Wstęp do statystyki matematycznej
Lista 9

1. Niech $X_1, ..., X_n$ będzie próbą z rozkładu $\mathcal{N}(\mu, 1)$. Wyznaczyć statystykę testu Neymana-Pearsona (najmocniejszego, w skrócie NP) dla testowania hipotezy $H_0 : \mu = 0$ przeciwko $H_1 : \mu = a,\ a > 0$. Przedstawić ja w dogodnej postaci (standaryzowany logarytm ilorazu wiarogodności). Wyznaczyć funkcję mocy dla ustalonego poziomu istotności $\alpha \in (0, 1/2)$. Wykreślić tę funkcję dla $\alpha = 0.05$ oraz $n = 4$ i $n = 16$ (na jednym wykresie). Dla $a = 1/2$ i $\alpha = 0.05$ oraz $\alpha = 0.01$ narysować wykres zależności mocy testu od liczebności próby n (na jednym wykresie), a dla $\alpha = 0.05$ i $a = 1/2$ oraz $a = 1$ na drugim wykresie. Przeprowadzić testowanie H_0 przeciwko H_1 dla $\alpha = 0.05$ i danych 0.48 0.77 2.41 0.87 0.75 0.32 0.37 -2.12 0.89 -1.54 -0.74 0.93 -1.22 -1.37 1.03 -0.32.

2. Niech $X_1, ..., X_n$ będzie próbą z rozkładu $\mathcal{N}(0, \sigma^2)$. Wyznaczyć statystykę testu NP (podobnie jak w zad. 1 w dogodnej postaci) dla testowania $H_0 : \sigma = 1$ przeciwko $H_1 : \sigma = b,\ b > 1$. Napisać funkcję mocy dla ustalonego $\alpha > 0$. Dla $\alpha = 0.05$ oraz $b = 2$ i $b = \sqrt{2}$ obliczyć moc dla kilku wartości n i wykreślić zależność mocy od n.

3. Niech $X_1, ..., X_n$ będzie próbą z rozkładu a) gamma $\mathcal{G}(\theta, 1),\ \theta > 0$; b) Pareto o gęstości $p_\theta(x) = \theta x^{-2}\mathbf{1}_{[\theta,\infty)}(x)$. Wyznaczyć statystykę testu NP dla testowania $H_0 : \theta = 1$ przeciwko $H_1 : \theta = \theta_1,\ \theta_1 > 1$. Wyznaczyć asymptotyczną wartość krytyczną i obliczyć ją dla $\alpha = 0.05$. Dla jakich n dokładna wartość krytyczna różni się o mniej niż 0.1 od wartości asymptotycznej. Dla b) wyznaczyc funkcję mocy testu.

4. Niech H_0 i H_1 będą hipotezami prostymi i ϕ^* najmocniejszym testem zrandomizowanym tych hipotez na poziomie $\alpha \in (0, 1)$. Udowodnić, że jeśli $\beta < 1$ jest mocą tego testu, to test $\phi^{**} = 1 - \phi^*$ jest najmocniejszym testem zrandomizowanym H_1 przeciwko H_0 na poziomie istotnosci $1 - \beta$.

5. Niech $X_1, ..., X_n$ będzie próbą z rozkładu dwupunktowego o gęstości $p_\theta(0) = 1 - \theta,\ p_\theta(1) = \theta,\ \theta \in (0, 1)$, względem miary liczącej na $\mathcal{X} = \{0, 1\}$. Uzasadnić, że rodzina rozkładów w rozważanym modelu ma monotoniczny iloraz wiarogodności względem pewnej statystyki. Wskazać tę statystykę i skonstruować test UMP dla hipotezy $H_0 : \theta \leqslant \theta_0$ przeciwko $H_1 : \theta > \theta_0$ na poziomie istotności α. Dla $\theta_0 = 0.25$ i $\alpha = 0.05$ wyznaczyć najmniejszą liczebność próby n, przy której moc testu dla $\theta = 0.45$ wynosi co najmniej 0.7, posługując się rozkładem dokładnym i asymptotycznym statystyki testowej.

6. Niech $X_1, ..., X_n$ będzie próbą z rozkładu

   a) normalnego $\mathcal{N}(0, \theta),\ \theta > 0$;

   b) beta $\mathcal{B}e(\theta, 1),\ \theta > 0$.

   Uzasadnić monotoniczność ilorazu wiarogodności względem odpowiedniej statystyki i skonstruować test UMP dla hipotezy $H_0 : \theta \leqslant \theta_0$ przeciwko $H_1 : \theta > \theta_0$. Wyznaczyć wartości krytyczne testów dla zadanego poziomu α.

7. Dla modelu hipergeometrycznego z parametrem θ skonstruować zrandomizowany test UMP dla hipotez jak w poprzednim zadaniu. Obliczyć wartość krytyczną dla $N = 20,\ n = 4,\ \theta_0 = 6$.

8. Niech $X_1, ..., X_n$ będzie próbą z przesuniętego rozkładu wykładniczego o gęstości $p_{a,b}(x) = ae^{-a(x+b)}\mathbf{1}_{(-b,\infty)}(x),\ a > 0, b \in R$. Pokazać, że transformacja $y = e^{-ax}$ przekształca tę rodzinę w rodzinę rozkładów jednostajnych na odcinku $[0, e^{ab}]$. Korzystając z przykładu omówionego na wykładzie, skonstruować test UMP dla hipotezy $H_0 : a \leqslant a_0, b \leqslant b_0$ przeciwko $H_1 : a > a_0, b > b_0,\ b_0 > 0$.

Wstęp do statystyki matematycznej
Lista 10

1. Niech $X_1, ..., X_n$ bedzie próbą z rozkładu Poissona z parametrem λ. Skonstruować test ilorazu wiarogodności dla hipotezy $H_0 : \lambda = \lambda_0$ przeciwko $H_1 : \lambda \neq \lambda_0$ na poziomie α. Wyznaczyć wartość krytyczną testu dla $n = 10,\ \lambda_0 = 1$ i $\alpha = 0.05$ oraz moc testu (niezrandomizowanego)ędla $\lambda = 0.6$. Wyznaczyć rozkład asymptotyczny statystyki testowej i asymptotyczną wartość krytyczną.

2. Niech $X_1, ..., X_n$ będzie próbą z rozkładu beta z parametrami θ i 1. Skonstruować test ilorazu wiarogodności dla hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta \neq \theta_0$ na poziomie istotności α.

3. Niech $X_1, ..., X_n$ będzie próbą z rozkładu $\mathcal{N}(\mu, \sigma^2)$ i $Y_1, ..., Y_m$ próbą z rozkładu $\mathcal{N}(\nu, \tau^2)$. Skonstruować test ilorazu wiarogodności dla hipotezy $H_0 : \tau = k\sigma$ przeciwko $H_1 : \tau \neq k\sigma$, gdzie k znana liczba (np. 1) na poziomie α. Dla $k = 1$, $n = 18$, $m = 12$, $\alpha = 0.05$ wyznaczyć wartość krytyczną, posługując się rozkładem Snedecora. Czy otrzymany test jest nieobciążony?

4. Niech $X_1, ..., X_n$ będzie próbą z (dwuparametrowego) rozkładu Pareto $\mathcal{P}a(\theta, a)$. Znaleźć test ilorazu wiarogodności dla hipotezy $H_0 : a = a_0$ przeciwko $H_1 : a \neq a_0$. Dlaczego trudno jest wyznaczyć wartość krytyczną testu?

5. Liczba niezrealizowanych zamówień N w $n = 260$ oddziałach pewnej firmy wyniosła

N	0	1	2	3	4	5	6	7	8	9
liczba oddziałów	9	18	36	53	54	41	27	14	5	3

Na poziomie istotności 0.05 zweryfikować hipotezę, że N ma rozkład a) Poissona z parametrem $\lambda = 4.5$; b) Poissona.

6. Stopa dochodów (względem miesiąca referencyjnego) hipermarketu w 20 wylosowanych miesiącach (w uporządkowaniu rosnącym) wyniosły 1.01 1.01 1.03 1.05 1.06 1.07 1.09 1.10 1.18 1.21 1.26 1.29 1.32 1.55 1.60 2.04 2.22 2.65 5.55 5.62. Zweryfikować hipotezę, żę dochody mają rozkład Pareto $\mathcal{P}a(1, 3)$ na poziomie 0.05 za pomocą testów Kołmogorowa, Craméra-von Misesa i chi-kwadrat. Obliczyć p-wartości tych testów.

7. Pewne detale można wytwarzać trzema metodami. Wylosowano 270 detali i zbadano ich jakość. Otrzymano wyniki

jakość	metoda I	metoda II	metoda III
dobra	40	90	60
zła	10	50	20

Na poziomie 0.05 zweryfikować hipotezę, że jakość detali nie zależy od metody wytwarzania. Użyć testu niezależności chi-kwadrat.

8. Udowodnić, że testy Kołmogorowa i Craméra-von Misesa są zgodne i nieobciążone.

9. Niech χ^2 oznacza statystykę testowa w teście zgodności chi-kwadrat hipotezy prostej, a $\widetilde{\chi}^2$ jej modyfikację postaci (Neyman, 1949)

$$\widetilde{\chi}^2 = \sum_{i=1}^{r} \frac{(N_i - np_i)^2}{N_i},$$

gdzie N_i liczebności empiryczne, a np_i odpowiednie liczebności teoretyczne. Udowodnić, że $\chi^2 - \widetilde{\chi}^2 \xrightarrow{P_0} 0$ tj. przy prawdziwości hipotezy zerowej.

TEST NEYMANA-PEARSONA DLA ROZKŁADU PARETO $p_\theta(x) = \theta x^{-2}\mathbf{1}_{[\theta,\infty)}(x)$

$X_1, ..., X_n$ próba
$H_0 : \theta = 1, \quad H_1 : \theta = \theta_1, \;\; \theta_1 > 1.$

Gęstość łączna próby przy H_0 $(x_1...x_n)^{-2}\mathbf{1}_{(1,\infty)}(x_{(1)})$, a przy H_1 $\theta_1^n(x_1...x_n)^{-2}\mathbf{1}_{(\theta_1,\infty)}(x_{(1)})$. Iloraz gęstości (statystyka) wynosi po opuszczeniu stałej

$$L(x) = \mathbf{1}_{(\theta_1,\infty)}(x_{(1)})$$

jest funkcją niemalejącą statystyki $X_{(1)}$. Zatem statystyką testu NP jest $T_n = X_{(1)}$. Rozkład $X_{(1)}$ przy H_0 $\quad P_0(X_{(1)} \leqslant x) = 1 - \frac{1}{x^n}$ dla $x > 1$, a przy H_1 $\quad P_1(X_{(1)} \leqslant x) = 1 - \left(\frac{\theta_1}{x}\right)^n$ dla $x > \theta_1$.

Zatem dokładna wartość krytyczna tego testu wynosi $t_{\alpha n} = 1/(\alpha)^{1/n}$. Wtedy moc wyraża się wzorem

$$\beta_1 = P_1(X_{(1)} > t_{\alpha n}) = \left(\frac{\theta_1}{t_{\alpha n}}\right)^n \wedge 1 = \alpha\theta_1^n \wedge 1.$$

Zauważmy, że przy H_0

$$n(T_n - 1) \xrightarrow{\mathcal{D}} W_1,$$

gdzie W_1 ma rozkład wykładniczy z parametrem $\lambda = 1$. Istotnie, dla $x \geqslant 0$

$$P_0(n(X_{(1)} - 1) \leqslant x) = P_0(X_{(1)} \leqslant 1 + \frac{x}{n}) = 1 - \frac{1}{(1 + \frac{x}{n})^n} \to 1 - e^{-x}.$$

Można zatem rozpatrywać test asymptotyczny określony przez statystykę $S = n(X_{(1)} - 1)$ z asymptotyczną wartością krytyczną $t_\alpha = \log(1/\alpha)$. Test S odrzuca hipotezę, gdy $X_{(1)}$ przekracza $s_{\alpha n} = 1 + \frac{1}{n}\log(1/\alpha)$. Ponieważ $t_{\alpha n} = \exp\{\frac{1}{n}\log(1/\alpha)\} > 1 + \frac{1}{n}\log(1/\alpha) = s_{\alpha n}$, to test asymptotyczny ma rozmiar większy od nominalnego α i zawyża moc w stosunku do testu dokładnego.

Dla $n = 5$ dokładna wartość krytyczna wynosi 1.82, a asymptotyczna 1.60. Podobnie dla $n = 10$ 1.35 i 1.30, dla $n = 20$ 1.16 i 1.15.