

Raport czwartej listy z laboratorium Statystyki

Kamil Zdancewicz

December 25, 2025

Wstęp

We wszystkich zadaniach będę korzystał z definicji i określeń prób, rang, hipotez, statystyki rangowej i testów tak jak we wstępie do Listy 4.

Oczekujemy, że statystyka Wilcoxona będzie dobrze mierzyć różnicę położenia między dwiema próbami.

Statystyka Ansari-Bradleya będzie się sprawdzać przy mierzeniu różnicy rozproszenia/wariancji między próbami.

Statystyka Lepage'a sumuje efekty położenia i skali, więc będzie dobra dla ogólnych przypadków, ale mało dokładna.

Statystyka Kołmogorowa-Smirnowa mierzy największą różnicę dystrybuant empirycznych, przez co powinna dobrze reagować dla dużych przesunięć i ciężkich ogonów.

Zadanie 1

Cel

Celem zadania było przetestowanie zachowania i określenie poprawności obliczania różnych statystyk z wstępu listy oraz wyznaczenie wartości krytycznych odpowiadających im testów prawostronnych.

Stosowane metody

Wygenerowałem 10 000 razy próbę wielkości $n = m = 20$. Dla każdej z tych prób obliczyłem wartości statystyk $W^{(k)}, AB^{(k)}, L^{(k)}, KS^{(k)}$. Następnie wyznaczyłem empiryczny kwantyl rzędu $1 - \alpha$:

$$c_T = \inf\{t : \widehat{P}(T \leq t) \geq 1 - \alpha\}$$

Czyli wartość, która spełnia w przybliżeniu:

$$\mathbb{P}_{H_0}(T > c_T) \approx \alpha$$

Której dokładne wyznaczenie znajduje się w [obliczeniach](#).

Obliczenia

Dla realizacji losowej statystyki T z naszych czterech, dla próby k , $T^{(k)} = T(Z_1^{(k)}, \dots, Z_N^{(k)})$

Definiujemy empiryczną dystrybuantę:

$$\widehat{F}_T(t) = \frac{1}{10000} \sum_{k=1}^{10000} \mathbf{1}\{T^{(k)} \leq t\}$$

Czyli estymator dystrybuanty $F_T(t) = \mathbb{P}(T \leq t)$ pod H_0 .

Z definicji wartości krytycznej c_α :

$$\mathbb{P}_{H_0}(T > c_\alpha) = \alpha$$

Czyli

$$F_T(c_\alpha) = 1 - \alpha$$

Wyznaczamy:

$$\inf\{t : \widehat{F}_T(t) \geq 1 - \alpha\}$$

Czyli dla $\alpha = 0.05$, i posortowanych wartości $T_{(1)} \leq \dots T_{(10000)}$:

$$\hat{c}_{0.05} = T_{(0.95 \cdot 10000)} = T_{(9500)}$$

Wyniki

Statystyka	Wynik
Wilcoxon	3.888
Ansari–Bradley	3.888
Lepage	6.123
Kołmogorow–Smirnow	1.265

Te wartości będą później używane w funkcjach mocy przy wszystkich rozkładach i parametrach, ponieważ nie zależą od rozkładu ani jego parametrów.

Wnioski

Wyniki pierwszych dwóch statystyk są bliskie teoretycznemu górnemu 5% kwantylowi χ_1^2 równemu 3.841 co potwierdza poprawność generowania wartości krytycznych w ten sposób.

Trzecia statystyka jest sumą dwóch pierwszych, poprawna dla teoretycznego kwantylu $\chi_2^2 = 5.991$ więc jeśli one są poprawne to ona też.

Czwarta statystyka jest trochę poniżej oczekiwanej wartości 1.358 5% górnego kwantylu rozkładu Kołmogorowa.

Wszystkie uzyskane wartości krytyczne są zgodne z rozkładami asymptotycznymi, co sugeruje, że symulacja Monte-Carlo poprawnie odzwierciedla rozkład statystyk pod hipotezą zerową.

Załączniki

Generacja danych, obliczanie przedziału i dalsza logika znajduje się w załączonym pliku `main.cpp`.

Zadanie 2

Cel

Celem zadania było obliczenie wartości statystyk W , AB , L i KS dla różnych rozkładów z różnymi parametrami **przesunięcia**. Oszacowanie wartości funkcji mocy analizowanych testów oraz narysowanie wyestymowanych funkcji mocy w zależności od tego parametru.

Stosowane metody

Dla każdego podpunktu (a) - (c) wygenerowałem 10 000 razy próby wielkości $n = m = 20$. Dla każdej z tych prób obliczyłem wartości wszystkich czterech statystyk. Następnie według wzoru z [obliczeń](#) obliczyłem wartość funkcji mocy dla danego pod-punktu (łącznie $4 \cdot 3 \cdot 7 = 84$ wartości). Następnie stworzyłem wykresy tych wartości.

Obliczenia

Dla danego testu o statystyce T i wartości krytycznej c_α funkcja mocy testu to:

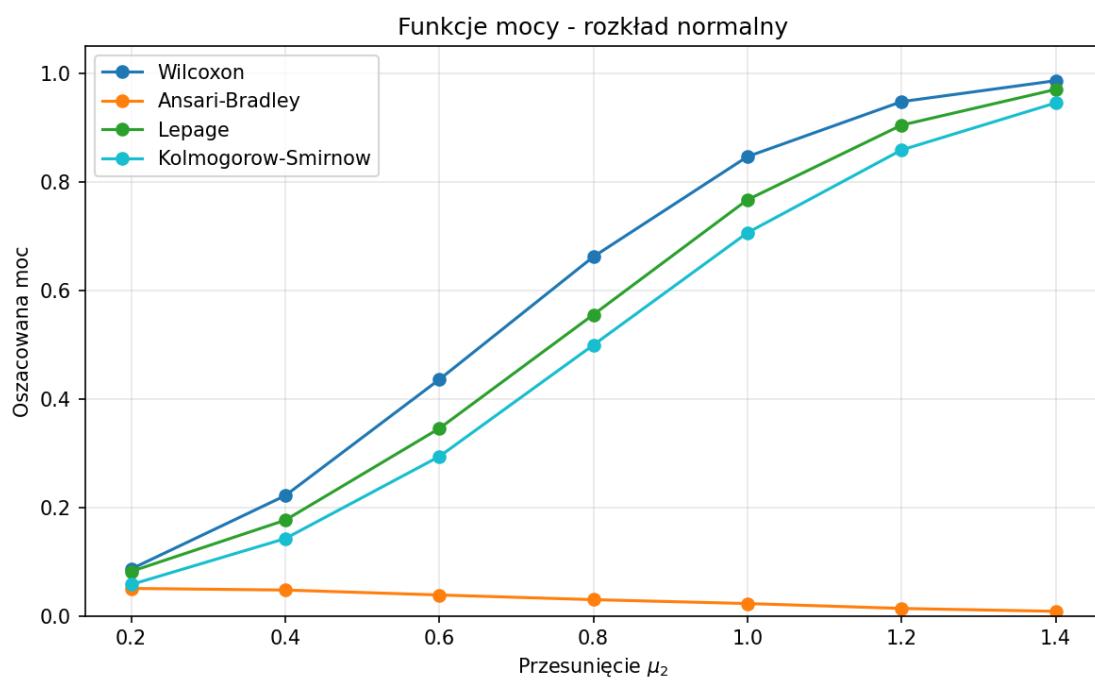
$$\pi(\mu_2) = \mathbb{P}(T > c_\alpha)$$

Czyli u nas przybliżeniem:

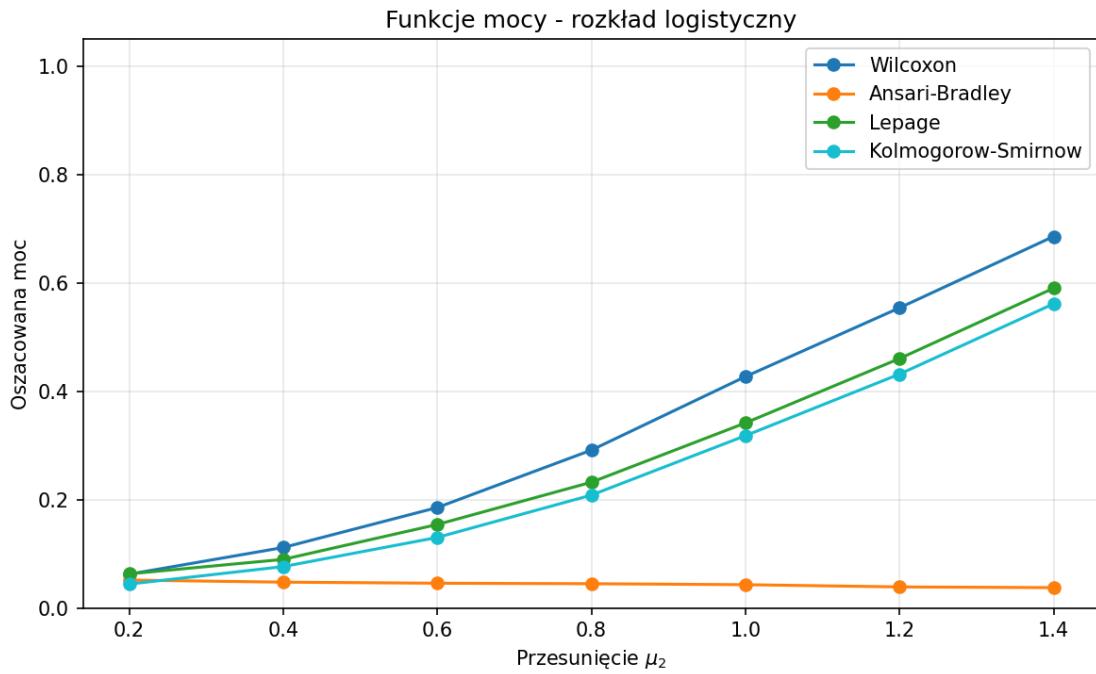
$$\hat{\pi}(\mu_2) = \frac{1}{10000} \sum_{i=1}^{10000} \mathbf{1}\{T^{(i)}(\mu_2) > c_\alpha\}$$

Wyniki

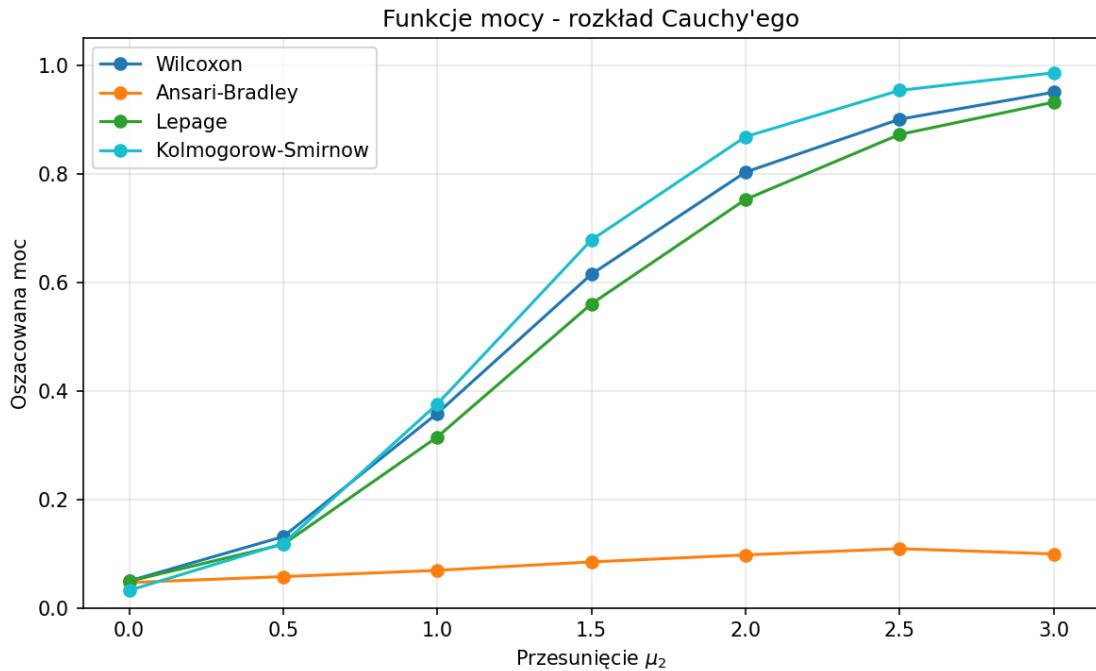
(a) Rozkład normalny



(b) Rozkład logistyczny



(c) Rozkład Cauchy'ego



Wnioski

Test W jest najbardziej czuły na różnice w położeniu, choć przy rozkładzie cauchy'ego i dużych przesunięć test KS go przegania.

Test AB działa dobrze tylko przy różnicach w wariancji/skali, w przypadku samego przesunięcia jego moc pozostaje niska a nawet się zmniejsza z przesunięciem.

Test L łączy zalety W i AB oraz daje nieco wyższą moc niż W w przypadku ciężkich ogonów rozkładu Cauchy'ego.

Test KS wykrywa różnice w dystrybuantach globalnie, więc czułość jest mniejsza niż W dla małych przesunięć, ale bardzo dobra przy dużych przesunięciach np. dla podpunktu (c). Pozostaje blisko testów W i L.

Wszystkie obserwacje zgadzają się z teorią we [wstępie](#).

Załączniki

Generacja danych, obliczanie przedziału i dalsza logika znajduje się w załączonym pliku `main.cpp`. Natomiast generowanie wykresów znajduje się w załączonym pliku `z2.py`

Zadanie 3

Cel

Celem zadania było obliczenie wartości statystyk W , AB , L i KS dla różnych rozkładów z różnymi parametrami skali. Oszacowanie wartości funkcji mocy analizowanych testów oraz narysowanie wyestymowanych funkcji mocy w zależności od tego parametru.

Stosowane metody

Dla każdego podpunktu (a) - (c) wygenerowałem 10 000 razy próby wielkości $n = m = 20$. Dla każdej z tych prób obliczyłem wartości wszystkich czterech statystyk. Następnie według wzoru z [obliczeń](#) obliczyłem wartość funkcji mocy dla danego pod-podpunktu (łącznie $4 \cdot 3 \cdot 7 = 84$ wartości). Następnie stworzyłem wykresy tych wartości.

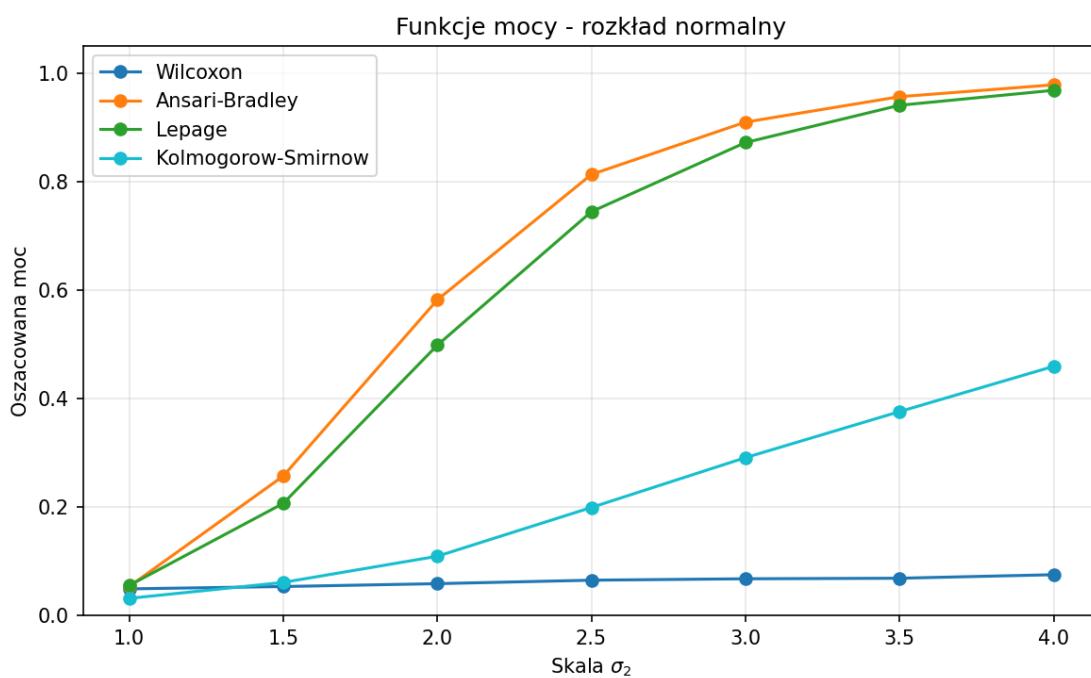
Obliczenia

Podobnie jak w poprzednim zadaniu, nasz wzór na funkcję mocy:

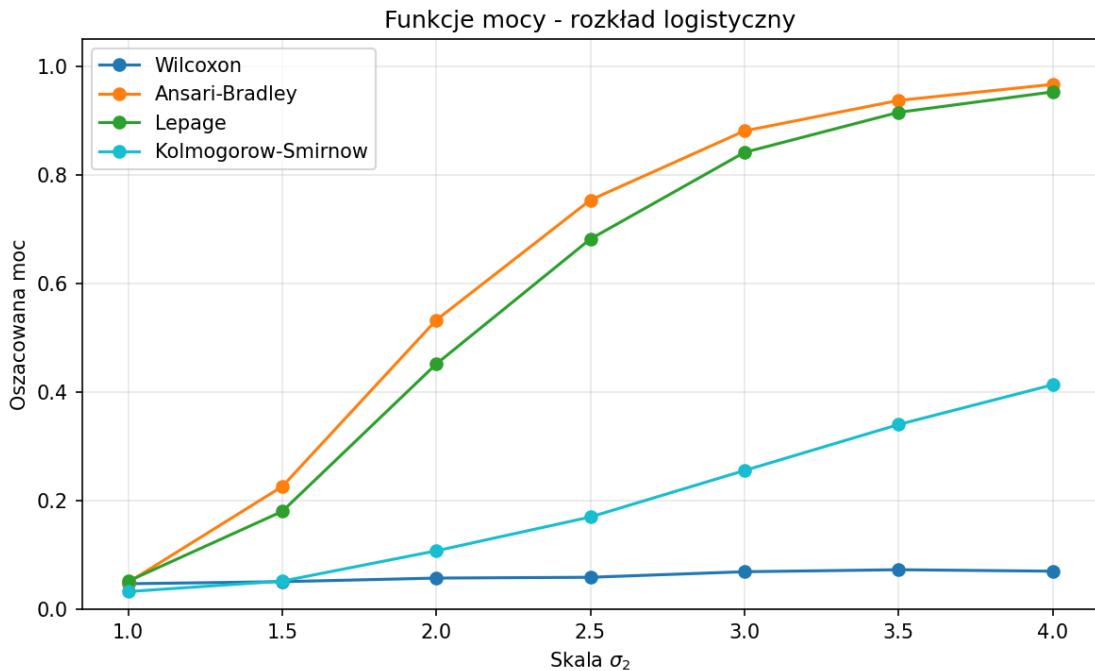
$$\hat{\pi}(\sigma_2) = \frac{1}{10000} \sum_{i=1}^{10000} \mathbf{1}\{T^{(i)}(\sigma_2) > c_\alpha\}$$

Wyniki

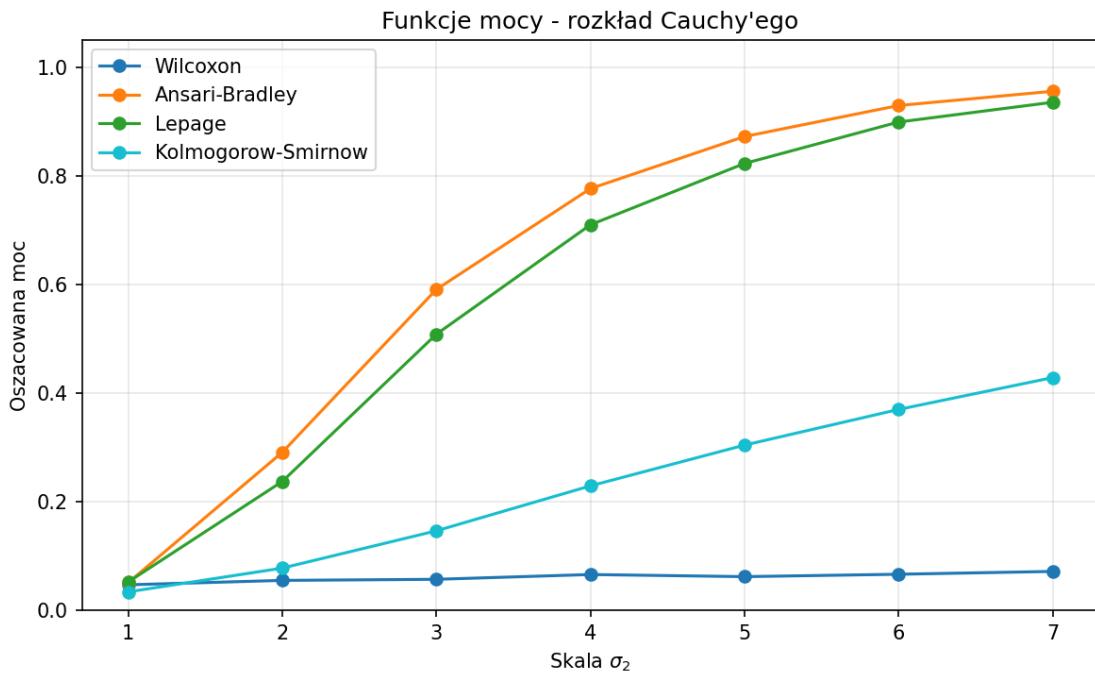
(a) Rozkład normalny



(b) Rozkład logistyczny



(c) Rozkład Cauchy'ego



Wnioski

Test W niezależnie od parametrów i rozkładu minimalnie rośnie ponad 0.05, praktycznie nie reaguje na alternatywę skalową.

Test AB jest zdecydowanie najbardziej efektywnym testem różnicę skali. Najszybciej zaczyna rosnąć i osiąga najwyższe wartości funkcji mocy.

Test L łączy zalety W i AB, przez traci część mocy względem AB, ponieważ agreguje informację o położeniu i skali co trochę pogarsza wynik.

Test KS, przez ogólność, jest wyraźnie mniej czuły od AB i L, chociaż ciągle reaguje na zmiany.

Wszystkie obserwacje zgadzają się z teorią we [wstępie](#).

Załączniki

Generacja danych, obliczanie przedziału i dalsza logika znajduje się w załączonym pliku `main.cpp`. Natomiast generowanie wykresów znajduje się w załączonym pliku `z3.py`

Zadanie 4

Cel

Celem zadania było obliczenie wartości statystyk W , AB , L i KS dla różnych rozkładów z różnymi parametrami **parametru i skali**. Oszacowanie wartości funkcji mocy analizowanych testów oraz narysowanie wyestymowanych funkcji mocy w zależności od tego parametru.

Stosowane metody

Dla każdego podpunktu (a) - (c) wygenerowałem 10 000 razy próby wielkości $n = m = 20$. Dla każdej z tych prób obliczyłem wartości wszystkich czterech statystyk. Następnie według wzoru z [obliczeń](#) obliczyłem wartość funkcji mocy dla danego pod-podpunktu (łącznie $4 \cdot 3 \cdot 7 = 84$ wartości). Następnie stworzyłem wykresy tych wartości.

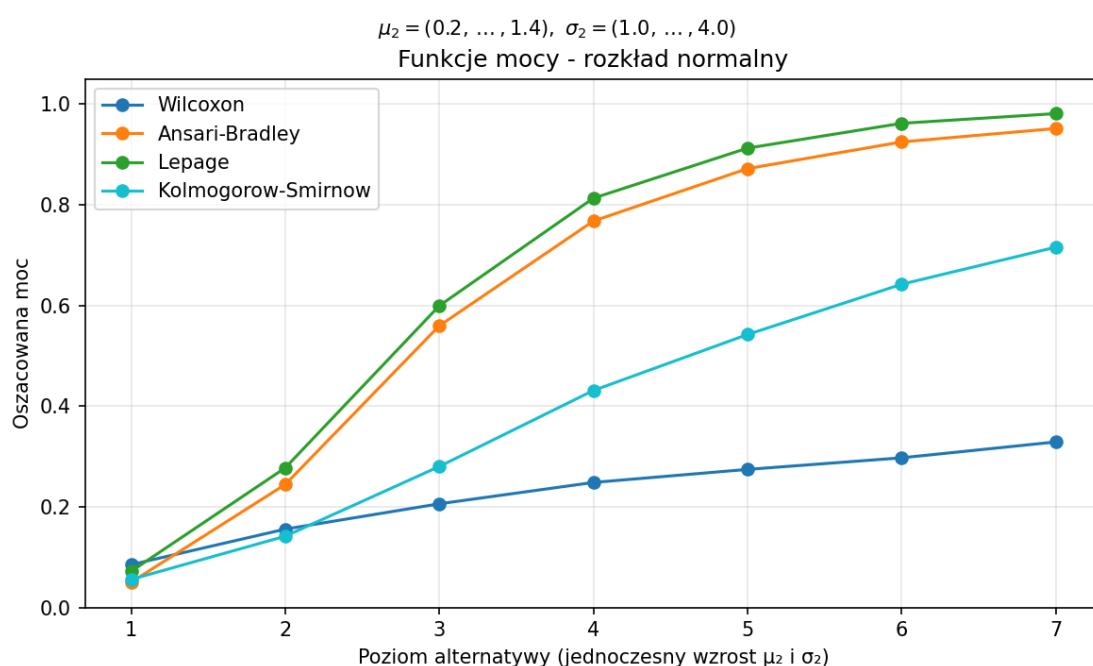
Obliczenia

Podobnie jak w poprzednich zadaniach, nasz wzór na funkcję mocy:

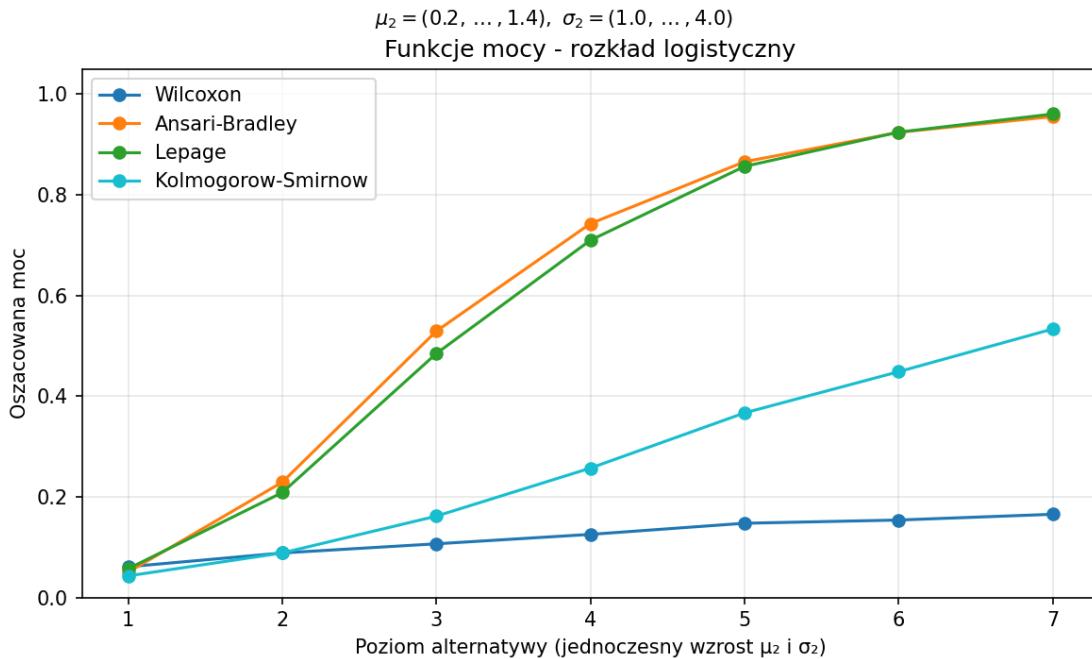
$$\hat{\pi}(\mu_2, \sigma_2) = \frac{1}{10000} \sum_{i=1}^{10000} \mathbf{1}\{T^{(i)}(\mu_2, \sigma_2) > c_\alpha\}$$

Wyniki

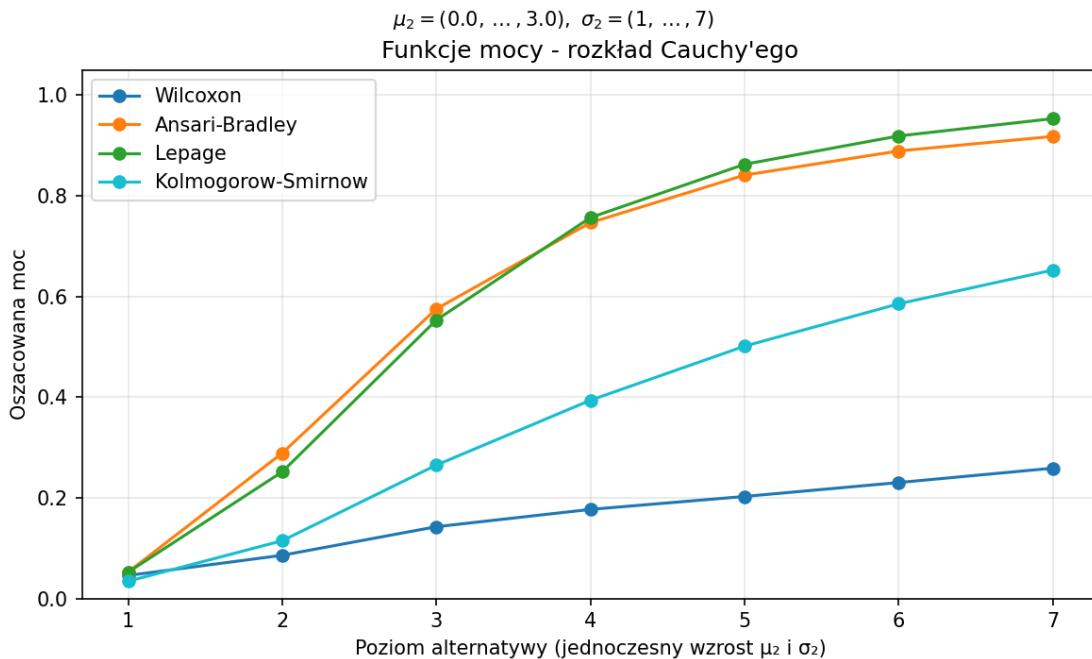
(a) Rozkład normalny



(b) Rozkład logistyczny



(c) Rozkład Cauchy'ego



Wnioski

Test W wykazuje ograniczoną skuteczność, mimo że alternatywa zawiera przesunięcia. Jednoczesny wzrost skali powoduje zwiększenie rozrzutu obserwacji, co osłabia zdolność testu, jest najgorszy z wszystkich czterech.

Test AB osiąga wysoką moc już dla umiarkowanych poziomów alternatywy, co potwierdza jego dużą wrażliwość na różnice skali. Jednoczesna zmiana położenia nie degraduje jego skuteczności, a wręcz sprzyja rozróżnieniu rozkładów.

Średnio jest minimalnie gorszy od testu L.

Test L jest się najbardziej uniwersalnym i efektywnym testem. Łącząc informację o położeniu i skali, reaguje jednocześnie na oba komponenty alternatywy, co skutkuje najszybszym wzrostem mocy.

Test KS, wykazuje umiarkowaną skuteczność w zadaniu 4. Choć reaguje na ogólne różnice dystrybuant, jego moc rośnie wolniej niż w testach rangowych. Jak poprzednio lepiej sobie radzi w przypadku rozkładu z ciężkimi ogonami (Cauchy).

Wszystkie obserwacje zgadzają się z teorią we [wstępie](#).

Załączniki

Generacja danych, obliczanie przedziału i dalsza logika znajduje się w załączonym pliku `main.cpp`. Natomiast generowanie wykresów znajduje się w załączonym pliku `z4.py`

Zadanie 5

Cel

Celem zadania było obliczenie różnych statystyk z wstępu listy oraz wyznaczenie wartości krytycznych odpowiadających im testów prawostronnych. Jedyną różnicą od [zadania 1](#) są wartości n i m .

Stosowane metody

Użyte zostały te same wzory i metody jak w [zadaniu 1](#), jednak ze zmienionymi wartościami n i m .

Wyniki

Statystyka	Wynik
Wilcoxon	3.817
Ansari–Bradley	3.871
Lepage	6.097
Kołmogorow–Smirnow	1.300

Wnioski

Wyniki pierwszych dwóch statystyk coraz bardziej zbliżają się do 5% kwantyla χ_1^2 równemu 3.841 co potwierdza zbieżność asymptotycznych przybliżeń.

Trzecia statystyka jest sumą dwóch pierwszych, poprawna dla teoretycznego kwantylu $\chi_2^2 = 5.991$ więc jeśli one są poprawne to ona też.

Czwarta statystyka wciąż nie dochodzi do teoretycznego 1.358 5% górnego kwantylu rozkładu Kołmogorowa, ale jest bliżej niż przy mniejszych próbach, czyli możemy wnioskować, że przybliżenie jest poprawne, choć zbiega ono wolno.

Porównanie zadań 1 i 5 empirycznie potwierdza, że wzrost liczebności próby prowadzi do stabilizacji rozkładów statystyk testowych oraz ich zbieżności do rozkładów asymptotycznych.

Załączniki

Generacja danych, obliczanie przedziału i dalsza logika znajduje się w załączonym pliku `main.cpp`. Należy zmienić wielkości prób w sekcji `#define`.

Zadanie 6

Cel

Celem zadania było powtórzenie zadań [2](#), [3](#) i [4](#) dla większych prób i zauważenie różnic między wynikami dla dla oryginalnych wielkości prób w poprzednich zadaniach.

Stosowane metody

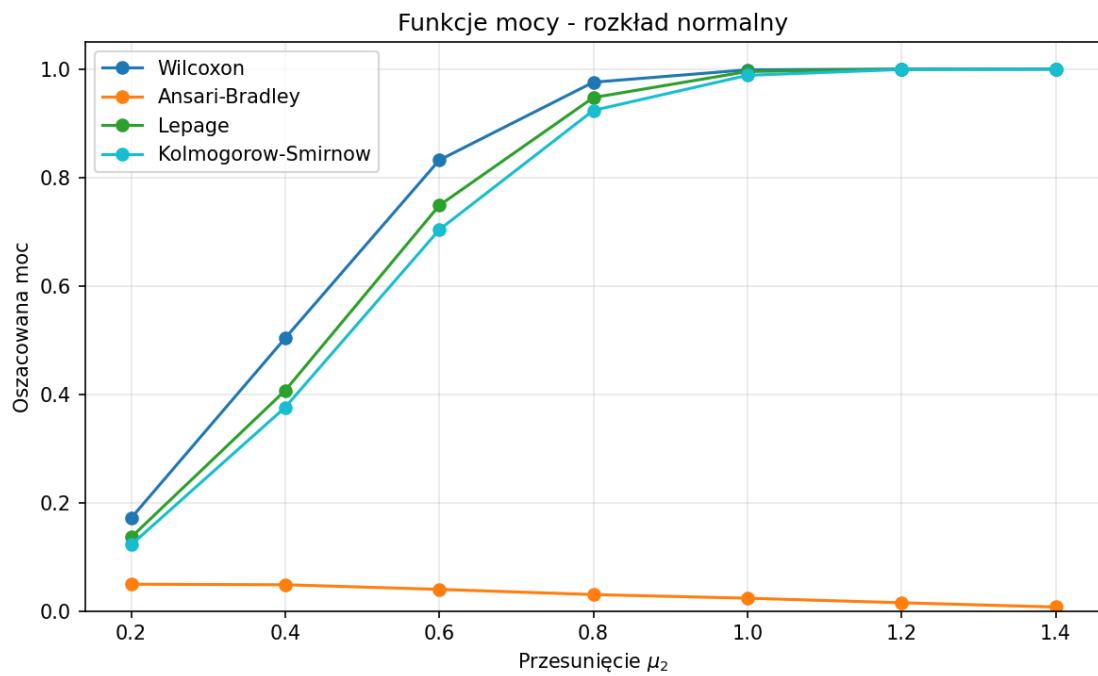
Użyte zostały wartości krytyczne wyznaczone w zadaniu 5 oraz wzory wyznaczone w oryginalnych zadaniach.

Wyniki

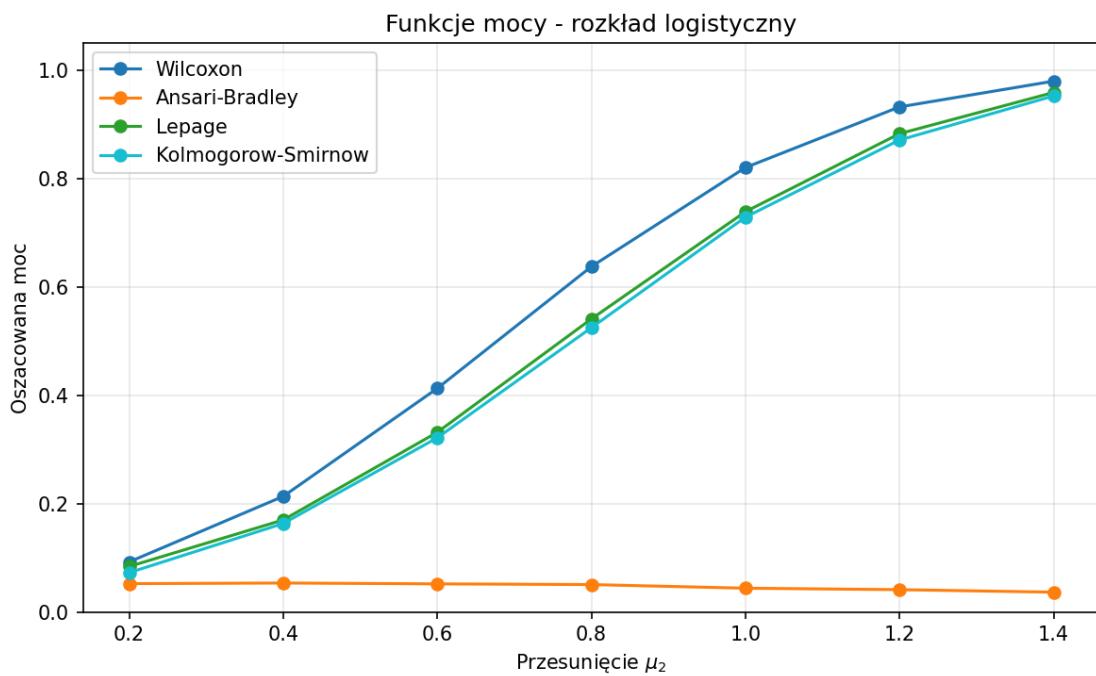
Klasycznie jak w ostatnim zadaniu na każdej liście, tu znajduje się bardzo dużo danych.

Zadanie 2

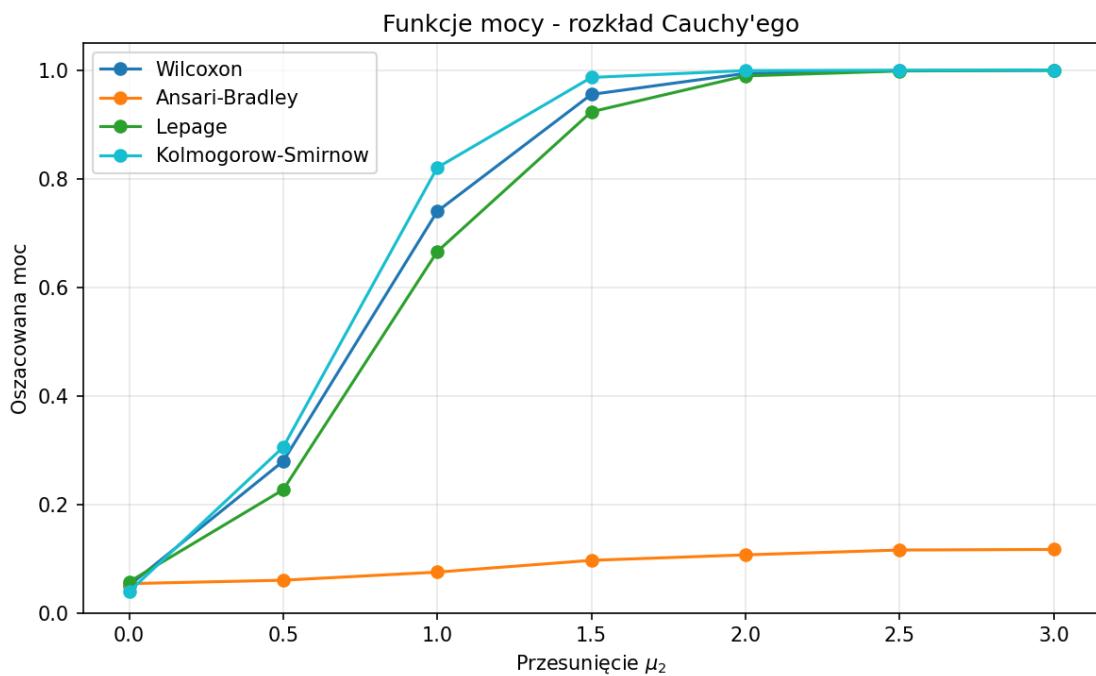
(a) Rozkład normalny



(b) Rozkład logistyczny

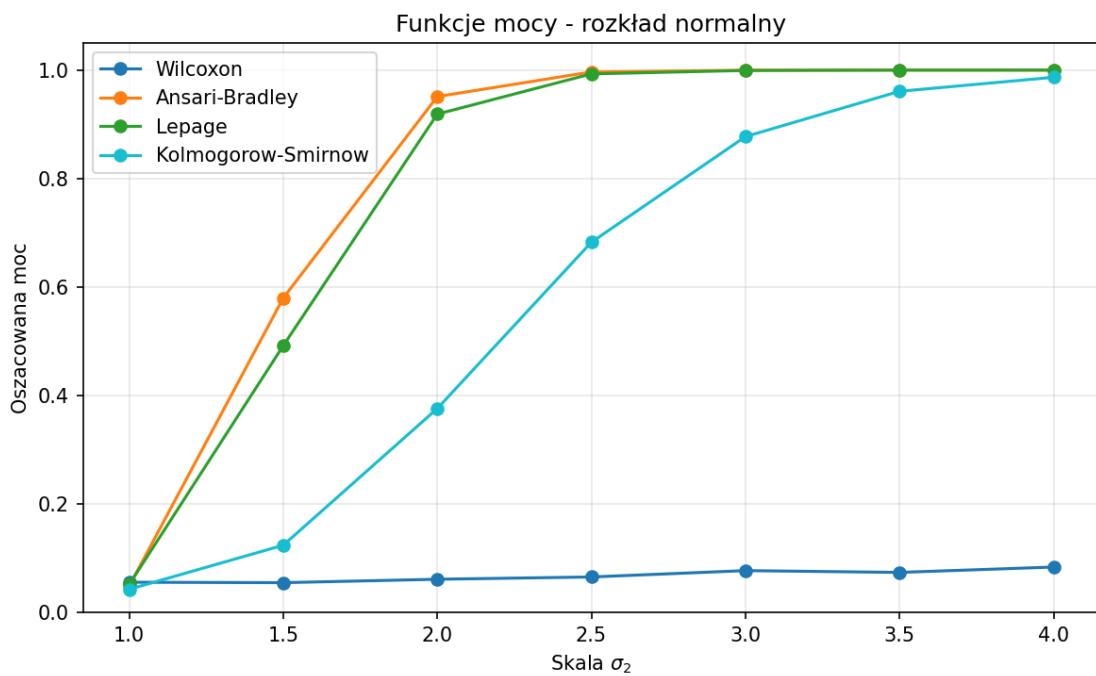


(c) Rozkład Cauchy'ego

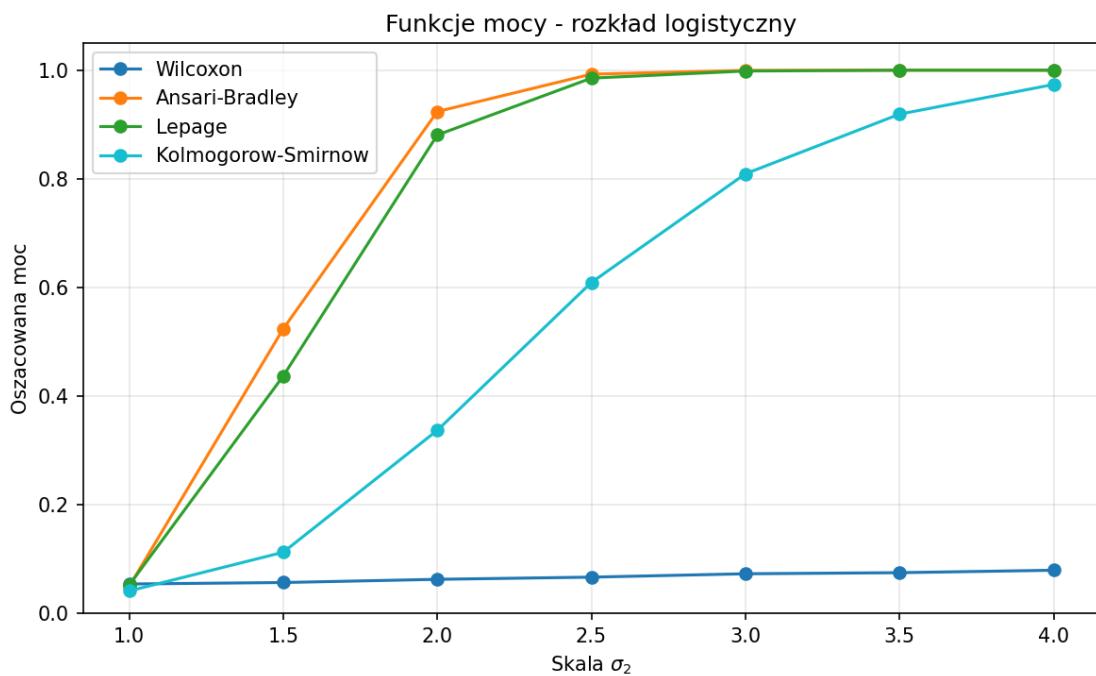


Zadanie 3

(a) Rozkład normalny

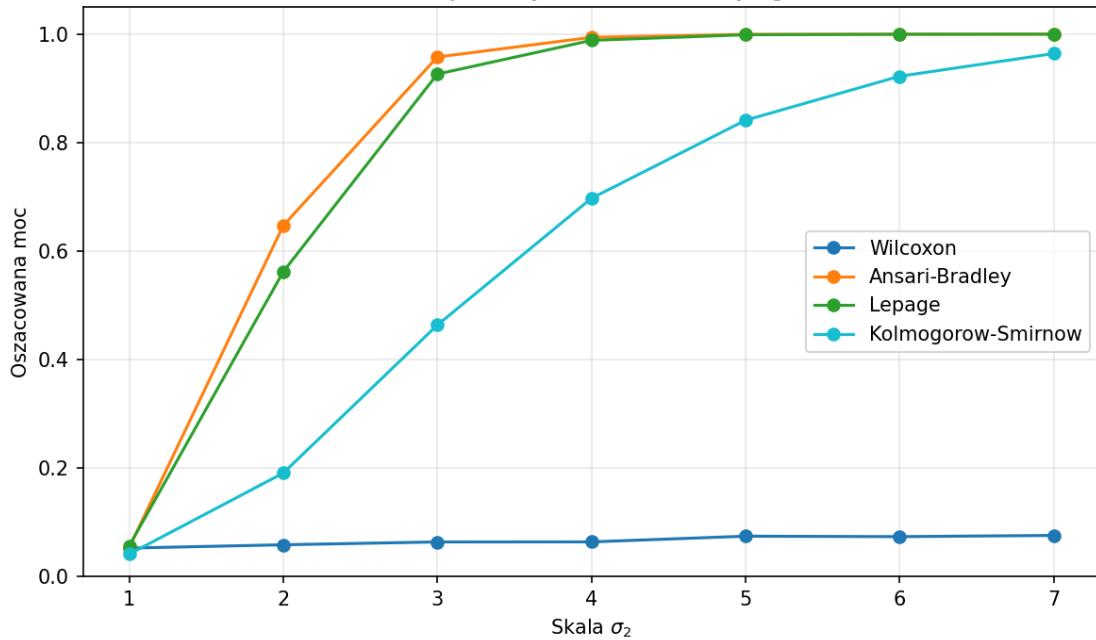


(b) Rozkład logistyczny



(c) Rozkład Cauchy'ego

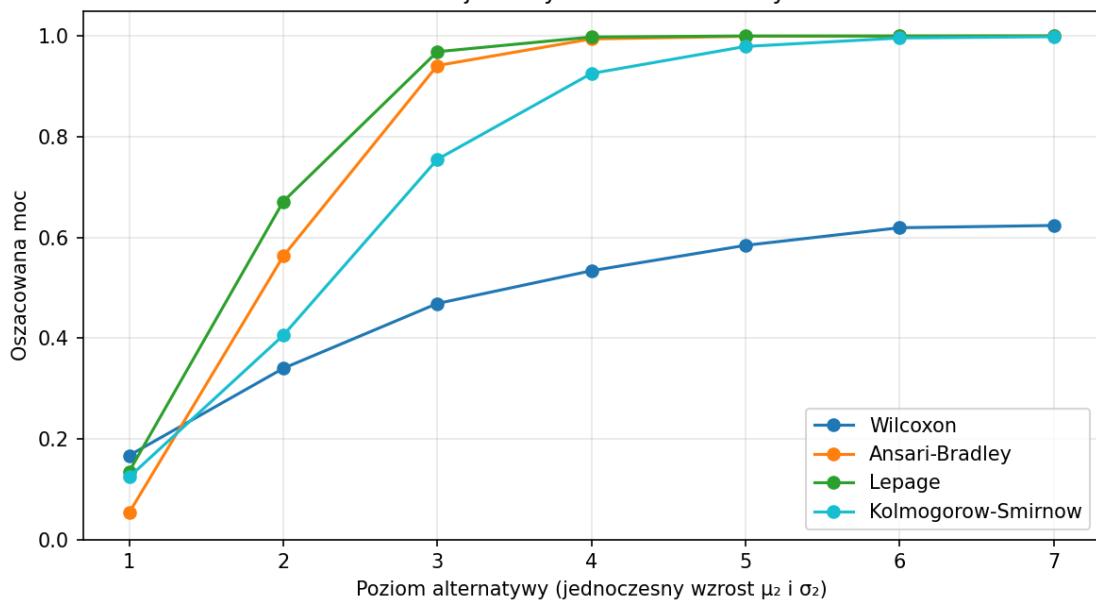
Funkcje mocy - rozkład Cauchy'ego



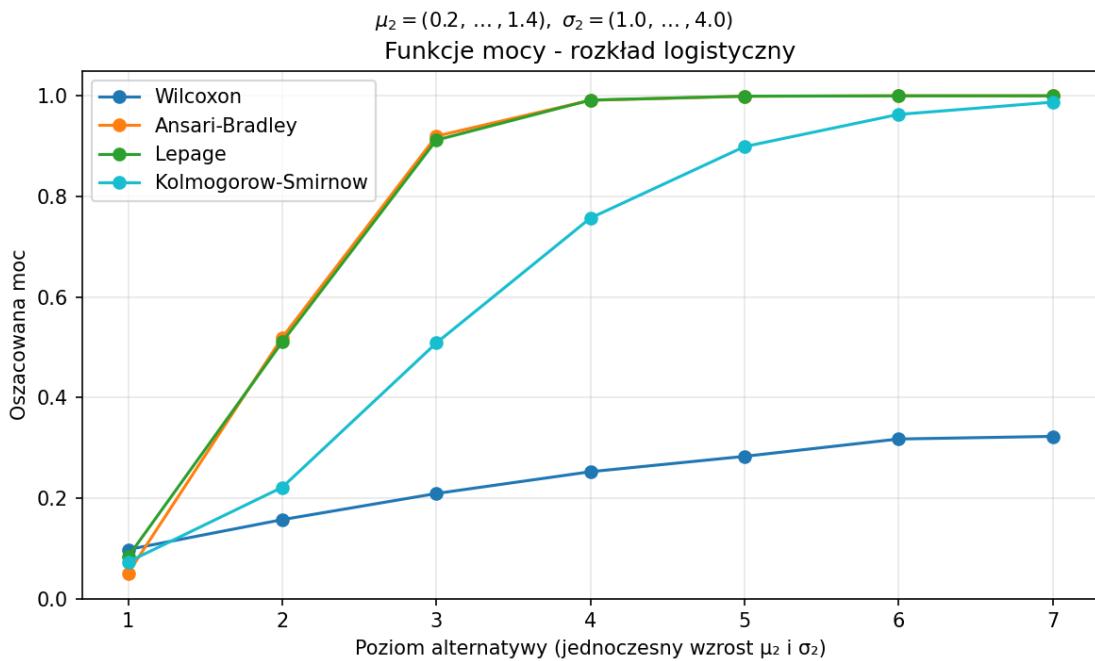
Zadanie 4

(a) Rozkład normalny

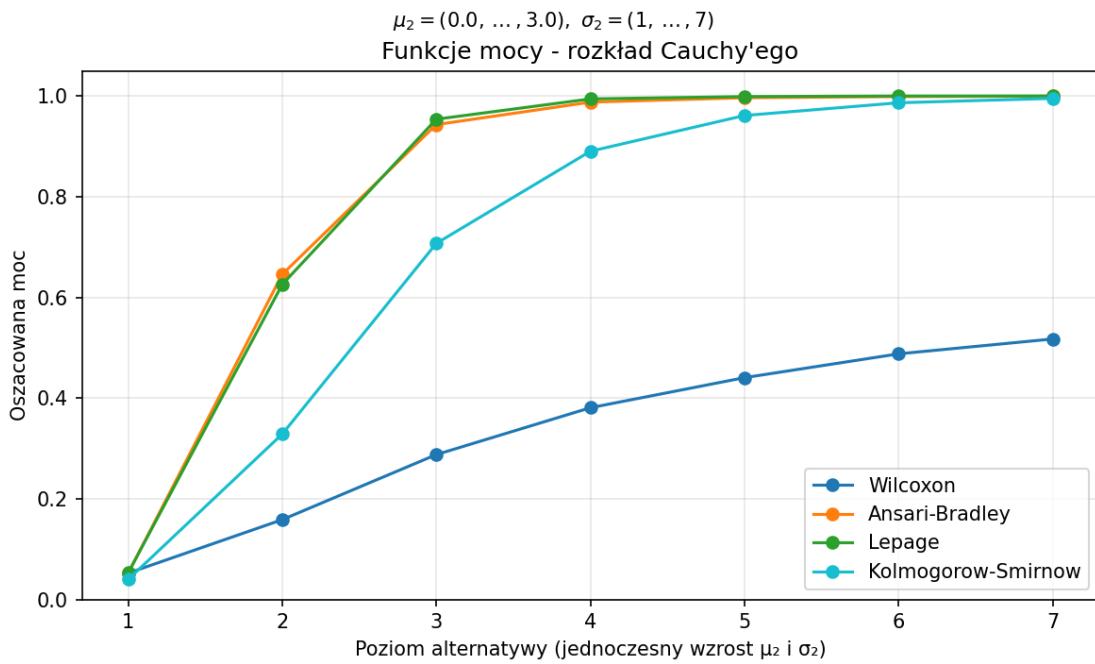
$\mu_2 = (0.2, \dots, 1.4)$, $\sigma_2 = (1.0, \dots, 4.0)$
Funkcje mocy - rozkład normalny



(b) Rozkład logistyczny



(c) Rozkład Cauchy'ego



Wnioski

Test W lepiej reaguje na zmiany przesunięcia μ_2 . Przy większych próbach różnica w medianach staje się wyraźniejsza, stąd szybszy wzrost mocy. W zadaniu 2 (a) i (c) ostatnie 3 wartości są równe 1.0.

Test AB dobrze reaguje na zmiany skali σ_2 . Większe próby zwiększą dokładność oszacowania rozrzutu w zadaniach 3 i 4 ostatnie 4 wartości osiągają 1.0.

Test L przez większą uniwersalność w kwestii samego przesunięcia μ_2 rośnie nieznacznie wolniej od testu W, jednak

osiąga 1.0 przy podobnym przesunięciu. W kwestii zadań ze skalą σ_2 ponownie jest trochę gorszy od wyspecjalizowanego testu AB, jednak znów osiąga 1.0 przy podobnych wartościach.

Test KS jako test ogólny, reaguje na każdą zmianę w rozkładzie, ale jego wzrost mocy jest zwykle bardziej umiarkowany. Przy największych wartościach μ_2 lub σ_2 wciąż zwykle osiąga 1.0. Warto zauważać, że w zadaniu 2 (c) radzi sobie najlepiej.

To dokładnie odpowiada oczekiwanej zachowaniu, większa próbka zwiększa precyzję rang i dystrybuant, co zwiększa moc testu.

Uwaga

Na wykresach zdecydowałem się uwzględnić wszystkie poziomy alternatyw dla wszystkich zadań, ponieważ moim zdaniem lepiej wizualizuje to jakość testu dla przypadku, oraz było to prostsze implementacyjnie. Dodatkowo we [wnioskach](#) zaznaczyłem główne przypadki, kiedy funkcja mocy staje się zdegenerowana (zwiększenie parametru nie daje nowych informacji).

Załączniki

Generacja danych, obliczanie przedziału i dalsza logika znajduje się w załączonym pliku `main.cpp`. Należy zmienić wielkości prób w sekcji `#define`. Natomiast generowanie wykresów znajduje się w załączonych plikach `z2.py`, `z3.py` i `z4.py` ze zmienionymi nazwami plików wyjściowych.

Źródła

- *Introduction to Mathematical Statistics (6th edition)* - Hogg, McKean, Craig
- *Statystyka dla studentów kierunków technicznych i przyrodniczych* Koronacki, J., Mielniczuk, J. (2009)