

Statystyka

Lista 6

Niech $X_1, \dots, X_n$ będą niezależnymi zmiennymi losowymi z rozkładu o ciągłej dystrybucji F . Rozważamy problem testowania hipotezy

$$H_0 : F = F_0 \quad \text{przeciwko} \quad \text{alternatywie} \quad H_1 : F \neq F_0, \quad (1)$$

gdzie F_0 jest znaną dystrybucją.

Definiujemy nowe zmienne $U_1 = F_0(X_1), \dots, U_n = F_0(X_n)$. Wtedy, problem testowania (H_0, H_1) sprowadza się do weryfikowania

$$H_0 : U_1 \sim U(0, 1) \quad \text{przeciwko} \quad H_1 : U_1 \sim U(0, 1), \quad (2)$$

gdzie $U(0, 1)$ oznacza rozkład jednostajny na odcinku $(0, 1)$.

Niech $A_1, \dots, A_k$ będzie partycją odcinka $(0, 1)$, tzn. $\cup_{j=1}^k A_j = (0, 1)$ oraz $A_j \cap A_l = \emptyset$ dla $j \neq l$, $j, l = 1, \dots, k$. Niech $N_j = \#\{U_i \in A_j : i = 1, \dots, n\}$, a $p_j = P_0(U_1 \in A_j)$, $j = 1, \dots, k$.

Klasyczny test chi-kwadrat Pearsona oparty jest na statystyce

$$P_k = \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j}. \quad (3)$$

Przy prawdziwości hipotezy zerowej, statystyka P_k ma asymptotyczny rozkład chi-kwadrat z $k - 1$ stopniami swobody. Hipotezę H_0 odrzucamy dla dużych wartości statystyki P_k .

Niech $\{b_j\}_{j \in \mathbb{N}}$ będzie układem ortonormalnym wielomianów Legendre'a na $(0, 1)$.

Gładki test Neymana z k komponentami oparty jest na statystyce

$$N_k = \sum_{j=1}^k \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n b_j(U_i) \right\}^2. \quad (4)$$

Przy prawdziwości hipotezy zerowej, statystyka N_k ma asymptotyczny rozkład chi-kwadrat z k stopniami swobody. Hipotezę H_0 odrzucamy dla dużych wartości statystyki N_k .

Test Kołmogorowa-Smirnowa oparty jest na statystyce

$$KS = \sqrt{n} \sup_{u \in (0, 1)} |G_n(u) - u|, \quad (5)$$

gdzie G_n jest dystrybucją empiryczną w próbie $U_1, \dots, U_n$. Przy prawdziwości hipotezy zerowej, statystyka KS ma asymptotyczny rozkład Kołmogorowa. Hipotezę H_0 odrzucamy dla dużych wartości statystyki KS .

Celem ćwiczenia będzie badanie własności wybranych rozwiązań problemu (1).

Dokładniej, będziemy analizować

- (i) test chi-kwadrat Pearsona oparty na statystyce P_4 oraz P_8 z równomierną partycją,
- (ii) gładki test Neymana z 1, 4 i 8 składowymi,
- (iii) test Kołmogorowa-Smirnowa oparty na statystyce KS .

Poziom istotności $\alpha = 0.05$.

Zadanie 1.

Wygeneruj $n = 10$ obserwacji z rozkładu $U(0, 1)$. Na ich podstawie oblicz wartość statystyki P_4 , P_8 , N_1 , N_4 , N_8 i KS . Doświadczenie powtórz 10 000 razy. Na tej podstawie wyznacz wartości krytyczne analizowanych testów. Otrzymane wyniki porównaj z kwantylami rzędu 0.95 odpowiednich rozkładów granicznych. Przedyskutuj uzyskane wyniki. Powtórz eksperyment dla $n = 20, 30, \dots, 100$.

Zadanie 2.

Omów algorytm generowania obserwacji z rozkładu o zadanej gęstości. Patrz np. rozdział 8.2.4, str. 433-434, Koronacki i Mielniczuk (2009).

Zadanie 3.

Wygeneruj $n = 10$ obserwacji z rozkładu o gęstości $C_1(u, 0.4) = 1 + 0.4 \cos(\pi u)$, $u \in (0, 1)$. Na ich podstawie oblicz wartość statystyki P_4 , P_8 , N_1 , N_4 , N_8 i KS . Doświadczenie powtórz 10 000 razy. Na tej podstawie dokonaj estymacji mocy testów. Powtórz eksperyment dla $n = 20, 30, \dots, 100$. Sporządź wykres funkcji mocy testów w zależności od n . Przedyskutuj uzyskane rezultaty.

Zadanie 4.

W zadaniu 3, wymień $C_1(u, 0.4) = 1 + 0.4 \cos(\pi u)$, $u \in (0, 1)$ na $C_j(u, \rho) = 1 + \rho \cos(j\pi u)$, $u \in (0, 1)$. Następnie, powtórz eksperyment numeryczny dla

- (i) $j = 2$, $\rho = 0.5$,
- (ii) $j = 3$, $\rho = 0.5$,
- (iii) $j = 4$, $\rho = 0.6$,
- (iv) $j = 5$, $\rho = 0.7$,
- (v) $j = 6$, $\rho = 0.7$.

Literatura

Koronacki, J., Mielniczuk, J. (2009). *Statystyka dla studentów kierunków technicznych i przyrodniczych*. Wydawnictwa Naukowo-Techniczne, Warszawa.

Rayner, J.C.W., Best, D.J. (1989). *Smooth Tests of Goodness of Fit*. Oxford University Press, New York.