frac{(y_k - \theta x_k)}{\sigma^2} \\ \frac{\partial \ell}{\partial \sigma} &= -\frac{n}{\sigma} + \sum_{k=1}^{n} \frac{(y_k - \theta x_k)^2}{\sigma^3 x_k} \end{aligned}$$

Izenačimo parcialna odvoda z 0*. Iz prve enačbe sledi*

$$\hat{\theta} = \frac{\sum_{k=1}^{n} y_k}{\sum_{k=1}^{n} x_k}$$

iz druge pa

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} \frac{(y_k - \hat{\theta} x_k)^2}{x_k}.$$

b. (10) Izračunajte Fisherjevo matriko informacije in navedite aproksimativni standardni napaki cenilk $\hat{\theta}$ in $\hat{\sigma}$.

Rešitev: Računamo za $n = 1$*:*

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \theta^2} &= -\frac{x}{\sigma^2}, \\ \frac{\partial^2 \ell}{\partial \theta\, \partial \sigma} &= -2\,\frac{y - \theta x}{\sigma^3}, \\ \frac{\partial^2 \ell}{\partial \sigma^2} &= \frac{1}{\sigma^2} - 3\,\frac{(y - \theta x)^2}{\sigma^4 x}. \end{aligned}$$

Nadomestimo x z X in y z Y. Iz prve naloge razberemo, da je $X \sim \exp(1)$, torej je

$$E\left[\frac{\partial^2 \ell}{\partial \theta^2}(\theta, \sigma|X, Y)\right] = -\frac{E(X)}{\sigma^2} = -\frac{1}{\sigma^2}\,.$$

Pri izračunu preostalih dveh pričakovanih vrednosti si pomagamo z namigom:

$$\begin{aligned} E\left[\frac{\partial^2 \ell}{\partial \theta\, \partial \sigma}(\theta, \sigma|X, Y)\right] &= -\frac{2E\big(Z\sqrt{X}\big)}{\sigma^3} = 0\,, \\ E\left[\frac{\partial^2 \ell}{\partial \sigma^2}(\theta, \sigma|X, Y)\right] &= \frac{1}{\sigma^2} - \frac{3E(Z^2)}{\sigma^4} = -\frac{2}{\sigma^2}\,. \end{aligned}$$

Fisherjeva matrika je tako enaka

$$I(\theta, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}\,,$$

ocenjeni standardni napaki pa sta enaki

$$\mathrm{se}(\hat\theta) = \frac{\hat\sigma}{\sqrt{n}} \quad \textit{in} \quad \mathrm{se}(\hat\sigma) = \frac{\hat\sigma}{\sqrt{2n}}\,.$$

c. (10) Izračunajte eksaktno standardno napako cenilke $\hat\theta$. Kot znano upoštevajte, da je porazdelitev $\sum_{k=1}^n Y_k$ pogojno na $\sum_{k=1}^n X_k = x$ normalna s parametroma θx in $\sigma^2 x$.

Rešitev: Iz namiga razberemo, da je

$$\begin{aligned} E\left(\hat\theta\right) &= E\left(\frac{\sum_{k=1}^n Y_k}{\sum_{k=1}^n X_k}\right) \\ &= E\left[E\left(\frac{\sum_{k=1}^n Y_k}{\sum_{k=1}^n X_k} \Big| \sum_{k=1}^n X_k\right)\right] \\ &= E\left[\frac{1}{\sum_{k=1}^n X_k}\theta \sum_{k=1}^n X_k\right] \\ &= \theta\,. \end{aligned}$$

in

$$\mathrm{var}(\hat\theta | \sum_{k=1}^n X_k) = \frac{\sigma^2}{\sum_{k=1}^n X_k}\,.$$

Po formuli za razcep variance je

$$\mathrm{var}(\hat\theta) = \sigma^2 E\left(\frac{1}{\sum_{k=1}^n X_k}\right)\,.$$

Vemo, da je $\sum_{k=1}^n X_k \sim \Gamma(n, 1)$. Sledi

$$E\left(\frac{1}{\sum_{k=1}^n X_k}\right) = \frac{1}{\Gamma(n)}\int_0^\infty \frac{1}{s}\cdot s^{n-1}e^{-s}ds = \frac{\Gamma(n-1)}{\Gamma(n)} = \frac{1}{n-1}\,.$$

Sledi

$$\mathrm{var}(\hat{\theta}) = \frac{\sigma^2}{n-1}\,.$$

4. (25) Privzemite regresijski model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\,,$$

kjer je $E(\boldsymbol{\epsilon}) = 0$ in $\mathrm{var}(\boldsymbol{\epsilon}) = \sigma^2(\mathbf{I} + a\mathbf{1}\mathbf{1}^T)$ za $a > 0$. Privzamemo, da je konstanta a znana in je $\mathbf{X}$ matrika velikosti $n \times m$ z rangom m.

a. (10) Najdite najboljšo nepristransko linearno cenilko $\boldsymbol{\beta}$.

Namig: poiščite $c \in \mathbb{R}$, za katerega je

$$\left(\mathbf{I} + a\mathbf{1}\mathbf{1}^T\right)\left(\mathbf{I} + c\mathbf{1}\mathbf{1}^T\right) = \mathbf{I}\,.$$

Rešitev: Naj bo $\tilde{\boldsymbol{\beta}}$ poljubna linearna nepristranska cenilka parametra $\boldsymbol{\beta}$. Iz te predpostavke sledi, da je

$$\tilde{\boldsymbol{\beta}} = \mathbf{L}\mathbf{Y}$$

za neko matriko $\mathbf{L}$, ki zadošča

$$\mathbf{L}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}\,.$$

Računamo

$$\begin{aligned}\mathrm{var}\big(\tilde{\boldsymbol{\beta}}\big) &= \mathrm{var}\big(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}\big)\\ &= \mathrm{var}\big(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}\big) + \mathrm{var}\big(\hat{\boldsymbol{\beta}}\big) + 2\,\mathrm{cov}\big(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}},\, \hat{\boldsymbol{\beta}}\big)\,.\end{aligned}$$

Označimo

$$\mathbf{A} = \left(\mathbf{I} + a\mathbf{1}\mathbf{1}^T\right) \quad \textit{in} \quad \mathbf{C} = \mathbf{I} + c\mathbf{1}\mathbf{1}^T$$

ter izračunamo

$$\mathbf{A}\mathbf{C} = \mathbf{I} + a\mathbf{1}\mathbf{1}^T + c\mathbf{1}\mathbf{1}^T + ac\mathbf{1}\mathbf{1}^T\mathbf{1}\mathbf{1}^T = \mathbf{I} + (a + c + nac)\mathbf{1}\mathbf{1}^T = \mathbf{I}\,.$$

Z upoštevanjem $\mathrm{cov}\,(\mathbf{Y}, \mathbf{Y}) = \sigma^2\mathbf{A}$ izračunamo še

$$\begin{aligned}\mathrm{cov}\left(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}\right) &= \mathrm{cov}\left(\left(\mathbf{L} - \left(\mathbf{X}^T\mathbf{C}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{C}\right)\mathbf{Y},\ \left(\mathbf{X}^T\mathbf{C}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{C}\mathbf{Y}\right)\\ &= \sigma^2\left(\mathbf{L} - \left(\mathbf{X}^T\mathbf{C}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{C}\right)\mathbf{A}\mathbf{C}\mathbf{X}\left(\mathbf{X}^T\mathbf{C}\mathbf{X}\right)^{-1}\\ &= \sigma^2\left(\mathbf{L}\mathbf{X} - \mathbf{I}\right)\left(\mathbf{X}^T\mathbf{C}\mathbf{X}\right)^{-1}\\ &= 0\,.\end{aligned}$$

Trditev sledi kot v izreku Gauss-Markova.

b. (15) Popravite cenilko $\sum_{k=1}^n \hat{\epsilon}_k^2$ tako, da bo nepristranska za parameter σ^2.

Namig: pomagajte si z lastnostmi sledi matrik.

Rešitev: Ena od možnosti je, da cenilka temelji na normi reziduala

$$\hat{\boldsymbol{\epsilon}} = \begin{pmatrix}\hat{\epsilon}_1\\ \hat{\epsilon}_2\\ \vdots\\ \hat{\epsilon}_n\end{pmatrix} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

Velja

$$\hat{\boldsymbol{\epsilon}} = \Big(\mathbf{I} - \mathbf{X}\big(\mathbf{X}^T\mathbf{C}\mathbf{X}\big)^{-1}\mathbf{X}^T\mathbf{C}\Big)\boldsymbol{\epsilon}$$

in

$$\begin{aligned}
\sum_{k=1}^{n} \hat{\epsilon}_k^2 &= \big(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\big)^T\big(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\big) \\
&= \boldsymbol{\epsilon}^T\Big(\mathbf{I} - \mathbf{C}\mathbf{X}\big(\mathbf{X}^T\mathbf{C}\mathbf{X}\big)^{-1}\mathbf{X}^T\Big)\Big(\mathbf{I} - \mathbf{X}\big(\mathbf{X}^T\mathbf{C}\mathbf{X}\big)^{-1}\mathbf{X}^T\mathbf{C}\Big)\boldsymbol{\epsilon} \\
&= \mathrm{Sl}\Big[\Big(\mathbf{I} - \mathbf{C}\mathbf{X}\big(\mathbf{X}^T\mathbf{C}\mathbf{X}\big)^{-1}\mathbf{X}^T\Big)\Big(\mathbf{I} - \mathbf{X}\big(\mathbf{X}^T\mathbf{C}\mathbf{X}\big)^{-1}\mathbf{X}^T\mathbf{C}\Big)\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\Big] .
\end{aligned}$$

Ker je $\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T = \sigma^2\mathbf{A}$, *je nadalje*

$$\begin{aligned}
E\left(\sum_{k=1}^{n} \hat{\epsilon}_k^2\right) &= \sigma^2\,\mathrm{Sl}\Big[\Big(\mathbf{I} - \mathbf{C}\mathbf{X}\big(\mathbf{X}^T\mathbf{C}\mathbf{X}\big)^{-1}\mathbf{X}^T\Big)\Big(\mathbf{I} - \mathbf{X}\big(\mathbf{X}^T\mathbf{C}\mathbf{X}\big)^{-1}\mathbf{X}^T\mathbf{C}\Big)\mathbf{A}\Big] \\
&= \sigma^2\,\mathrm{Sl}\big(\mathbf{A} - \mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\big) .
\end{aligned}$$

Sledi, da je

$$\hat{\sigma}^2 = \frac{1}{\mathrm{Sl}\big(\mathbf{A} - \mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\big)} \sum_{k=1}^{n} \hat{\epsilon}_k^2$$

nepristranska cenilka za σ^2.