Статистике слободне од расподеле

1. Нека су $X_{(1)} \leq \cdots \leq X_{(n)}$ статистике поретка случајног узорка са експоненцијалном $\mathcal{E}(\frac{1}{\theta}), 0 < \theta < \infty$, расподелом. Показати да је статистика $Z = \frac{nX_{(1)}}{\sum_{i=1}^{n} X_{(i)}}$ слободна од расподеле на класи независних и једнако расподељених случајних величина са експоненцијалном $\mathcal{E}(\frac{1}{\theta}), 0 < \theta < \infty$, расподелом.

2. Нека је $X_1, ..., X_n$ прост случајан узорак из логистичке расподеле центриране у 0 са параметром скалирања $\eta > 0$, односно расподела од X_i је

$$F(x;\eta) = \frac{1}{1 - e^{-\frac{x}{\eta}}},\ -\infty < x < \infty.$$

Показати да је статистика $V = \frac{med(X_1,...,X_n)}{\sum_{i=1}^{n} |X_i - med(X_1,...,X_n)|}$, где је $med(X_1, ..., X_n)$ медијана случајних величина, слободна од расподеле на класи наведених расподела.

3. Златни правоугаоник је правоугаоник код којег однос ширине (ω) и дужине (l) је златни пресек, што је приближно 0.618. У многим културама овакав правоугаоник је присутан у уметности и архитектури. Приликом проучавања културе племена Шошони проучаване су корпе са истканим правоугаоницима. Нека X представља количник ширине и дужине истканог материјала. Нека је θ медијана од X. Узет је узорак од 20 количника ширина и дужина материјала

0.553	0.570	0.576	0.601	0.606	0.606	0.609	0.611	0.615	0.628
0.654	0.662	0.668	0.670	0.672	0.690	0.693	0.749	0.844	0.933

Испитати тестом знака да ли ови симболи имају облик златног правоугаоника.

4. На основу претходних истраживања просечна старост девојака приликом првог састанка била је 14 година. Социолози тврде да се то променило. Скупљен је узорак од 15 девојака и постављено им је ово питање. Добијени су следећи резултати: 13.0, 12.5, 13.5, 14.2, 11.5, 12.5, 15.0, 15.5, 13.5, 13.0, 16.0, 15.5, 13.7, 12.0, 14.5. С прагом значајности 0.05 тестом знака испитати да ли су социолози у праву.

5. Превозник разматра промену трасе путовања коришћењем новог аутопута. Просечно време путовања на старој рути је 4 сата. Приликом 18 пробних вожњи на новој траси добијени су следећи резултати: 4.4, 3.9, 5.2, 4.6, 4.3, 3.6, 4.4, 5.5, 3.9, 4.7, 4.1, 3.8, 4.8, 4.5, 5.6, 4.1, 4.5, 4.0. Тестом знака испитати да ли је просечно време путовања на новој траси другачије.

6. Произвођач сапуна жели да покаже како његов нови сапут за лице не исушује кожу. Узет је узорак од 10 жена и свака је две недеље једном у два дана прала једну половину лица тим сапуном, а другу сапуном конкуренског произвођача. На крају је мерена сувоћа лица (на скали од 1 до 10, мањи број - већа сувоћа). Добијени су следећи резултати

нови сапун	5.0	4.3	7.3	2.1	9.8	6.9	10.0	1.5	8.2	7.3
конкурентски сапун	6.1	4.5	6.0	2.0	7.5	8.0	9.2	1.0	8.0	6.9

Са прагом значајности 0.05 тестирати тестом знака да ли је нови сапун бољи од конкурентског.

7. Развијен је нови тип средства за полирање аутомобила и компанија тврди да је потребно мање времена за полирање него са другим. Да би се то потврдило узето је 15 парова аутомобила и при истим условима полирања добијени су резултати:

Нови	2.1	1.0	3.6	2.5	4.0	1.7	2.9	3.0	4.5	3.1	3.1	1.5	1.6	1.0	0.8
Други	2.7	1.3	2.0	2.3	3.9	1.8	3.0	3.0	4.6	3.0	3.4	1.7	1.8	1.3	1.0

Користећи тест знака испитати да ли је тврђење тачно.

8. Осигуравајућа компанија тврди да агенти који су прошли специјални тренинг имају више успеха у успостављању односа са потенцијалним клијентима. Да би се ово тестирало, случајно су изабрана 22 млада агента и од њих је 10 похађало специјални тренинг. Затим су оцењивани приликом пробних интервјуа с клијентима. Добијени су следећи резултати (просечна оцена са три интервјуа)

специјални тренинг	8.1	7.9	9.0	4.3	7.0	9.1	7.2	8.0	3.1	9.0		
остали	9.1	6.3	2.5	6.0	0.0	2.0	7.0	9.0	5.5	9.7	5.1	1.0

С прагом значајности 0.05 испитати тврдњу компаније Вилкоксоновим тестом збира рангова.

9. Нови серум против леукемије тестиран је на 5 лабораторијских мишева, док 4 није третирано. Дата је дужина живота (у месецима) мишева након експеримента

третирани	2.1	5.3	1.4	4.6	0.9
нетретирани	1.9	0.5	2.8	3.1	

С прагом значајности 0.05 испитати Вилкоксоновим тестом збира рангова да ли серум има ефекта.

10. Проценат неписмених међу белцима у Америци 1969. године је била 0.7. Сматра се да је просечан проценат неписмених у великим градовима (преко пола милиона становника) већи. Да би се то испитало, изабрано је 20 градова и изведено је истраживање које је дало следеће податке: 0.6, 0.5, 0.62, 1.7, 0.75, 1.0, 0.69, 0.8, 0.8, 0.57, 0.9, 1.5, 0.95, 0.53, 1.1, 1.2, 2.0, 0.65, 0.79, 0.61. Претпостављајући симетричност расподеле, испитати наведену претпоставку Вилкоксоновим тестом означених рангова.

11. Постоји тврдња да је просечно трајање телефонског позива три минута. Званичници сматрају да је веће. Да би се то проверило узет је узорак од 10 случајно изабраних позива и добијено је: 2.7, 10.5, 3.8, 15.2, 5.7, 3.5, 2.1, 4.0, 3.7, 3.2. С прагом значајниости 0.01 Вилкоксоновим тестом означених рангова испитати наведену тврдњу.

12. Фармацеутска компанија има два метода за производњу лека за лечење уједа пчеле. Сматра се да метод A даје мање успешне резултате од метода B. Да би се ово испитало узет је узорак од 12 пацијената на којима су примењена оба третмана. Добијено је

метод A	1.5	1.4	1.4	1.0	1.1	0.9	1.3	1.2	1.1	0.9	0.7	1.8
метод B	2.0	1.8	0.7	1.3	1.2	0.5	1.1	0.9	1.5	1.7	0.9	0.9

Претпостављајући симетрију разлике ових података, Вилкоксоновим тестом означених рангова испитати наведену тврдњу за праг значајности 0.05.

13. Осигуравајућа компанија жели да покаже да су њихове годишње премије мање од конкурентских. Да би се потврдило, узет је узорак од 15 великих метропола и одговарајуће премије су:

Компанија	500	498	505	495	490	498	480	501
	496	478	520	513	506	497	482	
Други	515	495	500	510	500	502	483	500
	498	490	526	515	513	499	490	

Вилкоксоновим тестом означених рангова испитати да ли је компанија у праву.

$U-$статистике

15. Показати да је $\theta(F) = D_F(X)$ функционал очекивања.

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

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

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

19. Одредити непристрасну оцену са најмањом дисперзијом дисперзије σ^2.

20. Доказати да је диспезрија $U-$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 су $\sigma_1^2 = cov(\Phi(X_1, X_2), \Phi(X_1, X_3))$ и $\sigma_2^2 = cov(\Phi(X_1, X_2), \Phi(X_1, X_2))$.

21. 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 уз што мање претпоставки (без познавања расподеле) одредити $\sigma_1^2 = cov(\Phi(X_1, X_2), \Phi(X_1, X_3))$ и одредити његову вредност.

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

23. Нека је $X_1, X_2, ...$ низ независних случајних величина са функцијом расподеле F. Нека је $\Phi(x_1, x_2, x_3) = I\{x_1 + x_2 + x_3 < 0\}$. Одредити $i-$ту пројекцију φ_i и $i-$ту коваријацију $\sigma_i^2, i = 1, 2, 3$, језгра Φ дефинисане, у општем случају, са

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

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

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

25. Нека је $U_n = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} \Phi(X_{i_1}, ..., X_{i_a})$. Доказати да је $\sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_m^2$.

26. 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 $X_1, \ldots, X_n$, U-статистику U_n, са симетричним језгром Φ, 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 $EX = \mu$, $EX^{-1} = \frac{\mu+\lambda}{\mu\lambda}$, $DX = \frac{\mu^3}{\lambda}$, $DX^{-1} = \frac{2\mu+\lambda}{\mu\lambda^2}$.

27. 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.

28. Нека је $X_1, ..., X_n$ узорак из расподеле са функцијом расподеле $F(x)$. Нека је $p = P\{X_1 > 0\}$ и $\theta = p(1-p)$.

    а) Дефинисати $U-$статистику које је оцена од θ.

    б) Одредити тачну дисперзију од U_n у функцији од p.

    в) Одредити граничну расподелу од $\sqrt{n}(U_n - \theta)$.

29. Нека је $X_1, ..., X_n$ узорак из експоненцијалне $\mathcal{E}(\lambda)$ расподеле. Нека је

$$\theta = P\{2\min(X_1, X_2) < X_3\} - P\{X_1 < X_3\}.$$

    а) Дефинисати $U-$статистику које је оцена од θ.

    б) Одредити тачну дисперзију од U_n.

    в) Одредити граничну расподелу од $\sqrt{n}(U_n - \theta)$.

30. Нека је $X_1, ..., X_n$ узорак из расподеле са функцијом расподеле симетричном око θ. 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 које је оцена од γ.

    б) Одредити прву, φ_1, и другу, φ_2, пројекцију од $U_n - \gamma$.

    в) Одредити граничну дисперзију од $\sqrt{n}(U_n - \gamma)$.

31. Нека је W_n Вилкоксонова статистика, тј. специјалан случај U-статистике када је $a = 2$, а $\Phi(x_1, x_2) = I\{x_1 + x_2 > 0\}$.

    а) Наћи асимптотску расподелу $\sqrt{n}(W_n - \theta)$ у случају да је узорак $X_1, ..., X_n$ из униформне $U[-1, 1]$ расподеле.

    б) Наћи и граничну расподелу вектора $(W_n, \bar{X}_n)$, где је $\bar{X}_n$ узорачка средина.

32. 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 $X_1, ..., X_n$ из униформне $U[0, \theta]$ расподеле.

    б) Наћи и граничну расподелу вектора $(G_n, \bar{X}_n)$, где је $\bar{X}_n$ узорачка средина.

33. Када имамо два теста, заснована на независним тест статистикама U_1 и U_2 чије су велике вредности значајне, тестирање се може вршити тако што се израчунају нивои значајности оба теста и користи се онај тест који има већи ниво значајности. Показати да је у том случају прави ниво значајности $\alpha' = 1 - (1-\alpha)^2$, где је α ниво значајности сваког од тестова.

34. а) Нека је $X_1, ..., X_n$ узорак из експоненцијлне $\mathcal{E}(\lambda)$ 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 λ.

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 вектора $\sqrt{n}T_n$ за узорак из експоненцијлне $\varepsilon(\lambda)$ расподеле.

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

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

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 $\sigma_{10} = cov(\Phi(X_1;Y_1),\Phi(X_1;Y_1')), \sigma_{01} = cov(\Phi(X_1;Y_1),\Phi(X_1';Y_1)), \sigma_{00} = cov(\Phi(X_1;Y_1),\Phi(X_1';Y_1'))$.

36. 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 уз што мање претпоставки (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.

37. 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 и Y, редом, такве да важи $G(y) = F(y - \Delta)$ и F је

а) униформна расподела $U(0, 1)$

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

38. 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 $\Phi(x, y) = I\{x < y\}$.

39. 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.

40. 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 $P\{X_1 < Y_1, X_2 < Y_2\} = E(F^2(Y))$.

41. Нека су $X_1, ..., X_m$ и $Y_1, ..., Y_n$ 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 која је оцена од θ.

б) При $H_0(F(x) = G(x)$ за све $x)$, одредити θ.

в) Одредити граничну расподелу за $\sqrt{N}(U_{mn} - \theta), N = m + n$, при H_0.

г) Показати да при $H_0, \sigma_{10} = \sigma_{01}$.

42. Нека су $X_1, ..., X_m$ и $Y_1, ..., Y_n$ независни случајни узорци из непрекидних расподела $F(x - \theta_X)$ и $F(\frac{y-\theta_Y}{\eta})$, редом, где су θ_X и θ_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 $i = 1, ..., m$ и $Y_j^* = |Y_j - \theta_Y|$ за $j = 1, ..., n$. Тест за тестирање H_0 против H_1 базира се на U-статистици 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 η, за очекивање од U и асимптотску дисперзију од $\sqrt{N}(U_{mn} - \gamma)$.

43. Нека су $X_1, ..., X_m$ и $Y_1, ..., Y_n$ независни случајни узорци из непрекидних расподела са функцијом расподеле $F(x - \theta)$ и $F(\frac{x-\theta}{\eta})$, редом, где је непознати параметар θ тачка симетрије обе расподеле. Претпоставимо да се тестира $H_0(\eta = 1)$ против $H_1(\eta > 1)$. У ту сврху разматрамо параметар $\gamma = P\{Y_1 < X_1 < Y_2\}$.

   а) Дефинисати U-статистику која је оцена од γ.

   б) Одредити γ при H_0.

   в) Одредити граничну дисперзију за $\sqrt{N}(U - \gamma), N = m + n$.

   г) 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?

44. Нека су $X_1, ..., X_m$ и $Y_1, ..., Y_n$ 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 θ, редом. Претпоставимо да се тестира $H_0(\theta = 0)$ против $H_1(\theta > 0)$. У ту сврху разматрамо параметар $\gamma = P\{X < 0 < Y\}$.

   а) Дефинисати U-статистику која је оцена од γ.

   б) Одредити граничну расподелу за $\sqrt{N}(U_{mn} - \gamma), N = m + n$.

   в) Да ли је тест за тестирање H_0 против H_1 заснован на U из дела а) непараметарски слободан од расподеле или се може направити асимптотски непараметарски слободним од расподеле?

$V-$статистике

45. 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 је $D(\sqrt{n}(V_n - \theta)) \to 4\sigma_1^2$.

46. Нека је $\Phi(x_1, x_2) = x_1 x_2$ и $E(X^2) < \infty, E(X^4) = \infty$. Показати да је асимптотска дисперзија од $\sqrt{n}(V_n - \theta)$ коначна, док је $D(\sqrt{n}(V_n - \theta)) = \infty$ за све n.

Провера квалитета теста

47. Нека је $\mathbf{X}_4$ узорак из Пуасонове $\mathcal{P}(\lambda)$ расподеле. Претпоставимо да тестирамо $H_0(\lambda = \frac{1}{4})$ против $H_1(\lambda > \frac{1}{4})$ са тестом који одбацује H_0 ако је $\sum_{i=1}^{4} X_i \geq 3$. Одредити вероватноћу грешке прве врсте овог теста и одредити моћ теста за $\lambda = \frac{2}{5}$.

48. Размотримо класу расподела $\mathcal{F}$ која садржи расподеле симетричне око 0. Једна могућа мера расипености такве расподеле је $p = P\{|X| > 1\}$. Претпоставимо да желимо да тестирамо $H_0(p = \frac{1}{2})$ против $H_1(p > \frac{1}{2})$ користећи статистику $V = \sum_{i=1}^{n} I\{|X_i| > 1\}$, где је $\mathbf{X}$ узорак из

расподеле $F \in \mathcal{F}$. Нека је $n = 10$ и критична област $V \geq 8$. Одредити ниво значајности теста. Одредити функцију моћи теста ако је узорак из нормалне расподеле и $\sigma = 2$.

49. Нека је $\mathbf{X}$ узорак обима n. За тестирање хипотезе $H_0(X \in \mathcal{E}(\lambda))$ користи се тест статистика $T_n = \frac{1}{\bar{X}} \frac{1}{\binom{n}{2}} \sum\limits_{i<j} (2\min(X_i, X_j) - \frac{1}{2}(X_i + X_j))$ и критична област $W = \{|T_n| > c\}$.

   а) Ако је праг значајности теста 0.05, одредити емпиријску меру теста користећи и граничну расподелу и користећи емпријску нулту таблицу за $n = 20$ и $n = 50$. Образложити добијене резултате.

   б) Одредити емпиријску моћ теста против алтернативе $H_1(X \in \gamma(2, 1))$ користећи граничну расподелу и користећи емпријску нулту таблицу за $n = 20$ и $n = 50$.

50. Упоредити емпиријске нивое значајности и моћи тестова $T_n = \frac{1}{\bar{X}} \frac{1}{n^2} \sum\limits_{i,j} (2\min(X_i, X_j) - \frac{1}{2}(X_i + X_j))$ и $M_n = \overline{X} \frac{1}{n^2} \sum\limits_{i,j} (\frac{1}{1+2\min(X_i, X_j)} - \frac{1}{2}(\frac{1}{1+X_i} + \frac{1}{1+X_j}))$ за тестирање $H_0(X \in \mathcal{E}(\lambda))$ против $H_1(X \in \gamma(2, 1))$, ако су значајне велике вредности статистика и ако је ниво значајности $\alpha = 0.05$.

51. Нека је $\mathbf{X}$ узорак обима n. За тестирање хипотезе $H_0 : f(x) = 1, x \in [0, 1]$ против алтернативе $H_1 : f(x; \theta) = \frac{1}{\theta}, x \in [0, \theta], \theta > 1$, предлаже се тест функција $\varphi_T = I\{|T_n| > k\}$, где је тест статистика $T_n = \sqrt{n}(2\bar{X} - 1)$.

   а) У случају $n = 100$ одредити константу k тако да тест φ_T буде мере $\alpha = 0.05$.

   б) Одредити функцију моћи теста. Да ли је тест φ_T непристрасан? Да ли је функција моћи монотона функција?

   в) Да ли је овај тест постојан?

52. Нека је $\mathbf{X}$ узорак из непрекидне расподеле $F(x - \theta)$, где је θ медијана дате расподеле. За тестирање $H_0 : \theta = \theta_0$ против $H_1 : \theta > \theta_0$, користи се тест знака. Испитати непристрасност теста и показати да је тест постојаност против свих алтернатива.

53. Нека је $\mathbf{X}$ узорак из непрекидне расподеле симетричне око параметра положаја θ и има функцију расподеле $F(x - \theta)$. Да би се тестирало $H_0(\theta = 0)$ против $H_1(\theta > 0)$ може се користити једноузорачки Вилкоксонов тест означених рангова. Испитати непристрасност теста и показати да је тај тест постојан када $n \to \infty$.

54. Нека је $\mathbf{X}$ узорак обима n. За тестирање хипотезе $H_0 : f(x) = \lambda e^{-\lambda x}, \lambda > 0$ предложен је тест

$$T_n = \frac{1}{n(n-1)(n-2)} \sum_{1 \leq i_1 < i_2 < i_3 \leq n} \sum_{\pi(3)} (I\{2\min(X_{i_{\pi(1)}}, X_{i_{\pi(2)}}) < X_{i_{\pi(3)}}\} - I\{X_{i_{\pi(1)}} < X_{i_{\pi(3)}}\}).$$

Одредити Бахадуров нагиб теста теста за блиске алтернативе $g(x; \theta) = (1 + \theta)x^{\theta} e^{-x^{1+\theta}}$ и $g(x; \theta) = (1 + \theta x)e^{-x - \frac{\theta}{2}x^2}$.

55. Нека је $\mathbf{X}$ узорак обима n. За тестирање хипотезе $H_0 : F(x) = 1 - x^{-\alpha}, x \geq 1, \alpha > 0$ предложен је тест

$$T_n = \frac{1}{n(n-1)(n-2)} \sum_{1 \leq i_1 < i_2 < i_3 \leq n} \sum_{\pi(3)} \left(I\left\{ \max\left(\frac{X_{\pi(i)}}{X_{\pi(j)}}, \frac{X_{\pi(j)}}{X_{\pi(i)}} \right) < X_{i_{\pi(3)}} \right\} - I\{X_{i_{\pi(1)}} < X_{i_{\pi(3)}}\} \right).$$

Одредити Бахадуров нагиб теста теста за блиске алтернативе $G(x; \theta) = 1 - e^{-\alpha(\ln(x))^{\theta+1}}, x \geq 1, \alpha > 0, \theta \in (0, 1)$ и $G(x; \theta) = 1 - e^{-\alpha \ln(x) - \theta \ln^2 x}, x \geq 1, \alpha > 0, \theta \in (0, 1)$.

Статистички функционали

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

57. Одредити функционал θ дефинисан за све једнодимензионе расподеле са коначним првим моментом за које је $\theta(\widehat{F}_n) = \frac{1}{n}\sum\limits_{i=1}^{n} |X_i - \bar{X}|$.

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

59. 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 $\widehat{Q}_n = \widehat{F}_n^{-1}$.

60. За податке `nerv.txt` одредити емпиријску функцију расподеле и 95% траку поверења.

61. Нека је $\sigma(F)$ стандардна девијација случајне променљиве X са функцијом расподеле F. Одредити утицајну криву и граничну расподелу од $\sqrt{n}(\sigma(\widehat{F}_n) - \sigma(F))$, где је $\sigma(\widehat{F}_n)$ оцена стандардне девијације.

62. 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.

63. 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 $\widehat{F}_n(x)$ емпиријска функција расподеле и за $0 < t < 1$ дефинишемо $\widehat{F}_n^{-1}(t) = \inf\{x|\widehat{F}_n(x) \geq t\}$. Дефинишемо интерквартилни ранг са $\widehat{\tau}_n = \widehat{F}_n^{-1}(0.75) - \widehat{F}_n^{-1}(0.25)$. Наћи граничну расподелу од $\sqrt{n}(\widehat{\tau}_n - \tau(F))$, где је $\tau(F) = h_{\frac{3}{4}}(F) - h_{\frac{1}{4}}(F)$.

64. 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 $\sqrt{n}(h(\widehat{F}_n) - h(F))$.

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

66. Ако је $h(F) = \frac{h_1(F)}{h_2(F)}$, одредити утицајну криву од $h(F)$.

67. Нека је $h(F) = g(h_1(F))$, где је g нека диференцијабилна функција, а $h_1(F)$ функцијонал са утицајном кривом $\Phi_1(x; F)$. Ако је $\gamma_1(F) = \int \Phi^2(x; F)dF(x)$, доказати да је $\gamma(F) = \int \Phi^2(x; F)dF(x) = (g'(h_1(F)))^2\gamma_1(F)$.

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

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

Џекнајф и бутстреп

70. Нека су $X_1, ..., X_n$ независне и једнако расподељене случајне величине са расподелом која има очекивање μ и дисперзију σ^2. Одредитит кориговану џекнајф оцену од σ^2.

71. Нека су $X_1, ..., X_n$ независне и једнако расподељене случајне величине са густином $f(x;\theta) = \frac{1}{\theta}, 0 \leq x \leq \theta$, где је θ непознати параметар. Одредити кориговану џекнајф оцену параметра.

72. За податке *prihodi.txt* одредити кориговану џекнајф оцену и оцену стандардне грешке Ђини индекса $h(F) = 1 - 2\int\limits_0^1 q_F(t)dt$, где је $q_F(t) = \frac{\int\limits_0^t F^{-1}(s)ds}{\int\limits_0^1 F^{-1}(t)dt}$.

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

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

75. 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: $\frac{n-1}{n}\frac{\sum\limits_{i=1}^{n}\psi^2(X_i-\hat{\theta})}{(\sum\limits_{i=1}^{n}\psi'(X_i-\hat{\theta}))^2}$

76. 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 $\hat{c}_k = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^k, k = 2, 3, 4$.

77. 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 $\hat{c_k} = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^k, k = 2, 3, 4$.

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

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

80. 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 $n = 3$.

81. 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 четири 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 $\hat{F}_n$.

82. 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 $x_{i1}, ..., x_{in}$.

83. 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) буттреп.

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

85. 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)

86. 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.

87. 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.

88. Нека је $X_1, ..., X_m$ узорак из расподеле са очекивањем μ_X и дисперзијом σ_X^2, а $Y_1, ..., Y_n$ узорак из расподеле са очекивањем μ_Y и дисперзијом σ_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 $\{X_1, ..., X_m, Y_1, ..., Y_n\}$,

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

3) узети узорак $\{X_1^*, ..., X_m^*\}$ из узорка $\left\{\frac{X_1}{\widehat{\sigma}_X}, ..., \frac{X_m}{\widehat{\sigma}_X}\right\}$ и узети узорак $\{Y_1^*, ..., Y_n^*\}$ из узорка $\left\{\frac{Y_1}{\widehat{\sigma}_Y}, ..., \frac{Y_n}{\widehat{\sigma}_Y}\right\}$.

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

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

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

90. Подаци представљају резултате фармаколошког екперимента 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

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

92. Одредити 95% траку поверења за хистограм података `suicide.txt`.

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

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

95. 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.

96. 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.

97. 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.

98. 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 језра, ако је параметар глаткости одређен Силвермановом формулом. Која је разлика између случајева (а) и (б)?

99. Одредити оптималну глаткост оцене густине језгром за податке `SDSS1.txt`, користећи метод најмањег квадрата са унакрсном провером ако се користи

    а) Гаусово језгро,

    б) Фуријеова трансформација.

100. Одредити оптималну глаткост оцене густине језгром за узорак из нормалне расподеле, користећи метод најмањег квадрата са унакрсном провером ако се користи

    а) Гаусово језгро,

    б) Фуријеова трансформација.

101. Одредити опрималну глаткост оцене густине језгром за податке `SDSS1.txt`, користећи метод функције веродостојности са унакрсном провером.

102. Одредити опрималну глаткост оцене густине језгром за узорак из нормалне расподеле, користећи метод функције веродостојности са унакрсном провером.

103. 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 $p = (0.2, 0.3, 0.5)$.

104. Оценити густину дводимензионог узорка (X, Y) из базе `geyser` из пакета `MASS`.

105. Оценити густину узорка из експоненцијалне $\mathcal{E}(1)$ расподеле

    а) рефлексијом података,

    б) користећи асиметрично језгро.

106. Оценити густину за податке `suicide.txt` користећи асиметрично језгро.

107. Оценити густину узорка из скалиране бета $\beta(3, 2)$ расподеле на интервалу $[0, 5]$.