
Testy oparte na danych dla jednowymiarowej symetrii względem nieznanej mediany

*Praca pisana pod kierunkiem
Dr. Grzegorza Wyłupka*

Układ pracy

- Określenie problemu testowania i opisanie statystyki wynikowej
- Specyfikacja testów
- Część symulacyjna
- Wnioski

Testowanie hipotez statystycznych

Testowanie hipotez statystycznych to formalna procedura podejmowania decyzji na podstawie danych z próby. Celem jest sprawdzenie, czy dostępne dane dostarczają wystarczających dowodów przeciwko przyjętej **hipotezie zerowej(H_0)**

Proces testowania hipotez:

- Sformułowanie hipotez
- Wybór statystyki wynikowej
- Obliczenie statystyki testowej
- Podjęcie decyzji

Statystyka wynikowa

Statystyka wynikowa to funkcja określona na próbie danych, która służy do wyciągania wniosków o populacji, z której pochodzi ta próba. Najczęściej statystyka wynikowa przyjmuje postać jednej liczby — ułatwia to porównywanie wyników, testowanie hipotez, czy szacowanie parametrów. Jej wartość porównujemy z rozkładem teoretycznym (np. rozkładem normalnym, t-Studenta, chi-kwadrat), aby ocenić, jak „nietypowy” jest nasz wynik.

Testowane zagadnienie

Zakładamy, że $X_1, \dots, X_n$ są losowymi, niezależnymi zmiennymi o nieznanej medianie μ , ciągłej dystrybuancie $F(x - \mu)$ i gęstości $f(x - \mu)$, gdzie $F(0) = \frac{1}{2}$. Naszym celem jest sprawdzenie hipotezy

$$H_0 : F(x) = 1 - F(-x), x \in R,$$

czyli zbadanie symetryczności F wokół zera (równoważnie: symetryczności rozkładu X_i wokół μ).

Statystyki wynikowe(score statistics)

$$W_k = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \ell_*(X_i) \right)^T I_*^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \ell_*(X_i) \right)$$

$$\ell_*(x) = \Psi(F_s(x - \mu)) + \frac{v}{J} \frac{f'_s}{f_s}(x - \mu)$$

$$\bar{I}_* = I_k - J^{-1} v v^T$$

$$J = J_{f_s} = \int_R \frac{(f'_s(x))^2}{f_s(x)} dx$$

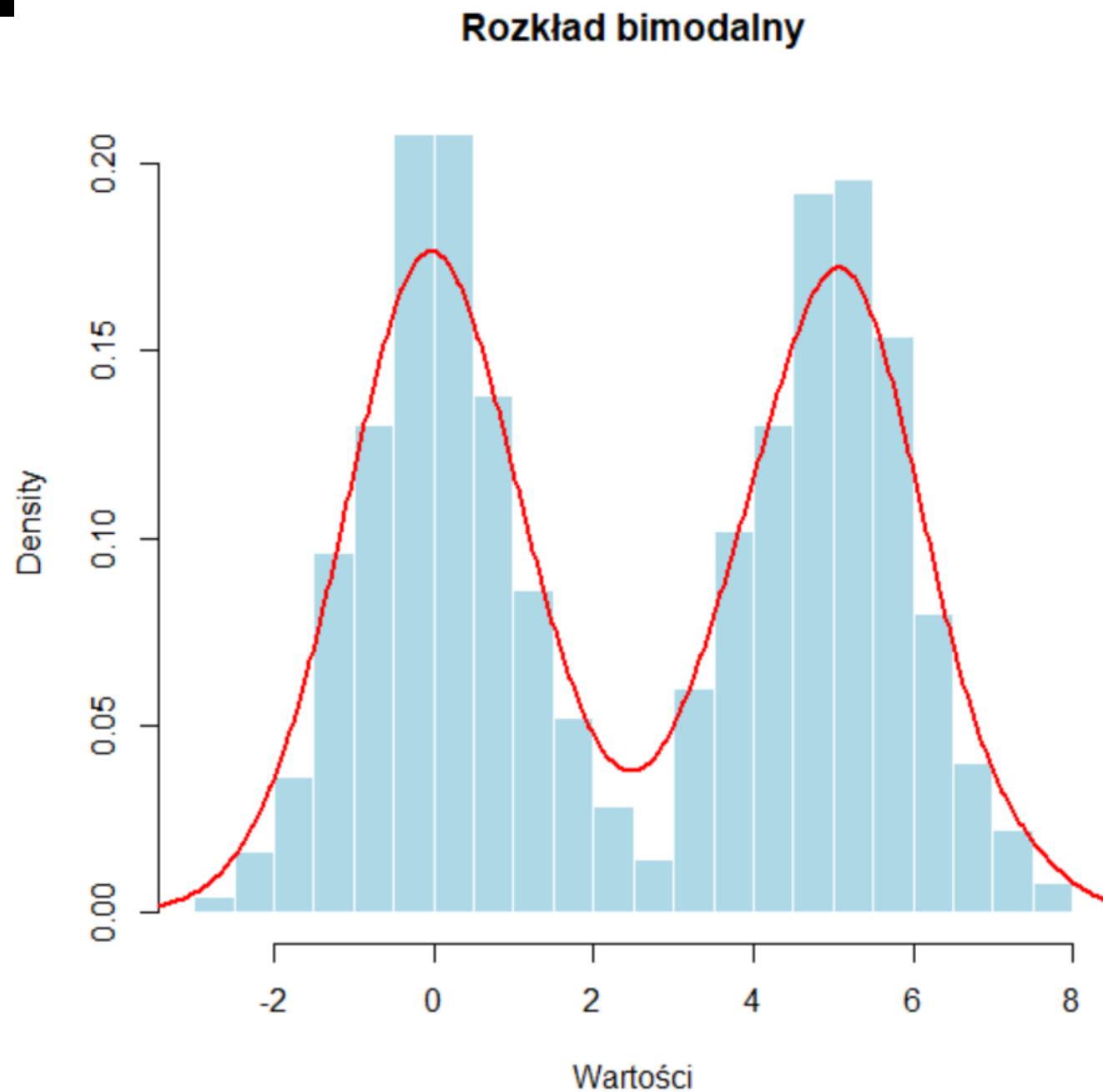
Reguły selekcji Schwarza

$$S = \min \left\{ 1 \leq k \leq d(n) : \hat{W}_k - k \log n = \max_{1 \leq j \leq d(n)} \left(\hat{W}_j - j \log n \right) \right\}$$

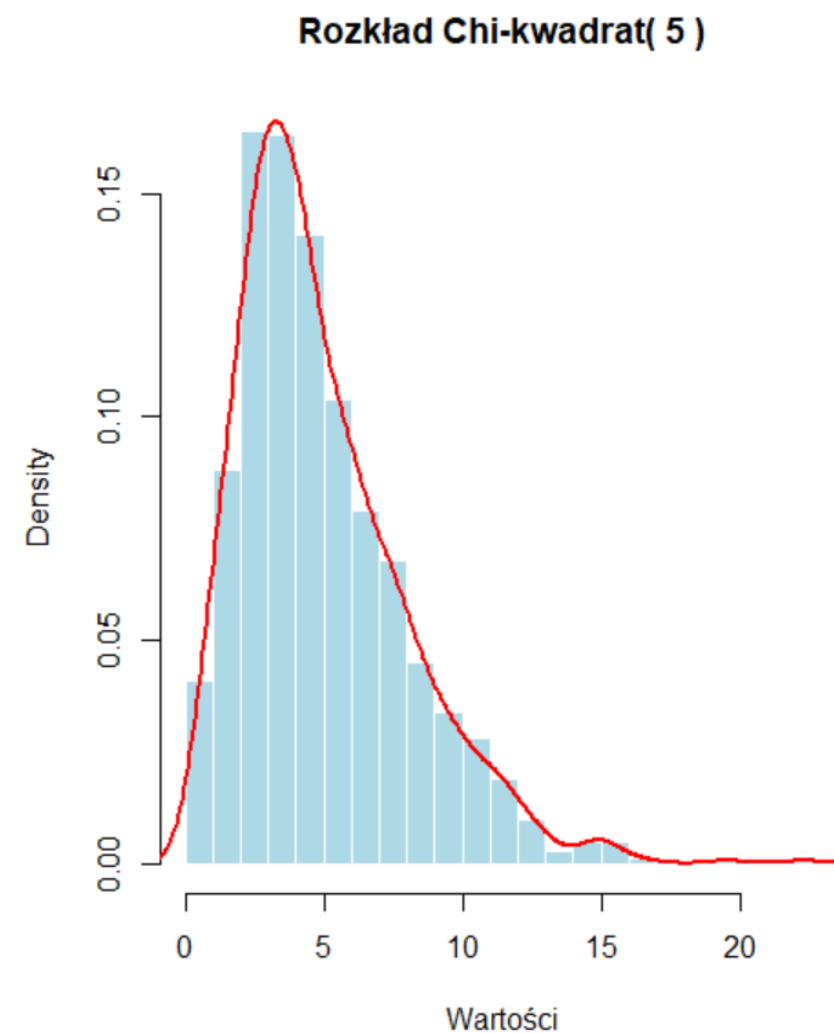
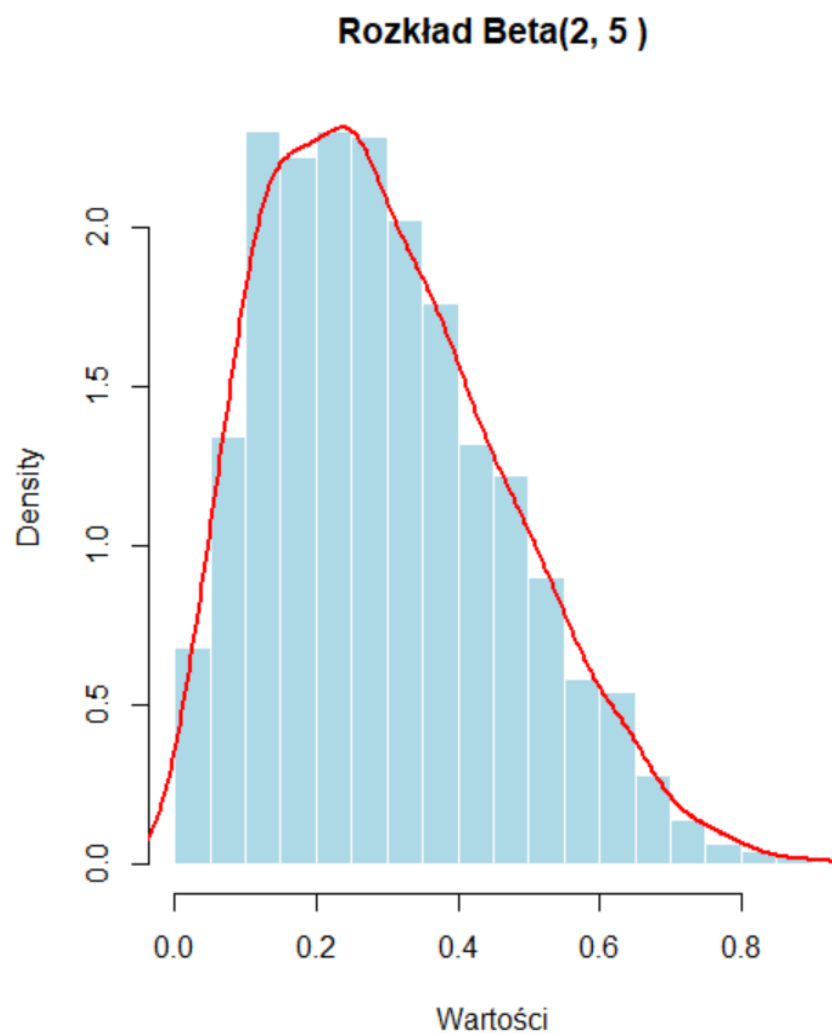
$$L = \min \left\{ 1 \leq k \leq d(n) : \hat{W}_k - \pi(k, n) = \max_{1 \leq j \leq d(n)} \left(\hat{W}_j - \pi(j, n) \right) \right\}$$

Estymator środka

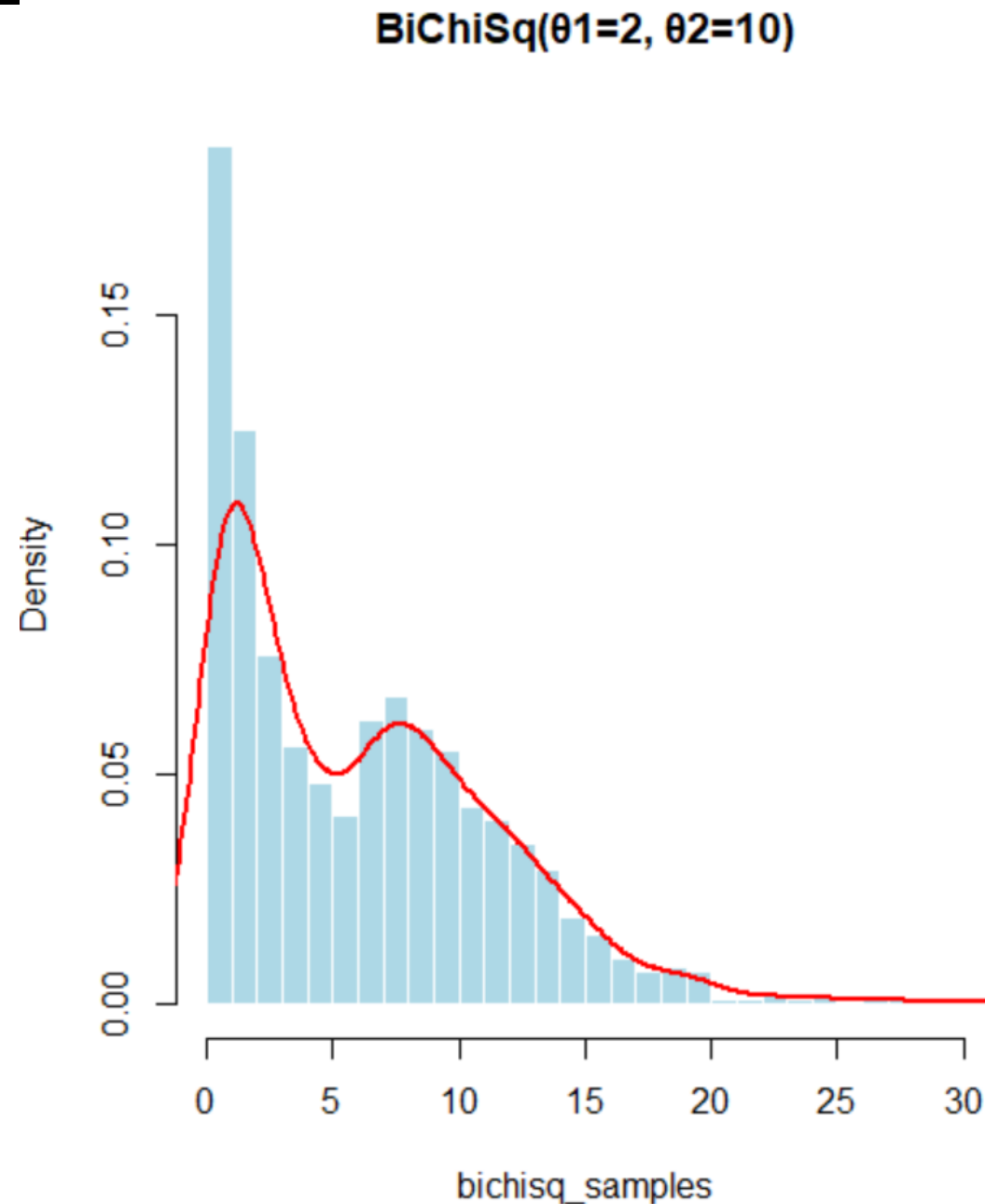
$$\hat{\mu} = (X_{(\lfloor nq \rfloor)} + X_{(n - \lfloor nq \rfloor + 1)})/2$$



Dominująca asymetria w ogonach rozkładu



Asymetria zarówno w
ogonach, jak i w
centrum rozkładu



Asymetria tylko w centrum rozkładu

