

# Statystyka - Laboratorium 5

Martyna Konopacka

## Wprowadzenie i cele

Rozważmy problem testowania hipotezy  $H_0 : F = G$  przeciwko  $H_1 : F \neq G$ , gdzie  $F, G$  są ciągłymi dystrybuantami nieznanymi rozkładów, z których pochodzą próby  $X_1, \dots, X_m; Y_1, \dots, Y_n$  - na naszej liście  $n = m$ . Przyjmujemy, że testy będą prawostronne, na poziomie istotności  $\alpha$ . Rozważane testy są nieparametryczne - opierają się na statystykach rangowych, przez co mają zastosowanie również do danych innych niż liczbowe, ale dających się uporządkować. Drugą zaletą tego rodzaju testów jest możliwość stosowania ich bez założeń o ich rozkładzie cechy w populacji.

*Rangą*  $R_i$  nazwiemy liczbę przyporządkowaną obserwacji poprzez posortowanie wybranej zmiennej. Przykładowo dla  $X = (1, 7, 3, 10)$  rangi to  $(1, 3, 2, 4)$ . Uwaga: dla rosnącego przekształcenia zmiennej losowej rangi pozostają bez zmian - to przyda się w późniejszych rozważaniach. Przykładowo  $2X + 7 = (9, 21, 13, 27)$  - rangi pozostają takie same.

*Testem* nazwiemy funkcję próby  $\varphi(X)$ . Na podstawie wartości  $\varphi(X)$  podejmujemy decyzję: odrzucamy  $H_0$  dla dużych wartości funkcji.

*Moc testu* definiujemy jako prawdopodobieństwo słusznego odrzucenia hipotezy  $H_1$  w zależności od parametru.  $\gamma(\theta_0) = 1 - P(S(X) \leq c | H_1, \theta = \theta_0)$

*Wartość krytyczna* to wartość, która dzieli przestrzeń możliwych wartości statystyki testowej  $S(X)$  na dwa podzbiory - dla jednego z nich odrzucamy hipotezę zerową. W tym przypadku będą to przedziały na osi liczb rzeczywistych i hipotezę odrzucimy, gdy  $S(X) \geq c$ .

## Jak wyznaczamy wartości krytyczne?

Liczba  $\alpha$  reprezentuje prawdopodobieństwo popełnienia błędu typu I, czyli odrzucenia  $H_0$ , gdy jest ona prawdziwa.

$$\alpha = P(S(X) \geq c | H_0) = 1 - P(S(X) \leq c | H_0)$$

$$P(S(X) \leq c | H_0) = 1 - \alpha$$

zatem  $c = F_S^{-1}(1 - \alpha)$  jest kwantylem rzędu  $1 - \alpha$  rozkładu  $S(X)$  przy  $H_0$ . Uwaga: nie znamy prawdziwego rozkładu statystyki testowej przy  $H_0$ , a jedynie rozkład asymptotyczny.

Celem poniższego raportu jest wyznaczenie wartości krytycznych oraz zbadanie zachowania funkcji mocy w zależności od parametrów populacji dla testów: Wilcoxona (W), Ansari-Bradley'a (AB), Lepage'a (L) i Kołmogorowa-Smirnowa (KS). W zadaniach rozważałam od razu obie liczebności prób: 20 i 50.

## Zadanie 1 + Zadanie 5

Zadanie 1 stanowi wstęp do kolejnych zadań - problem został zawężony do rozkładu normalnego. W tym zadaniu zajmujemy się tylko obliczaniem wartości krytycznych. Na początek zdefiniujemy odpowiednie statystyki:

```

# implementacja wzoru (2) z listy; oblicza wartość statystyki T dla wybranej funkcji (testu) phi i połą
# funkcja rank(X) zwraca wektor rang obliczonych dla wektora X
T_phi <- function(phi, X, Y){
  Z <- c(X, Y)
  m <- length(X)
  n <- length(Y)
  N <- n + m
  s <- sqrt(m * n / N)
  m1 <- mean(phi((rank(Z)[1:m] - 0.5)/N))
  m2 <- mean(phi((rank(Z)[(m+1):N] - 0.5)/N))
  return (s * (m1 - m2))
}

# implementacja funkcji (testów) phi1, phi2 z listy
phi1 <- function(u){sqrt(3)*(2*u - 1)}
phi2 <- function(u){sqrt(48)*(0.25 - abs(u - 0.5))}

# statystyki zdefiniowane na liście
W <- function(X, Y){T_phi(phi1, X, Y)^2}
AB <- function(x, y){T_phi(phi2, X, Y)^2}
L <- function(X, Y){W(X, Y) + AB(X, Y)}
KS <- function(X, Y){
  Z <- c(X, Y)
  m <- length(X)
  n <- length(Y)
  N <- n + m
  s <- sqrt(m*n/N)
  # return (ks.test(X,Y)$statistic) # to jest bez czynnika z pierwiastkiem
  return (s * ks.test(X, Y)$statistic) # zwróć wartość obliczoną przez funkcję ks.test odpowiadającą su
}

```

Z użyciem powyższych funkcji generujemy wyniki.

Table 1: Wartości krytyczne;  $n = m = 20$

	x
W.95%	3.888000
AB.95%	3.888000
L.95%	6.027337
KS.95%	1.264911

Table 2: Wartości krytyczne;  $n = m = 50$

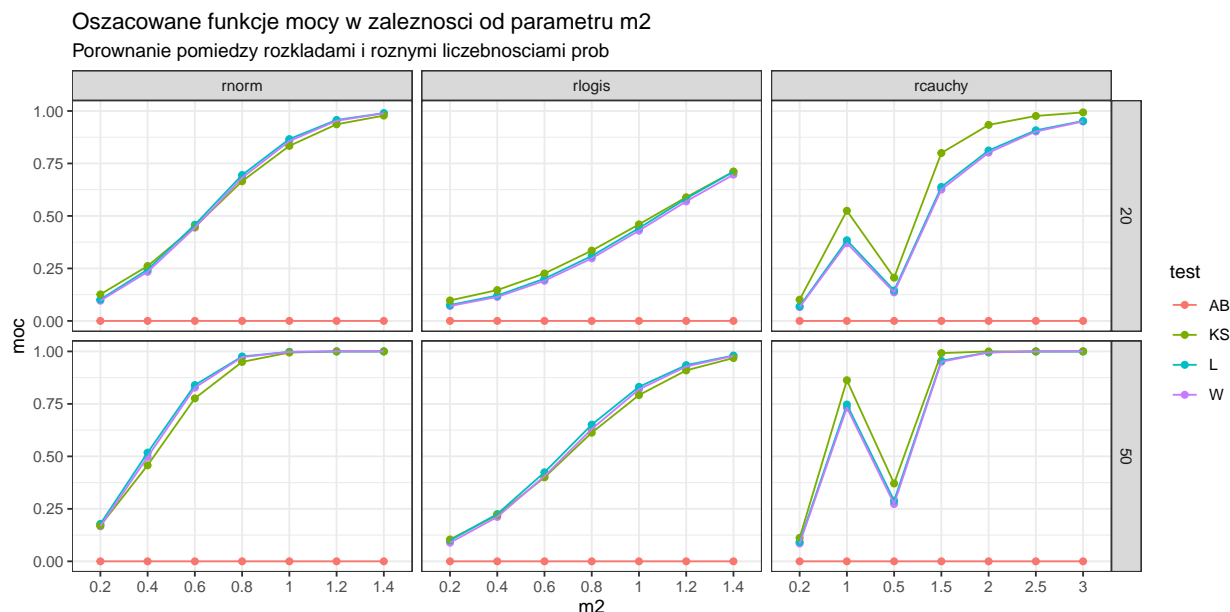
	x
W.95%	3.844272
AB.95%	3.926208
L.95%	5.954227
KS.95%	1.300000

**Komentarz:**

1. Wydaje się, że wartości zostały wyznaczone prawidłowo. W celu sprawdzenia tego porównałam je z kwantylami rzędu 95% z rozkładu chi-kwadrat (pierwsze trzy testy), jednak nie potrafię wyjaśnić dlaczego są one zbliżone.
2. Wyniki nie różnią się znacząco dla prób o liczebności 20 i 50.
3. Korzystając z własności, że przekształcenie rosnące nie zmienia rang, można udowodnić, że użyty w zadaniu sposób wyznaczania wartości krytycznych jest poprawny dla dowolnych dwóch rozkładów  $F = G$ .

## Zadanie 2

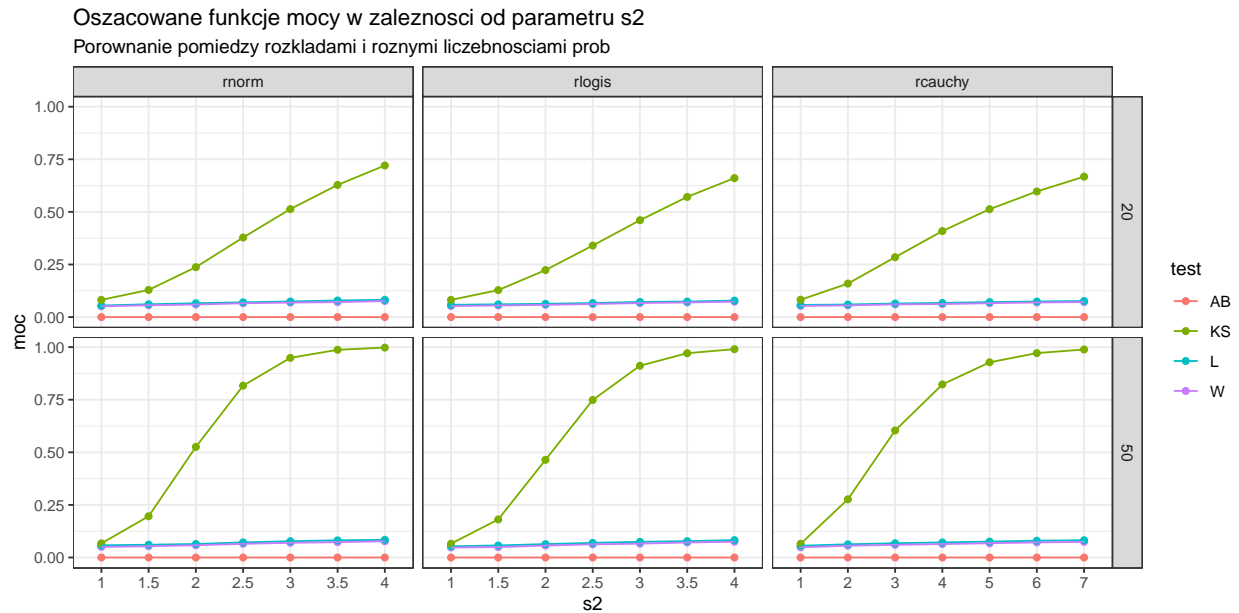
Funkcję mocy oszacujemy jako odsetek obserwacji (obserwacjami są wyniki 10000 powtórzeń eksperymentu) dla których nie popełniliśmy błędu typu II - a więc dla których na podstawie statystyki i wartości krytycznej odrzucamy hipotezę  $H_0$ . Sprowadza się to do obliczenia dla jakiego odsetka iteracji  $S(X) \geq c$ .



## Komentarz

1. Funkcje mocy dla testów Wilcoxona, Lepage'a i Kołmogorowa-Smirnowa wyglądają bardzo podobnie. Dla rozkładu normalnego i logistycznego są one rosnące. Intuicyjnie ma to sens, bo ze wzrostem  $\mu_2$  ('m2') rośnie różnica pomiędzy rozkładami, z których generujemy próby, więc powinny być one łatwiejsze do rozróżnienia a ryzyko popełnienia błędu mniejsze. Jest to dobrze widoczne, gdy naszkicujemy rozkłady o różnych parametrach i zaznaczymy obszar związany z błędem typu II.
2. W przypadku rozkładu Cauchy'ego wykres jest bardziej chaotyczny, co może mieć związek z dziwną naturą tego rozkładu - mimo, że testy nieparametryczne powinny działać dobrze niezależnie od rozkładu.
3. Można zauważyć, że testy mają większą moc, gdy zwiększymy liczebność próby. Intuicyjnie, więcej danych powinno dawać mniejsze ryzyko popełnienia błędu.
4. Test Ansari-Bradley'a odznacza się niską mocą - we wszystkich eksperymentach uzyskałam moc równą 0, co wydaje się nie mieć sensu i może być spowodowane złym sposobem wyliczania statystyki lub innym błędem.

### Zadanie 3

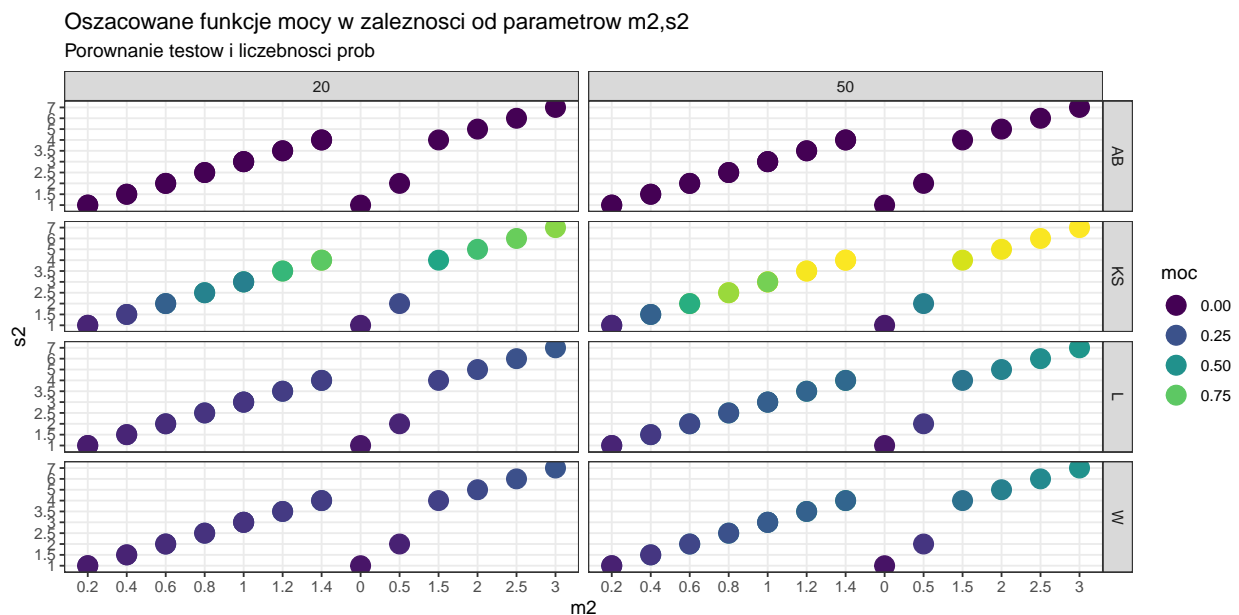


#### Komentarz:

1. Z wykresów wynika, że gdy rozkłady mają taki sam parametr  $\mu$ , a różnią się tylko parametrem skali, wszystkie testy oprócz testu KS mają moc równą lub bliską zero.
2. Dla testu Kołmogorowa-Smirnowa widać wzrost mocy wraz ze wzrostem zarówno  $n$ , jak i różnicy pomiędzy parametrami.

### Zadanie 4

W tym zadaniu badamy zależność mocy od dwóch parametrów na raz, dlatego wybrany został inny sposób prezentacji wyników. Rozkłady różnią się równocześnie parametrami przesunięcia i skali.



**Komentarz:**

1. Z wykresu i dodatkowych analiz wynika, że największą moc w przypadku populacji o różniących się obu parametrach uzyskamy dla testu KS, jak największej liczebności i jak największej różnicy parametrów.
2. Patrząc na powyższe wyniki ciężko wnioskować, czy większe znaczenie ma tutaj różnica  $\sigma$ , czy  $\mu$ .
3. Dla testu AB wyniki ponownie są bliskie zero, a dla pozostałych dwóch testów moc jest mniejsza niż w przypadku testu KS, ale widać, że rośnie wraz ze wzrostem różnicy parametrów i rozmiarem próby.

**Podsumowanie**

Eksperymenty wykazały, że wybór testu o największej mocy jest inny, gdy rozkłady różnią się parametrem przesunięcia lub skali. Ponadto, dla testów dla których moc nie jest bardzo mała lub zerowa, występuje wzrost mocy testu wraz ze zwiększaniem  $n$  i różnicy pomiędzy rozkładami.