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

FAKULTETA ZA MATEMATIKO IN FIZIKO

Oddelek za matematiko

Statistika

Pisni izpit

28. junij 2019

Navodila

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

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

1. (25) Slučajne spremenljivke $X, Y_1, Y_2, \ldots Y_{100}$ so neodvisne, pri čemer je $P(X = 1) = 2/3$, $P(X = 2) = 1/3$, $E(Y_i) = 1$ in $\mathrm{var}(Y_i) = 100$. Slučajne spremenljivke $Y_1, \ldots Y_{100}$ so tudi enako porazdeljene. Označimo $S = X(Y_1+Y_2+\cdots+Y_{100})$. Približno izračunajte $P(S > 150)$.

Rešitev: Centralnega limitnega izreka ne moremo uporabiti neposredno na S, ker produkt slučajne spremenljivke X in normalno porazdeljene slučajne spremenljivke, četudi neodvisne od X, ni nujno normalno porazdeljen. Prav tako centralni limitni izrek ne velja za vsoto $XY_1 + XY_2 + \cdots XY_{100}$, saj so seštevanci odvisni (centralni limitni izrek se sicer da posplošiti tudi na vsote slučajnih spremenljivk z določeno vrsto odvisnosti, vendar pa je odvisnost prej omenjenih seštevancev premočna). Pravilno pa bo iskano verjetnost računati s pomočjo pogojnih verjetnosti glede na X. Če pišemo $T := Y_1 + Y_2 + \cdots + Y_{100}$, po izreku o polni verjetnosti velja:

$$\begin{aligned} P(S > 150) &= P(X = 1)P(S > 150 \mid X = 1) + P(X = 2)P(S > 150 \mid X = 2) = \\ &= \frac{2}{3}P(T > 150) + \frac{1}{3}P(T > 75)\,. \end{aligned}$$

Za slučajno spremenljivko T pa centralni limitni izrek velja: iz $E(T) = 100$ in $\mathrm{var}(T) = 10000$ dobimo:

$$P(T > a) \approx 1 - \Phi\left(\frac{a - 100}{100}\right)\,,$$

torej je:

$$P(S > 150) \approx \frac{2}{3}\left(1 - \Phi\left(\frac{1}{2}\right)\right) + \frac{1}{3}\,\Phi\left(\frac{1}{4}\right) \doteq 0{,}405\,.$$

2. (25) Vsaki enoti populacije velikosti N pripadata dve vrednosti statistične spremenljivke. Vrednosti označimo z $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$. Oceniti želimo povprečje povprečij obeh spremenljivk

$$\frac{\mu+\nu}{2} = \frac{1}{2N}\left(\sum_{k=1}^{N} x_k + \sum_{k=1}^{N} y_k\right).$$

Izberemo enostavni slučajni vzorec velikosti n. Od enot, izbranih v vzorec, dobimo samo eno od vrednosti, ki jo enota izbere naključno, in sicer x z verjetnostjo $1/2$ in y z verjetnostjo $1/2$. Enote izbirajo odgovore neodvisno druga od druge in neodvisno od postopka izbire vzorca. Označimo z I_k indikator, da bo enota k izbrana v vzorec, in z J_k indikator, da bo enota za odgovor izbrala x. Po predpostavkah je vektor $(I_1, \ldots, I_N)$ neodvisen od vektorja $(J_1, \ldots, J_N)$, indikatorji $J_1, \ldots, J_N$ pa so med sabo neodvisni. S temi oznakami je vzorčno povprečje enako

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{N} I_k\big(x_k J_k + y_k(1-J_k)\big).$$

a. (5) Utemeljite, da je vzorčno povprečje nepristranska cenilka količine $\theta = \frac{1}{2}(\mu+\nu)$.

Rešitev: Zapišemo

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{N} I_k\big(x_k J_k + y_k(1-J_k)\big).$$

Vemo, da je $E(I_k) = \frac{n}{N}$ *in* $E(J_k) = \frac{1}{2}$. *Po predpostavki so* $J_1, \ldots, J_n$ *neodvisne od* $I_1, \ldots, I_N$. *Uporabimo linearnost in sledi*

$$E(\bar{X}) = \frac{1}{n}\sum_{k=1}^{N} \frac{n}{N}\left(\frac{x_k}{2} + \frac{y_k}{2}\right) = \frac{\mu+\nu}{2}.$$

b. (10) Izračunajte $\mathrm{var}(\bar{X}|J_1, J_2, \ldots, J_N)$.

Rešitev: Po predpostavkah o neodvisnosti lahko člene $x_k J_k + y_k(1-J_k)$ *obravnavamo kot konstante. Nadaljujemo lahko na vsaj tri načine.*

Prvi način: pišemo

$$\mathrm{var}(\bar{X}|J_1, J_2, \ldots, J_N) = E(\bar{X}^2|J_1, J_2, \ldots, J_N) - \big[E(\bar{X}|J_1, J_2, \ldots, J_N)\big]^2.$$

Velja

$$E(\bar{X}|J_1, \ldots, J_N) = \frac{1}{n}\sum_{k=1}^{N} E(I_k|J_1, \ldots, J_N)\big(x_k J_k + y_k(1-J_k)\big).$$

Zaradi neodvisnosti je $E(I_k|J_1, \ldots, J_N) = E(I_k) = \frac{n}{N}$, *torej je*

$$E(\bar{X}|J_1, \ldots, J_N) = \frac{1}{N}\sum_{k=1}^{N}\big(x_k J_k + y_k(1-J_k)\big).$$

Kvadriramo:

$$\begin{aligned}\left[E(\bar{X}|J_1,\ldots,J_N)\right]^2 &= \frac{1}{N^2}\sum_{k=1}^{N}\big(x_kJ_k + y_k(1-J_k)\big)^2 \\ &\quad + \frac{1}{N^2}\sum_{\substack{k,l=1\\k\neq l}}^{N}\big(x_kJ_k + y_k(1-J_k)\big)\big(x_lJ_l + y_l(1-J_l)\big)\,.\end{aligned}$$

Ker je $J_k^2 = J_k$, $(1-J_k)^2 = 1-J_k$ *in* $J_k(1-J_k)=0$, *pa je tudi*

$$\begin{aligned}\left[E(\bar{X}|J_1,\ldots,J_N)\right]^2 &= \frac{1}{N^2}\sum_{k=1}^{N}\big(x_k^2J_k + y_k^2(1-J_k)\big) \\ &\quad + \frac{1}{N^2}\sum_{\substack{k,l=1\\k\neq l}}^{N}\big(x_kJ_k + y_k(1-J_k)\big)\big(x_lJ_l + y_l(1-J_l)\big)\,.\end{aligned}$$

Nadalje je

$$\begin{aligned}&E(\bar{X}^2|J_1,\ldots,J_N) \\ &\quad = \frac{1}{n^2}\sum_{k=1}^{N}\sum_{l=1}^{N}E(I_kI_l|J_1,\ldots,J_N)\big(x_kJ_k + y_k(1-J_k)\big)\big(x_lJ_l + y_l(1-J_l)\big) \\ &\quad = \frac{1}{n^2}\sum_{k=1}^{N}\sum_{l=1}^{N}E(I_k|J_1,\ldots,J_N)\big(x_kJ_k + y_k(1-J_k)\big)^2 \\ &\qquad + \frac{1}{n^2}\sum_{k=1}^{N}\sum_{\substack{k,l=1\\k\neq l}}^{N}E(I_kI_l|J_1,\ldots,J_N)\big(x_kJ_k + y_k(1-J_k)\big)\big(x_lJ_l + y_l(1-J_l)\big) \\ &\quad = \frac{1}{Nn}\sum_{k=1}^{N}\big(x_kJ_k + y_k(1-J_k)\big)^2 \\ &\qquad + \frac{n-1}{N(N-1)n}\sum_{k=1}^{N}\sum_{\substack{k,l=1\\k\neq l}}^{N}\big(x_kJ_k + y_k(1-J_k)\big)\big(x_lJ_l + y_l(1-J_l)\big) \\ &\quad = \frac{1}{Nn}\sum_{k=1}^{N}\big(x_k^2J_k + y_k^2(1-J_k)\big) \\ &\qquad + \frac{n-1}{N(N-1)n}\sum_{k=1}^{N}\sum_{\substack{k,l=1\\k\neq l}}^{N}\big(x_kJ_k + y_k(1-J_k)\big)\big(x_lJ_l + y_l(1-J_l)\big)\,.\end{aligned}$$

Odštejemo in dobimo

$$
\begin{aligned}
&\operatorname{var}(\bar{X}|J_1, J_2, \ldots, J_N) \\
&\quad = \frac{N-n}{N^2 n} \sum_{k=1}^{N} \bigl(x_k^2 J_k + y_k^2 (1 - J_k)\bigr) \\
&\qquad\qquad - \frac{N-n}{N^2 (N-1) n} \sum_{\substack{k,l=1 \\ k \neq l}}^{N} \bigl(x_k J_k + y_k (1 - J_k)\bigr)\bigl(x_l J_l + y_l (1 - J_l)\bigr) \,.
\end{aligned}
$$

Drugi način: uporabimo običajne formule za varianco linearne kombinacije. *Računamo*

$$
\begin{aligned}
&\operatorname{var}(\bar{X}|J_1, J_2, \ldots, J_N) \\
&\quad = \frac{1}{n^2} \sum_{k=1}^{N} \bigl(x_k J_k + y_k (1 - J_k)\bigr)^2 \operatorname{var}(I_k) \\
&\qquad\quad + \frac{1}{n^2} \sum_{\substack{k,l=1 \\ k \neq l}}^{N} \bigl(x_k J_k + y_k (1 - J_k)\bigr)\bigl(x_l J_l + y_l (1 - J_l)\bigr) \operatorname{cov}(I_k, I_l) \\
&\quad = \frac{N-n}{N^2 n} \sum_{k=1}^{N} \bigl(x_k^2 J_k + y_k^2 (1 - J_k)\bigr) \\
&\qquad\quad - \frac{N-n}{N^2 (N-1) n} \sum_{\substack{k,l=1 \\ k \neq l}}^{N} \bigl(x_k J_k + y_k (1 - J_k)\bigr)\bigl(x_l J_l + y_l (1 - J_l)\bigr) \,,
\end{aligned}
$$

ker je isto kot pri prvem načinu.

Tretji način: uporabimo kar formulo za standardno napako pri enostavnem slučajnem vzorčenju za spremenljivko, ki je na k-ti enoti v populaciji enaka

$x_kJ_k + y_k(1 - J_k)$. *Dobimo*

$$
\begin{aligned}
&\operatorname{var}(\bar{X} \mid J_1, J_2, \ldots, J_N) \\
&\quad = \frac{N-n}{N-1}\,\frac{1}{Nn} \sum_{k=1}^{N} \left(x_kJ_k + y_k(1-J_k) - \frac{1}{N} \sum_{l=1}^{N} \big(x_lJ_l + y_l(1-J_l)\big) \right)^2 \\
&\quad = \frac{N-n}{N-1}\,\frac{1}{Nn} \sum_{k=1}^{N} \big(x_kJ_k + y_k(1-J_k)\big)^2 \\
&\qquad\quad - \frac{N-n}{N-1}\,\frac{1}{N^2 n} \left(\sum_{l=1}^{N} \big(x_lJ_l + y_l(1-J_l)\big) \right)^2 \\
&\quad = \frac{N-n}{N^2 n} \sum_{k=1}^{N} \big(x_kJ_k + y_k(1-J_k)\big)^2 \\
&\qquad\quad - \frac{N-n}{N^2(N-1)n} \sum_{\substack{k,l=1 \\ k \neq l}}^{N} \big(x_kJ_k + y_k(1-J_k)\big)\big(x_lJ_l + y_l(1-J_l)\big) \\
&\quad = \frac{N-n}{N^2 n} \sum_{k=1}^{N} \big(x_k^2 J_k + y_k^2(1-J_k)\big) \\
&\qquad\quad - \frac{N-n}{N^2(N-1)n} \sum_{\substack{k,l=1 \\ k \neq l}}^{N} \big(x_kJ_k + y_k(1-J_k)\big)\big(x_lJ_l + y_l(1-J_l)\big)\,,
\end{aligned}
$$

kar je isto kot pri prvih dveh načinih.

c. (10) Pokažite, da je

$$
\begin{aligned}
\operatorname{var}(\bar{X}) = \frac{(N-n)}{nN^2} &\left[\sum_{k=1}^{N} \frac{x_k^2 + y_k^2}{2} - \frac{1}{N-1} \sum_{\substack{k,l=1 \\ k \neq l}}^{N} \frac{(x_k + y_k)(x_l + y_l)}{4} \right] \\
&+ \frac{1}{N^2} \sum_{k=1}^{N} \frac{(x_k - y_k)^2}{4}\,.
\end{aligned}
$$

Namig: formula za razcep variance.

Rešitev: Iz prvega načina rešitve točke b. se spomnimo, da je

$$E(\bar{X} \mid J_1, \ldots, J_N) = \frac{1}{N} \sum_{k=1}^{N} \big(x_kJ_k + y_k(1-J_k)\big)\,.$$

Velja $\operatorname{var}\big(x_kJ_k + y_k(1-J_k)\big) = \operatorname{var}\big((x_k - y_k)J_k + y_k\big) = (x_k - y_k)^2 \operatorname{var}(J_k) = \frac{1}{4}(x_k - y_k)^2$. *Zaradi neodvisnosti je*

$$\operatorname{var}\big(E(\bar{X} \mid J_1, \ldots, J_N)\big) = \frac{1}{4N^2} \sum_{k=1}^{N} (x_k - y_k)^2\,.$$

Nadalje je

$$
\begin{aligned}
E\big(\mathrm{var}(\bar{X}|J_1,\ldots,J_N)\big) = {} & \frac{N-n}{2N^2 n}\sum_{k=1}^{N}(x_k^2+y_k^2) \\
& - \frac{N-n}{4N^2(N-1)n}\sum_{\substack{k,l=1\\k\neq l}}^{N}(x_k+y_k)(x_l+y_l)\,.
\end{aligned}
$$

Zdaj pa v formulo za razcep variance

$$
\mathrm{var}(\bar{X}) = E\big(\mathrm{var}(\bar{X}|J_1,J_2,\ldots,J_N)\big) + \mathrm{var}\big(E(\bar{X}|J_1,\ldots,J_N)\big)
$$

vstavimo dobljene izraze in dobimo želeni rezultat.

3. (25) Naj bodo opazovane vrednosti $x_1, x_2, \ldots, x_n$ kot vzorec $X_1, X_2, \ldots, X_n$ med sabo neodvisnih slučajnih spremenljivk z *inverzno Gaussovo* porazdelitvijo, dano z gostoto

$$f(x) = \sqrt{\frac{\tau}{2\pi x^3}} \exp\left\{-\frac{\tau}{2x\mu^2}(x-\mu)^2\right\}, \quad x > 0,\ \tau > 0,\ \mu > 0.$$

Kot znano upoštevajte, da je $E(X_1) = \mu$. Privzemite tudi, da je verjetje največje v stacionarni točki in da so izpolnjene vse predpostavke za uporabo asimptotskih rezultatov za cenilke po metodi največjega verjetja.

a. (10) Poiščite cenilki za parametra μ in τ po metodi največjega verjetja.

Rešitev: Zapišemo

$$\begin{aligned}
\ell(\mu, \tau|\mathbf{X}) &= \frac{n}{2}\log\tau - \frac{n}{2}\log(2\pi) - \frac{3}{2}\sum_{k=1}^{n}\log x_k - \frac{\tau}{2\mu^2}\sum_{k=1}^{n}\frac{(x_k-\mu)^2}{x_k} \\
&= \frac{n}{2}\log\tau - \frac{n}{2}\log(2\pi) - \frac{3}{2}\sum_{k=1}^{n}\log x_k - \frac{\tau}{2\mu^2}\sum_{k=1}^{n}x_k + \frac{n\tau}{\mu} - \frac{\tau}{2}\sum_{k=1}^{n}\frac{1}{x_k}\,.
\end{aligned}$$

Parcialno odvajamo in izenačimo z 0. *Dobimo*

$$\begin{aligned}
\frac{\partial\ell}{\partial\mu} &= \frac{\tau}{\mu^3}\sum_{k=1}^{n}x_k - \frac{n\tau}{\mu^2} = 0\,, \\
\frac{\partial\ell}{\partial\tau} &= \frac{n}{2\tau} - \frac{1}{2\mu^2}\sum_{k=1}^{n}\frac{(x_k-\mu)^2}{x_k} = 0
\end{aligned}$$

ter od tod cenilki

$$\hat{\mu} = \bar{x} \qquad \text{in} \qquad \hat{\tau} = \frac{n\bar{x}^2}{\sum_{k=1}^{n}(x_k-\bar{x})^2/x_k} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n}\frac{1}{x_k} - \frac{1}{\bar{x}}}\,.$$

Cenilki obstajata, brž ko izmerjene vrednosti $x_1, \ldots, x_n$ niso vse enake. Če pa so vse enake, recimo x, je pri $\mu = x$ logaritem verjetja strogo naraščajoč v τ in zato ni navzgor omejen. V tem primeru torej cenilka po metodi največjega verjetja ni definirana.

b. (10) Izračunajte Fisherjevo matriko informacije $I(\mu, \tau)$.

Rešitev: Izračunamo ustrezne druge odvode $\ell(\mu, \tau|\mathbf{X})$ za $n = 1$.

$$\begin{aligned}
\frac{\partial^2\ell}{\partial\mu^2} &= -\frac{3\tau x_1}{\mu^4} + \frac{2\tau}{\mu^3}\,, \\
\frac{\partial^2\ell}{\partial\tau^2} &= -\frac{1}{2\tau^2}\,, \\
\frac{\partial^2}{\partial\mu\partial\tau} &= \frac{x_1}{\mu^3} - \frac{1}{\mu^2}\,.
\end{aligned}$$

Upoštevamo še namig in dobimo

$$\mathrm{I}(\mu,\tau) = \begin{pmatrix} \tau/\mu^3 & 0 \\ 0 & 1/2\tau^2 \end{pmatrix}$$

c. (5) Zapišite interval zaupanja pri stopnji zaupanja $1-\alpha = 0{,}95$ za parameter τ na osnovi podatkov $X_1, X_2, \ldots, X_n$.

Rešitev: Interval zaupanja je

$$\hat{\tau} \pm 1{,}96 \cdot \frac{\hat{\tau}\sqrt{2}}{\sqrt{n}}\,.$$

4. (25) Predpostavite, da velja naslednji regresijski model:

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

z $E(\boldsymbol{\epsilon}) = 0$ in

$$\mathrm{var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{V}\,,$$

kjer je

$$v_{ij} = \frac{\rho^{|i-j|}}{1-\rho^2}\,.$$

Privzemite, da je σ^2 neznana konstanta, $\rho \in (-1,1)$ pa znan.

a. (15) Naj bodo komponente $Z_1, Z_2, \ldots, Z_n$ vektorja $\mathbf{Z}$ dane z *Cochran-Orcuttovo* transformacijo

$$Z_1 = \sqrt{1-\rho^2}\, Y_1 \quad \text{in} \quad Z_i = Y_i - \rho Y_{i-1}$$

za $i = 2, 3, \ldots, n$. Izračunajte $\mathrm{var}(\mathbf{Z})$ in navedite najboljšo linearno nepristansko cenilko parametra $\boldsymbol{\beta}$.

Rešitev: Najprej računamo

$$\mathrm{var}(Z_1) = \sigma^2$$

in za $i = 2, 3, \ldots n$

$$\begin{aligned}
\mathrm{cov}(Z_1, Z_i) &= \sqrt{1-\rho^2}\,\mathrm{cov}\left(Y_1, Y_i - \rho Y_{i-1}\right) \\
&= \frac{\sigma^2\sqrt{1-\rho^2}}{1-\rho^2}\left(\rho^{i-1} - \rho\cdot\rho^{i-2}\right) \\
&= 0\,.
\end{aligned}$$

Nadaljujemo za $1 < i \leq n$*:*

$$\begin{aligned}
\mathrm{var}(Z_i) &= \mathrm{var}(Y_i - \rho Y_{i-1}) \\
&= \mathrm{var}(Y_i) - 2\rho\,\mathrm{cov}(Y_i, Y_{i-1}) + \rho^2\mathrm{var}(Y_{i-1}) \\
&= \frac{\sigma^2}{1-\rho^2} - 2\,\frac{\rho^2\sigma}{1-\rho^2} + \frac{\rho^2\sigma^2}{1-\rho^2} \\
&= \sigma^2\,,
\end{aligned}$$

ter

$$\begin{aligned}
\mathrm{cov}(Z_i, Z_j) &= \mathrm{cov}(Y_i - \rho Y_{i-1}, Y_j - \rho Y_{j-1}) \\
&= \frac{\sigma^2}{1-\rho^2}\left(\rho^{j-i} - \rho^{j-i+2} - \rho^{j-i} + \rho^{j-i+2}\right) \\
&= 0\,.
\end{aligned}$$

Če ustrezno popravimo še vrstice $\mathbf{X}_i$ *matrike* $\mathbf{X}$ *v*

$$\tilde{\mathbf{X}}_1 = \sqrt{1-\rho^2}\,\mathbf{X}_1 \quad \textit{in} \quad \tilde{\mathbf{X}}_i = \mathbf{X}_i - \rho\mathbf{X}_{i-1}$$

ter slučajne vplive v

$$\eta_1 = \sqrt{1-\rho^2}\,\epsilon_1 \quad \textit{in} \quad \eta_i = \epsilon_i - \rho\epsilon_{i-1}\,,$$

model

$$\mathbf{Z} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \boldsymbol{\eta}$$

ustreza predpostavkam izreka Gauss-Markova in je najboljša nepristranska cenilka parametra $\boldsymbol{\beta}$ enaka

$$\hat{\boldsymbol{\beta}} = \left(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^T\mathbf{Z}\,.$$

b. (10) Naj bo $\hat{\boldsymbol{\beta}}$ cenilka $\boldsymbol{\beta}$ po običajni metodi najmanjših kvadratov, torej

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}\,.$$

Naj bo

$$\hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

vektor ostankov. Pokažite, da je

$$E\left(\sum_{i=1}^n \hat{Y}_i^2\right) = \sigma^2 \mathrm{Sl}\left(\mathbf{V} - \mathbf{V}\mathbf{H}\right)\,.$$

Rešitev: Naj bo

$$\mathbf{H} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\,.$$

Velja $\hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ in posledično

$$\mathrm{var}(\hat{\mathbf{Y}}) = \sigma^2(\mathbf{I} - \mathbf{H})\mathbf{V}(\mathbf{I} - \mathbf{H})\,.$$

Ker je $\mathbf{H}\mathbf{X} = \mathbf{X}$, je $E(\hat{\mathbf{Y}}) = 0$. Sledi

$$E(\hat{Y}_i^2) = \mathrm{var}(\hat{Y}_i)$$

in posledično

$$E(\hat{Y}_i^2) = \sigma^2\big((\mathbf{I} - \mathbf{H})\mathbf{V}(\mathbf{I} - \mathbf{H})\big)_{ii}\,.$$

Ko seštejemo, dobimo sled, ta pa se ne spremeni, če spremenimo vrstni red množenja matrik in upoštevamo, da je $\mathbf{I} - \mathbf{H}$ idempotentna matrika.