FAKULTETA ZA MATEMATIKO IN FIZIKO

Oddelek za matematiko

Statistika

Pisni izpit

10. junij 2014

Ime in priimek: ______________________ Vpisna št.: ☐☐☐☐☐☐☐☐

Navodila

Pazljivo preberite besedilo naloge, preden se lotite reševanja. Nalog je 6, 5 rešenih nalog pa je že 100%. Na razpolago imate 2 uri.

Naloga	a.	b.	c.	d.	
1.			•	•	
2.					
3.				•	
4.			•	•	
5.					
6.			•	•	
Skupaj					

1. (20) Naj bosta $\mathbf{X}$ in $\mathbf{Y}$ slučajna vektorja dimenzije p in taka, da je

$$\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix} \sim N\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{I} & \rho\mathbf{1}\mathbf{1}^T\\ \rho\mathbf{1}\mathbf{1}^T & \mathbf{I}\end{pmatrix}\right) .$$

Pri tem je $\rho \in (-1,1)$ in $\mathbf{1}^T = (1,1,\ldots,1)$.

a. (10) Utemeljite, da velja

$$E\left(\mathbf{Y}^T\mathbf{Y}|\mathbf{X}\right) = E\left(\mathrm{Sl}(\mathbf{Y}\mathbf{Y}^T)|\mathbf{X}\right) = \mathrm{Sl}\Big(\mathrm{var}(\mathbf{Y}|\mathbf{X}) + E(\mathbf{Y}|\mathbf{X})E(\mathbf{Y}|\mathbf{X})^T\Big) .$$

Rešitev: Vemo, da velja

$$\mathbf{Y}^T\mathbf{Y} = \mathrm{Sl}(\mathbf{Y}\mathbf{Y}^T) .$$

Zaradi linearnosti lahko zamenjamo E in Sl. *Po definiciji je*

$$\mathrm{var}(\mathbf{Y}|\mathbf{X}) = E(\mathbf{Y}\mathbf{Y}^T|\mathbf{X}) + E(\mathbf{Y}|\mathbf{X})E(\mathbf{Y}|\mathbf{X})^T$$

b. (10) Izračunajte

$$E\left(\mathbf{Y}^T\mathbf{Y}|\mathbf{X}\right) .$$

Rešitev: Poznamo pogojno porazdelitev $\mathbf{Y}$ glede ne $\mathbf{X}$, ki je

$$\mathbf{Y}|_{\mathbf{X}=\mathbf{x}} \sim N\left(\rho\mathbf{1}\mathbf{1}^T\mathbf{x}, \mathbf{I} - p\rho^2\mathbf{1}\mathbf{1}^T\right) .$$

Računamo

$$\begin{aligned}
E\left(\mathbf{Y}^T\mathbf{Y}|\mathbf{X}\right) &= E\left(\mathrm{Sl}(\mathbf{Y}\mathbf{Y}^T)|\mathbf{X}\right)\\
&= \mathrm{Sl}\Big(E\left(\mathbf{Y}\mathbf{Y}^T|\mathbf{X}\right)\Big)\\
&= \mathrm{Sl}\Big(\mathrm{var}\left(\mathbf{Y}|\mathbf{X}\right) + E(\mathbf{Y}|\mathbf{X})E(\mathbf{Y}|\mathbf{X})^T\Big)\\
&= \mathrm{Sl}\left(\mathbf{I} - p\rho^2\mathbf{1}\mathbf{1}^T + \rho^2\mathbf{1}\mathbf{1}^T\mathbf{X}\mathbf{X}^T\mathbf{1}\mathbf{1}^T\right)\\
&= p - p^2\rho^2 + p\rho^2\left(\sum_{k=1}^{p} X_k\right)^2 .
\end{aligned}$$

2. (20) Privzemite, da je populacija velikosti N razdeljena na K stratumov velikosti $N_1, N_2, \ldots, N_K$. Enostavni slučajni vzorci so velikosti $n_1, n_2, \ldots, n_K$. Z μ označite populacijsko povprečje in z σ^2 populacijsko varianco za celotno populacijo, z μ_k in σ_k^2 označite populacijsko povprečje in populacijsko varianco za k-ti stratum.

a. (5) Pokažite, da je

$$\sigma^2 = \sum_{k=1}^{K} w_k \sigma_k^2 + \sum_{k=1}^{K} w_k (\mu_k - \mu)^2$$

kjer je $w_k = \frac{N_k}{N}$ za $k = 1, 2, \ldots, K$.

Rešitev: Po definiciji je

$$\sigma^2 = \frac{1}{N} \left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} (y_{ki} - \mu)^2 \right)$$

kjer je y_{ki} vrednost i-te enote v k-tem stratumu. Opazimo, da

$$\begin{aligned}
&\sum_{i=1}^{N_k} (y_{ki} - \mu)^2 = \\
&\quad = \sum_{i=1}^{N_k} (y_{ki} - \mu_k + \mu_k - \mu)^2 \\
&\quad = \sum_{i=1}^{N_k} (y_{ki} - \mu_k)^2 + \sum_{i=1}^{N_k} (\mu_k - \mu)^2 + 2(\mu_k - \mu) \sum_{i=1}^{N_k} (y_{ki} - \mu) \\
&\quad = \sum_{i=1}^{N_k} (y_{ki} - \mu_k)^2 + \sum_{i=1}^{N_k} (\mu_k - \mu)^2 \\
&\quad = N_k \sigma_k^2 + N_k (\mu_k - \mu)^2 .
\end{aligned}$$

Z uporabo tega v zgornji vsoti dobimo ustrezen rezultat.

b. (10) Naj bo $\bar{Y}_k$ vzorčno povprečje v k-tem stratumu za $k = 1, 2, \ldots, K$ in $\bar{Y} = \sum_{k=1}^{K} w_k \bar{Y}_k$ nepristranska cenilka za populacijsko povprečje. Privzemite, da so $\bar{Y}_1, \ldots, \bar{Y}_n$ neodvisne cenilke. Za oceno σ^2 moramo oceniti

$$\sigma_b^2 = \sum_{k=1}^{K} w_k (\mu_k - \mu)^2 = \sum_{k=1}^{K} w_k \mu_k^2 - \mu^2 .$$

Predlagana je cenilka

$$\hat{\sigma}_b^2 = \sum_{k=1}^{K} w_k \bar{Y}_k^2 - \bar{Y}^2 .$$

Pokažite, da je

$$E(\hat{\sigma}_b^2) = \sum_{k=1}^{K} w_k(1-w_k)\mathrm{var}(\bar{Y}_k) + \sum_{k=1}^{K} w_k\mu_k^2 - \mu^2\,.$$

Rešitev: Vemo, da

$$E(\bar{Y}_k^2) = \mathrm{var}(\bar{Y}_k^2) + \mu_k^2$$

in

$$E(\bar{Y}^2) = \mathrm{var}(\bar{Y}) + \mu^2\,.$$

Imamo

$$E(\hat{\sigma}_b^2) = \sum_{k=1}^{K} w_k\left(\mathrm{var}(\bar{Y}_k^2) + \mu_k^2\right) - \mathrm{var}(\bar{Y}) - \mu^2\,.$$

Z upoštevanjem

$$\mathrm{var}(\bar{Y}) = \sum_{k=1}^{K} w_k^2\mathrm{var}(\bar{Y}_k)$$

rezultat sledi.

c. (5) Ali obstaja nepristranska cenilka za σ^2? Utemeljite svoj odgovor.

Rešitev: Vemo, da

$$\sigma^2 = \sum_{k=1}^{K} w_k\sigma_k^2 + \sum_{k=1}^{K} w_k(\mu_k-\mu)^2$$

Nepristransko cenilko za σ_k^2 imamo. Torej nam manjka še nepristranska cenilka za drugi člen, torej za $\sum_{k=1}^K w_k(\mu_k-\mu)^2$. Malo popravimo cenilko iz b. dela naloge. In sicer tako, da bo veljalo $E(\tilde{\sigma}_b^2) = \sigma_b^2$. Za cenilko vzamemo

$$\tilde{\sigma}_b^2 = \sum_{k=1}^{K} w_k\bar{Y}_k^2 - \bar{Y}^2 - \sum_{k=1}^{K} w_k(1-w_k)\mathrm{var}(\bar{Y}_k)\,.$$

Ker znotraj stratuma vzorčimo enostavno, lahko $\mathrm{var}(\bar{Y}_k)$ nadomestimo z $\frac{\hat{\sigma}_k^2}{n_k}\cdot\frac{N_k-n_k}{N_k-1}$. Dobljena cenilka bo nepristranska za parameter σ_b^2. Torej nepristranska cenilka za σ^2 obstaja.

3. (20) Predpostavite, da opazovane vrednosti $x_1, x_2, \ldots, x_n$ nastanejo kot neodvisne, enako porazdeljene slučajne spremenljivke $X_1, X_2, \ldots, X_n$ s porazdelitvijo

$$P(X_1 = x) = \frac{(\theta-1)^{x-1}}{\theta^x}$$

za $x = 1, 2, 3, \ldots$ in $\theta > 1$.

a. (10) Poiščite cenilko parametra θ po metodi največjega verjetja na podlagi opazovanih vrednosti.

Rešitev: Zapišemo logaritemsko funkcijo verjetja

$$\ell(\theta, \mathbf{x}) = \left(\sum_{k=1}^{n} x_k - n\right) \log(\theta - 1) - \left(\sum_{k=1}^{n} x_k\right) \log\theta\,.$$

Odvajamo in dobimo

$$\ell'(\theta, \mathbf{x}) = \frac{\sum_{k=1}^{n} x_k - n}{\theta - 1} - \frac{\sum_{k=1}^{n} x_k}{\theta} = 0\,.$$

Sledi

$$\hat{\theta} = \frac{1}{n}\sum_{k=1}^{n} x_k = \bar{x}\,.$$

b. (10) Zapišite aproksimativni 99% interval zaupanja za parameter θ na podlagi opazovanih vrednosti. Kot znano privzemite

$$\sum_{x=1}^{\infty} x a^{x-1} = \frac{1}{(1-a)^2}$$

za $|a| < 1$.

Rešitev: Imamo

$$\ell''(\theta, x) = -\frac{x-1}{(\theta-1)^2} + \frac{x}{\theta^2}\,.$$

Za Fisherjevo informacijo rabimo

$$E(X_1) = \sum_{x=1}^{\infty} x \frac{(\theta-1)^{x-1}}{\theta^x}\,.$$

Z uporabo namiga dobimo

$$E(X_1) = \frac{1}{\theta} \cdot \left(1 - \frac{\theta-1}{\theta}\right)^{-2} = \theta\,.$$

Torej

$$I(\theta) = \frac{1}{\theta(\theta - 1)} .$$

Aproksimativni 99% interval zaupanja je

$$\hat{\theta} \pm 2.56 \cdot \sqrt{\frac{\hat{\theta}(\hat{\theta} - 1)}{n}} .$$

4. (20) Na razpolago imamo opazovane vrednosti $X_1, X_2, \ldots, X_n$ iz normalne porazdelitve $N(\mu, \sigma^2)$. Predpostavljamo, da so opazovane vrednosti med sabo neodvisne, enako porazdeljene slučajne spremenljivke. Preizkusiti želimo domnevo $H_0 : \mu = 0$ proti $H_1 : \mu \neq 0$. **Privzemite, da je parameter σ^2 znan**.

a. (10) Domnevo H_0 lahko preizkusimo pri dani stopnji tveganja α na dva načina:

- H_0 zavrnemo, če je $|\bar{X}| > c$ za primerno izbran c, tako da bo tveganje ravno α.
- Iz opazovanih vrednosti ocenimo μ in izračunamo zgornjo in spodnjo mejo intervala zaupanja pri stopnji tveganja α. Če interval ne pokriva 0, domnevo H_0 zavržemo.

Ali sta zgornja testa popolnoma enaka? Komentirajte.

Rešitev: Ker je σ^2 znan, je širina intervala zaupanja fiksna in enaka c. Testa sta popolnoma enaka.

b. (10) Izračunajte še testno statistiko za zgornjo situacijo po metodi kvocienta verjetij.

Rešitev: Ko izračunamo Λ, se konstanti $(1/\sqrt{2\pi\sigma^2})^n$ pokrajšata, tako da je

$$\Lambda = \exp\left(-\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2 - x_i^2}{2\sigma^2}\right) .$$

Izraz nekoliko preoblikujemo in dobimo

$$\Lambda = \exp\left(\frac{n\bar{x}^2}{2\sigma^2}\right) .$$

Ker je σ^2 znan, bomo ničelno domnevo zavrnili, ko bo $|\bar{x}| > c$ za primerno izbran c.

5. (20) 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 ϵ_i in η_j med sabo nekorelirani. Predpostavljamo tudi, da so vsi ϵ_i in η_j nekorelirani.

a. (5) Utemeljite, da lahko modela združimo v

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \\ Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & x_1 \\ 1 & 0 & x_2 \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_m \\ 1 & 1 & w_1 \\ 1 & 1 & w_2 \\ \vdots & \vdots & \vdots \\ 1 & 1 & w_n \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \beta \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \\ \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix}.$$

Utemeljite, da ta model ustreza standardnim predpostavkam modela linearne regresije.

Rešitev: Z množenjem matrik ugotovimo, da dobimo povsem iste enačbe. Da ta novi model ustreza vsem predpostavkam standardnega modela sledi iz navedenih predpostavk.

b. (5) 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 cenilko parametra β. Izračunajte njeno standardno napako.

Rešitev: Ugotovimo, da je

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} m+n & n & 0 \\ n & n & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Matriko obrnemo in dobimo

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

Ugotovimo še

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

Sledi

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

Za standardno napako dobimo po formuli

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

c. (10) Podajte najboljšo nepristransko cenilko parametra α_2. Podajte njeno standardno napako.

Rešitev: Če imamo najboljšo cenilko $\alpha_2 - \alpha_1$ in najboljšo cenilko α_1, je po izreku Gauss-Markova njuna vsota najboljša linearna cenilka α_2. Z množenjem matrik dobimo, da je $\hat{\alpha}_1 = \bar{Y}$ in $\hat{\alpha}_2 - \hat{\alpha}_1 = \bar{Z} - \bar{Y}$. Torej je $\hat{\alpha}_2 = \bar{Z}$. Standardna napaka je $\sigma/\sqrt{n}$.

6. (20) 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. (10) 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:

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

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

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

Rešitev: Po centralnem limitnem izreku je:

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