Ime in priimek: _____ Vpisna št: | | | | | | | | |

# FAKULTETA ZA MATEMATIKO IN FIZIKO

## Oddelek za matematiko

## Statistika

## Pisni izpit

15. junij 2018

### 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) Verjetnost nekega dogodka je 0,1706, a to je težko izračunati. Verjetnostnik Karl in statističarka Monika, ki omenjene eksaktne vrednosti ne poznata, se lotita ocenjevanja vsak na svoj način. Karl s pomočjo centralnega limitnega izreka dobi približek 0,1702, Monika pa gre z metodo Monte Carlo: pri vsaki izvedbi simulira poskus, pri katerem se lahko zgodi ta dogodek, in njena ocena za verjetnost dogodka je delež izvedb, pri katerih se je dejansko zgodil dani dogodek.

a. (15) Monika želi doseči natančnost, primerljivo s Karlovo. Pri koliko izvedbah bo standardna napaka metode Monte Carlo (približno) enaka napaki, ki jo je naredil Karl?

*Opomba.* Monika tega izračuna ne more narediti, ker ne pozna točne verjetnosti, torej tudi ne napake, ki jo je naredil Karl. Lahko pa pri vse več izvedbah primerja razliko med svojim in Karlovim približkom ter ocenjeno standardno napako za svoj približek. Lahko se npr. odloči, da s simulacijo zaključi, brž ko je ocenjena standardna napaka manjša od opažene razlike.

*Rešitev: Označimo s* $p = 0{,}1706$ *dejansko verjetnost danega dogodka. Če Monika naredi* $n$ *simulacij, standardna napaka metode Monte Carlo znaša* $\sqrt{p(1-p)/n}$. *Če s* $\hat{p}_K = 0{,}1702$ *označimo Karlov približek, dobimo*

$$\sqrt{\frac{p(1-p)}{n}} = |\hat{p}_K - p|$$

*oziroma*

$$n = \frac{p(1-p)}{(\hat{p}_K - p)^2} \doteq 884{,}348\,.$$

b. (10) Recimo, da je Monika naredila dvakrat toliko izvedb, kot je opisano v prejšnji točki. Približno kolikšna je verjetnost, da je dobila boljši približek kot Karl?

*Rešitev: Boljši približek pomeni vrednost med* 0,1702 *in* 0,1710. *Če z* $\hat{p}_M$ *označimo Monikin približek, je torej iskana verjetnost enaka*

$$\begin{aligned} P(0{,}1702 < \hat{p}_M < 0{,}1710) &\approx \Phi\left(\frac{0{,}1710 - p}{\sqrt{p(1-p)}}\sqrt{2n}\right) - \Phi\left(\frac{0{,}1702 - p}{\sqrt{p(1-p)}}\sqrt{2n}\right) \\ &= \Phi\big(\sqrt{2}\,\big) - \Phi\big(-\sqrt{2}\,\big) \\ &= 2\,\Phi\big(\sqrt{2}\,\big) - 1 \\ &\doteq 0{,}8427\,. \end{aligned}$$

**2.** (25) Statistik si ogleduje impresivno bibliografijo svojega uspešnega kolega. Iz izpiska hitro razbere, da je kolega napisal $N$ člankov, bodisi samostojnih bodisi v soavtorstvu. Statistika pa zanima prispevek dotičnega kolega k tem člankom, pri čemer privzame, da k posameznemu članku vsi avtorji prispevajo enako. Članki se torej delijo s številom avtorjev. Primer: če je $N = 4$ in je od tega en samostojen članek, en članek, pri katerem sta avtorja dva, in dva članka, pri katerih so avtorji trije, ustrezni prispevek znaša $1 + 1/2 + 2/3 = 2\,1/6$ članka.

Statistik bi sicer lahko za vse članke iz izpiska pogledal število avtorjev in ustrezno seštel, a ker je njegov kolega zelo uspešen in ima zelo veliko člankov, raje vzame enostavni slučajni vzorec $n$ člankov in na podlagi le-tega oceni ustrezni prispevek.

a. (10) Predlagajte nepristransko cenilko za zahtevani prispevek, ki bo vse članke iz vzorca obravnavala enakovredno. Utemeljite nepristranskost.

*Rešitev: Če z $x_k$ označimo število avtorjev $k$-tega članka iz bibliografije, ocenjujemo količino:*

$$\theta = \sum_{k=1}^{N} \frac{1}{x_k}\,.$$

*Cenilko nastavimo v obliki:*

$$\hat{\theta} = c \sum_{i=1}^{n} \frac{1}{Y_i}\,,$$

*kjer je $Y_i$ število avtorjev $i$-tega članka iz vzorca. Ker je vsaka slučajna spremenljivka $Y_i$ z enakimi verjetnostmi enaka $x_1, x_2, \ldots, x_N$, velja*
$E\left(\frac{1}{Y_i}\right) = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{x_k}$, *torej mora biti* $c = \frac{N}{n}$.

b. (15) Zadevo pogledamo malo širše: $N$ in $n$ naj ostaneta fiksna, še vedno vzamemo enostavni slučajni vzorec $n$ enot, število avtorjev posameznega članka iz bibliografije pa naj bo slučajna spremenljivka. Če je $X_k$ število avtorjev $k$-tega članka iz bibliografije, naj bo:

$$P(X_k = l) = (1-q)^2\, l q^{l-1}\,; \quad l = 1, 2, 3, \ldots\,,$$

kjer je $q \in (0, 1)$ fiksno število. Privzamemo tudi, da so slučajne spremenljivke $X_1, \ldots, X_N$ med seboj neodvisne in da je slučajni vektor $(X_1, \ldots, X_N)$ neodvisen od tega, katere enote so izbrane v vzorec. Izračunajte standardno napako cenilke iz prejšnje točke.

*Namig: pogojujte na vzorec.*

*Rešitev: Standardna napaka je kvadratni koren variance, slednjo pa lahko izračunamo na vsaj dva načina.*

Prvi način: *v skladu z namigom pogojujemo na slučajni vektor $\mathbf{V} = (V_1, V_2, \ldots, V_n)$, kjer je $V_i$ številka enote v populaciji, ki je izbrana v vzorec kot $i$-ta: tako je $Y_i = X_{V_i}$. Ker je $\mathbf{V}$ neodvisen od $(X_1, X_2, \ldots, X_N)$, so tudi pogojno na dogodek $V_1 = k_1, V_2 = k_2, \ldots, V_n = k_n$ slučajne spremenljivke $X_1, \ldots, X_N$ še vedno neodvisne in velja*

$$P(X_k = l|\mathbf{V}) = (1-q)^2\, l q^{l-1}\,; \quad l = 1, 2, 3, \ldots$$

*Enako velja za slučajne spremenljivke $Y_1 = X_{k_1}, Y_2 = X_{k_2}, \ldots, Y_n = X_{k_n}$: pogojno na $\mathbf{V}$ so neodvisne in velja*

$$P(Y_i = l|\mathbf{V}) = (1-q)^2\, l q^{l-1}\,; \quad l = 1, 2, 3, \ldots$$

*Ker ima slučajni vektor $(Y_1, \ldots, Y_n)$ pogojno na $V_1 = k_1, V_2 = k_2, \ldots, V_n = k_n$ enako porazdelitev ne glede na $k_1, k_2, \ldots, k_n$, vse skupaj velja tudi brezpogojno: slučajne spremenljivke $Y_1, Y_2, \ldots, Y_n$ so neodvisne in velja*

$$P(Y_i = l) = (1-q)^2\, l q^{l-1}\,; \quad l = 1, 2, 3, \ldots$$

*Izračunajmo*

$$\begin{aligned} E\left(\frac{1}{Y_i}\right) &= (1-q)^2 \sum_{l=1}^{\infty} q^{l-1} = 1-q\,, \\ E\left(\frac{1}{Y_i}\right) &= (1-q)^2 \sum_{l=1}^{\infty} \frac{q^{l-1}}{l} = -\frac{(1-q)^2 \ln(1-q)}{q}\,, \\ \sigma^2 := \operatorname{var}\left(\frac{1}{Y_i}\right) &= (1-q)^2 \left(-\frac{\ln(1-q)}{q} - 1\right). \end{aligned}$$

*Iz neodvisnosti sledi*

$$\operatorname{var}(\hat{\theta}) = \frac{N^2}{n^2} \sum_{i=1}^{n} \operatorname{var}\left(\frac{1}{Y_i}\right) = \frac{N^2}{n}\,\sigma^2\,.$$

Drugi način: *pogojujemo na vrednosti $\mathbf{X} = (X_1, X_2, \ldots, X_N)$. Tokrat varianco razcepimo:*

$$\operatorname{var}(\hat{\theta}) = \operatorname{var}\big(E(\hat{\theta}|\mathbf{X})\big) + E\big(\operatorname{var}(\hat{\theta}|\mathbf{X})\big)\,.$$

*Ker je vsaka slučajna spremenljivka $Y_i$ pogojno na $\mathbf{X}$ z enakimi verjetnostmi enaka $X_1, X_2, \ldots, X_N$, velja*

$$E\left(\frac{1}{Y_i}\;\middle|\;\mathbf{X}\right) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{X_k}\,,$$

*od koder sledi*

$$E(\hat{\theta}|\mathbf{X}) = \sum_{k=1}^{N}\frac{1}{X_k}$$

*in zaradi neodvisnosti*

$$\operatorname{var}\big(E(\hat{\theta}|\mathbf{X})\big) = \sum_{k=1}^{N} \operatorname{var}\left(\frac{1}{X_k}\right) = N\sigma^2\,.$$

*Nadalje iz teorije vzorčenja, uporabljene pogojno na $\mathbf{X}$, sledi*

$$\operatorname{var}(\hat{\theta}|\mathbf{X}) = N^2\,\frac{N-n}{N-1}\,\frac{1}{Nn}\sum_{k=1}^{N}\left(\frac{1}{X_k} - \frac{1}{N}\sum_{l=1}^{N}\frac{1}{X_l}\right)^2.$$

*Iz znanega rezultata za pričakovano vrednost empirične variance dobimo*

$$E\big(\mathrm{var}(\hat{\theta}|\mathbf{X})\big) = \frac{N(N-n)}{n}\,\sigma^2\,.$$

*Seštejemo in dobimo*

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

*kar je isto kot prej.*

**Opomba**. *Glede na to, da je tisto, kar ocenjujemo, namreč* $\theta = \sum_{k=1}^{N} \frac{1}{X_k}$, *zdaj slučajna količina, standardna napaka kot koren variance ni več merodajna. Merodajen je koren srednje kvadratične napake* $E\big[(\hat{\theta}-\theta)^2\big]$, *slednja pa je zaradi slučajnosti količine* $\theta$ *različna od* $\mathrm{var}(\theta)$ *– znaša* $\frac{N^2}{n}\big(1-\frac{n}{N}\big)\sigma^2$.

**3.** (25) Opazovane vrednosti naj bodo pari $(x_1, y_1), \ldots, (x_n, y_n)$, $n \geq 2$, za katere privzamemo, da so vzorec neodvisnih realizacij neizrojene dvorazsežne normalne porazdelitve $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

a. (15) Poiščite cenilko za $\boldsymbol{\Sigma}$ po metodi največjega verjetja.

*Namig*: matriko parametrizirajte v obliki:

$$\boldsymbol{\Sigma} = \left( \frac{a}{\cos\theta} \begin{bmatrix} q & \sin\theta \\ \sin\theta & 1/q \end{bmatrix} \right)^{-1} .$$

Kot znano lahko privzamete, da se da vsaka pozitivno definitna matrika parametrizirati na ta način.

*Rešitev: Najprej izračunamo* $\det(\boldsymbol{\Sigma}) = 1/a^2$. *Označimo še* $\mathbf{x} = (x_1, \ldots, x_n)$ *in* $\mathbf{y} = (y_1, \ldots, y_n)$. *Funkcijo verjetja tako lahko zapišemo kot*

$$L(a, q, \theta | \mathbf{x}, \mathbf{y}) = \left(\frac{a}{2\pi}\right)^n \exp\left( -\frac{aq}{2\cos\theta} \sum_{k=1}^n x_k^2 - a \operatorname{tg}\theta \sum_{k=1}^n x_k y_k - \frac{a}{2q\cos\theta} \sum_{k=1}^n y_k^2 \right) .$$

*Če označimo*

$$m_{xx} = \frac{1}{n}\sum_{k=1}^n x_k^2 , \qquad m_{xy} = \frac{1}{n}\sum_{k=1}^n x_k y_k , \qquad m_{yy} = \frac{1}{n}\sum_{k=1}^n y_k^2 ,$$

*lahko logaritem verjetja zapišemo v obliki*

$$\ell(a, q, \theta | \mathbf{x}, \mathbf{y}) = n\left( \log a - \log(2\pi) - \frac{aq}{2\cos\theta} m_{xx} - a \operatorname{tg}\theta\, m_{xy} - \frac{a}{2q\cos\theta} m_{yy} \right) .$$

*Odvajamo:*

$$\begin{aligned} \frac{\partial \ell}{\partial a} &= n\left( \frac{1}{a} - \frac{q}{2\cos\theta} m_{xx} - \operatorname{tg}\theta\, m_{xy} - \frac{1}{2q\cos\theta} m_{yy} \right) , \\ \frac{\partial \ell}{\partial q} &= n\left( -\frac{a}{2\cos\theta} m_{xx} + \frac{a}{2q^2\cos\theta} m_{yy} \right) , \\ \frac{\partial \ell}{\partial \theta} &= n\left( -\frac{aq\sin\theta}{2\cos^2\theta} m_{xx} - \frac{a}{\cos^2\theta} m_{xy} - \frac{a\sin\theta}{2q\cos^2\theta} m_{yy} \right) . \end{aligned}$$

*Ko izenačimo z nič, po nekaj računanja dobimo ustrezne cenilke:*

$$\hat{a} = \frac{1}{\sqrt{m_{xx}m_{yy} - m_{xy}^2}} , \qquad \hat{q} = \left(\frac{m_{yy}}{m_{xx}}\right)^{1/2} , \qquad \hat{\theta} = -\arcsin\frac{m_{xy}}{\sqrt{m_{xx}m_{yy}}}$$

*in spet po nekaj računanja:*

$$\hat{\boldsymbol{\Sigma}} = \left( \frac{\hat{a}}{\cos\hat{\theta}} \begin{bmatrix} \hat{q} & \sin\hat{\theta} \\ \sin\hat{\theta} & 1/\hat{q} \end{bmatrix} \right)^{-1} = \begin{bmatrix} m_{xx} & m_{xy} \\ m_{xy} & m_{yy} \end{bmatrix} .$$

b. (10) Preizkusiti želimo domnevo, da je matrika $\mathbf{\Sigma}$ diagonalna. Poiščite testno statistiko po metodi kvocienta verjetij in navedite njeno aproksimativno porazdelitev pri velikem vzorcu.

*Rešitev: Domnevo $H_0$, da je matrika $\mathbf{\Sigma}$ diagonalna, lahko izrazimo tudi tako, $\mathbf{\Sigma}$ parametriziramo tako kot v prejšnji točki, pri čemer postavimo $\theta = 0$. Velja*

$$\ell(a, q, 0|\mathbf{x}, \mathbf{y}) = n\left(\log a - \log(2\pi) - \frac{aq}{2}\, m_{xx} - \frac{a}{2q}\, m_{yy}\right)$$

*in*

$$\begin{aligned}\frac{\partial \ell}{\partial a} &= n\left(\frac{1}{a} - \frac{q}{2}\, m_{xx} - \frac{1}{2q}\, m_{yy}\right), \\ \frac{\partial \ell}{\partial q} &= n\left(-\frac{a}{2}\, m_{xx} + \frac{a}{2q^2}\, m_{yy}\right).\end{aligned}$$

*Izenačimo z nič in dobimo ustrezni cenilki:*

$$\tilde{a} = \frac{1}{\sqrt{m_{xx} m_{yy}}}, \qquad \tilde{q} = \left(\frac{m_{yy}}{m_{xx}}\right)^{1/2}.$$

*Za izračun testne statistike potrebujemo maksimuma logaritmov verjetij:*

$$\begin{aligned}\ell(\hat{a}, \hat{q}, \hat{\theta}|\mathbf{x}, \mathbf{y}) &= -n\Big(\tfrac{1}{2}\log\big(m_{xx} m_{yy} - m_{xy}^2\big) + 1 + \log(2\pi)\Big), \\ \ell(\tilde{a}, \tilde{q}, 0|\mathbf{x}, \mathbf{y}) &= -n\Big(\tfrac{1}{2}\log(m_{xx} m_{yy}) + 1 + \log(2\pi)\Big).\end{aligned}$$

*Odštejemo, pomnožimo z 2 in dobimo testno statistiko:*

$$\begin{aligned}\lambda &= 2\,\ell(\hat{a}, \hat{q}, \hat{\theta}|\mathbf{x}, \mathbf{y}) - 2\,\ell(\tilde{a}, 1, 0|\mathbf{x}, \mathbf{y}) \\ &= n \log \frac{m_{xx} m_{yy}}{m_{xx} m_{yy} - m_{xy}^2} \\ &= n \log \frac{1}{1 - \frac{m_{xy}^2}{m_{xx} m_{yy}}}.\end{aligned}$$

*Ker ima širši model dimenzijo 3, ožji pa dimenzijo 2, je pri velikih vzorcih približno $\lambda \sim \chi^2(1)$.*

*Ničelno domnevo zavrnemo, če je testna statistika $\lambda$ dovolj velika. Vidimo, da je to takrat, ko je dovolj velika absolutna vrednost vzorčne korelacije $\frac{m_{xy}}{\sqrt{m_{xx} m_{yy}}}$: tudi slednjo bi lahko vzeli za testno statistiko.*

**4.** (25) Privzemite regresijski model

$$\begin{aligned} Y_1 &= \alpha + \beta x_1 + \epsilon_1 \\ Y_2 &= \alpha + \beta x_2 + \epsilon_1 + \epsilon_2 \\ \cdots &= \quad \cdots\cdots\cdots \\ Y_n &= \alpha + \beta x_n + \epsilon_1 + \epsilon_2 + \cdots + \epsilon_n \,, \end{aligned}$$

kjer predpostavljamo $E(\epsilon_k) = 0$, $\mathrm{var}(\epsilon_k) = \sigma^2$ za vse $k = 1, 2, \ldots, n$ in $\mathrm{cov}(\epsilon_k, \epsilon_l) = 0$ za $k \neq l$. Predpostavite, da so vsi $x_1, x_2, \ldots, x_n$ med sabo različni.

a. (10) Eksplicitno poiščite najboljši linearni nepristranski cenilki za $\alpha$ in $\beta$.

*Rešitev:*

Prvi način: *model prevedemo na standardno linearno regresijo, tako da uvedemo*

$$\boldsymbol{U} = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \\ \vdots \\ U_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 - Y_1 \\ Y_3 - Y_2 \\ \vdots \\ Y_n - Y_{n-1} \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} 1 & x_1 \\ 0 & x_2 - x_1 \\ 0 & x_3 - x_2 \\ \vdots & \vdots \\ 0 & x_n - x_{n-1} \end{bmatrix}, \quad \boldsymbol{\gamma} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

*Tedaj namreč velja* $\mathbf{U} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$*, slučajni vektor* $\boldsymbol{\epsilon}$ *pa ima pričakovano vrednost nič in kovariančno matriko* $\sigma^2 \mathbf{I}$*. Najboljša nepristranska linearna cenilka za* $\boldsymbol{\gamma}$ *v novem modelu sovpada z najboljšo linearno nepristransko cenilko v izvirnem modelu, saj je* $\mathbf{U}$ *dobljen iz vektorja* $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ *z linearnim izomorfizmom. Ta predstavlja bijektivno korespondenco med linearnimi funkcionali na* $\mathbf{Y}$ *in linearnimi funkcionali na* $\mathbf{U}$*.*

*Označimo*

$$\begin{aligned} Q_1 &:= \sum_{k=2}^{n} (x_k - x_{k-1})^2 = x_1^2 + 2\sum_{k=2}^{n-1} x_k^2 + x_n^2 + 2\sum_{k=2}^{n} x_{k-1} x_k \,, \\ S_1 &:= \sum_{k=2}^{n} (x_k - x_{k-1})(Y_k - Y_{k-1}) \\ &= x_1 Y_1 + 2\sum_{k=2}^{n-1} x_k Y_k + x_n Y_n + \sum_{k=2}^{n} \big(x_{k-1} Y_k + x_k Y_{k-1}\big) \end{aligned}$$

*in izračunamo*

$$\mathbf{Z}^T\mathbf{Z} = \begin{bmatrix} 1 & x_1 \\ x_1 & x_1^2 + Q_1^2 \end{bmatrix}, \qquad \big(\mathbf{Z}^T\mathbf{Z}\big)^{-1} = \frac{1}{Q_1} \begin{bmatrix} x_1^2 + Q_1 & -x_1 \\ -x_1 & 1 \end{bmatrix},$$

$$\mathbf{Z}^T\boldsymbol{U} = \begin{bmatrix} Y_1 \\ x_1 Y_1 + S_1 \end{bmatrix}.$$

*Po izreku Gaussa in Markova je najboljša nepristranska linearna cenilka za* $\boldsymbol{\gamma}$ *vektor*

$$\hat{\boldsymbol{\gamma}} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\boldsymbol{U} = \frac{1}{Q_1}\begin{bmatrix} Q_1Y_1 - x_1S_1 \\ S_1 \end{bmatrix},$$

*se pravi, da sta najboljši nepristranski linearni cenilki za* $\alpha$ *in* $\beta$*:*

$$\hat{\alpha} = Y_1 - \frac{x_1S_1}{Q_1}, \qquad \hat{\beta} = \frac{S_1}{Q_1}.$$

**Opomba**. *Ker so vsi* $x_1, x_2, \ldots, x_n$ *med seboj različni, je tudi* $Q_1 > 0$*, zato smemo deliti.*

Drugi način: *uporabimo posplošeni izrek Gaussa in Markova za primer, ko je kovariančna matrika šumov večkratnik znane matrike* $\boldsymbol{\Sigma}$*. Izvirni model zapišemo v vektorski obliki kot* $\boldsymbol{Y} = \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\eta}$*, kjer sta* $\boldsymbol{Y}$ *in* $\boldsymbol{\gamma}$ *kot pri prvem načinu ter*

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad in \quad \boldsymbol{\eta} = \begin{bmatrix} \epsilon_1 \\ \epsilon_1 + \epsilon_2 \\ \vdots \\ \epsilon_1 + \epsilon_2 + \cdots + \epsilon_n \end{bmatrix}.$$

*Po posplošenem izreku Gaussa in Markova je najboljša nepristranska linearna cenilka za* $\boldsymbol{\gamma}$ *enaka* $\left(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$*. V našem primeru je kovariančna matrika slučajnega vektorja* $\boldsymbol{\eta}$ *enaka* $\sigma^2\boldsymbol{\Sigma}$*, kjer je*

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 2 & 2 & \cdots & 2 & 2 \\ 1 & 2 & 3 & \cdots & 3 & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2 & 3 & \cdots & n-1 & n-1 \\ 1 & 2 & 3 & \cdots & n-1 & n \end{bmatrix}.$$

*Izračunamo*

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix},$$

$$\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ x_1 & x_1^2 + Q_1^2 \end{bmatrix}, \qquad \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ x_1Y_1 + S_1 \end{bmatrix}$$

*in vidimo, da res dobimo isto cenilko kot pri prvem načinu.*

b. (5) Eksplicitno navedite standardni napaki za najdeni cenilki parametrov $\alpha$ in $\beta$.

*Rešitev:* Prvi način: neposredno. *Iz*

$$\operatorname{var}(Y_k - Y_{k-1}) = \operatorname{var}(\epsilon_k) = \sigma^2; \quad k = 2, 3, \ldots, n$$

*dobimo*

$$\operatorname{var}(\hat{\beta}) = \frac{\sigma^2}{Q_1^2}\sum_{k=2}^{n}(x_k - x_{k-1})^2 = \frac{\sigma^2}{Q_1}\,.$$

*Nadalje opazimo, da je* $\hat{\alpha} = Y_1 - x_1\hat{\beta}$. *Slučajni spremenljivki* $Y_1$ *in* $\hat{\beta}$ *sta neodvisni, saj je* $Y_1$ *deterministično odvisna le od* $\epsilon_1$, $\hat{\beta}$ *pa le od* $\epsilon_2, \ldots, \epsilon_n$. *Sledi*

$$\operatorname{var}(\hat{\alpha}) = \operatorname{var}(Y_1) + x_1^2 \operatorname{var}(\hat{\beta}) = \sigma^2\left(1 + \frac{x_1^2}{Q_1}\right)\,.$$

*S korenjenjem varianc dobimo standardni napaki.*

Drugi način: s pomočjo kovariančne matrike. *Znano je namreč, da je*

$$\mathbf{var}(\hat{\boldsymbol{\gamma}}) = \sigma^2\big(\mathbf{Z}^T\mathbf{Z}\big)^{-1} = \frac{\sigma^2}{Q_1}\begin{bmatrix} x_1^2 + Q_1 & -x_1 \\ -x_1 & 1 \end{bmatrix}$$

*Varianci cenilk sta diagonalca matrike in se ujemata z rezultatoma iz prvega načina.*

c. (5) Predlagajte nepristransko cenilko za $\sigma^2$.

*Rešitev: Cenilko spet dobimo iz standardne linearne regresije:*

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\|\boldsymbol{U} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\|^2 \\
&= \frac{1}{n-2}\big(\boldsymbol{U} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\big)^T\big(\boldsymbol{U} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\big) \\
&= \frac{1}{n-2}\big(\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\gamma}}\big)^T\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\gamma}}\big) \\
&= \frac{1}{n-2}\left[(Y_1 - \hat{\alpha} - \hat{\beta}x_1)^2 + \sum_{k=2}^{n}\big(Y_k - Y_{k-1} - \hat{\beta}(x_k - x_{k-1})\big)^2\right] \\
&= \frac{1}{n-2}\sum_{k=2}^{n}\left(Y_k - Y_{k-1} - \frac{S_1}{Q_1}(x_k - x_{k-1})\right)^2\,.
\end{aligned}$$

d. (5) Naj bosta $\tilde{\alpha}$ in $\tilde{\beta}$ cenilki parametrov $\alpha$ in $\beta$ po metodi najmanjših kvadratov. Pokažite, da sta cenilki nepristranski, in eksplicitno izračunajte njuni standardni napaki.

*Rešitev: Cenilki po metodi najmanjših kvadratov tvorita vektor* $\tilde{\boldsymbol{\gamma}} = \begin{bmatrix}\tilde{\alpha}\\ \tilde{\beta}\end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y} = \boldsymbol{\gamma} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\eta}$. *Ker je* $E(\boldsymbol{\eta}) = 0$, *je tudi* $E(\tilde{\boldsymbol{\gamma}}) = \boldsymbol{\gamma}$, *torej sta cenilki res nepristranski.*

*Standardni napaki lahko izrazimo s pomočjo matrik – to sta diagonalca matrike* $\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$. *Eksplicitno izražavo bi lahko dobili z množenjem, a gre lažje, če že cenilki sami izrazimo eksplicitno. Če definiramo*

$$S_x := \sum_{k=1}^{n} x_k\,, \qquad S_{xx} := \sum_{k=1}^{n} x_k^2\,, \qquad S_{xY} := \sum_{k=1}^{n} x_k Y_k\,, \qquad \Delta := n\,S_{xx} - S_x^2\,,$$

*velja*

$$\tilde{\alpha} = \frac{S_{xx}S_Y - S_x S_{xY}}{\Delta} = \frac{1}{\Delta}\sum_{k=1}^{n}(S_{xx} - S_x x_k)Y_k\,,$$
$$\tilde{\beta} = \frac{n\,S_{xY} - S_x S_Y}{\Delta} = \frac{1}{\Delta}\sum_{k=1}^{n}(n\,x_k - S_x)Y_k\,.$$

*Glede na to, da so slučajne spremenljivke $U_1, U_2, \ldots, U_n$ neodvisne, bo za izračun variance priročnejša izražava z njimi. Velja*

$$\tilde{\alpha} = \frac{1}{\Delta}\sum_{k=1}^{n}\sum_{l=1}^{k}(S_{xx} - S_x x_k)U_l = \frac{1}{\Delta}\sum_{l=1}^{n}\sum_{k=l}^{n}(S_{xx} - S_x x_k)U_l\,,$$
$$\tilde{\beta} = \frac{1}{\Delta}\sum_{k=1}^{n}\sum_{l=1}^{k}(nx_k - S_x)U_l = \frac{1}{\Delta}\sum_{l=1}^{n}\sum_{k=l}^{n}(nx_k - S_x)U_l$$

*in varianci sta enaki*

$$\begin{aligned}\operatorname{var}(\tilde{\alpha}) &= \frac{\sigma^2}{\Delta^2}\sum_{l=1}^{n}\left(\sum_{k=l}^{n}(S_{xx} - S_x x_k)\right)^2 = \frac{\sigma^2}{\Delta^2}\sum_{l=1}^{n}\sum_{j=l}^{n}\sum_{k=l}^{n}(S_{xx} - S_x x_j)(S_{xx} - S_x x_k)\\ &= \frac{\sigma^2}{\Delta^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\min\{j,k\}(S_{xx} - S_x x_j)(S_{xx} - S_x x_k)\,,\\ \operatorname{var}(\tilde{\beta}) &= \frac{\sigma^2}{\Delta^2}\sum_{l=1}^{n}\left(\sum_{k=l}^{n}(nx_k - S_x)\right)^2 = \frac{\sigma^2}{\Delta^2}\sum_{l=1}^{n}\sum_{j=l}^{n}\sum_{k=l}^{n}(nx_j - S_x)(nx_k - S_x)\\ &= \frac{\sigma^2}{\Delta^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\min\{j,k\}(nx_j - S_x)(nx_k - S_x)\,.\end{aligned}$$

*S korenjenjem varianc dobimo standardni napaki.*