Одабрана поглавља математичке статистике

1. Показати да је $h(F) = D_F(X)$ функционал очекивања.

2. Нека су (X_i, Y_i) независне и једнако расподељене случајне величине из дводимензионе расподеле F. Показати да је $h(F) = cov(X_i, Y_i)$ функционал очекивања.

3. Ако су θ_1 и θ_2 функционали очекивања, онда је и $\theta_1 + \theta_2$ функционал очекивања. Доказати.

4. Ако су θ_1 и θ_2 функционали очекивања, онда је и $\theta_1 \cdot \theta_2$ функционал очекивања. Доказати.

5. Одредити оцену (преко U статистике) од σ^2.

6. Доказати за $a = 2$ да важи: $DU = \frac{1}{\binom{n}{2}} \sum\limits_{i=1}^{2} \binom{2}{i}\binom{n-2}{2-i}\sigma_i^2$, где су $\sigma_1^2 = cov(\Phi(X_1, , X_2), \Phi(X_1, X_3))$ и $\sigma_2^2 = cov(\Phi(X_1, , X_2), \Phi(X_1, X_2))$.

7. Нека је $a = 2$ и нека су Φ и θ дати на следећи начин:

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

Навести два примера у којима је могуће (без познавања расподеле) одредити $\sigma_1^2 = cov(\Phi(X_1, X_2), \Phi(X_1, X_3))$ и одредити његову вредност.

8. Израчунати $\theta = P\{X_1 + X_2 > 0\}$ ако су X_1 и X_2 из нормалне мешавине $F = (1-\epsilon)N(0,1) + \epsilon N(m, \sigma^2)$.

9. Нека су $X_1, ..., X_n$ 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.

10. Нека је $\Phi(x_1, x_2, x_3) = I\{x_1 + x_2 + x_3 < 0\}$. Одредити Φ_i и $\sigma_i^2, i = 1, 2, 3$, дате са

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

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

за U статистику са језгром Φ на основу независних случајних величина $X_1, X_2, ...$ са функцијом расподеле F.

11. Нека је $X_1, ..., X_n$ низ независних случајних величина са коначним моментима EX_1^2 и EX_1^{-2}. Нека је $\mu = EX_1$ и $\bar{\mu} = EX_1^{-1}$. Одредити U статистику која је непристрасна оцена од $\mu\bar{\mu}$, одредити њену дисперзију и асимптотску расподелу.

12. Нека су X_1, X_2, X_3 независне и једнако расподељене случајне величине са нормалном $N(0,1)$ расподелом и нека је $\Phi(x_1, x_2) = e^{\lambda x_1 x_2}$ са $\frac{1}{2} \leq \lambda < \frac{1}{\sqrt{2}}$. Доказати да је тада, за $a = 2$, $\sigma_1^2 < \infty$ и $\sigma_2^2 = \infty$.

13. Нека је $U = \frac{1}{\binom{n}{2}} \sum\limits_{1 \leq i_1 < i_2 < \cdots < i_a \leq n} \Phi(X_{i_1}, ..., X_{i_m})$. Доказати да је $\sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_a^2$.

14. За $s=2, a=b=1$ и одговарајућу $U-$статистику, показати да важи

$$DU = \frac{1}{mn}\sum_{i=0}^{1}\sum_{j=0}^{1}\binom{m-1}{1-i}\binom{n-1}{1-j}\sigma_{ij}^2,$$

где су $\sigma_{10}^2 = cov(\Phi(X_1;Y_1),\Phi(X_1;Y_1'))$, $\sigma_{01}^2 = cov(\Phi(X_1;Y_1),\Phi(X_1';Y_1))$ и $\sigma_{00}^2 = cov(\Phi(X_1;Y_1),\Phi(X_1';Y_1'))$

15. Нека је $a=b=1$ и нека је Φ дато са:

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

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

16. Нека је $a=b=1$ и нека су Φ и θ дати на следећи начин:

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

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

   а) униформна расподела $U(0,1)$,

   б) дупла експоненцијална расподела $DE(0,1)$.

17. 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 и језгро дефинисано као у претходном задатку.

18. Показати да $\sigma_{10}^2 = cov(\Phi(X_1;Y_1),\Phi(X_1;Y_1'))$ и $\sigma_{01}^2 = cov(\Phi(X_1;Y_1),\Phi(X_1';Y_1))$ могу узети вредност различиту од $\frac{1}{12}$ када су расподеле F и G од X, односно Y, различите али обе симетричне око 0.

19. Ако су X_1 и X_2 независне и $P\{X_i \geq u_\alpha\} = \alpha, i=1,2$. Доказати да је

$$P\{\max\{X_1,X_2\}\geq u_\alpha\} = 1-(1-\alpha)^2.$$

20. Одредити оцену функционала $h(F)=E(X-EX)^k$ и испитати непристрасност тако добијене оцене када је $k=2$.

21. Одредити функционал h дефинисан за све једнодимензионе расподеле са коначним првим моментом за које је $h(\hat{F}_n) = \frac{1}{n}\sum_{i=1}^{n}|X_i-\bar{X}|$.

22. Одредити оцену функционала корелације између X и Y са заједничком дводимензионом расподелом $F(x,y)$.

23. 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.

24. У узорку величине n из расподеле из претходног задатка, нека су X и Y бројеви опсервација једнаки a и b, редом. Одредити $\hat{Q}_n = \hat{F_n}^{-1}$.

25. Испитати да ли су следеће функције непрекидне у F

    а) $h(F) = F(a)$,

    б) $h(F)$ растојање између F и фиксиране расподеле F_0 дато са $h(F) = \int (F - F_0)^2 dF_0$,

    в) $h(F)$ је очекивање од F дефинисано за све F за које очекивање постоји.

26. Нека је F^{-1} непрекидна у околини p и нека је $h(F)$ $p-$ти квантил од F. Показати да је h непрекидна за свако такво F.

27. 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.

28. За дати функционал h и функцију расподеле F_0, нека је $h^*(F) = h(F) - h(F_0)$. Доказати да је $\Phi(x; F) = \Phi^*(x; F)$ за све x (са Φ је означена утицајна крива).

29. Одредити утицајну криву стандардизованог трећег момента случајне величине X са расподелом F датог са
$$h(F) = \frac{E(X - EX)^3}{(E(X - EX)^2)^{\frac{3}{2}}}.$$

30. Одредити утицајну криву Ђини индекса $h(F) = 1 - 2\int\limits_0^t q_F(t)dt$, где је $q_F(t) = \frac{\int\limits_0^t F^{-1}(s)ds}{\int\limits_0^1 F^{-1}(t)dt}$.

31. 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 $\hat{\tau}_n = \hat{F_n}(0.75) - \hat{F_n}(0.25)$. Наћи граничну расподелу од $\sqrt{n}(\hat{\tau}_n - \tau(F))$, где је $\tau(F) = h_{\frac{3}{4}}(F) - h_{\frac{1}{4}}(F)$.

32. Нека је $X_1, X_2, ...$ прост случајан узорак са ненегативним вредностима, функцијом расподеле F и параметром
$$h(F) = \frac{\left(\int\limits_0^\infty x dF(x)\right)^2}{\int\limits_0^\infty x^2 dF(x)}.$$

    а) Одредити утицајну криву параметра $h(F)$.

    б) Користећи $X_1, ..., X_n$ одредити оцену методом замене за $h(F)$ и граничну расподелу за $\sqrt{n}(h(\hat{F_n}) - h(F))$.

33. Одредити функцију утицаја $k-$тог централног момента μ_k и граничну расподелу од $\sqrt{n}(M_k - \mu_k)$, где је M_k оцена од μ_k.

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

35. Нека је $X_1, ..., X_{30}$ узорак од 30 прихода пре опорезивања. Одредити стандардну грешку оцене Ђини индекса $h(F) = 1 - 2\int\limits_0^t q_F(t)dt$, где је $q_F(t) = \frac{\int_0^t F^{-1}(s)ds}{\int_0^1 F^{-1}(t)dt}$.

36. За податке *prihodi.txt* одредити џекнајф оцену и оцену стандардне грешке оцене Ђини индекса.

37. За податке *nerv.txt* одредити џекнајф оцену стандардне грешке оцене коефицијента асиметрије.

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

   а) Нека су $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ статистике поретка. Показати да је џекнајф 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 $\varepsilon(1)$ расподелом. Показати да $m(X_{(m+1)} - X_{(m)})$ има $\varepsilon(1)$.

   в) У општем случају, показати да за $n = 2m$,

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

   кад $n \to \infty$, где је $Z \in \varepsilon(1)$.

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

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

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

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

   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:

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

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

40. Нека су $X_1, ..., X_n$ случајне променљиве и нека је $\hat{\theta} = \bar{X}_n^2$. Показати да је џекнајф 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 $\hat{c}_k = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2, k = 2, 3, 4$.

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

42. Оценити бутстрепом грешку расподеле параметра очекивања, када је $n = 3$.

43. 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 $n = 3$.

44. Одредити бутстреп оцену стандардне девијације медијене нормалне расподеле (у $R-$у).

45. 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 $Y_i = LSTAT_i$ и $Z_i = GPA_i$. Наћи оцену коефицијента корелације. Оценити стандардну грешку оцене користећи: (i) џекнајф, (ii) бутстреп.

46. 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 $F(x) = (1-\epsilon)\Phi(x) + \epsilon\Phi(\frac{x}{\tau})$, за четри комбинације $\epsilon = 0.1, 0.2, \tau = 2, 3$, узимањем узорака $n = 50$ из дате расподеле и онда узимањем 10 бутстреп узорака величине $B = 100$ из $\hat{F}_n$.

47. Решити претходни задатак када 10 вредности $\lambda_n(\hat{F}_n)$ су добијене узимањем новог узорка $(X_{i1}, ..., X_{in}), i = 1, ..., 10$ за сваки од 10 случајева и једног бутстрепа $B = 100$ за сваки узорак $x_{i1}, ..., x_{in}$.

48. За податке *nerv.txt* одредити бутстреп оцену стандардне девијације и интервале поверења (нормални, перцентилни, стожерни и студентизовани) оцене коефицијента асиметрије.

49. Нека је $n = 50$ и нека је $h(F)$ коефицијент асиметрије. Одредити 95% бутстреп интервале поверења за $h(F)$, ако је узорак $X_1, ..., X_n$ из логнормалне расподеле са параметрима 0 и 1. (За добијање узорка из логнормалне расподеле користити нормалну расподелу)

50. Нека је $X_1, ..., X_{25}$ 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.

51. Претпоставимо да је 50 људи добило плацебо, а 50 људи добило нови лек. Тридесет плацебо пацијената је показало побољшање, док 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 τ и 90% интервал поверења користећи бутстреп.

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

53. Приказати хистограм (применом Стурџисовог правила) за податке $oldFaithful.txt$. Одредити наивну оцену густине података и приказати је графички.

54. 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.

55. Оценити густину узорка из нормалне $N(0,1)$ расподеле оценама на основу различитих врста језра, ако је параметар глаткости одређен формулом на основу претпоставке о стандардној расподели.

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

    а) униформне $U(0,1)$ расподеле,

    б) Студентове t расподеле,

    в) логнормалне расподеле.

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

    а) $a = \frac{1}{2}$

    б) $a = 2$

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

58. Одредити оптималну глаткост оцене густине језгром за податке $SDSS1.txt$, користећи унакрсну проверу најмањег квадрата уз помоћ Фуријеове трансформације.

59. Одредити оптималну глаткост оцене густине језгром за узорак из нормалне расподеле, користећи унакрсну проверу најмањег квадрата ако се користи

    а) Гаусово језгро,

    б) Фуријеова трансформација.

60. Одредити опрималну глаткост оцене густине језгром за податке $SDSS1.txt$, користећи унакрсну проверу функције веродостојности.

61. Одредити опрималну глаткост оцене густине језгром за узорак из нормалне расподеле, користећи унакрсну проверу функције веродостојности.

62. Одредити 95% траку поверења за хистограм података *suicide.txt*.

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

64. Оценити густину дводимензионог узорка (X,Y) из дводимензионе нормалне мешавине са векторима очекивања

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

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

65. Оценити густину дводимензионог узорка (X,Y) из базе *geyser* из пакета $MASS$.