Badanie funkcji mocy testu

Helena Sękowska-Słoka, nr indeksu 321531

2024-01-18

SPIS TREŚCI

Cel raportu

Celem tego raportu będzie badanie funkcji mocy testu. Będziemy testować hipotezę:

$$H_0 : F = G$$

Przeciwko:

$$H_1 = F \neq G$$

Gdzie F to dystrybuanta próby losowej $X_1, X_2, ..., X_m$, zaś G to dystrybuanta $Y_1, Y_2, ..., Y_n$, czyli drugiej próby losowej z pary. Będziemy zatem testować równość dystrybuant, a co za tym idzie - równość rozkładów.

Wyznaczenie wartości krytycznych

Zaczynamy od wygenerowania 10000 razy po dwóch prób o liczebnościach $n = m = 20$ i dwóch prób o liczebnościach $n = m = 50$. Dla każdego powtórzenia obliczamy wartości statystyk:

1. Wilcoxona (W)
2. Ansari-Bradleya (AB)
3. Lepage'a (L)
4. Kołmogorowa-Smirnova (KS)

Robimy to zgodnie ze wzorami podanymi na liście, osobno dla par o liczebności $n = 20$ i par o liczebności $n = 50$.
Następnie na tej podstawie wyznaczamy wartości krytyczne odpowiadające testom prawostronnym dla powyższych statystyk. Robimy to w oparciu o dystrybuantę empiryczną, czyli oszacowaną na podstawie przeprowadzonego eksperymentu. Za poziom istotności przyjmujemy standardowo $\alpha = 0.05$. Wartości krytyczne będą zatem kwantylami rzędu 0.95 z rozkładów empirycznych.
Taki sposób generowania wartości krytycznych jest poprawny, ponieważ wartości statystyk nie zależą od rozkładu, z którego generujemy. Mamy tę pewność, ponieważ warunek całkowy jest spełniony dla zadanych funkcji φ_1, φ_2, czyli przy prawdziwości hipotezy zerowej statystyki T_φ mają rozkład asymptotyczny normalny. Co za tym idzie, statystyki W i AB będące ich kwadratami mają asymptotyczny rozkład χ^2 z jednym stopniem swobody, natomiast statystyka L, będąca sumą dwóch zmiennych z asymptotycznego rozkładu χ^2 z jednym stopniem swobody, będzie miała rozkład asymptotyczny χ^2 z dwoma stopniami swobody. Z kolei statystyka KS ma rozkład Kołmogorowa, który również nie zależy od rozkładów próbek, dzięki czemu test KS też jest nieparametryczny. Ponadto rozpatrywane rozkłady, to znaczy rozkład normalny, logistyczny i Cauchy'ego są ciągłe. Zatem postępując w wyżej opisany sposób, uzyskaliśmy wiarygodne empiryczne wartości krytyczne.

Tabela wartości krytycznych

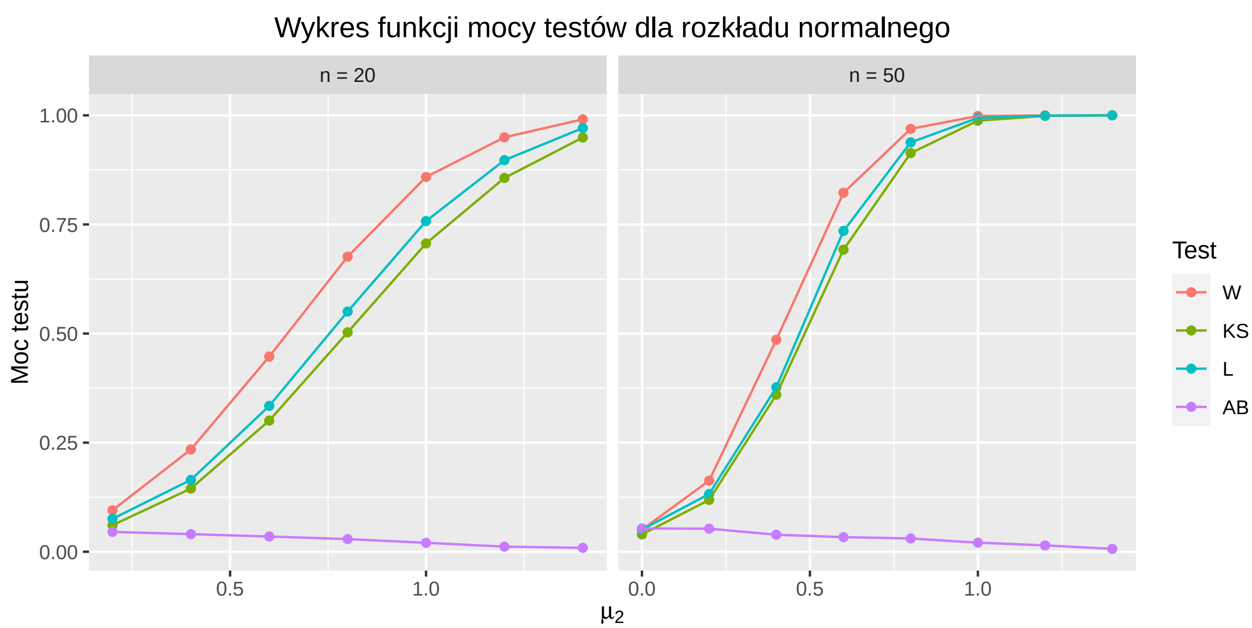
Table 1: Wartości krytyczne dla zadanych testów prawostronnych

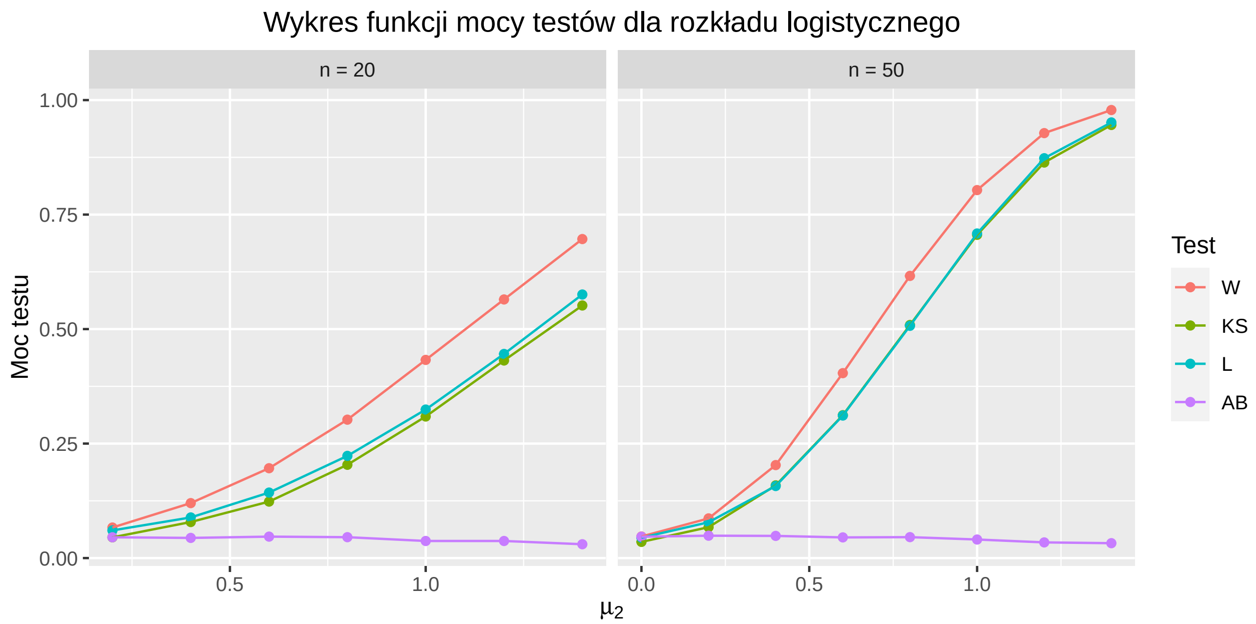
	n = 20	n = 50
test	wartość krytyczna	wartość krytyczna
W	3.7808	3.9002
AB	3.8880	3.9262
L	6.0750	6.0722
KS	1.2649	1.3000

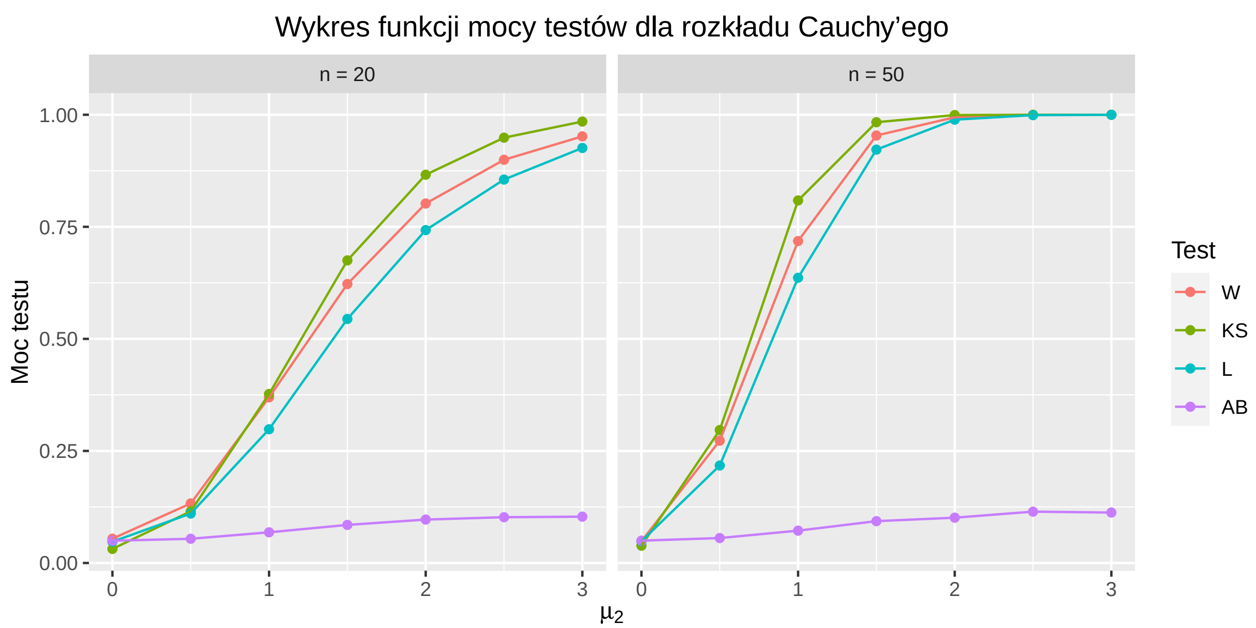
Funkcja mocy w zależności od parametru μ_2

Będziemy przybliżać funkcję mocy w zależności od parametru μ_2 empirycznie, to znaczy dla danego μ_2 będziemy losować 10000 razy próby analogiczne jak w poprzednim podpunkcie, przy czym dla każdej z nich będziemy określać moc, to znaczy sprawdzać, ile razy została odrzucona hipoteza zerowa w stosunku do liczby wszystkich par. Hipotezę tę odrzucać będziemy na podstawie wyliczonych wcześniej wartości krytycznych, to znaczy w sytuacji, gdy wartość statystyki jest większa od wartości krytycznej.
Dla $n = 20$ zakres parametru μ_2 przyjmujemy taki, jak zadany, natomiast dla $n = 50$ dobieramy go tak, żeby zobaczyć zmianę mocy w pełnym zakresie. Wobec tego dla prób pięćdziesięcioelementowych bierzemy μ_2 w zakresie od 0 do 1.4, idąc co 0.2, za wyjątkiem rozkładu Cauchy'ego, gdzie przyjmujemy je jak na liście od 0 do 3 co 0.5.
Warto jednak wspomnieć, że dla $n = 20$ w przypadku rozkładu logistycznego nie zobaczymy pełnego zakresu mocy, jednakże dla tej liczebności próby zostały na liście zadane takie, a nie inne wartości μ_2, zatem zostawiamy je bez zmian, żeby wykonać zadanie zgodnie z poleceniem.

Wykresy funkcji mocy







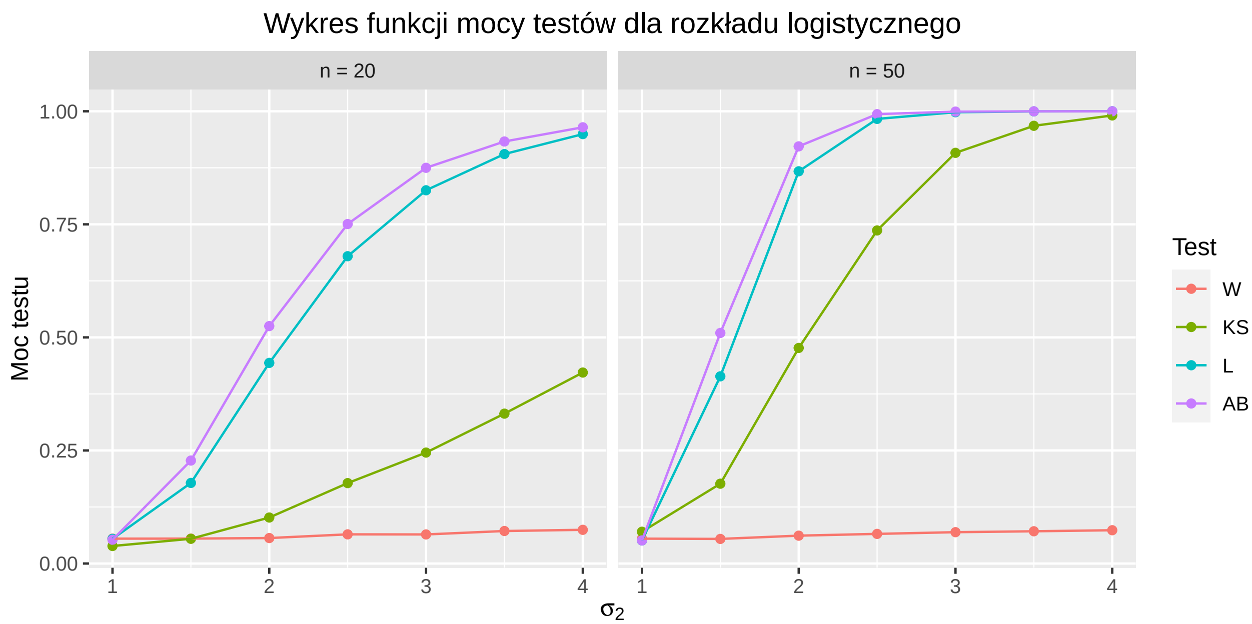
Widzimy, że dla zadanych parametrów z ruchomym μ_2 zdecydowanie najgorzej wypada test Ansari-Bradleya - dla obu liczebności i wszystkich rozpatrywanych rozkładów próbek. Wyniki pozostałych testów były dość zbliżone. Dla rozkładu normalnego i logistycznego największe moce miał test Wilcoxona, z kolei dla rozkładu Cauchy'ego - test Kołmogorowa-Smirnova. Rozkład Cauchy'ego różni się znacząco od dwóch pozostałych rozkładów, między innymi tym, że ma nieokreśloną wariancję i wartość oczekiwaną. Test KS, który w jego przypadku okazał się najlepszy, oparty jest na porównywaniu dystrybuant empirycznych.
Test AB daje tak złe wyniki, ponieważ bada on równość rozkładów poprzez porównywanie parametrów skali, a tutaj parametry te dla obu próbek są takie same. Manipulowaliśmy natomiast parametrem położenia w drugiej próbce, zatem testy, które biorą pod uwagę takie różnice, sprawdzają się znacznie lepiej.
Dla większej liczebności próbek wyniki były lepsze niż dla mniejszych, co jest zrozumiałe ze względu na fakt, że mamy do czynienia z rozkładami asymptotycznymi, zatem w miarę, jak n dąży do nieskończoności, testy będą działać coraz lepiej (o ile znajdują zastosowanie w podanym przypadku). Różnica ta jest bardziej widoczna w przypadku rozkładów innych niż normalny.

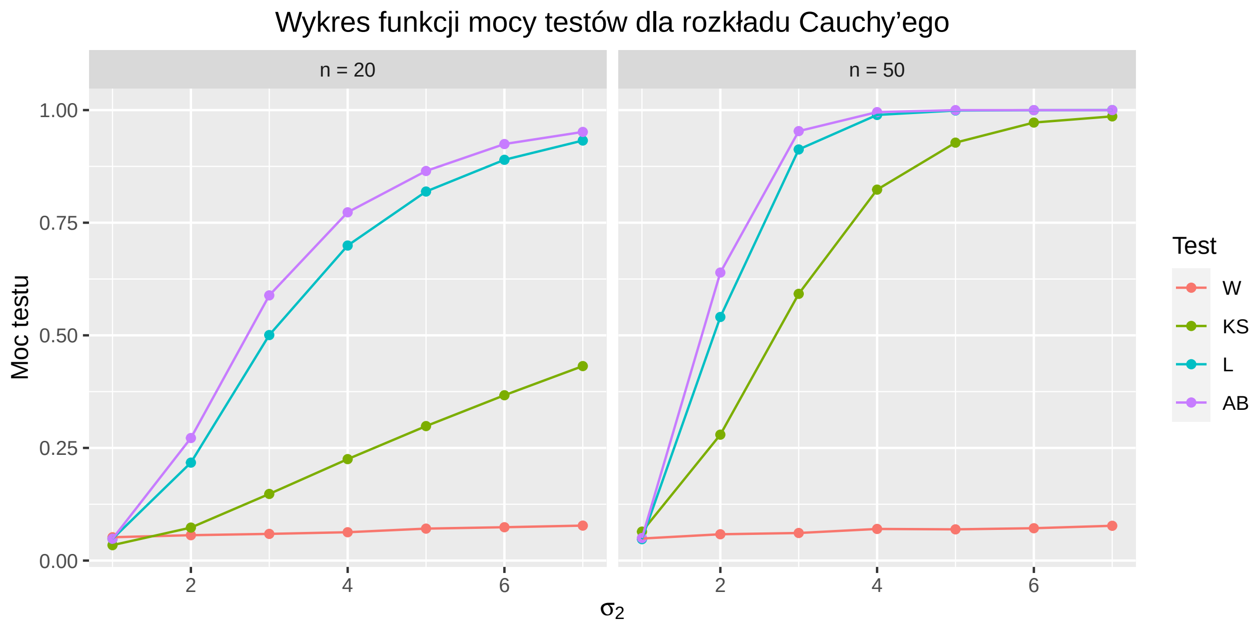
Spośród trzech rozkładów dla tak zadanych zakresów μ_2 najbardziej wyróżnia się wykres dla rozkładu logistycznego - widac, że przy $n = 20$ potrzebowaliśmy większych wartości parametru, żeby osiągnąć wartości mocy bliskie 1.
Podsumowując, test Ansari-Bradleya nie powinien być stosowany do porównywania rozkładów różniących się jedynie parametrem położenia. Pozostałe testy zaś sprawdzają się coraz lepiej wraz ze zwiększaniem się różnicy między μ_1 i μ_2. Jeśli chcemy, żeby nasze moce testów były większe, powinniśmy wybrać większą próbkę.

Funkcja mocy w zależności od parametru σ_2

Funkcję mocy będziemy wyznaczać empirycznie, analogicznie jak w poprzednim przypadku, z tą różnicą, że będziemy manipulować parametrem σ_2, a μ_2 pozostanie tym razem bez zmian.
W tym wypadku zakresy σ_2 dla $n = 50$ pozostawiamy takie, jak zostały zadane dla $n = 20$.

Wykresy funkcji mocy



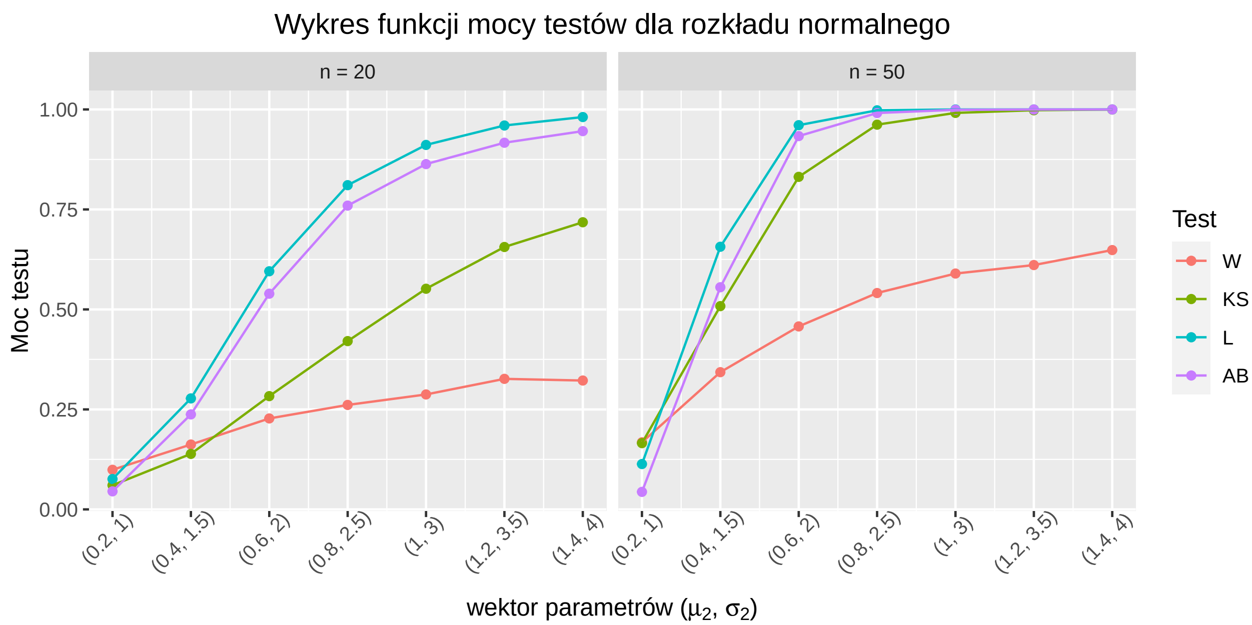


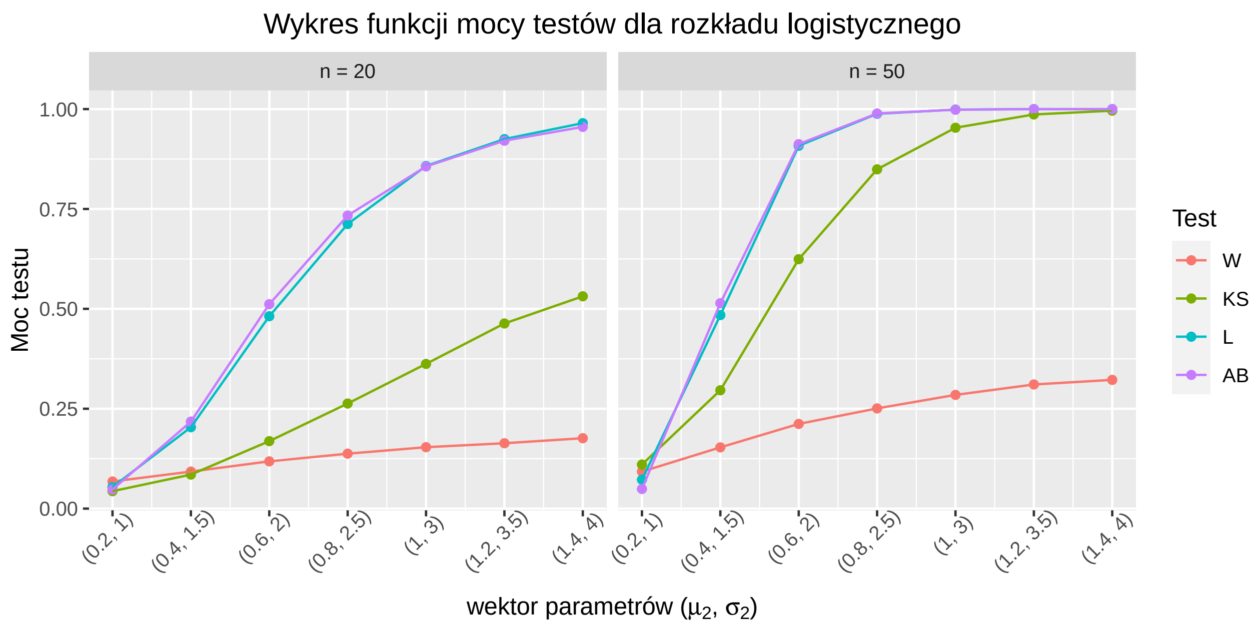
W przypadku manipulacji parametrem σ_2 wyniki uległy znaczącej zmianie. Tym razem zdecydowanie najgorszy okazał się test Wilcoxona. Stało się tak, ponieważ test ten bada równość rozkładów w oparciu o porównywanie ich parametrów przesunięcia, a te były sobie równe dla wszystkich próbek. Efekt zatem okazał się analogiczny jak dla testu AB w przypadku takich samych parametrów skali, a różnych przesunięć.
Tutaj test AB sprawdza się najlepiej ze wszystkich, bowiem został on skonstruowany do wykrywania zmian parametru skali, a to właśnie tym parametrem manipulujemy. Obserwujemy natomiast inne niż w przypadku zmiany μ_2 zachowanie funkcji mocy dla testu KS - wypada on tu o wiele gorzej dla próbek o mniejszej liczebności.
Pary wykresów dla wszystkich rozkładów wyglądają stosunkowo podobne, w przeciwieństwie do poprzedniego przypadku, gdzie wykresy dla rozkładu logistycznego odróżniały się od pozostałych.

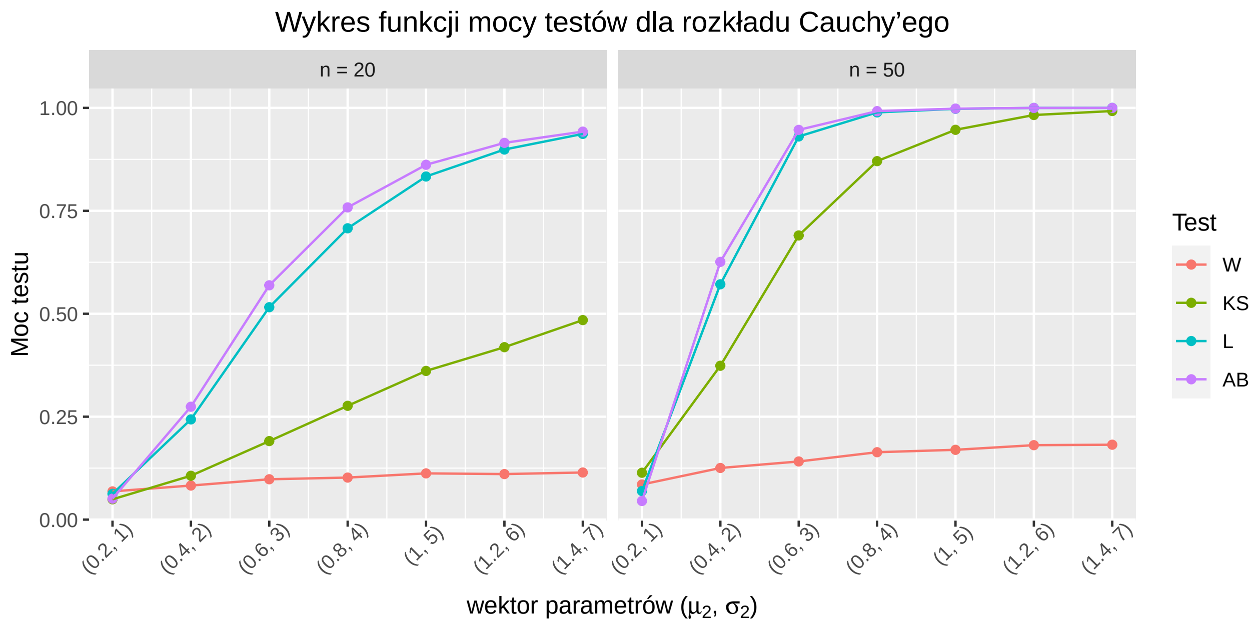
Funkcja mocy w zależności od wektora parametrów (μ_2, σ_2)

Funkcję mocy będziemy wyznaczać empirycznie, analogicznie jak w poprzednich przypadkach, z tą różnicą, że będziemy manipulować obydwoma parametrami, tj. σ_2 oraz μ_2.
W tym zadaniu zostawiamy dla $n = 50$ takie zakresy, jak dla $n = 20$ ze względu na zachowanie oryginalnych par i niepowtarzanie wartości σ_2, a jednocześnie pozostawienie funkcji mocy jako niezdegenerowanej, rosnącej.

Wykresy funkcji mocy







We wszystkich przypadkach najniższe wartości funkcji mocy uzyskaliśmy dla testu Wilcoxona. Co więcej, są one coraz niższe w zależnosci od tego, jak bardzo rozkład różni się od normalnego - w rozkładzie logistycznych test ten wypada mniej więcej dwukrotnie gorzej, za to w rozkładzie Cauchy'ego - około czterokrotnie. Obserwujemy również podobne zjawisko, jak w poprzednim podpunkcie w przypadku testu Kołmogorowa-Smirnova - wyniki znacząco się poprawiają przy zwiększeniu liczebności próby. Dla rozkładu normalnego najlepiej wypada test Lepage'a, natomiast dla pozostałych rozkładów porównywalnie dobre, a nawet nieco wyższe wyniki daje również test Ansari-Bradley'a.

Dobre wyniki testu L tłumaczy budowa statystyki, na której jest on oparty, bowiem jest ona sumą statystyk W i AB, zatem uwzględnia i zmianę parametru położenia, i zmianę parametru skali. Z kolei dobre wyniki testu AB, natomiast złe wyniki testu W wskazują na to, że przy zmianie obu parametrów lepiej rozpoznawalna jest różnica w skali aniżeli w położeniu, a przynajmniej tak stało się przy parametrach dobranych jak w tym zadaniu. Zaznaczmy, że różnice między kolejnymi σ_2 były znacząco większe niż między kolejnymi μ_2.

Podsumowanie

Jeśli chcemy badać różnice między dwoma rozkładami, dobrze jest mieć pewne przypuszczenia co do tego, czy różnią się skalą, położeniem czy też obydwoma tymi parametrami, ponieważ w zależności od tego inny test będzie odpowiedni do tego badania.
W przypadku, gdy spodziewamy się, że nasze próbki pochodzą z rozkładów o różnych parametrach położenia, najprawdopodobniej najlepiej będzie użyć testu Wilcoxona, natomiast nie należy stosować testu Ansari-Bradley'a.
Jeśli przypuszczamy, że mamy do czynienia z różnymi skalami, będzie odwrotnie - najskuteczniejszym testem będzie test Ansari-Bradley'a, z kolei nie powinniśmy stosować testu Wilcoxona.
Natomiast w przypadku oczekiwań, że oba parametry są różne, najprawdopodobniej powinniśmy wybrać test Lepage'a. Wyjątek stanowi sytuacja, gdzie jeden parametr różni się w parze znacznie bardziej niż drugi, wówczas lepszy może okazać się test nakierowany na ten parametr (jak test Ansari-Bradley'a w przypadku znacząco większych różnic w skali, a mniejszych w położeniu).
W sytuacji, gdy nic nie wiemy o rozkładach, z których pochodzą próbki (za wyjątkiem założenia, że rozkłady te są ciągłe), najbezpieczniej będzie użyć testu AB lub KS, ponieważ uwzględniają one oba parametry. W przypadku powyższych rozkładów średnio lepiej wypadał test AB.