

Badanie zachowania funkcji mocy testów w zależności od wybieranych statystyk testowych

Antoni Bieniasz

2024-01-12

Wstęp

Celem tej pracy jest testowanie hipotez statystycznych o równości rozkładów za pomocą różnych statystyk testowych. Wyznaczane będą wartości krytyczne testów oraz badane będzie zachowanie funkcji mocy analizowanych testów.

Badany problem

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

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

na poziomie istotności $\alpha = 0.05$.

Konstrukcja poszczególnych statystyk testowych

Niech $N = m + n$, a $Z = (X_1, \dots, X_m; Y_1, \dots, Y_n)$ będzie wektorem łączącym próby. Dalej niech R_i będzie rangą Z_i w próbie Z , $i = 1, \dots, N$. Wtedy klasyczna liniowa funkcja rangowa związana z funkcją wynikową $\varphi \in L_2(0, 1)$ ma postać:

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

Wybór funkcji φ determinuje czułość testu opartego na T_φ . Rozważamy testy Wilcoxona, Ansari-Bradley'a, Lepage'a oraz Kolmogorowa-Smirnowa. Ten ostatni związany jest ze statystyką:

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

Wyznaczanie wartości statystyk na podstawie obserwacji z rozkładu normalnego

Wygenerujemy $m = n = 20$ z rozkładu $N(0, 1)$. Na ich podstawie obliczymy wartości statystyk W, AB, L oraz KS. Doświadczenie powtórzmy 10000 razy. Wyznamy wartości krytyczne odpowiadających im testów prawostronnych.

Z przeprowadzonej symulacji otrzymujemy, że wartości krytyczne wynoszą odpowiednio:

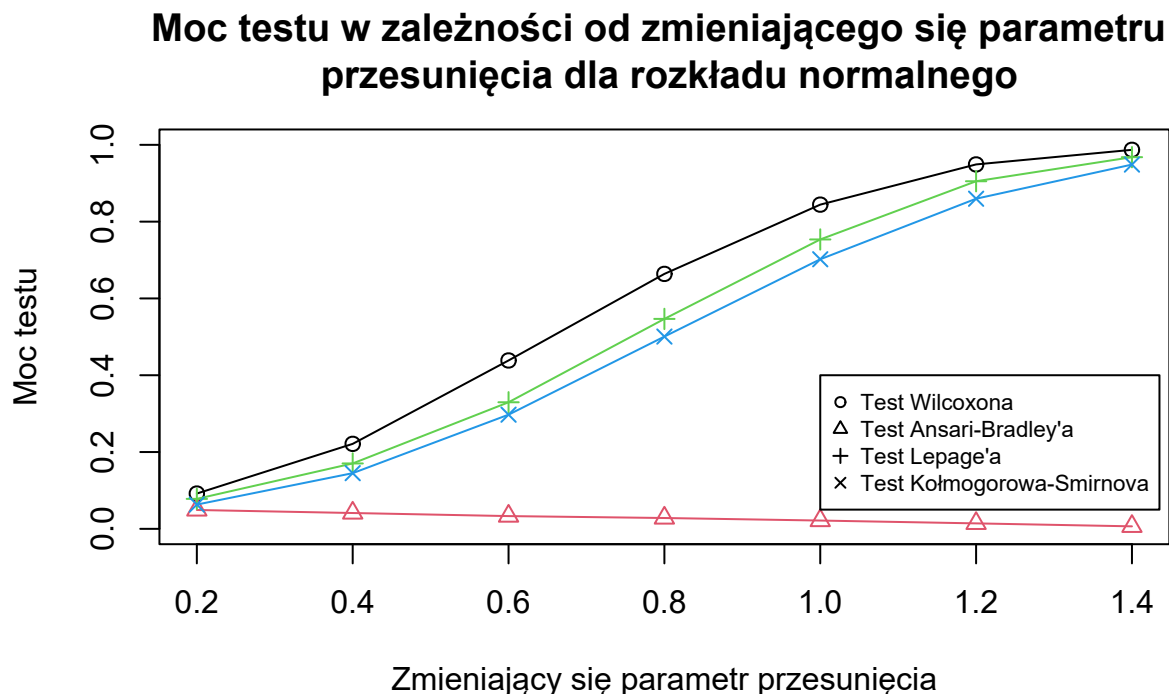
- 3.888 dla testu Wilcoxona,
- 3.888 dla testu Ansari-Bradley'a,
- 6.07575 dla testu Lepage'a,
- 1.2649 dla testu Kołmogorowa-Smirnowa.

Wiemy, że przy prawdziwości H_0 statystyki W, AB mają asymptotyczny rozkład chi-kwadrat z jednym stopniem swobody, a statystyka L ma asymptotyczny rozkład chi-kwadrat z 2 stopniami swobody. Teoretyczne wartości krytyczne wynoszą w tym wypadku 3,841 dla pierwszych dwóch z tych testów i 5,991 dla trzeciego z tych rozkładów. Zatem symulacja względnie dobrze wyznacza szukane wartości krytyczne.

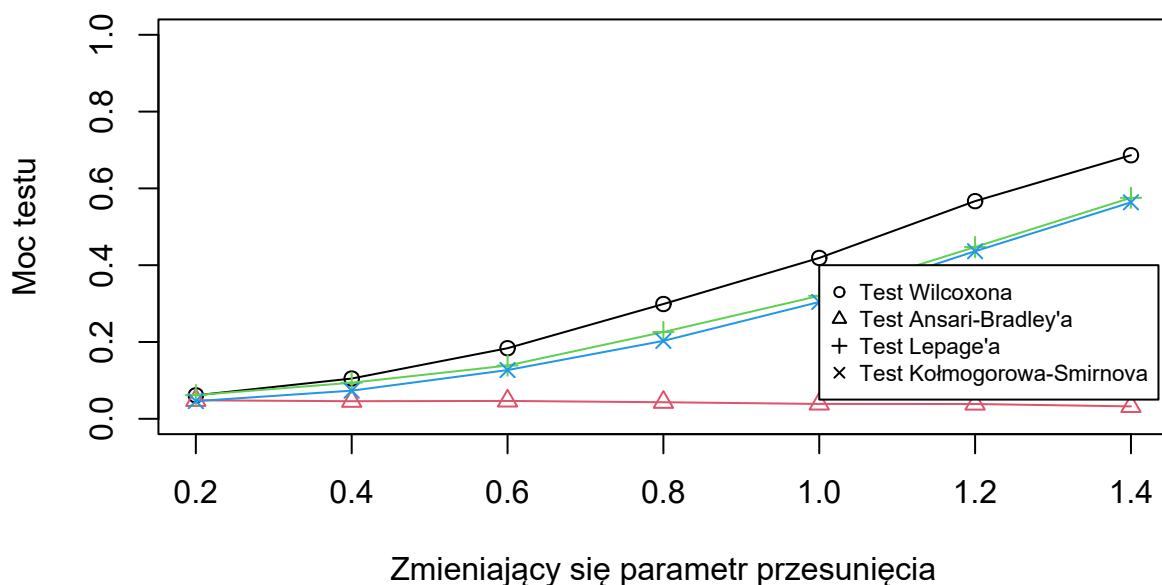
Zastanówmy się czy taki sposób generowania jest poprawny. Możemy stwierdzić, że tak patrząc na konstrukcję statystyki rangowej oraz statystyki Kołmogorowa-Smirnowa. W tej pierwszej na zmienność wpływają jedynie rangi. Ponieważ rangi możemy przypisywać obserwacjom z różnych rozkładów, nie tylko tym z rozkładu normalnego to możemy zastosować ten sposób generowania wartości krytycznych w przypadku innych rozkładów. Dla drugiej z tych statystyk bierzemy supremum po $x \in \mathbb{R}$ dla dystrybuant rozkładów F oraz G przemnożone przez skalar. Każdy rozkład ma dystrybuante, więc również jest to właściwe podejście.

Szacowanie wartości funkcji mocy analizowanych testów w zależności od zmieniającego się parametru przesunięcia

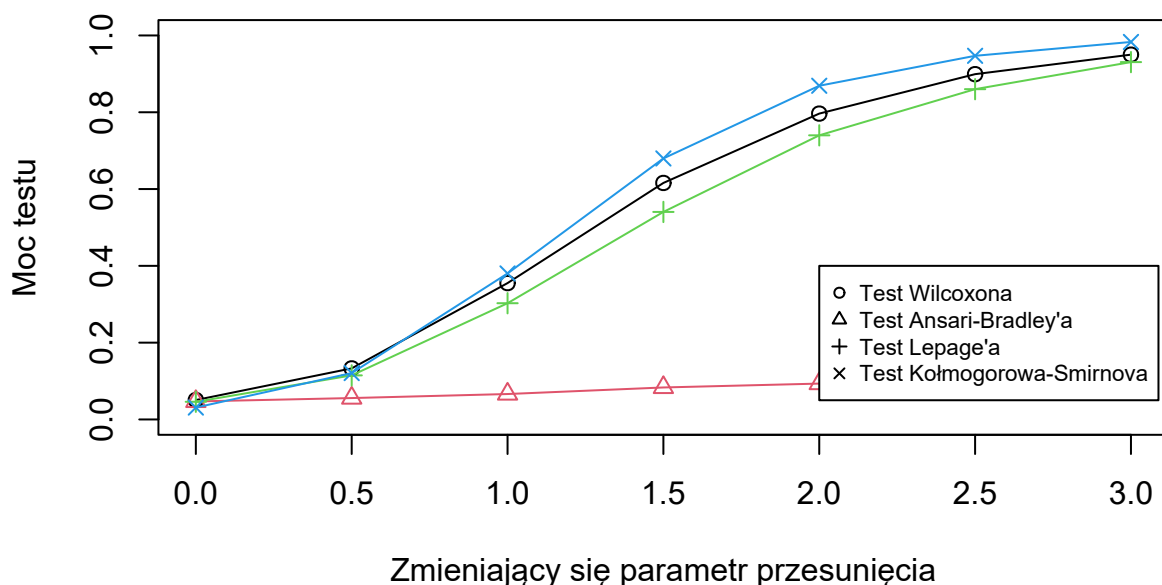
Wygenerujemy $m = n = 20$ obserwacji z rozkładu normalnego, logistycznego, Cauchy'ego z parametrami przesunięcia μ_1, μ_2 oraz skali σ_1, σ_2 , odpowiednio. Zmianom podlegać będzie parametr μ_2 (w zakresie opisanym na poniższych wykresach) Na podstawie uzyskanych obserwacji obliczymy wartość statystyki W , AB , L oraz KS . Doświadczenie powtórzymy 10000 razy. Oszacujemy wartości funkcji mocy analizowanych testów a następnie narysujemy wyestymowane funkcje mocy w zależności od parametru μ_2



Moc testu w zależności od zmieniającego się parametru przesunięcia dla rozkładu logistycznego



Moc testu w zależności od zmieniającego się parametru przesunięcia dla rozkładu Cauchy'ego

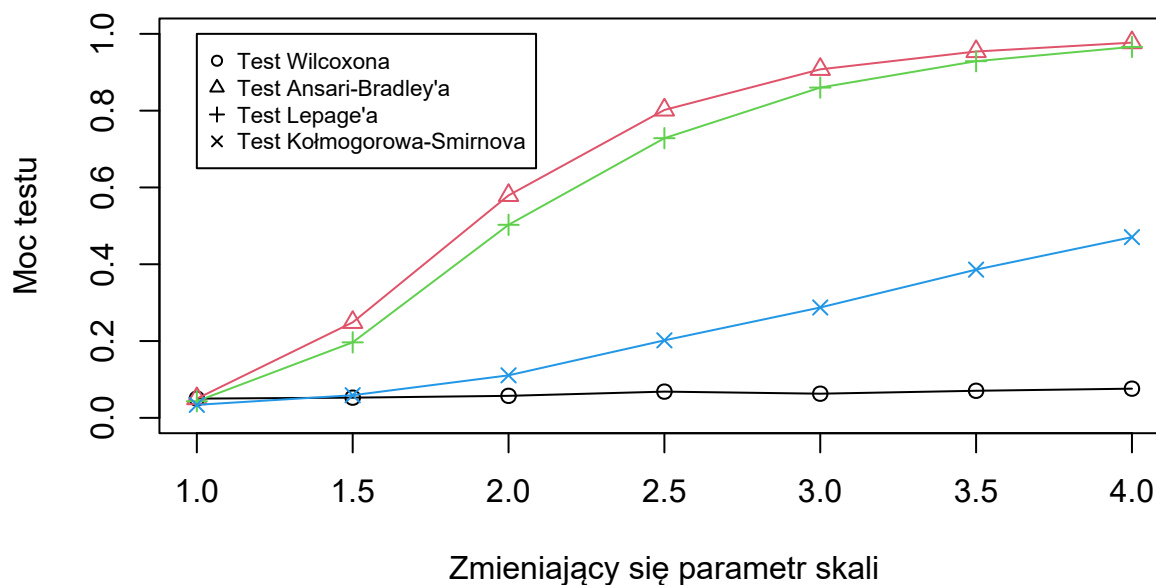


Na podstawie powyższych wykresów możemy stwierdzić, że moc powyższych testów rośnie wraz z rosnącym parametrem przesunięcia dla drugiej próby z rozkładu normalnego. Jest to spodziewany rezultat, jest coraz większa różnica między parametrami przesunięcia dla tych rozkładów zatem statystyka T_φ przyjmuje coraz większe wartości, często przekraczając tą krytyczną. Największą wartość mocy uzyskujemy ogólnie dla testu Lepage'a.

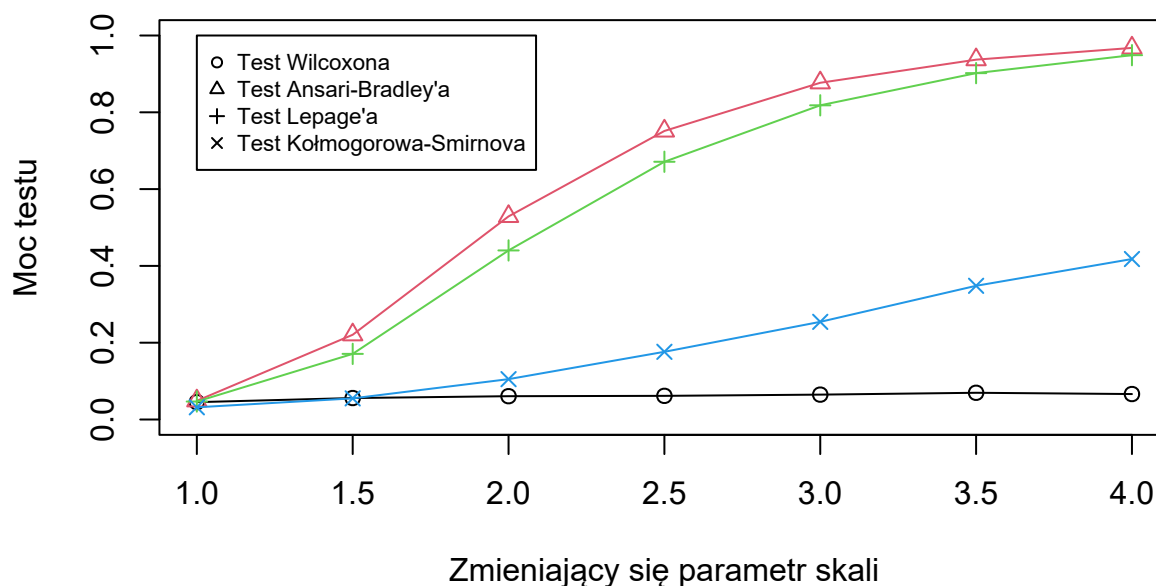
Szacowanie wartości funkcji mocy analizowanych testów w zależności od zmieniającego się parametru skali

W tej części wykonamy podobne czynności co w poprzedniej części, ale będziemy manipulować parametrem skali.

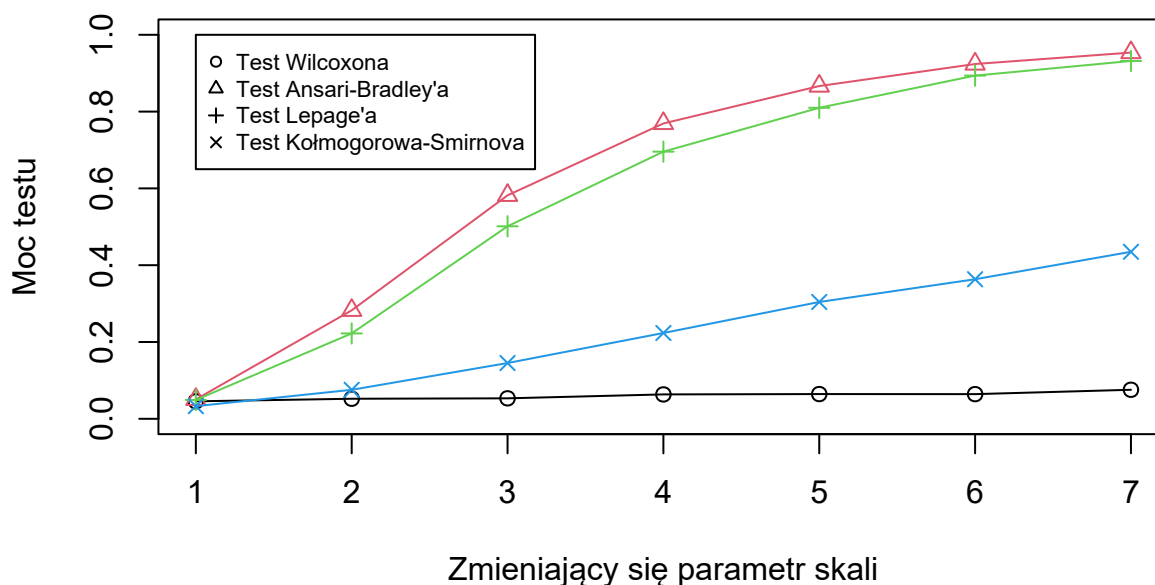
Moc testu w zależności od zmieniającego się parametru skali dla rozkładu normalnego



Moc testu w zależności od zmieniającego się parametru skali dla rozkładu logistycznego



Moc testu w zależności od zmieniającego się parametru skali dla rozkładu Cauchy'ego

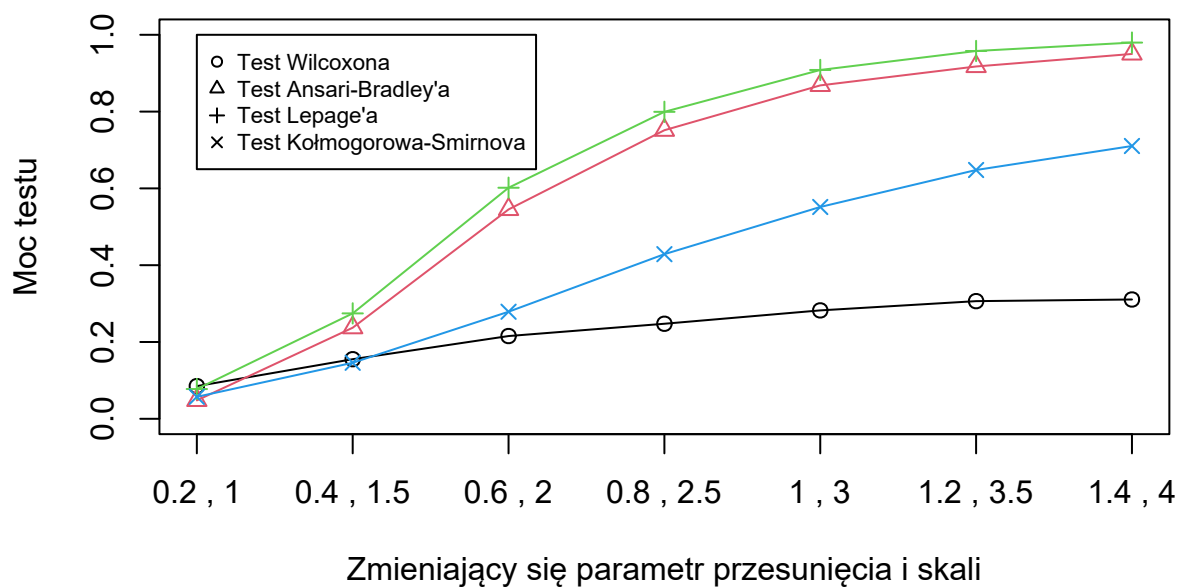


W tym przypadku moce analizowanych testów są dla każdego rozkładu znacznie mniejsze, poniżej 0.1, jedynie dla testu Kołmogorowa - Smirnowa osiągamy wartości zbliżone do 0.4. Widzimy, że wartości mocy rosną wraz ze zwiększającym się parametrem skali jednak nie są to bardzo gwałtowne przyrosty. Mniejszą częstotliwość odrzucania H_1 można tłumaczyć faktem, że rozkłady z obu prób mają te same parametry przesunięcia, więc najwięcej obserwacji dla ich obydwu pochodzi z tego samego obszaru. Większe wartości statystyki KS można tłumaczyć inną konstrukcją tej statystyki w stosunku do innych (bierzemy supremum po x , dla obydwu dystrybuant empirycznych).

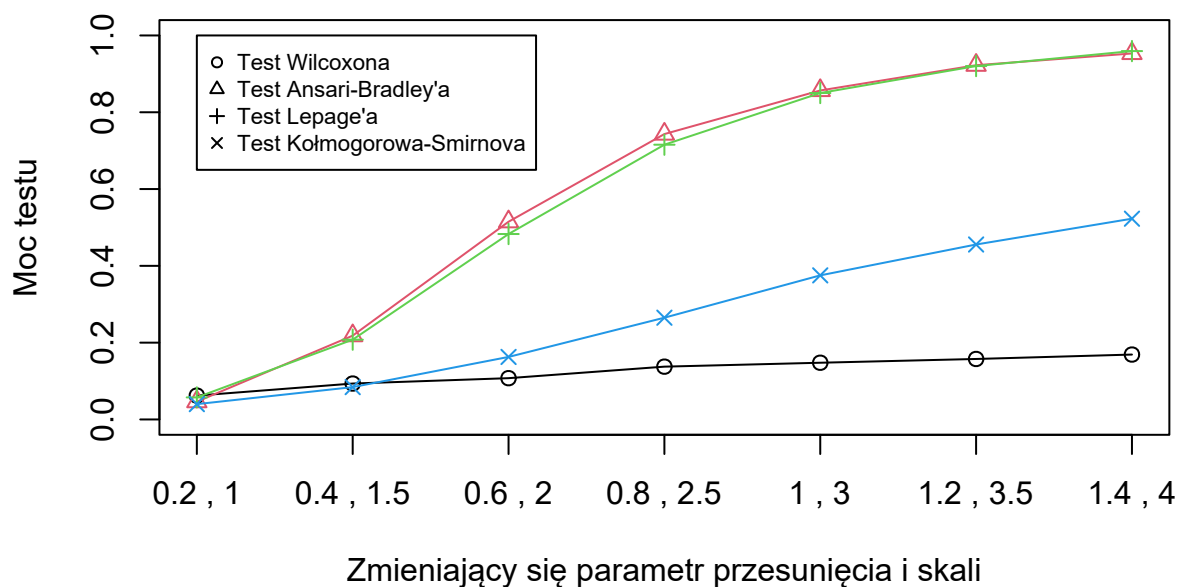
Szacowanie wartości funkcji mocy analizowanych testów w zależności od zmieniającego się naraz parametru przesunięcia oraz skali

Teraz będziemy dokonywać zmian w wartości parametru przesunięcia oraz skali jednocześnie.

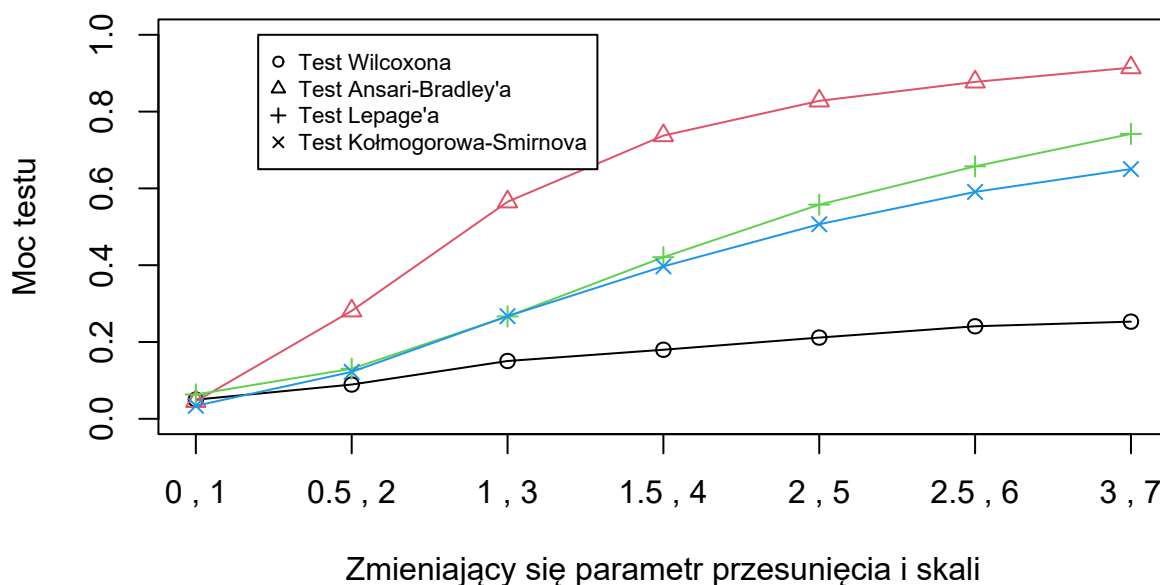
Moc testu w zależności od zmieniającego się parametru przesunięcia oraz skali dla rozkładu normalnego



Moc testu w zależności od zmieniającego się parametru przesunięcia oraz skali dla rozkładu logistycznego



Moc testu w zależności od zmieniającego się parametru przesunięcia oraz skali dla rozkładu Cauchy'ego



Przy manipulowaniu oboma parametrami uzyskujemy wartości mocy większe niż tylko dla zmiany parametru skali, ale mniejsze niż dla zmiany tylko parametru przesunięcia. Takie pośrednie wartości można tłumaczyć tym, że zwiększający się parametr przesunięcia sprawia, że najwięcej obserwacji z tej drugiej próby oddala się od obszaru, w którym jest najwięcej obserwacji dla tej pierwszej próby, ale zwiększający się parametr skali sprawia, że więcej obserwacji może znajdować się dalej od średniej rozkładu. Ponownie statystyka Kołmogorowa-Smirnowa daje największe wartości mocy.

Wyznaczanie wartości statystyk na podstawie obserwacji z rozkładu normalnego dla $n=m=50$

Wygenerujemy tym razem $m = n = 50$ obserwacji z rozkładu $N(0, 1)$. Na ich podstawie obliczymy wartości statystyk W, AB, L oraz KS. Doświadczenie powtórzymy 10000 razy. Wyznamy wartości krytyczne odpowiadających im testów prawostronnych.

Z przeprowadzonej symulacji otrzymujemy, że wartości krytyczne wynoszą odpowiednio:

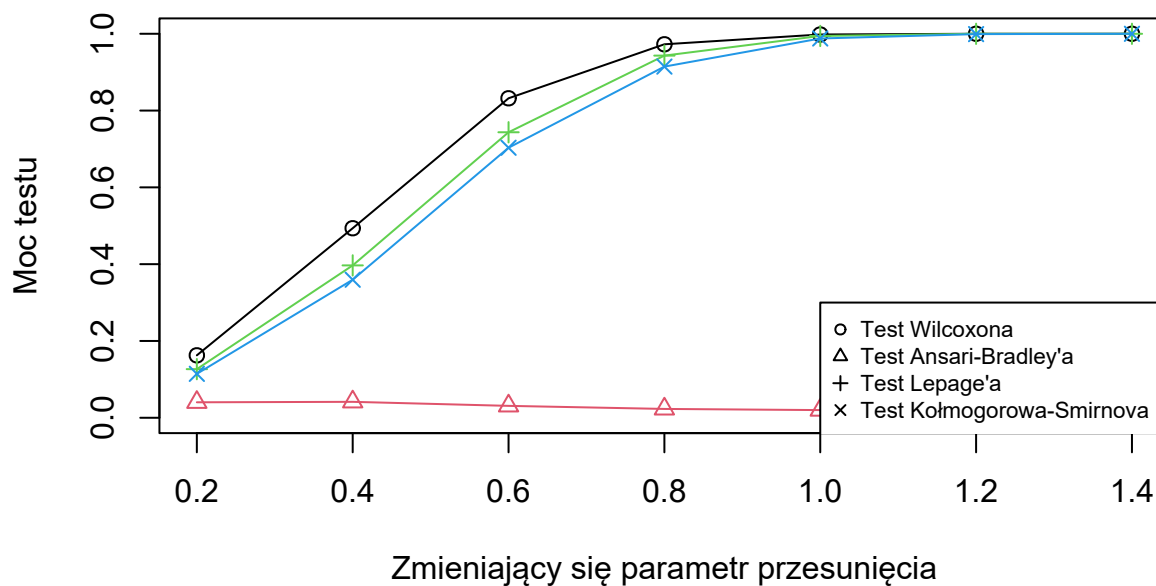
- 3.844272 dla testu Wilcoxona,
- 4.14893 dla testu Ansari-Bradley'a,
- 5.96165 dla testu Lepage'a,
- 1.3 dla testu Kołmogorowa-Smirnowa.

Powtórzenie poprzednich części dla $n=m=50$

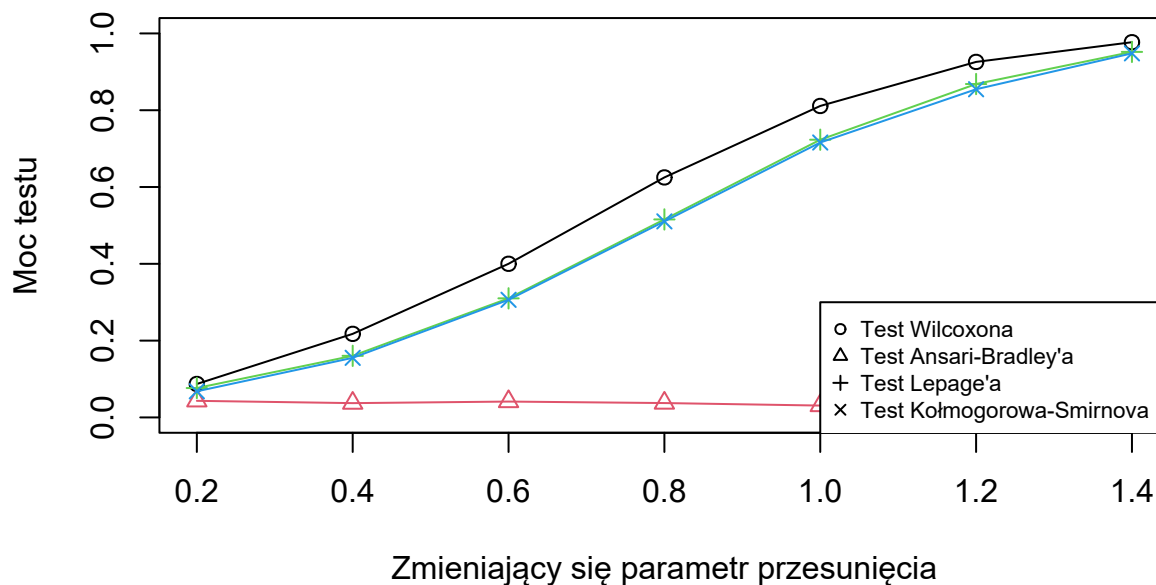
Powtórzmy teraz wykonywane wcześniej czynności dla większych prób.

Zmieniający się parametr przesunięcia

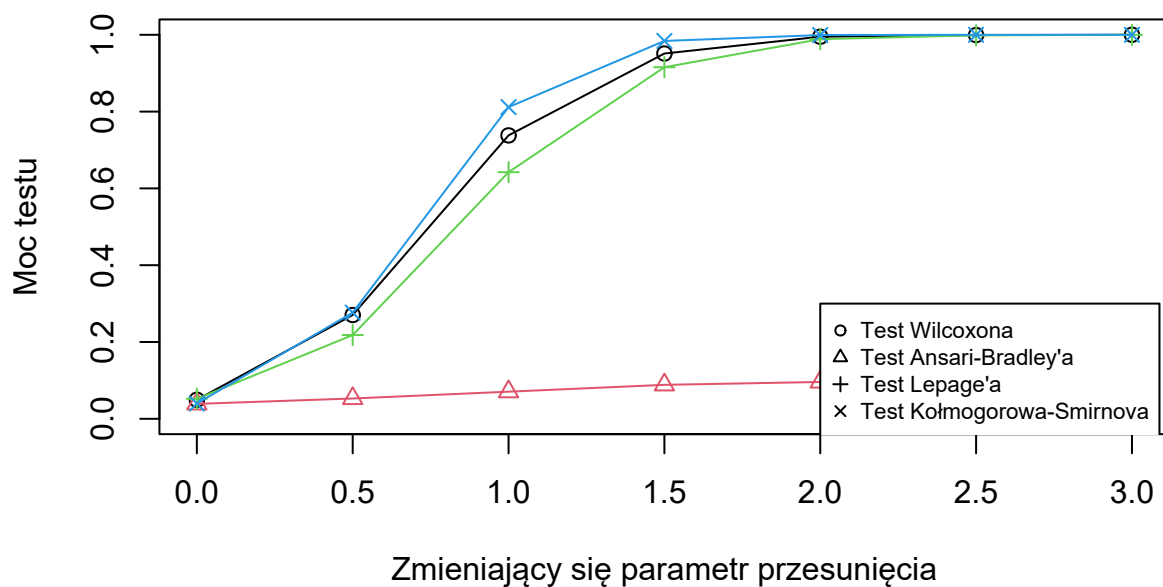
Moc testu w zależności od zmieniającego się parametru przesunięcia dla rozkładu normalnego



Moc testu w zależności od zmieniającego się parametru przesunięcia dla rozkładu logistycznego



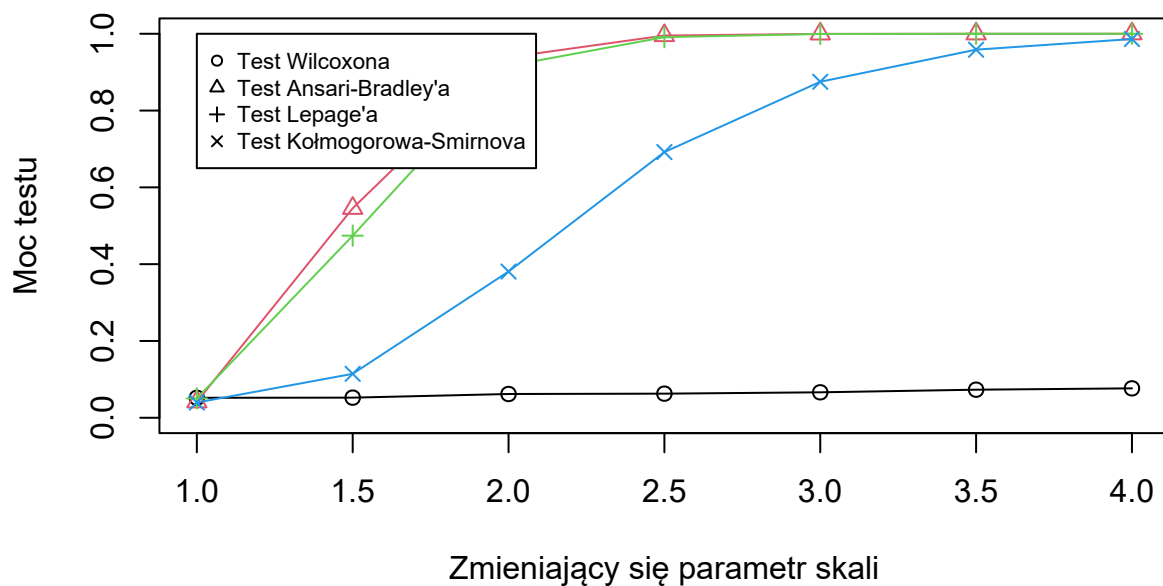
Moc testu w zależności od zmieniającego się parametru przesunięcia dla rozkładu Cauchy'ego



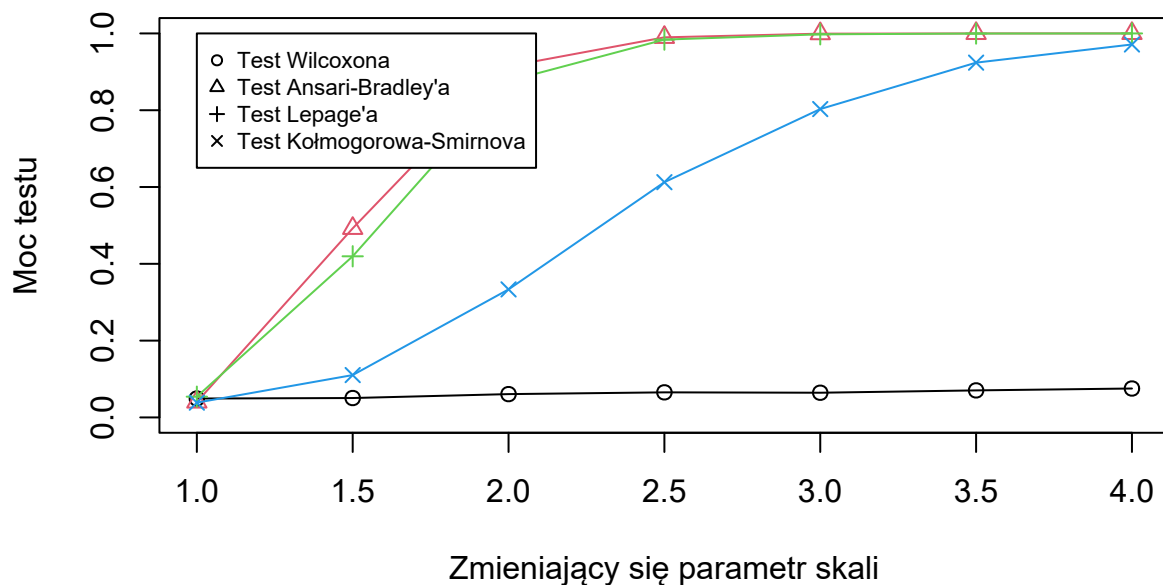
W porównaniu do $n = m = 20$ mamy większe wartości funkcji mocy dla odpowiednich parametrów przesunięcia. Jest prawdopodobnie spowodowane faktem, że dla $n = m = 50$ odpowiednia różnica w statystyce T_φ staje się większa jeżeli dane rozkłady są różne, a więc wartość T_φ rośnie.

Zmieniający się parametr skali

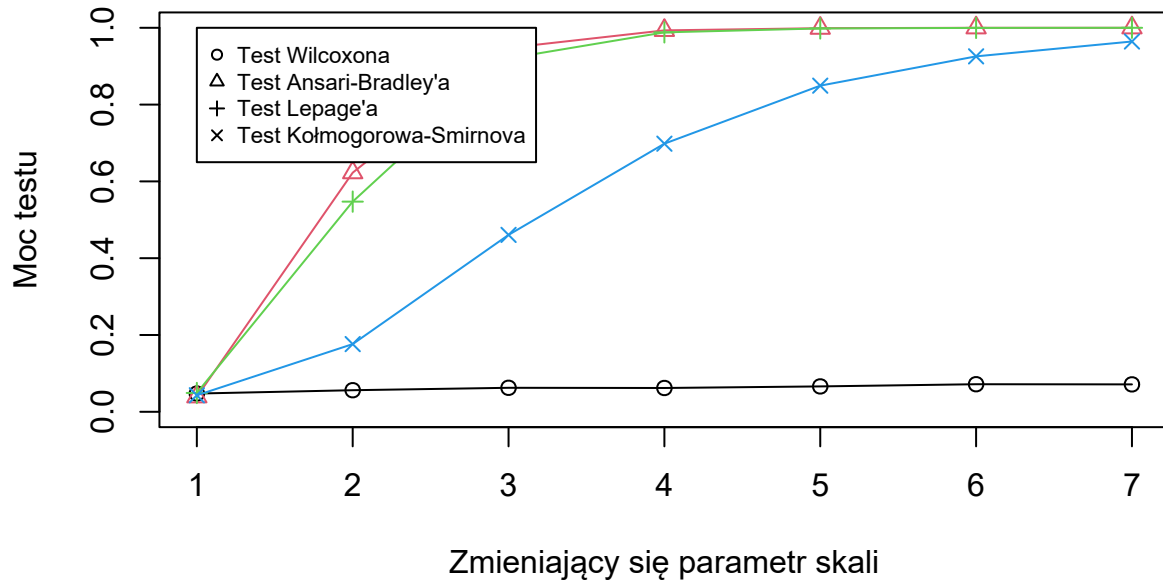
Moc testu w zależności od zmieniającego się parametru skali dla rozkładu normalnego



Moc testu w zależności od zmieniającego się parametru skali dla rozkładu logistycznego



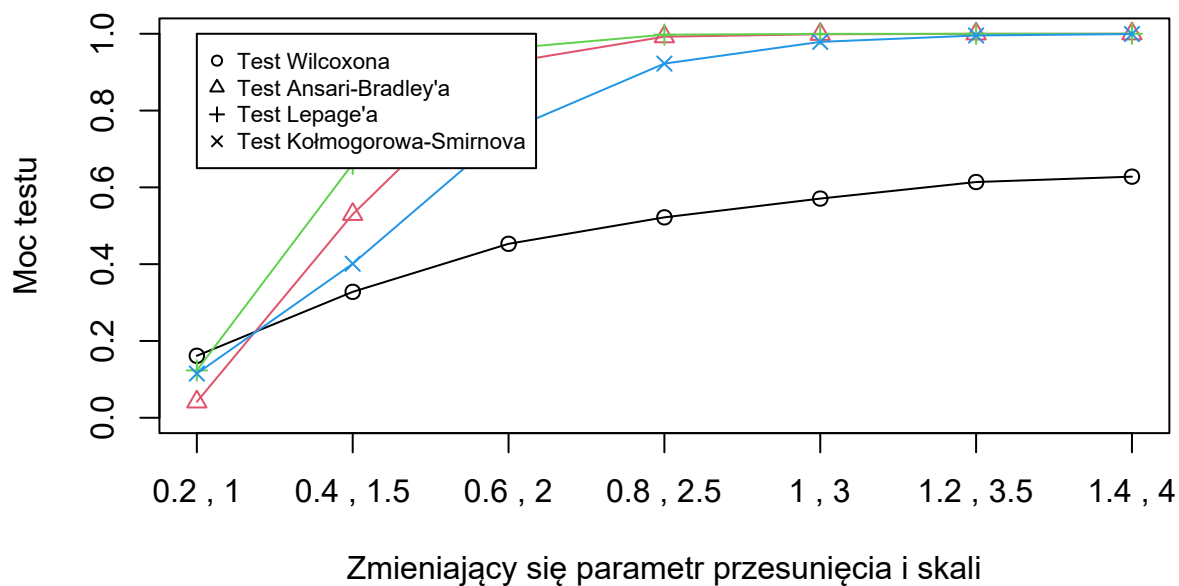
Moc testu w zależności od zmieniającego się parametru skali dla rozkładu Cauchy'ego



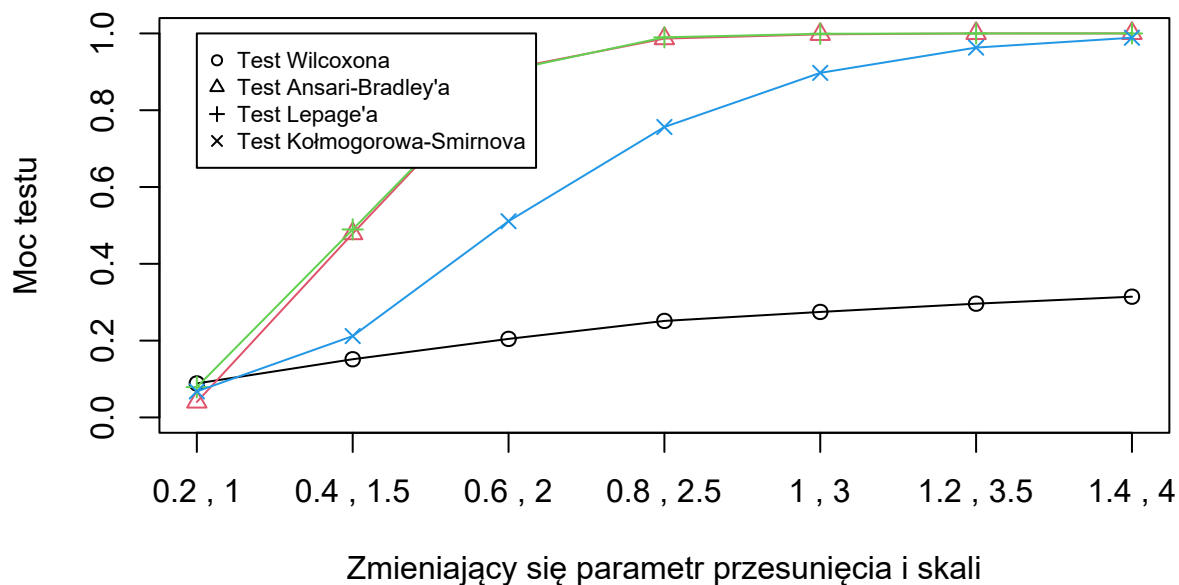
Dla zmieniającego się samego parametru skali przy zwiększeniu liczebności próby moc testu dla statystyki KS dąży do 1 przy coraz większych wartościach σ_2 . Dla innych testów utrzymuje się na poziomie niewiele większym od zera. Widać tu bardzo jak inny typ konstrukcji statystyk testowych wpływa na wyniki. Jednak w rzeczywistości wiemy, że badane rozkłady są różne, stąd możemy stwierdzić, że statystyka KS w tym przypadku jest bardzo skuteczna.

Zmieniający się parametr przesunięcia oraz skali

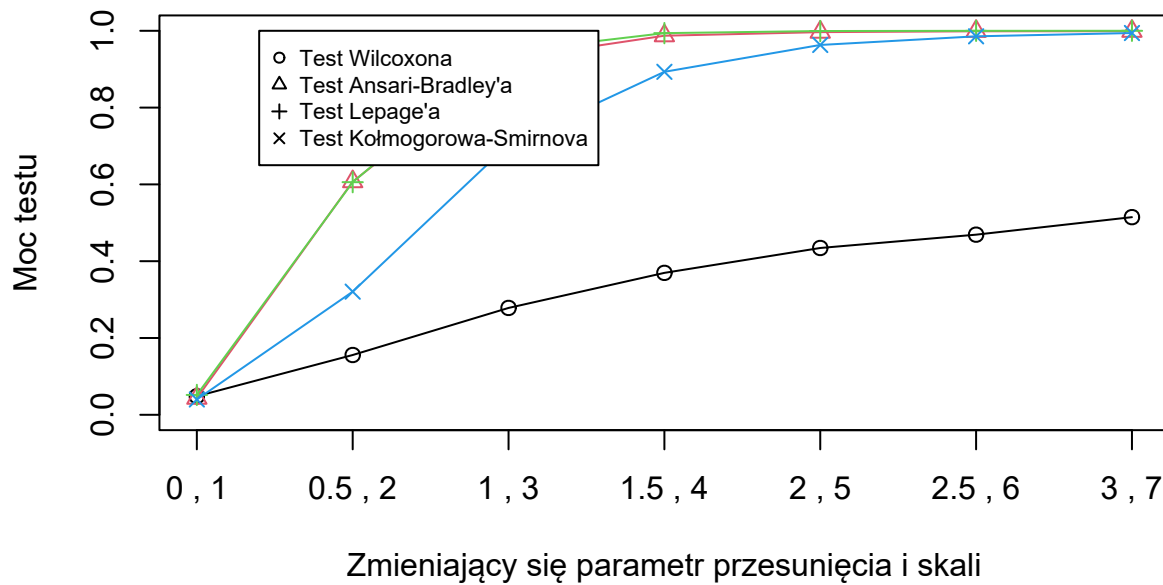
Moc testu w zależności od zmieniającego się parametru przesunięcia oraz skali dla rozkładu normalnego



Moc testu w zależności od zmieniającego się parametru przesunięcia oraz skali dla rozkładu logistycznego



Moc testu w zależności od zmieniającego się parametru przesunięcia oraz skali dla rozkładu Cauchy'ego



Przy zmianach obydwu parametrów moce testu największe są dla statystyk Kołmogorowa-Smirnowa, potem dla Lepage'a, a mniejsze dla statystyk Wilcoxona oraz Ansari-Bradley'a. Ogólnie moce testów ponownie są mniejsze niż dla zmieniającego się samego parametru przesunięcia, ale większe niż dla tylko zmieniającego się parametru skali.