Statystyka

Lista 5

Niech $X_1, \ldots, X_m$ będą niezależnymi zmiennymi losowymi z rozkładu o ciągłej dystrybuancie F. Niech $Y_1, \ldots, Y_n$ będą niezależnymi zmiennymi losowymi z rozkładu o ciągłej dystrybuancie G. Zakładamy, że wszystkie zmienne są niezależne. Rozważamy problem testowania hipotezy

$$H_0 : F = G \quad \text{przeciwko alternatywie} \quad H_1 : F \neq G \tag{1}$$

na poziomie istotności $\alpha = 0.05$.

Niech $N = m+n$, a $\boldsymbol{Z} = (Z_1, \ldots, Z_N) = (X_1, \ldots, X_m; Y_1, \ldots, Y_n)$ będzie wektorem połączonych prób. Niech R_i będzie rangą Z_i w próbie $\boldsymbol{Z}$, $i = 1, \ldots, N$. Klasyczna liniowa statystyka rangowa związana z funkcją wynikową $\varphi \in L_2(0,1)$ ma postać

$$T_\varphi = \sqrt{\frac{mn}{N}} \left\{ \frac{1}{m} \sum_{i=1}^{m} \varphi\Big(\frac{R_i - 0.5}{N}\Big) - \frac{1}{n} \sum_{i=m+1}^{N} \varphi\Big(\frac{R_i - 0.5}{N}\Big) \right\}, \tag{2}$$

przy czym wybór funkcji φ determinuje czułość testu opartego na statystyce T_φ. Jeżeli $\varphi(u) = \varphi_1(u) = \sqrt{3}(2u - 1)$, to otrzymujemy statystykę Wilcoxona. Wybór $\varphi(u) = \varphi_2(u) = \sqrt{48}(0.25 - |u - 0.5|)$ prowadzi do statystyki Ansari-Bradley'a. Jeżeli $\int_0^1 \varphi(u)du = 0$, a $\int_0^1 \varphi^2(u)du = 1$, to przy prawdziwości hipotezy zerowej statystyka T_φ ma asymptotyczny rozkład standardowy normalny. Ponadto, H_0 odrzucamy na korzyść H_1 dla dużych wartości T_φ^2.

W problemie testowania (H_0, H_1) innym klasycznym rozwiązaniem jest, na przykład, test Kołmogorowa-Smirnowa odrzucający H_0 dla dużych wartości statystyki

$$KS = \sqrt{\frac{mn}{N}} \sup_{x \in \mathbb{R}} |F_m(x) - G_n(x)|, \tag{3}$$

gdzie F_m oraz G_n są dystrybuantami empirycznymi w próbie X-ów i Y-ów, odpowiednio.

Celem ćwiczenia będzie badanie zachowania funkcji mocy wybranych rozwiązań problemu (1). Dokładniej, będziemy analizować

(i) test Wilcoxona oparty na statystyce $W = T_{\varphi_1}^2$,

(ii) test Ansari-Bradley'a oparty na statystyce $AB = T_{\varphi_2}^2$,

(iii) test Lepage'a oparty na statystyce $L = W + AB$,

(iv) test Kołmogorowa-Smirnowa oparty na statystyce KS.

Zadanie 1.

Wygeneruj $m = n = 20$ obserwacji z rozkładu $N(0,1)$. Na ich podstawie oblicz wartość statystyki W, AB, L i KS. Doświadczenie powtórz 10 000 razy. Wyznacz wartości krytyczne odpowiadających im testów prawostronnych. Czy taki sposób generowania wartości krytycznych jest poprawny? Odpowiedź uzasadnij.

Zadanie 2.

Wygeneruj $m = n = 20$ obserwacji z rozkładu

(a) normalnego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2, odpowiednio,

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.2$, $\sigma_2 = 1$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.4$, $\sigma_2 = 1$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.6$, $\sigma_2 = 1$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.8$, $\sigma_2 = 1$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 1$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.2$, $\sigma_2 = 1$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.4$, $\sigma_2 = 1$,

(b) logistycznego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2, odpowiednio,

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.2$, $\sigma_2 = 1$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.4$, $\sigma_2 = 1$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.6$, $\sigma_2 = 1$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.8$, $\sigma_2 = 1$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 1$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.2$, $\sigma_2 = 1$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.4$, $\sigma_2 = 1$,

(c) Cauchy'ego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2, odpowiednio,

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.0$, $\sigma_2 = 1$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.5$, $\sigma_2 = 1$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 1$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.5$, $\sigma_2 = 1$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 2.0$, $\sigma_2 = 1$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 2.5$, $\sigma_2 = 1$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 3.0$, $\sigma_2 = 1$.

Na ich podstawie oblicz wartość statystyki W, AB, L i KS. Doświadczenie powtórz 10 000 razy. Oszacuj wartość funkcji mocy analizowanych testów. Narysuj wyestymowane funkcje mocy w zależności od parametru μ_2. Przedyskutuj uzyskane wyniki.

Zadanie 3.

Wygeneruj $m = n = 20$ obserwacji z rozkładu

(a) normalnego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2, odpowiednio,

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 1.0$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 1.5$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 2.0$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 2.5$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 3.0$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 3.5$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 4.0$,

(b) logistycznego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2, odpowiednio,

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 1.0$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 1.5$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 2.0$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 2.5$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 3.0$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 3.5$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 4.0$,

(c) Cauchy'ego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2, odpowiednio,

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 1.0$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 2.0$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 3.0$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 4.0$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 5.0$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 6.0$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 7.0$.

Na ich podstawie oblicz wartość statystyki W, AB, L i KS. Doświadczenie powtórz 10 000 razy. Oszacuj wartość funkcji mocy analizowanych testów. Narysuj wyestymowane funkcje mocy w zależności od parametru σ_2. Przedyskutuj uzyskane wyniki.

Zadanie 4.

Wygeneruj $m = n = 20$ obserwacji z rozkładu

(a) normalnego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2, odpowiednio,

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.2$, $\sigma_2 = 1.0$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.4$, $\sigma_2 = 1.5$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.6$, $\sigma_2 = 2.0$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.8$, $\sigma_2 = 2.5$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 3.0$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.2$, $\sigma_2 = 3.5$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.4$, $\sigma_2 = 4.0$,

(b) logistycznego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2, odpowiednio,

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.2$, $\sigma_2 = 1.0$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.4$, $\sigma_2 = 1.5$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.6$, $\sigma_2 = 2.0$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.8$, $\sigma_2 = 2.5$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 3.0$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.2$, $\sigma_2 = 3.5$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.4$, $\sigma_2 = 4.0$,

(c) Cauchy'ego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2, odpowiednio,

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.0$, $\sigma_2 = 1.0$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.5$, $\sigma_2 = 2.0$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 3.0$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.5$, $\sigma_2 = 4.0$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 2.0$, $\sigma_2 = 5.0$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 2.5$, $\sigma_2 = 6.0$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 3.0$, $\sigma_2 = 7.0$.

Na ich podstawie oblicz wartość statystyki W, AB, L i KS. Doświadczenie powtórz 10 000 razy. Oszacuj wartość funkcji mocy analizowanych testów. Narysuj wyestymowane funkcje mocy w zależności od wektora parametrów (μ_2, σ_2). Przedyskutuj uzyskane wyniki.

Zadanie 5.

Wygeneruj $m = n = 50$ obserwacji z rozkładu $N(0, 1)$. Na ich podstawie oblicz wartość statystyki W, AB, L i KS. Doświadczenie powtórz 10 000 razy. Wyznacz wartości krytyczne analizowanych testów prawostronnych.

Zadanie 6.

Wygeneruj $m = n = 50$ obserwacji z rozkładu

(a) normalnego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2,

(b) logistycznego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2,

(c) Cauchy'ego z parametrem przesunięcia μ_1 i μ_2 oraz skali σ_1 i σ_2.

W każdym przypadku dobierz parametry μ_1 i μ_2 oraz skali σ_1 i σ_2, analogicznie jak w zadaniach 2, 3, 4, tak aby uzyskać moce w pełnym zakresie, ale nie były one zdegenerowane. Sporządź wykresy funkcji mocy w zależności od μ_2, σ_2 oraz (μ_2, σ_2), odpowiednio. Przedyskutuj uzyskane rezultaty.