

## Statystyka

### Lista 5

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

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

na poziomie istotności  $\alpha = 0.05$ .

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

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

przy czym wybór funkcji  $\varphi$  determinuje czułość testu opartego na statystyce  $T_\varphi$ . Jeżeli  $\varphi(u) = \varphi_1(u) = \sqrt{3}(2u - 1)$ , to otrzymujemy statystykę Wilcoxa. 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  $T_\varphi$  ma asymptotyczny rozkład standardowy normalny. Ponadto,  $H_0$  odrzucamy na korzyść  $H_1$  dla dużych wartości  $|T_\varphi|$ .

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

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

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 Wilcoxa oparty na statystyce  $W = T_{\varphi_1}$ ,
- (ii) test Ansari-Bradley'a oparty na statystyce  $AB = T_{\varphi_2}$ ,
- (iii) test Lepage'a oparty na statystyce  $L = W^2 + AB^2$ ,
- (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. Na tej podstawie wyznacz wartości krytyczne analizowanych 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  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ , 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  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ , 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  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ , 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. Na tej podstawie oszacuj wartość funkcji mocy analizowanych testów. Narysuj wyestymowane funkcje mocy w zależności od parametru  $\mu_2$ . Przedyskutuj uzyskane wyniki.

### Zadanie 3.

Wygeneruj  $m = n = 20$  obserwacji z rozkładu

- (a) normalnego z parametrem przesunięcia  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ , 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  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ , 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  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ , 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. Na tej podstawie oszacuj wartość funkcji mocy analizowanych testów. Narysuj wyestymowane funkcje mocy w zależności od parametru  $\sigma_2$ . Przedyskutuj uzyskane wyniki.

#### Zadanie 4.

Wygeneruj  $m = n = 20$  obserwacji z rozkładu

(a) normalnego z parametrem przesunięcia  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ , 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  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ , 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  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ , 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. Na tej podstawie oszacuj wartość funkcji mocy analizowanych testów. Narysuj wyestymowane funkcje mocy w zależności od wektora parametrów  $(\mu_2, \sigma_2)$ . 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. Na tej podstawie wyznacz wartości krytyczne analizowanych testów prawostronnych.

### Zadanie 6.

Wygeneruj  $m = n = 50$  obserwacji z rozkładu

- (a) normalnego z parametrem przesunięcia  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ ,
- (b) logistycznego z parametrem przesunięcia  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ ,
- (c) Cauchy'ego z parametrem przesunięcia  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ .

W każdym przypadku dobierz parametry  $\mu_1$  i  $\mu_2$  oraz skali  $\sigma_1$  i  $\sigma_2$ , 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  $\mu_2, \sigma_2$  oraz  $(\mu_2, \sigma_2)$ , odpowiednio. Przedyskutuj uzyskane rezultaty.