

Statystyka - raport 5

Łukasz Rębisz

17.01.2023

Implementacja wzorów teoretycznych

Statystyka rangowa

Niech X_1, \dots, X_m - i.i.d z rozkładu o dystrybuancie F , Y_1, \dots, Y_n - i.i.d z rozkładu o dystrybuancie G . Zakładamy, że wszystkie zmienne są niezależne. Rozważmy problem testowania hipotezy

$$H_0 : F = G \quad \text{vs} \quad H_1 : F \neq G$$

na poziomie istotności $\alpha = 0.05$.

Niech $Z = (Z_1, \dots, Z_N) = (X_1, \dots, X_m; Y_1, \dots, Y_n)$ będzie wektorem połączonych prób, R_i będzie rangą Z_i w próbie Z .

Klasyczna **liniowa statystyka rangowa** związana z funkcją wynikową $\varphi \in L_2(0, 1)$ ma postać

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

```
T_phi <- function(X, Y, phi){
  m <- length(X)
  n <- length(Y)
  N <- m + n

  Z <- c(X, Y)
  ranks <- rank(Z)
  r1 <- ranks[1:m]
  r2 <- ranks[m+1:N]
  r2 <- r2[1:n]

  T_phi <- sqrt(m*n/N)*(mean(phi((r1-0.5)/N)) - mean(phi((r2-0.5)/N)))
  return(T_phi)
}
```

Test Wilcoxona (prawostronny)

Test oparty jest na statystyce $T_{\varphi_1}^2$, gdzie $\varphi_1(u) = \sqrt{3}(2u - 1)$. Funkcja φ_1 spełnia:

$$\int_0^1 \varphi_1(u) du = 0, \quad \int_0^1 \varphi_1^2(u) du = 1,$$

zatem statystyka T_{φ_1} ma asymptotyczny rozkład $N(0, 1)$ przy prawdziwości hipotezy zerowej. Stąd statystyka $T_{\varphi_1}^2$ ma asymptotyczny rozkład $\chi^2(1)$.

```
phi_1 <- function(u){return(sqrt(3)*(2*u-1))}

W <- function(X, Y){
  T_phi <- T_phi(X, Y, phi_1)
  return(T_phi^2)
}
```

Test Ansari-Bredley'a (prawostronny)

Test oparty jest na statystyce $T_{\varphi_2}^2$, gdzie $\varphi_2(u) = \sqrt{48}(0.25 - |u - 0.5|)$. Funkcja φ_2 spełnia:

$$\int_0^1 \varphi_2(u)du = 0, \quad \int_0^1 \varphi_2^2(u)du = 1,$$

zatem statystyka T_{φ_2} ma asymptotyczny rozkład $N(0, 1)$ przy prawdziwości hipotezy zerowej. Stąd statystyka $T_{\varphi_2}^2$ ma asymptotyczny rozkład $\chi^2(1)$.

```
phi_2 <- function(u){return(sqrt(48)*(0.25 - abs(u-0.5))))}

AB <- function(X, Y){
  T_phi <- T_phi(X, Y, phi_2)
  return(T_phi^2)
}
```

Test Lepage'a (prawostronny)

Test oparty jest na statystyce $L = W + AB = T_{\varphi_1}^2 + T_{\varphi_2}^2$.

Skoro $\langle \varphi_1, \varphi_2 \rangle = \int_0^1 \varphi_1(u)\varphi_2(u)du = 0$, to $\varphi_1 \perp \varphi_2$. Statystyka L ma więc asymptotyczny rozkład $\chi^2(2)$.

```
L <- function(X, Y){
  L <- T_phi(X, Y, phi_1)^2 + T_phi(X, Y, phi_2)^2
  return(L)
}
```

Test Kołmogorowa-Smirnowa (prawostronny)

Test odrzuca hipotezę zerową dla dużych wartości statystyki

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

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

Dla prób odpowiednio dużych rozmiarów wartość krytyczna wyraża się wzorem

$$c(\alpha) = \sqrt{-\ln\left(\frac{\alpha}{2}\right) \cdot \frac{1}{2}}.$$

```
distribution <- function(a, X){
  return(mean(X <= a))
}

KS <- function(X, Y){
  f <- function(a){return(distribution(a, X))}
  g <- function(a){return(distribution(a, Y))}
}
```

```

m <- length(X)
n <- length(Y)
N <- n + m
Z <- c(X,Y)

KS <- sqrt(m*n/N)*max(sapply(Z, function(x) abs(f(x) - g(x))))
return(KS)
}

```

Celem raportu będzie **zbadanie zachowania funkcji mocy** problemu testowego

$$H_0 : F = G \text{ vs } H_1 : F \neq G$$

dla powyższych testów (W , AB , L , KS), po przeanalizowaniu wyników dla rozkładu normalnego, logistycznego oraz rozkładu Cauchy'ego.

Zadanie 1

Na podstawie wartości testów W , AB , L , KS otrzymanych w 10 000 powtórzeń eksperymentu wyznaczmy **wartości krytyczne dla testów prawostronnych** na poziomie istotności $\alpha = 0.05$.

```

ex_1 <- function(m, n){
  X <- rnorm(m)
  Y <- rnorm(n)
  W <- T_phi(X, Y, phi_1)^2
  AB <- T_phi(X, Y, phi_2)^2
  L <- W + AB
  KS <- KS(X, Y)
  return(c(W, AB, L, KS))
}

M <- 10000
m <- 20
n <- 20
alpha <- 0.05

set.seed(1)
results1 <- replicate(M, ex_1(m,n))
W20 <- sort(results1[,1])
qW <- W20[floor(0.95*M)]
diff_W <- qW - qchisq(1-alpha, 1)

AB20 <- sort(results1[,2])
qAB <- AB20[floor(0.95*M)]
diff_AB <- qAB - qchisq(1-alpha, 1)

L20 <- sort(results1[,3])
qL <- L20[floor(0.95*M)]
diff_L <- qL - qchisq(1-alpha, 2)

KS20 <- sort(results1[,4])
qKS <- KS20[floor(0.95*M)]
diff_KS <- qKS - sqrt(-log(alpha/2)/2)

```

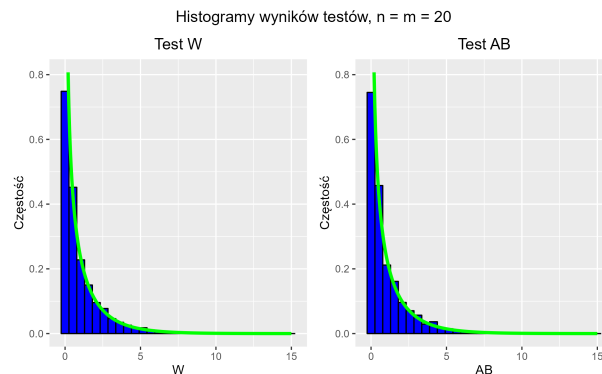
```
q_estimators <- c(qW, qAB, qL, qKS)
teoretic <- c(qchisq(1-alpha, 1), qchisq(1-alpha, 1), qchisq(1-alpha, 2), sqrt(-log(alpha/2)/2))
diffs <- c(diff_W, diff_AB, diff_L, diff_KS)
```

	Asymptotyczny kwantyl teoretyczny	Kwantyl empiryczny, n=20	Różnica, n=20	Kwantyl empiryczny, n=50	Różnica, n=50
W	3.841459	3.888000	0.0465412	3.844272	0.0028132
AB	3.841459	3.888000	0.0465412	3.926208	0.0847492
L	5.991465	6.027000	0.0355355	5.954160	-0.0373045
KS	1.358102	1.264911	-0.0931905	1.300000	-0.0581015

Analiza powyższych wyników pokazuje, że:

- dla testów W oraz AB otrzymane wartości kwantyli empirycznych zarówno dla prób rozmiaru 20 jak i 50 są bliskie asymptotycznemu kwantylowi pochodzącemu z rozkładu $\chi^2(1)$.
- Test L wykazuje się również dużą zgodnością z kwantylem asymptotycznym (rozkład $\chi^2(2)$) dla prób obu rozmiarów.
- Kwantyle empiryczne dla testu KS również nie odbiegają znacząco od kwantylu asymptotycznego. Dla tego testu wyraźniejsza jest poprawa dokładności wraz ze wzrostem rozmiaru próby.

Sprawdźmy, czy rzeczywiście rozkłady wyników testów W i AB mają w przybliżeniu (asymptotyczny) rozkład $\chi^2(1)$ już dla prób rozmiarów $n = m = 20$.



Powyższe histogramy jasno obrazują, jak blisko rozkładu $\chi^2(1)$ (zielona krzywa - krzywa gęstości tego rozkładu) są wyniki dla testów W oraz AB już dla prób rozmiarów $n = m = 20$.

Mimo tego w kolejnych obliczeniach wykorzystamy wyznaczone powyżej **kwantyle empiryczne**.

Zadanie 2

Naszym zadaniem jest wyestymowanie mocy testów W , AB , L , KS dla prób pochodzących z rozkładów: normalnego, logistycznego i rozkładu Cauchy'ego w zależności od **parametru** μ_2 (ustalmy, że parametr $\sigma = \sigma_1 = \sigma_2 = 1$ oraz $\mu_1 = 0$ w każdym podpunkcie). Liczba powtórzeń eksperymentu $M = 10\,000$.

```
p <- function(n1, mu1, sigma1, n2, mu2, sigma2, alpha, distr, vec_q){
  if(distr=="norm"){
    X <- rnorm(n1, mu1, sigma1)
    Y <- rnorm(n2, mu2, sigma2)
  }
}
```

```

if(distr=="logis"){
  X <- rlogis(n1, mu1, sigma1)
  Y <- rlogis(n2, mu2, sigma2)
}

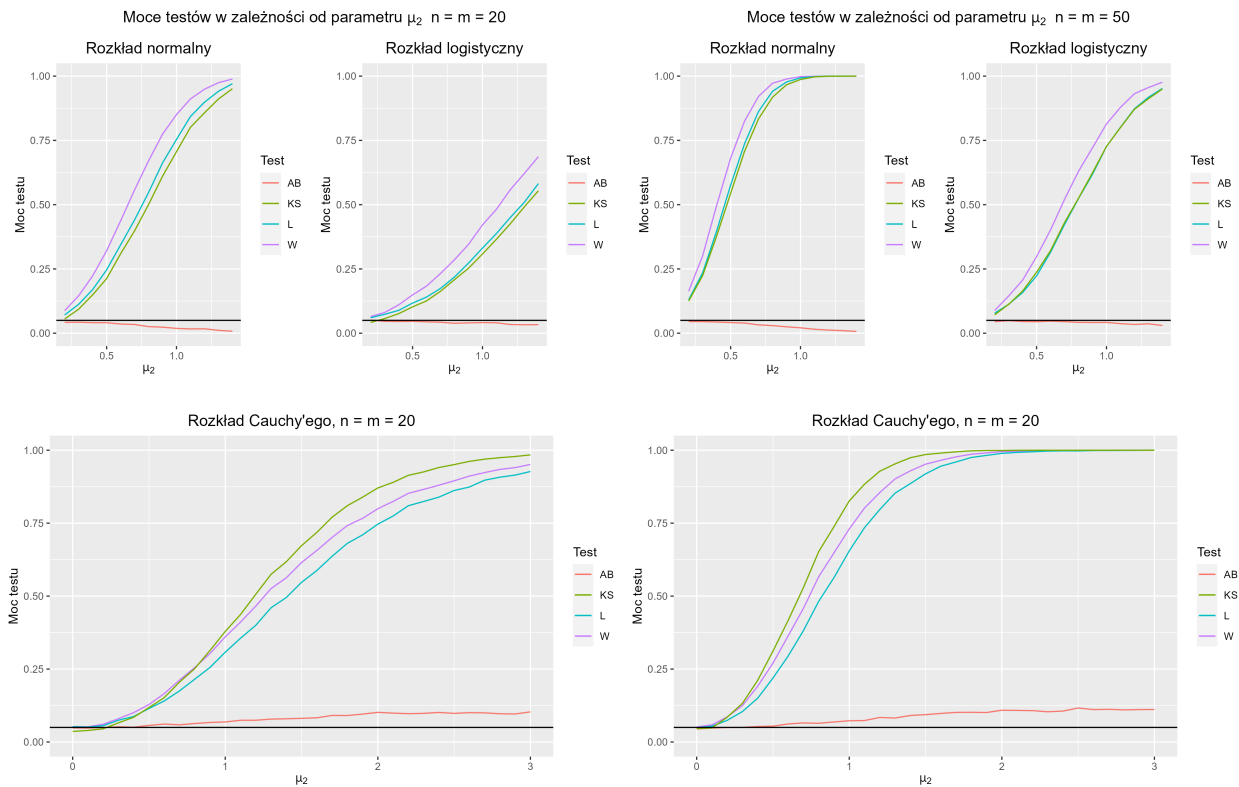
if(distr=="cauchy"){
  X <- rcauchy(n1, mu1, sigma1)
  Y <- rcauchy(n2, mu2, sigma2)
}

W <- W(X,Y)
AB <- AB(X,Y)
L <- W + AB
KS <- KS(X, Y)

statistics <- c(W, AB, L, KS)

return(statistics>vec_q)
}

```



Analiza powyższych wykresów pozwala wysunąć następujące wnioski:

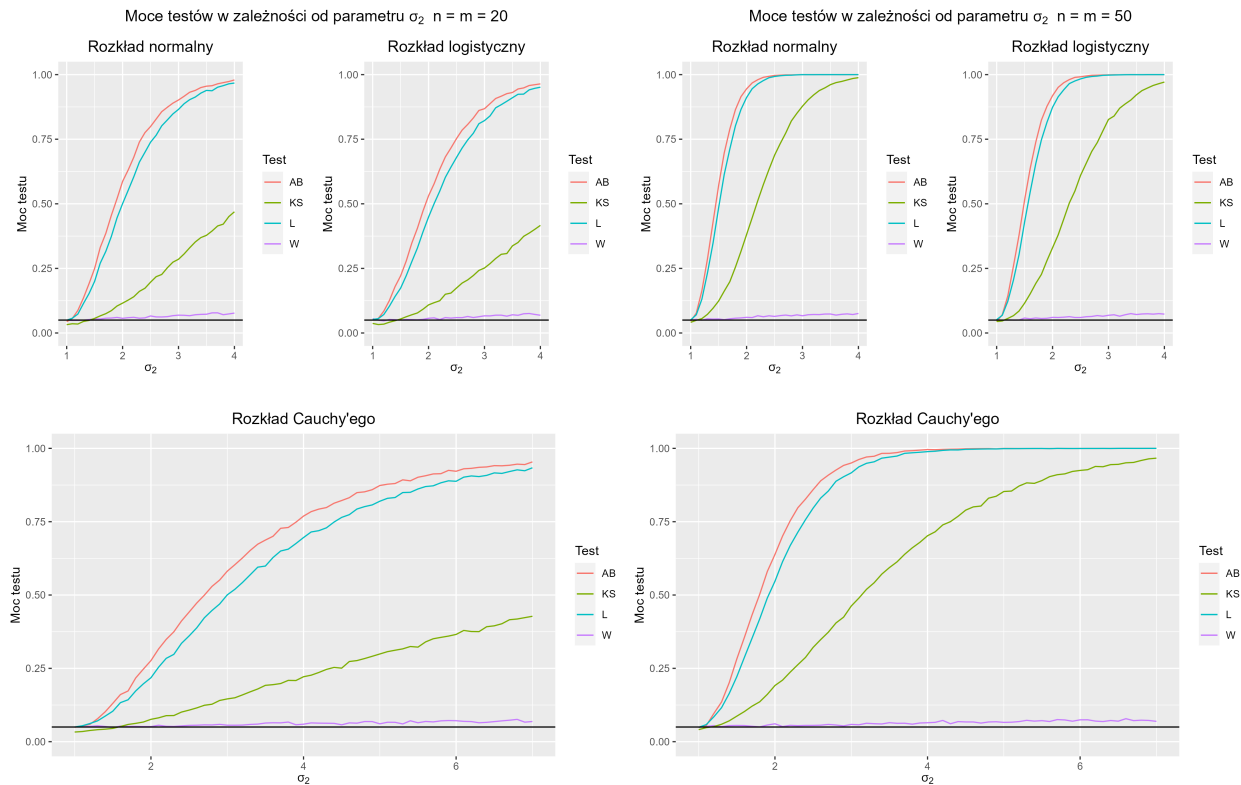
- testy W , L , KS zachowują się w tym przypadku (tzn. zależności mocy od parametru przesunięcia jednego z rozkładów) w sposób, którego oczekujemy - moc rośnie od wartości równej w przybliżeniu poziomowi istotności $\alpha = 0.5$ do poziomu 1 wraz ze wzrostem μ_2 , czyli wraz ze wzrostem różnicy pomiędzy rozkładami badanych zmiennych.
- Test AB nie reaguje w spodziewany sposób na zmianę parametru μ_2 - test ten nie rozpoznaje różnicy pomiędzy badanymi rozkładami. Przyjmuje wartości równe w przybliżeniu 0 bez względu na zmianę

μ_2 . Dla rozkładów normalnego i logistycznego moc zmniejsza się wraz ze wzrostem μ_2 . Otrzymane wyniki świadczą o tym, że test AB jest nieczuły na zmianę parametru przesunięcia. Być może lepiej wykrywa zmianę parametru skali.

- Zwiększenie wielkości prób z $n = m = 20$ na $n = m = 50$ spowodowało osiągnięcie pełni mocy testów przy mniejszej różnicy parametrów μ_1, μ_2 , tzn. moc równa 1 jest osiągana dla mniejszych wartości μ_2 dla prób rozmiarów 50.
- Spośród badanych rozkładów pełną moc najszybciej osiąga rozkład normalny (dla najmniejszej wartości μ_2), natomiast najpóźniej rozkład Cauchy'ego.
- W przypadku rozkładu normalnego i logistycznego największą moc posiada test W , tzn. dla danego μ_2 najsilniej rozpoznaje różnicę pomiędzy badanymi rozkładami.
- Dla rozkładu Cauchy'ego największą moc osiąga z kolei test KS .

Zadanie 3

Naszym zadaniem jest wyestymowanie mocy testów W, AB, L, KS dla prób pochodzących z rozkładów: normalnego, logistycznego i rozkładu Cauchy'ego w zależności od **parametru** σ_2 (ustalmy, że parametr $\sigma_1 = 1$ oraz $\mu_1 = \mu_2 = 0$ w każdym podpunkcie). Liczba powtórzeń eksperymentu $M = 10\,000$.



Analiza powyższych wykresów pozwala wysunąć następujące wnioski:

- W przypadku analizy mocy testów w zależności od zmiany parametru skali σ_2 jedynie test W okazał się nieczuły na zmianę tego parametru (tak jak test AB na zmianę parametru przesunięcia μ_2). Moc testu W wynosi w przybliżeniu 0 bez względu na wartość parametru σ_2 .
- Pozostałe testy zachowują się w pożądanym sposób - moc testu rośnie wraz ze wzrostem parametru μ_2 (czyli wraz ze wzrostem różnicy pomiędzy badanymi rozkładami) od wartości równej w przybliżeniu $\alpha = 0.5$ do pełnej mocy (równiej 1).

- Pełna moc testów jest oczywiście osiągana dla prób rozmiarów $n = m = 50$.
- Największą moc dla wszystkich trzech badanych rozkładów osiąga test AB .
- Poza testem W nieczułego na zmianę parametru skali najmniejszą moc osiąga w tym przypadku test KS .

Wniosek: optymalnym testem okazuje się test L - osiąga wysokie wartości mocy zarówno w przypadku zmiany parametru skali jak i przesunięcia. Nie jest to zaskakujące, gdyż test L jest sumą testów W i AB , które dobrze wykrywają zmianę parametru skali (test AB) i parametru przesunięcia (test W), ale są nieczułe na zmianę drugiego parametru. Zastosowanie testu KS również niesie informację o różnicach pomiędzy rozkładami. Test ten osiąga jednak mniejszą moc w przypadku zmiany parametru skali (moc ta rośnie wolniej wraz ze wzrostem σ_2 , ale nie jest zdegenerowana).

Zadanie 4

Naszym zadaniem jest zbadanie mocy danych testów w zależności od wektora parametrów (μ_2, σ_2) . Wykonajmy 1000 powtórzeń eksperymentu polegającego na obliczeniu mocy testów W , AB , L , KS dla rozkładów normalnego i logistycznego z parametrami $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 \in [0.2, 1.4]$, $\sigma_2 \in [1, 4]$ oraz rozkładu Cauchy'ego z parametrami $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 \in [0, 3]$, $\sigma_2 \in [1, 7]$.

```
ex4 <- function(mu2, sigma2){

  if(distr=="norm"){
    X <- rnorm(n1, mu1, sigma1)
    Y <- rnorm(n2, mu2, sigma2)
  }

  if(distr=="logis"){
    X <- rlogis(n1, mu1, sigma1)
    Y <- rlogis(n2, mu2, sigma2)
  }

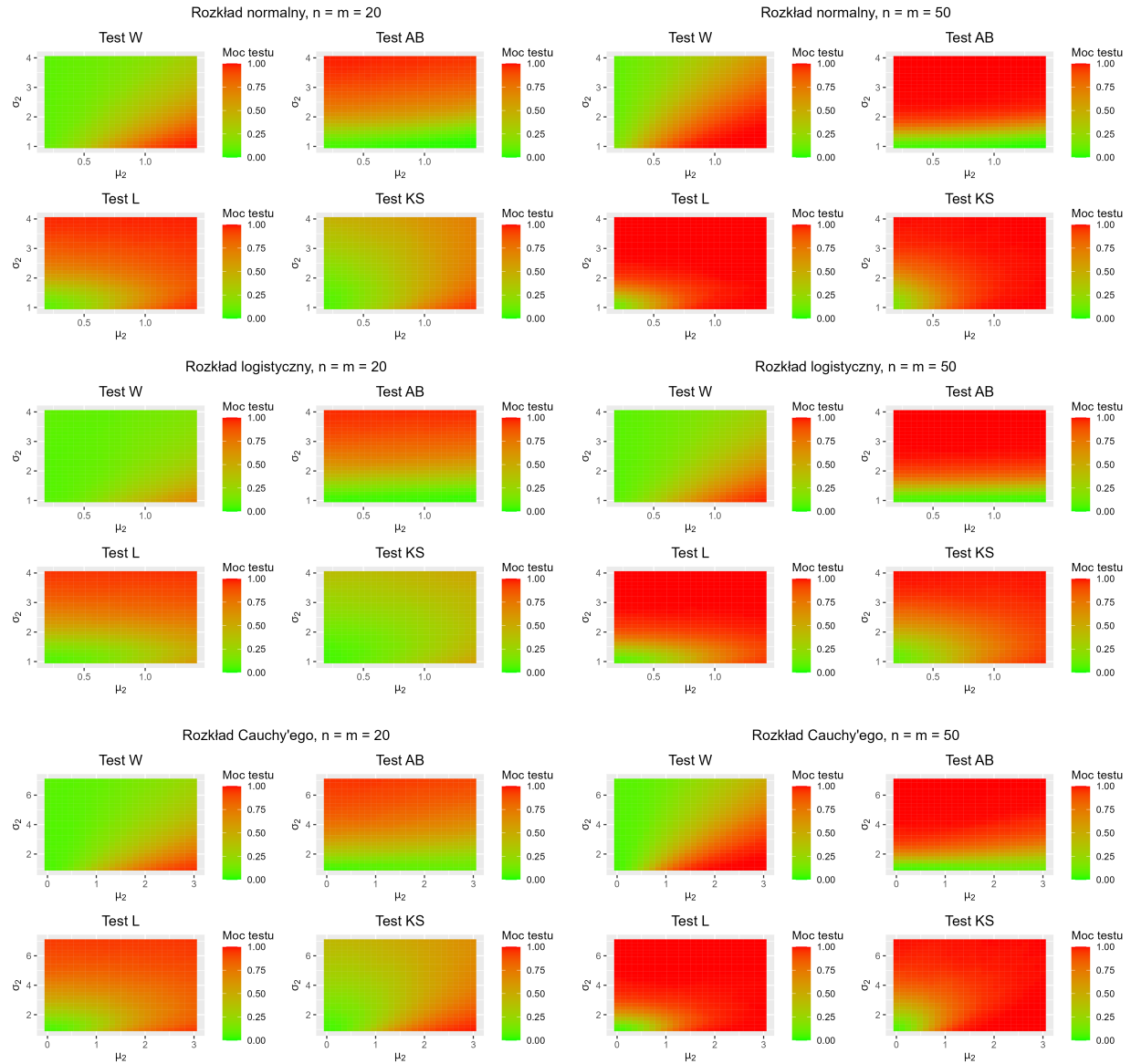
  if(distr=="cauchy"){
    X <- rcauchy(n1, mu1, sigma1)
    Y <- rcauchy(n2, mu2, sigma2)
  }

  if(tt == "W") return(W(X,Y)>vec_q[1])
  if(tt == "AB") return(AB(X,Y)>vec_q[2])
  if(tt == "L") return(W(X,Y)+AB(X,Y)>vec_q[3])
  if(tt == "KS") return(KS(X, Y)>vec_q[4])
}

ex4_p <- function(mu2, sigma2){
  set.seed(1)
  results <- replicate(M, ex4(mu2, sigma2))
  stat <- mean(results)

  return(stat)
}

vec_ex4_p <- Vectorize(ex4_p)
```



Analiza powyższych wykresów pokazuje, że:

- Zwiększenie wielkości prób z $n = m = 20$ na $n = m = 50$ znacząco wpłynęło na zwiększenie mocy testów. W przeciwieństwie do wykresów dla prób mniejszych rozmiarów pełna moc przy $n = m = 50$ jest osiągnięta dla prawie wszystkich parametrów (μ_2, σ_2) (kolor czerwony).
- Wszystkie trzy badane rozkłady charakteryzują się analogicznymi kształtami rozkładów mocy dla danych testów. Dla rozkładu normalnego uzyskujemy pełną moc najszybciej (tzn. dla najmniejszych wartości parametrów (μ_2, σ_2)). Z kolei rozkład Cauchy'ego charakteryzuje się najwolniejszym wzrostem funkcji mocy testów (skale parametrów są większe dla tego rozkładu w celu zaobserwowania pełni mocy).
- Test W okazuje się rzeczywiście testem, który nie wykrywa zmian parametru skali σ_2 . Moc tego testu wzrasta tylko wraz ze wzrostem parametru przesunięcia μ_2 .
- Z kolei test AB jest rzeczywiście nieczuły na zmianę parametru przesunięcia. Moc tego testu wzrasta tylko wraz ze wzrostem parametru skali.
- Moc testu KS wzrasta proporcjonalnie dla obu parametrów - test wykrywa zarówno zmiany parametru przesunięcia jak i skali. Dla prób rozmiarów $n = m = 20$ moc tego testu jest jednak niewielka. Dopiero

próby większych rozmiarów powodują dość szybkie uzyskanie pełnej mocy dla tego testu. Duża zależność mocy testu KS od rozmiarów prób sprawia, że test ten nie jest optymalny.

- Najbardziej optymalnym testem (zgodnie z wcześniejszymi obserwacjami) okazuje się test

$$L = W + AB.$$

Test ten wykrywa zarówno zmiany parametru przesunięcia jak i skali (i to w równomiernym stopniu). Jest to spowodowane budową testu - suma dwóch testów, z których każdy odpowiada za wykrycie tylko jednego z parametrów. Test L charakteryzuje się również dużą mocą osiąganą nawet przy próbach rozmiarów $n = m = 20$. Z tych powodów otrzymujemy następujący wniosek: **test L jest najbardziej optymalnym spośród badanych testów** - charakteryzuje się dużą mocą i wykrywa różnice pomiędzy rozkładami zarówno dla różnic parametru przesunięcia jak i skali.