

Kolokvij 1 - Osnove teoretične statistike

6. februar, 2020

1. Odgovorite na spodnja vprašanja in vsakega kratko utemeljite (v enem stavku)

- (a) Naj bo $\hat{\theta}$ cenilka za parameter θ . Srednjo kvadratno napako označimo z $MSE(\hat{\theta})$. Pokažite, kako je srednja kvadratna napaka odvisna od pristranosti in variance cenilke.
- (b) Dva raziskovalca sta merila hitrost plavanja japonskih krapov v ribniku. Raziskovalec A je ujel 100 krapov in izmeril vse naenkrat, raziskovalec B pa je preprosto meril hitrosti mimo-plavajočih krapov. Oba sta povprečno hitrost ocenila kot $\hat{\mu} = \bar{X}$ in $\widehat{var}(\hat{\mu}) = \frac{s^2}{n}$, kjer je $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Komentirajte lastnosti cenilk za vsakega raziskovalca posebej.
- (c) Ali za X_1, \dots, X_n i.i.d, $X_i \sim N(3,1)$ velja

$$P(\bar{X} \leq 2) > P(X \leq 2)?$$

- (d) Raziskovalec naredi poskus n enotah. Njegove meritve so porazdeljene po $U(0, a)$. Izračuna standardno napako za povprečje meritev. Narediti namerava še en tak poskus, v katerem pa bi rad velikost standardne napake prepolovil. Kako velik naj bo njegov nov vzorec?

2. Spremenljivka X je porazdeljena po *Paretovi porazdelitvi*, ki je zvezna porazdelitev z gostoto

$$f(x|a) = \begin{cases} \frac{c}{x^a} & ; \quad x > 1 \\ 0 & ; \quad \text{sicer.} \end{cases}$$

- (a) Utemeljite, zakaj je porazdelitev smiselna le za $a > 1$ in $c = a - 1$. *Namig: Kaj mora veljati za kumulativno porazdelitveno funkcijo?*
- (b) Ocenite a po metodi največjega verjetja.
- (c) Izpeljite cenilko za varianco a .
- (d) Kako bi s simulacijo preverili svojo izpeljano formulo za varianco? Zapišite psevdokodo.

3. Naj bodo X_1, X_2, \dots, X_n neodvisne enako porazdeljene spremenljivke iz $N(\mu, \sigma^2)$. Označimo $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Pokažite, da velja $\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$ (dovolj so ključni koraki tega dokaza).

OBRNI LIST

4. Naj bodo X_1, X_2, \dots, X_n neodvisne enako porazdeljene spremenljivke iz $N(\mu, \sigma^2)$. Poznamo nepristranski cenilki za oba parametra, in sicer sta \bar{X} za pričakovano vrednost in s^2 za varianco. Za cenilko po metodi največjega verjetja

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

vemo, da je pristranska cenilka za σ^2 .

- Pokažite, da velja $\text{var}(s^2) = \frac{2\sigma^4}{(n-1)}$. *Namig:* Uporabite rezultat prejšnje naloge in lastnosti porazdelitve χ^2 : če $Y \sim \chi_k^2$, potem velja $E(Y) = k$ in $\text{var}(Y) = 2k$.
 - Zapišite srednjo kvadratno napako (MSE) cenilke s^2 .
 - Izračunajte pristranost, varianco in srednjo kvadratno napako cenilke $\hat{\sigma}^2$.
 - Pokažite, da velja $MSE_{s^2} > MSE_{\hat{\sigma}^2}$.
 - Izkaže se, da dejstvo iz prejšnje točke ne velja pri vsaki porazdelitvi X_i . Kje v naši izpeljavi je bila predpostavka o normalnosti X_i nujna?
5. Naj bodo X_1, X_2, \dots, X_n neodvisne enako porazdeljene spremenljivke iz $N(\mu, \sigma^2)$. Naj bo s^2 nepristranska cenilka variance σ^2 , $\hat{\sigma}^2$ pa cenilka po metodi največjega verjetja. Velja: $MSE_{s^2} > MSE_{\hat{\sigma}^2}$
- Kaj to dejstvo pomeni glede izbire cenilke? Utemeljite, kdaj bi izbrali katero izmed cenilk za varianco normalne porazdelitve.
 - Kako je to dejstvo skladno z asimptotskimi lastnostmi cenilk po metodi največjega verjetja?