

## Одабрана поглавља математичке статистике

### Статистички функционали

1. За податке *peru* одредити емпиријску функцију расподеле и 95% траку поверења.
2. а) Показати да расподела тест статистике  $T_n$  Крамер-фон Мизес теста за тестирање хипотезе  $H_0 : F = F_0$ , где је  $F_0$  строго растућа функција, не зависи од  $F_0$ .  
 б) Тестира се хипотеза  $H_0 : X \in \mathcal{E}(\lambda)$ , где је  $\lambda$  непознат параметар. Показати да расподела  $T_n$  не зависи од  $\lambda$ .  
 в) Оценити моћ за алтернативу  $\gamma(1, 2)$ , ако се тестира хипотеза из дела б).
3. Одредити оцену функционала  $h(F) = E(X - EX)^k$  и испитати непристрасност тако добијене оцене када је  $k = 2$ .
4. Одредити оцену коефицијента корелације између  $X$  и  $Y$  са заједничком дводимензионом расподелом  $F(x, y)$ .
5. а) Одредити  $Q(y) = F^{-1}(y)$ , где је  $F$  функција расподеле дискретне случајне величине која узима вредности  $a < b < c$  са вероватноћама  $p, q, r = 1 - p - q$ , редом.  
 б) Доказати да ни  $F(Q(y)) = y$  ни  $Q(F(x)) = x$  не важи за све  $x$  и  $y$ .  
 в) Нека је узет узорак величине  $n$  из расподеле из дела а) и нека су  $X$  и  $Y$  бројеви опсервација једнаких  $a$  и  $b$ , редом. Одредити  $\hat{Q}_n = \hat{F}_n^{-1}$ .
6. Нека је  $\sigma(F)$  стандардна девијација случајне променљиве  $X$  са функцијом расподеле  $F$ . Одредити утицајну криву и граничну расподелу од  $\sqrt{n}(\sigma(\hat{F}_n) - \sigma(F))$ , где је  $\sigma(\hat{F}_n)$  оцена стандардне девијације.
7. а) Нека је  $F$  функција расподеле непрекидне случајне величине са густином  $f = F'$ . Наћи утицајну криву параметра  $h_p(F)$  датог са  $F(h_p(F)) = p$  за неко  $p \in (0, 1)$  ( $h_p(F)$  је  $p$ -ти квантил од  $F$ ).  
 б) Нека је  $\hat{F}_n(x)$  емпиријска функција расподеле и за  $0 < t < 1$  дефинишемо  $\hat{F}_n^{-1}(t) = \inf\{x | \hat{F}_n(x) \geq t\}$ . Дефинишемо интерквартилни ранг са  $\hat{\tau}_n = \hat{F}_n^{-1}(0.75) - \hat{F}_n^{-1}(0.25)$ . Наћи граничну расподелу од  $\sqrt{n}(\hat{\tau}_n - \tau(F))$ , где је  $\tau(F) = h_{\frac{3}{4}}(F) - h_{\frac{1}{4}}(F)$ .
8. Нека је  $X_1, X_2, \dots, X_n$  прост случајан узорак са ненегативним вредностима, функцијом расподеле  $F$  и нека је
 
$$h(F) = \frac{\left(\int_0^\infty x dF(x)\right)^2}{\int_0^\infty x^2 dF(x)}.$$
 а) Одредити утицајну криву параметра  $h(F)$ .  
 б) Користећи  $X_1, \dots, X_n$  одредити оцену методом замене за  $h(F)$  и граничну расподелу за  $\sqrt{n}(h(\hat{F}_n) - h(F))$ .
9. Одредити утицајну криву  $k$ -тог централног момента  $\mu_k$  и граничну расподелу од  $\sqrt{n}(M_k - \mu_k)$ , где је  $M_k$  оцена од  $\mu_k$ .
10. Ако је  $h(F) = \frac{h_1(F)}{h_2(F)}$ , одредити утицајну криву од  $h(F)$ .

11. Одредити утицајну криву стандардизованог трећег момента случајне величине  $X$  са расподелом  $F$  датог са

$$h(F) = \frac{E(X - EX)^3}{(E(X - EX)^2)^{\frac{3}{2}}}.$$

12. Нацртати у  $R$ -у утицајну криву за дисперзију нормалне расподеле.

### $U$ -статистике

13. Показати да је  $h(F) = D_F(X)$  функционал очекивања.
14. Нека су  $(X, Y)$  случајне величине са дводимензионом расподелом  $F$ . Показати да је  $h(F) = cov(X, Y)$  функционал очекивања.
15. Ако су  $\theta_1$  и  $\theta_2$  функционали очекивања онда је и  $\theta_1 + \theta_2$  функционал очекивања. Доказати.
16. Ако су  $\theta_1$  и  $\theta_2$  функционали очекивања, онда је и  $\theta_1\theta_2$  функционал очекивања. Доказати.
17. На основу узорка  $X_1, \dots, X_n$  из расподеле  $F$ , одредити  $U$ -статистику која је оцена дисперзије  $DX$ .
18. Нека су  $\Phi$  и  $\theta$  дати на следећи начин:

$$\Phi(x_1, x_2) = \begin{cases} 1, & x_1 + x_2 > 0; \\ 0, & \text{иначе.} \end{cases}, \quad \theta = P\{X_1 + X_2 > 0\}.$$

Навести два примера у којима је могуће уз што мање претпоставки (без познавања расподеле) одредити  $1^2 = cov(\Phi(X_1, X_2), \Phi(X_1, X_3))$  и одредити његову вредност.

19. Израчунати  $\theta = P\{X_1 + X_2 > 0\}$  ако су  $X_1$  и  $X_2$  из нормалне мешавине одређене са  $F = (1 - \varepsilon)\mathcal{N}(0, 1) + \varepsilon\mathcal{N}(m, \sigma^2)$ .
20. Нека је  $X_1, X_2, \dots, X_n$  низ независних случајних величина са функцијом расподеле  $F$  и нека је  $\Phi(x_1, x_2, x_3) = I\{x_1 + x_2 + x_3 < 0\}$ . Одредити  $i$ -ту пројекцију  $\varphi_i$  и  $i$ -ту коваријацију  $\sigma_i^2, i = 1, 2, 3$ , језгра  $\Phi$  дефинисане, у општем случају, са

$$\Phi_i(x_1, \dots, x_i) = E(\Phi(x_1, \dots, x_i, X_{i+1}, \dots, X_m)),$$

$$\sigma_i^2 = cov(\Phi(X_1, \dots, X_i, X_{i+1}, \dots, X_m), \Phi(X_1, \dots, X_i, X'_{i+1}, \dots, X'_m)).$$

21. Нека су  $X_1, X_2, X_3$  независне и једнако расподељене случајне величине са нормалном  $\mathcal{N}(0, 1)$  расподелом и нека је  $\Phi(x_1, x_2) = e^{\lambda x_1 x_2}$  за  $\frac{1}{2} \leq \lambda < \frac{1}{\sqrt{2}}$ . Доказати да је тада  $\sigma_1^2 < \infty$  и  $\sigma_2^2 = \infty$ .
22. Нека је  $U_n = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < i_2 < \dots < i_m \leq n} \Phi(X_{i_1}, \dots, X_{i_m})$ . Доказати да је  $\sigma_1^2 \leq \sigma_2^2 \leq \dots \leq \sigma_m^2$ .
23. Доказати да је дисперзија  $U$ -статистике реда 2 једнака  $DU = \frac{1}{\binom{n}{2}} \sum_{i=1}^2 \binom{2}{i} \binom{n-2}{2-i} \sigma_i^2$ , где је  $\sigma_1^2 = cov(\Phi(X_1, X_2), \Phi(X_1, X_3))$  и  $\sigma_2^2 = cov(\Phi(X_1, X_2), \Phi(X_1, X_2))$ .
24. Нека је  $X$  случајна величина с расподелом  $F$  таква да важи да је  $EX^2 < \infty$  и  $EX^{-2} < \infty$  и нека је

$$\theta(F) = \int x dF(x) \cdot \int \frac{1}{x} dF(x)$$

непознати параметар.

- а) Одредити, на основу узорка  $X_1, \dots, X_n$ ,  $U$ -статистику  $U_n$ , са симетричним језгром  $\Phi$ , која је његова непристрасна оцена.
- б) Израчунати дисперзију  $DU_n$  статистике  $U_n$ .
- в) Одредити асимптотску расподелу  $\sqrt{n}(U_n - EU_n)$  у случају да  $X$  има инверзну Гаусову расподелу с густином  $f(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}}$ ,  $x > 0$ ,  $\mu, \lambda > 0$ . За ову расподелу важи да је  $EX = \mu$ ,  $EX^{-1} = \frac{\mu+\lambda}{\mu\lambda}$ ,  $DX = \frac{\mu^3}{\lambda}$ ,  $DX^{-1} = \frac{2\mu+\lambda}{\mu\lambda^2}$ .
25. Нека су  $X_1, \dots, X_n$  независне и једнако расподељене случајне величине. Нека је  $W$  број парова  $i < j$  за које је  $X_i + X_j > 0$ , а  $V_s$  број парова  $i \leq j$  за које је  $X_i + X_j > 0$ . Показати да  $\frac{W-EW}{\sqrt{DW}}$  и  $\frac{V_s-EV_s}{\sqrt{DV_s}}$  имају исту граничну расподелу.
26. Нека је  $X_1, \dots, X_n$  узорак из расподеле са функцијом расподеле  $F(x)$ . Нека је  $p = P\{X_1 > 0\}$  и  $\theta = p(1-p)$ .
- а) Дефинисати  $U$ -статистику  $U_n$  које је оцена од  $\theta$ .
- б) Одредити тачну дисперзију од  $U_n$  у функцији од  $p$ .
- в) Одредити граничну расподелу од  $\sqrt{n}(U_n - \theta)$ .
27. Нека је  $X_1, \dots, X_n$  узорак из расподеле са функцијом расподеле симетричном око  $\theta$ . Претпоставимо да желимо да тестирамо да је  $\theta = \theta_0$  (позната вредност). У ту сврху разматрамо параметар  $\gamma = P\{X_1 + X_2 + X_3 > 3\theta_0\}$ .
- а) Дефинисати  $U$ -статистику  $U_n$  која је оцена од  $\gamma$ .
- б) Одредити прву,  $\varphi_1$ , и другу,  $\varphi_2$ , пројекцију од  $U_n - \gamma$ .
- в) Одредити граничну дисперзију од  $\sqrt{n}(U_n - \gamma)$ .
28. Дата је функција  $\Phi(x_1, x_2) = |x_1 - x_2|$ . Ђинијева средња разлика  $G_n$  је  $U$ -статистика, добијена на основу узорка обима  $n$ , чије је језгро  $\Phi(x_1, x_2)$ .
- а) Наћи асимптотску расподелу Ђинијеве средње разлике у случају да је узорак  $X_1, \dots, X_n$  из униформне  $U[0, \theta]$  расподеле.
- б) Наћи и граничну расподелу вектора  $(G_n, \bar{X}_n)$ , где је  $\bar{X}_n$  узорачка средина.
29. Ако су  $X_1$  и  $X_2$  независне и  $P\{X_i \geq u_\alpha\} = \alpha, i = 1, 2$ . Доказати да је
- $$P\{\max\{X_1, X_2\} \geq u_\alpha\} = 1 - (1 - \alpha)^2.$$
30. а) Нека је  $X_1, \dots, X_n$  узорак из експоненцијалне  $\mathcal{E}(\lambda)$  расподеле. Показати да је статистика
- $$T_n = \frac{1}{\bar{X}} \frac{1}{\binom{n}{2}} \sum_{i < j} (2 \min(X_i, X_j) - \frac{1}{2}(X_i + X_j))$$
- другоразредна за параметар  $\lambda$ .
- б) Ако је  $U_n = \bar{X}T_n$  показати да је, у случају  $\lambda = 1$ , гранична расподела  $\sqrt{n}U_n$  нормална  $\mathcal{N}(0, \frac{1}{3})$ .
- в) Наћи граничну расподелу  $\sqrt{n}T_n$  за узорак из експоненцијалне  $\mathcal{E}(\lambda)$  расподеле.

31. Доказати да је дисперзија дводимензионе  $U$ -статистике одређене са

$$U_{mn} = \frac{1}{mn} \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} \Phi(X_i; Y_j),$$

једнака

$$DU = \frac{1}{mn} \sum_{i=0}^1 \sum_{j=0}^1 \binom{m-1}{1-i} \binom{n-1}{1-j} \sigma_{ij},$$

где су  $\sigma_{10} = \text{cov}(\Phi(X_1; Y_1), \Phi(X_1; Y'_1))$ ,  $\sigma_{01} = \text{cov}(\Phi(X_1; Y_1), \Phi(X'_1; Y_1))$  и  $\sigma_{00} = \text{cov}(\Phi(X_1; Y_1), \Phi(X'_1; Y'_1))$ .

32. Нека је  $\Phi$  дато са:

$$\Phi(x, y) = \begin{cases} 1, & x < y; \\ 0, & \text{иначе.} \end{cases}$$

Навести два примера у којима је могуће уз што мање претпоставки (без познавања расподеле) одредити  $\sigma_{10}^2 = \text{cov}(\Phi(X_1; Y_1), \Phi(X_1; Y'_1))$ , односно  $\sigma_{01}^2 = \text{cov}(\Phi(X_1; Y_1), \Phi(X'_1; Y_1))$  и одредити њихову вредност.

33. Нека су  $\Phi$  и  $\theta$  дати на следећи начин:

$$\Phi(x, y) = \begin{cases} 1, & x < y; \\ 0, & \text{иначе.} \end{cases}, \quad \theta = P\{X < Y\}.$$

Одредити  $\theta$  ако су  $F$  и  $G$  расподеле од  $X$  и  $Y$ , редом, такве да важи  $G(y) = F(y - \Delta)$  и  $F$  је

- а) униформна расподела  $U(0, 1)$ ;
- б) дупла експоненцијална расподела  $DE(0, 1)$ .

34. Нека су  $X$  и  $Y$  случајне величине са расподелама  $F$  и  $G$ , редом. Ако је  $G(y) = F(y - \Delta)$ , показати да је  $\sigma_{10}^2 = \sigma_{01}^2$  када је  $F$  симетрична расподела, ако је  $\Phi(x, y) = I\{x < y\}$ .

35. Показати да  $\sigma_{10}^2$  и  $\sigma_{01}^2$  могу узети вредност различиту од  $\frac{1}{12}$  када су расподеле  $F$  и  $G$  од  $X$ , односно  $Y$ , различите али обе симетричне око 0.

36. Нека су  $X$  и  $Y$  случајне величине са расподелама  $F$  и  $G$ , редом. Показати да је:

- а)  $P\{X < Y\} = E(1 - G(X))$ ;
- б)  $P\{X_1 < Y_1, X_1 < Y_2\} = E(1 - G(X))^2$ ;
- в) ако је  $F$  непрекидна, онда  $P\{X_1 < Y_1, X_2 < Y_2\} = E(F^2(Y))$ .

37. Нека су  $X_1, \dots, X_m$  и  $Y_1, \dots, Y_n$  независни случајни узорци из непрекидних расподела  $F(x)$  и  $G(x)$ , редом. Као мера разлике у положају разматра се  $\theta = P\{X_1 + X_2 < Y_1 + Y_2\}$ .

- а) Дефинисати  $U$ -статистику  $U_{mn}$  која је оцена од  $\theta$ .
- б) При  $H_0(F(x) = G(x) \text{ за све } x)$ , одредити  $\theta$ .
- в) Одредити граничну расподелу за  $\sqrt{N}(U_{mn} - \theta)$ ,  $N = m + n$ , при  $H_0$ .
- г) Показати да при  $H_0$ ,  $\sigma_{10}^2 = \sigma_{01}^2$ .

38. Нека су  $X_1, \dots, X_m$  и  $Y_1, \dots, Y_n$  независни случајни узорци из непрекидних расподела  $F(x - \theta_X)$  и  $F(\frac{y - \theta_Y}{\eta})$ , редом, где су  $\theta_X$  и  $\theta_Y$  познати и важи да је  $F(x) = 1 - F(-x)$  за све  $x$ . Желимо да тестирамо  $H_0(\eta = 1)$  против  $H_1(\eta > 1)$ . Нека је  $X_i^* = |X_i - \theta_X|$  за  $i = 1, \dots, m$  и  $Y_j^* = |Y_j - \theta_Y|$  за  $j = 1, \dots, n$ . Тест за тестирање  $H_0$  против  $H_1$  базира се на  $U$ -статистици  $U_{mn}$  која је оцена од  $\gamma = P\{X_i^* < Y_j^*\}$ . Показати да је овај тест слободан од параметра и одредити општи израз, у функцији од  $F$  и  $\eta$ , за очекивање од  $U_{mn}$  и асимптотску дисперзију од  $\sqrt{N}(U_{mn} - \gamma)$ .
39. Нека су  $X_1, \dots, X_m$  и  $Y_1, \dots, Y_n$  независни случајни узорци из непрекидних расподела са функцијом расподеле  $F(x - \theta)$  и  $F(\frac{x - \theta}{\eta})$ , редом, где је непознати параметар  $\theta$  тачка симетрије обе расподеле. Претпоставимо да се тестира  $H_0(\eta = 1)$  против  $H_1(\eta > 1)$ . У ту сврху разматрамо параметар  $\gamma = P\{Y_1 < X_1 < Y_2\}$ .
- Дефинисати  $U$ -статистику  $U_{mn}$  која је оцена од  $\gamma$ .
  - Одредити  $\gamma$  при  $H_0$ .
  - Одредити граничну дисперзију за  $\sqrt{N}(U_{mn} - \gamma)$ ,  $N = m + n$ .
  - Да ли је тест за тестирање  $H_0$  против  $H_1$  заснована на  $U_{mn}$  из дела а) непараметарски слободан од расподеле или се може направити асимптотски непараметарски слободним од расподеле?
40. Нека су  $X_1, \dots, X_m$  и  $Y_1, \dots, Y_n$  независни случајни узорци из непрекидних расподела са функцијом расподеле  $F(x + \theta)$  и  $F(y - \theta)$ , редом. Додатно, претпоставити да је област дефинисаности ових расподела интервал на реалној правој симетричан око  $-\theta$  и  $\theta$ , редом. Претпоставимо да се тестира  $H_0(\theta = 0)$  против  $H_1(\theta > 0)$ . У ту сврху разматрамо параметар  $\gamma = P\{X < 0 < Y\}$ .
- Дефинисати  $U$ -статистику  $U_{mn}$  која је оцена од  $\gamma$ .
  - Одредити граничну расподелу за  $\sqrt{N}(U_{mn} - \gamma)$ ,  $N = m + n$ .
  - Да ли је тест за тестирање  $H_0$  против  $H_1$  заснован на  $U_{mn}$  из дела а) непараметарски слободан од расподеле или се може направити асимптотски непараметарски слободним од расподеле?

### V-статистике

41. Нека је  $V_n = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \Phi(X_i, X_j)$   $V$ -статистика. Ако је  $\sigma_2^2 < \infty$  и  $E(\Phi^2(X_i, X_i)) < \infty$ , доказати да  $D(\sqrt{n}(V_n - \theta)) \rightarrow 4\sigma_1^2$ .
42. Нека је  $\Phi(x_1, x_2) = x_1 x_2$  и  $E(X^2) < \infty$ ,  $E(X^4) = \infty$ . Ако је  $V_n$   $V$ -статистика са језгром  $\Phi$ , показати да је асимптотска дисперзија од  $\sqrt{n}(V_n - \theta)$  коначна, док је  $D(\sqrt{n}(V_n - \theta)) = \infty$  за све  $n$ .

### Џекнајф и бутстреп

43. Нека су  $X_1, \dots, X_n$  независне и једнако расподељене случајне величине са расподелом која има очекивање  $\mu$  и дисперзију  $\sigma^2$ . Одредити кориговану џекнајф оцену од  $\sigma^2$ .
44. Нека су  $X_1, \dots, X_n$  независне и једнако расподељене случајне величине са густином  $f(x; \theta) = \frac{1}{\theta}$ ,  $0 \leq x \leq \theta$ , где је  $\theta$  непознати параметар. Одредити кориговану џекнајф оцену параметра  $\theta$ .

45. За податке *prihodi* одредити кориговану цекнајф оцену и оцену стандардне грешке Ђини индекса  $h(F) = 1 - 2 \int_0^1 q_F(t) dt$ , где је  $q_F(t) = \frac{\int_0^t F^{-1}(s) ds}{\int_0^1 F^{-1}(t) dt}$ .

46. За податке *nerv* одредити цекнајф оцену стандардне грешке оценое коефицијента асиметрије.

47. Нека су  $X_1, \dots, X_n$  независне и једнако расподељене случајне величине са функцијом расподеле  $F$  и медијаном  $\theta$  и нека је  $F'(\theta) = \lambda > 0$  (претпоставимо да је  $n = 2m$ ). Дефинишемо  $\hat{\theta}_n$  да буде узорачка медијана.

- а) Нека су  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  статистике поретка. Показати да је цекнајф оцена од  $\hat{\theta}_n$ :  $\hat{D}(\hat{\theta}_n) = \frac{n-1}{4}(X_{(m+1)} - X_{(m)})^2$ .
- б) Претпоставимо да су  $X_i$  независне и једнако расподељене са експоненцијалном  $\mathcal{E}(1)$  расподелом. Показати да  $m(X_{(m+1)} - X_{(m)})$  има  $\mathcal{E}(1)$  расподелу.
- в) У општем случају, показати да за  $n = 2m$ ,

$$m(X_{(m+1)} - X_{(m)}) \rightarrow \frac{Z}{\lambda},$$

кад  $n \rightarrow \infty$ , где је  $Z \in \mathcal{E}(1)$ .

- г) Показати да  $\hat{D}(\hat{\theta}_n) \rightarrow \frac{1}{4\lambda^2} Z^2$ .

48. Нека су  $X_1, \dots, X_n$  независне и једнако расподељене случајне величине. Дефинишемо оцену  $\hat{\theta}$  са

$$\sum_{i=1}^n \psi(X_i - \hat{\theta}) = 0,$$

где је  $\psi$  парна функција ( $\psi(x) = -\psi(-x)$ ) са изводом  $\psi'$ .

- а) Нека је  $\hat{\theta}_{-j}$  оцена на основу свих  $X_i$  сем  $X_j$ . Показати да важи:

$$\sum_{i=1}^n \psi(X_i - \hat{\theta}_{-j}) = \psi(X_j - \hat{\theta}_{-j}).$$

- б) Користећи апроксимацију  $\psi(X_i - \hat{\theta}_{-j}) \approx \psi(X_i - \hat{\theta}) + (\hat{\theta} - \hat{\theta}_{-j})\psi'(X_i - \hat{\theta})$ , доказати да је:

$$\hat{\theta}_{-j} \approx \hat{\theta} - \frac{\psi(X_j - \hat{\theta})}{\sum_{i=1}^n \psi'(X_i - \hat{\theta})}.$$

- в) Показати да је цекнајф оцена дисперзије приближно  $\frac{n-1}{n} \frac{\sum_{i=1}^n \psi^2(X_i - \hat{\theta})}{(\sum_{i=1}^n \psi'(X_i - \hat{\theta}))^2}$ .

49. Нека су  $X_1, \dots, X_n$  случајне величине и нека је  $\hat{\theta} = \bar{X}_n^2$ . Показати да је цекнајф оцена дисперзије  $\hat{D}(\hat{\theta}) = \frac{4\bar{X}^2\hat{c}_2}{n-1} + \frac{4\bar{X}\hat{c}_3}{(n-1)^2} + \frac{\hat{c}_4 - \hat{c}_2^2}{(n-1)^3}$ , где је  $\hat{c}_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$ ,  $k = 2, 3, 4$ .

50. Нека су  $X_1, \dots, X_n$  случајне величине и нека је  $\hat{\theta} = \bar{X}_n^2$ . Показати да је бутстреп оцена дисперзије  $v_{boot} = \frac{4\bar{X}^2\hat{c}_2}{n} + \frac{4\bar{X}\hat{c}_3}{n^2} + \frac{\hat{c}_4 - \hat{c}_2^2}{n^3}$ , где је  $\hat{c}_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$ ,  $k = 2, 3, 4$ .

51. За податке  $dataN$  одредити бутстреп оцену стандардне девијације медијане.
52. Оценити бутстрепом грешку расподеле параметра очекивања, када је  $n = 3$ .
53. Нека је  $\theta = h(F)$  медијана од  $F$  и нека је  $\lambda_n(F)$  дисперзија узорачке медијане. Одредити  $\lambda_n(\hat{F}_n)$ , када је  $n = 3$ .
54. Израчунати 10 вредности оцене  $\lambda_n(\hat{F}_n) = P \left\{ \frac{\sqrt{n}(\bar{X}_n^* - \bar{X}_n)}{\sqrt{\frac{1}{n} \sum (X_i - \bar{X}_n)^2}} \leq a \right\}$  за нормалне мешавине облика  $F(x) = (1 - \varepsilon)\Phi(x) + \varepsilon\Phi(\frac{x}{\tau})$ , за  $\varepsilon = 0.1, 0.2$ ,  $\tau = 2, 3$  и  $a = 0, 0.5, 0.7$ , узимањем узорача обима  $n = 50$  из дате расподеле и узимањем бутстреп узорача величине  $B = 100$  из  $\hat{F}_n$ .
55. Решити претходни задатак када су 10 вредности  $\lambda_n(\hat{F}_n)$  добијене узимањем новог узорка  $(X_{i1}, \dots, X_{in})$ ,  $i = 1, \dots, 10$  за сваки од 10 случајева и једног бутстрепа  $B = 100$  за сваки узорак  $x_{i1}, \dots, x_{in}$ .
56. Нека су дати подаци:

|        |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|--------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| $LSAT$ | 576  | 635  | 558  | 578  | 666  | 580  | 555  | 661  | 651  | 605  | 653  | 575  | 545  | 572  | 594  |
| $GPA$  | 3.39 | 3.30 | 2.81 | 3.03 | 3.44 | 3.07 | 3.00 | 3.43 | 3.36 | 3.13 | 3.12 | 2.74 | 2.76 | 2.88 | 3.96 |

Сваки податак је облика  $X_i = (Y_i, Z_i)$ , где је  $Y_i = LSAT_i$  и  $Z_i = GPA_i$ . Наћи оцену коефицијента корелације. Оценити стандардну грешку оцене користећи: (i) џекнајф, (ii) бутстреп.

57. За податке  $peru$  одредити бутстреп оцену стандардне девијације и интервале поверења (нормални, перцентилни, стојерни и студентизовани) оцене коефицијента асиметрије.
58. Нека је  $n = 50$  и нека је  $h(F)$  коефицијент асиметрије. Одредити 95% бутстреп интервале поверења за  $h(F)$ , ако је узорак  $X_1, \dots, X_n$  из логнормалне расподеле са параметрима 0 и 1. (За добијање узорка из логнормалне расподеле користити нормалну расподелу).
59. Нека је  $X_1, \dots, X_{25}$  узорак из Студентове  $t_5$  расподеле и нека је  $h(F) = \frac{q_{0.75} - q_{0.25}}{1.34}$ , где  $q_p$  означава  $p$ -ти квантил. Упоредити дужине нормалног и перцентилног бутстреп интервала поверења.
60. Претпоставимо да је 50 људи добило плацебо, а 50 људи нови лек. Тридесет плацебо пацијената је показало побољшање, док је 40 пацијената којима је дат нови лек показало побољшање. Нека је  $\tau = p_2 - p_1$  где је  $p_2$  вероватноћа побољшања при третману леком и  $p_1$  вероватноћа побољшања при плацебу. Наћи стандардну грешку за  $\tau$  и 90% интервал поверења користећи бутстреп.
61. Нека је  $X_1, \dots, X_m$  узорак из расподеле са очекивањем  $\mu_X$  и дисперзијом  $\sigma_X^2$ , а  $Y_1, \dots, Y_n$  узорак из расподеле са очекивањем  $\mu_Y$  и дисперзијом  $\sigma_Y^2$ . Тестом

$$T_{mn} = \sqrt{\frac{mn}{m+n}} (\log s_X^2 - \log s_Y^2)$$

тестирати хипотезу  $H_0 : \sigma_X = \sigma_Y$  користећи бутстреп метод. Упоредити следеће методе реузорковања:

- 1) узети два независна узорка са понављањем из узорка  $\{X_1, \dots, X_m, Y_1, \dots, Y_n\}$ ,
- 2) узети два независна узорка са понављањем из узорка  $\{X_1 - \bar{X}, \dots, X_m - \bar{X}, Y_1 - \bar{Y}, \dots, Y_n - \bar{Y}\}$ ,

3) узети узорак  $\{X_1^*, \dots, X_m^*\}$  из узорка  $\left\{\frac{X_1}{\sigma_X}, \dots, \frac{X_m}{\sigma_X}\right\}$  и узети узорак  $\{Y_1^*, \dots, Y_n^*\}$  из узорка  $\left\{\frac{Y_1}{\sigma_Y}, \dots, \frac{Y_n}{\sigma_Y}\right\}$ .

62. Дате су вредности мерења две серије убрзања услед гравитације чије су вредности изражене као одступање од  $980000 \times 10^{-3} \text{ cm} \cdot \text{s}^{-2}$  у јединицама  $\text{cm} \cdot \text{s}^{-2} \times 10^{-3}$ :

|          |    |    |    |    |    |    |    |    |    |    |    |    |    |
|----------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| узорак 1 | 82 | 79 | 81 | 79 | 77 | 79 | 79 | 78 | 79 | 82 | 76 | 73 | 64 |
| узорак 2 | 84 | 86 | 85 | 82 | 77 | 76 | 77 | 80 | 83 | 81 | 78 | 78 | 78 |

Тестом  $T = \bar{X} - \bar{Y}$  тестирати хипотезу  $H_0 : \mu_X = \mu_Y$  против алтернативе  $H_1 : \mu_X < \mu_Y$  користећи бутстреп тест.

63. Подаци представљају резултате фармаколошког експеримента на псима. Дате су две променљиве: количина кисеоника ( $MVO$ ) и притисак леве коморе ( $LVP$ ). Добијени резултати су

|       |    |    |     |    |     |    |    |
|-------|----|----|-----|----|-----|----|----|
| $MVO$ | 78 | 92 | 116 | 90 | 106 | 78 | 99 |
| $LVP$ | 32 | 33 | 45  | 30 | 38  | 24 | 44 |

Бутстреп тестом тестирати хипотезу да су променљиве некорелисане.

### Оцењивање густине

64. За податке *suicide* одредити, применом различитих правила, ширине стубова хистограма и нацртати их.

65. Одредити 95% траку поверења за хистограм података *suicide*.

66. Одредити 95% траку поверења за хистограм података *SDSS1* ако је број подеока  $m = 308$ .

67. Приказати хистограм (применом Стурџисовог правила) за податке *oldFaithful*. Одредити наивну оцену густине података и приказати је графички, ако је  $h = 0.25$ .

68. Одредити оцену густине Гаусовим језгром у свакој тачки узорка: -0.77, -0.60, -0.25, 0.14, 0.45, 0.64, 0.65, 1.19, 1.71, 1.74, као и оцену густине за цео узорак, ако су параметри глаткости 0.25, 0.4, 0.6, 1.

69. Оценити густину узорка из нормалне  $\mathcal{N}(0, 1)$  расподеле оценама на основу различитих врста језгара, ако је параметар глаткости одређен формулом на основу претпоставке о стандардној расподели.

70. Оценити густину узорка оценама на основу различитих врста језгара, за различите параметре глаткости, ако је узорак из:

- а) униформне  $U(0, 1)$  расподеле,
- б) Студентове  $t$  расподеле,
- в) логнормалне расподеле.

71. Оценити густину узорка из  $F = \frac{1}{2}\mathcal{N}(a, 1) + \frac{1}{2}\mathcal{N}(-a, 1)$ , где је

- а)  $a = \frac{1}{2}$



б)  $a = 2$

оценама на основу различитих врста језгара, ако је параметар глаткости одређен Силвермановом формулом. Која је разлика између случајева (а) и (б)?

72. Одредити оптималну глаткост оцене густине језгром за узорак из нормалне расподеле, користећи метод најмањег квадрата са унакрсном провером ако се користи Гаусово језгро.
73. Одредити оптималну глаткост оцене густине језгром за податке *SDSS1*, користећи метод најмањег квадрата са унакрсном провером ако се користи Гаусово језгро.
74. Одредити оптималну глаткост оцене густине језгром за узорак из нормалне расподеле, користећи метод функције веродостојности са унакрсном провером.
75. Одредити оптималну глаткост оцене густине језгром за податке *SDSS1*, користећи метод функције веродостојности са унакрсном провером.
76. Оценити густину дводимензионог узорка  $(X, Y)$  из дводимензионе нормалне мешавине са векторима очекивања

$$\mu_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mu_2 = \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \quad \mu_3 = \begin{pmatrix} 3 \\ -1 \end{pmatrix},$$

коваријационом матрицом  $\Sigma = I_2$  и вероватноћама  $p = (0.2, 0.3, 0.5)$ .

77. Оценити густину дводимензионог узорка  $(X, Y)$  из базе *geyser* из пакета *MASS*.
78. Оценити густину узорка из експоненцијалне  $\mathcal{E}(1)$  расподеле
  - а) рефлексijом података,
  - б) користећи асиметрично језгро.
79. Оценити густину за податке *suicide* користећи асиметрично језгро.
80. Оценити густину узорка из скалиране бета  $\beta(3, 2)$  расподеле на интервалу  $[0, 5]$ .

### Непараметарска регресија

81. Претпоставимо да су подаци генерисани према моделу  $Y = m(X) + \sigma(X)\varepsilon$ , где је  $m(x) = 0.5x^2 + \sin(3x)$ ,  $\sigma(x) = 0.1 + 0.05x$  и  $\varepsilon \in \mathcal{N}(0, 1)$ . Узимајући узорак обима 100 из стандардне нормалне расподеле за  $X$ , оценити регресиону функцију локално константном оценом за:
  - а)  $h = 0.1, 0.25, 0.5, 0.75$ , користећи Гаусово језгро;
  - б)  $h = 0.5$ , користећи различите врсте језгара.
82. Претпоставимо да су подаци генерисани према моделу  $Y = m(X) + \sigma(X)\varepsilon$ , где је  $m(x) = \sin(2x) + 0.3 \cos(0.5x) + 0.08x$ ,  $\sigma(x) = 0.12 + 0.10|\sin(x)|$  и  $\varepsilon \in \mathcal{N}(0, 1)$ . Узимајући узорак обима 100 из  $\mathcal{U}(0, 10)$  расподеле за  $X$ , оценити регресиону функцију локално константном, локално линеарном и локално квадратном оценом за  $h = 0.5$ .
83. Претпоставимо да су подаци генерисани према моделу  $Y = m(X) + \sigma(X)\varepsilon$ , где је  $m(x) = \sin(x) + I(x \geq 5)$ ,  $\sigma(x) = 0.1 + 0.05x$  и  $\varepsilon \in \mathcal{N}(0, 1)$ . Узимајући узорак обима 100 из стандардне нормалне расподеле за  $X$ , оценити регресиону функцију локално константном и локално линеарном оценом за  $h$  добијено методом замене и за  $h$  добијено методом унакрсне провере.

84. Користећи податке из датотеке *flat*, оценити регресиону функцију локално константном и локално линеарном оценом, користећи метод унакрсне провере за избор  $h$ .
85. У бази *airquality* дати су дневни подаци о квалитету ваздуха у Њујорку током периода од маја до септембра 1973. године. Посматра се зависност концентрације озона (*Ozone*) од температуре ваздуха (*Temp*).
- а) Оценити регресиону функцију параметарском линеарном регресијом.
- б) Оценити регресиону функцију непараметарском регресијом, за  $p = 0, 1, 2$  и  $h = 0.75$ .

### Анализа преживљавања

86. Одредити функцију преживљавања и стопу хазарда за случајну величину  $T$  која има Вејбулову расподелу са густином

$$f(t) = \frac{k}{\lambda} \left( \frac{t}{\lambda} \right)^{k-1} e^{-\left(\frac{t}{\lambda}\right)^k}, \quad t > 0, \lambda > 0, k > 0.$$

Нацртати графике ових функција за  $k = 3$  и  $\lambda = 4$ .

87. Одредити оцене максималне веродостојности параметара  $k$  и  $\lambda$  Вејбулове расподеле на основу узорка  $(X_1, \delta_1), \dots, (X_n, \delta_n)$ .
88. Дати су подаци о времену рада осам сијалица (у данима) и индикатори цензурисања ( $\delta_i$ ), где  $\delta_i = 1$  означава да је сијалица прегорела, а  $\delta_i = 0$  да је цензурисана (није прегорела до краја испитивања):

| Сијалица | Време рада (дани) | Индикатор цензурисања ( $\delta_i$ ) |
|----------|-------------------|--------------------------------------|
| 1        | 200               | 1                                    |
| 2        | 450               | 1                                    |
| 3        | 600               | 0                                    |
| 4        | 700               | 1                                    |
| 5        | 800               | 1                                    |
| 6        | 1000              | 0                                    |
| 7        | 1000              | 0                                    |
| 8        | 1000              | 0                                    |

Оценити функцију преживљавања Каплан-Мајеровом оценом.

89. Посматране су још две сијалице са следећим временима рада и индикаторима. Оценити функцију преживљавања на основу посматраних 10 сијалица Каплан-Мајеровом оценом.

| Сијалица | Време рада (дани) | Индикатор цензурисања ( $\delta_i$ ) |
|----------|-------------------|--------------------------------------|
| 9        | 450               | 1                                    |
| 10       | 600               | 1                                    |

90. Одредити Нелсон-Аленову оцену функције преживљавања за податке из претходног задатка.
91. У бази *LeukSurv* из пакета *spBayesSurv* налазе се подаци о умирању од акутне мијелоидне леукемије за 1043 посматрана пацијента. Подаци укључују информације о локацији

становања пацијената, као и административне регионе (24 регије које обухватају цео анализирани простор) и Таунсенд индекс ( $tpi$ ), при чему веће вредности указују на мање развијена подручја. Подаци су прикупљени у сврху испитивања утицаја географског положаја на преживљавање, узимајући у обзир прогностичке факторе специфичне за испитанике: године старости, пол и број леукоцита ( $wbc$ ) при дијагнози. Одредити Каплан-Мајерову и Нелсон-Аленову оцену функције преживљавања за ове податке.

92. Симулирати 100 времена догађаја из униформне  $\mathcal{U}(0, 100)$  расподеле, тако да је 20% цензурирано. Одредити глатку оцену стопе хазарда користећи Гаусово језгро за параметре глаткости  $h = 2, 4, 12$ .
93. Симулирати 100 времена догађаја из експоненцијалне  $\mathcal{E}(\frac{1}{10})$  расподеле, тако да је 30% цензурирано. Одредити глатку оцену стопе хазарда користећи рефлексiju података и Гаусово језгро и параметар глаткости одређен Силвермановом формулом.
94. Коришћењем различитих језгара, одредити глатке оцене стопе хазарда за податке из задатка 101, ако је параметар глаткости одређен Силвермановом формулом.
95. Дати су подаци о времену до стабилизације хормона штитне жлезде код пацијената који су примали две различите терапије. У табели је дато време (у данима) до побољшања, док статус означава да ли је дошло до стабилизације пре краја посматрања (1) или није (0). Пацијенти су подељени у две групе у зависности од тога да ли су примали терапију  $A$  или терапију  $B$ . Тестирати једнакост функција преживљавања за две терапије коришћењем лог-ранк теста.

|                     |                                                                                                                            |
|---------------------|----------------------------------------------------------------------------------------------------------------------------|
| <b>Време (дани)</b> | 13, 15, 18, 10, 12, 14, 18, 20, 22, 25, 27, 30, 32, 35, 37, 40, 42, 45, 48, 50, 53, 55, 58, 60, 62, 65, 68, 70, 72, 75, 78 |
| <b>Статус</b>       | 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0                                   |
| <b>Група</b>        | $A, A, A, A, A, A, B, B, B, B, B, B, A, A, A, A, A, B, B, B, B, B, B, A, A, A, A, A, B, B, B$                              |

96. Поделити податке из 101. задатка у две групе, на групу са позитивним  $tpi$  индексом (пацијенти из подручја са вишим нивоом депривације (виша стопа незапослености, пренатрпаност, нижи приступ здравственој заштити итд., дакле пацијенти из вероватно сиромашнијих заједница)) и на групу са негативним  $tpi$  индексом (пацијенти из области са мањим нивоом депривације (боље социо-економске прилике, виша стопа запослености, бољи животни услови, итд., дакле пацијенти из вероватно богатијих заједница)). Тестирати хипотезу о једнакости функција преживљавања за ове групе лог-ранк тестом.
97. У бази *lung* из пакета *survival* налазе се подаци о умирању пацијената са узнатредовалим карциномом плућа из Северне централне групе за лечење карцинома. Тестирати хипотезу о једнакости функција преживљавања код мушкараца и жена лог-ранк тестом.