

# Porównanie testów dla parametru proporcji w rozkładzie dwumianowym

Szymon Malec, Michał Wiktorowski

## 1. Wstęp

Niniejsza praca poświęcona jest przedstawieniu trzech testów dla parametru proporcji (prawdopodobieństwa sukcesu w próbie Bernoulliego) rozkładu dwumianowego. Dla pewnej realizacji zmiennej losowej  $S \sim \mathcal{B}(n, p)$  rozważmy hipotezy:

- $H_0: p = p_0 = 0.5$
- $H_1: p \neq p_0 = 0.5,$

gdzie  $H_0$  i  $H_1$  są odpowiednio hipotezą zerową i alternatywną. Sprawdźmy ich poprawność przy użyciu trzech testów:

- testu opartego o **przedział Wilsona**,
- testu opartego o **przedział Cloppera-Pearsona**,
- testu opartego o **przedział Jeffreysa**,

na poziomie istotności  $\alpha = 0.05$ . Następnie porównamy wspomniane testy pod kątem mocy, aby odpowiedzieć na pytanie, czy można wyłonić wśród nich test jednostajnie najmocniejszy.

## 2. Test oparty o przedział Wilsona

Do konstrukcji przedziału wykorzystamy fakt, że rozkład dwumianowy można przybliżać rozkładem normalnym. Przyjmuje się, że przybliżenie to jest dobre, gdy  $np > 5$ ,  $n(1-p) > 5$  oraz wartość  $p$  jest bliska 0.5. Wartość oczekiwana  $S$  to  $ES = np$ , a odchylenie standardowe równe jest  $\text{Std}(S) = \sqrt{np(1-p)}$ , zatem dla pewnej zmiennej losowej  $Z \sim \mathcal{N}(0, 1)$  możemy powiedzieć, że

$$\frac{S - np_0}{\sqrt{np_0(1 - p_0)}} \stackrel{d}{\approx} Z$$

pod warunkiem, że  $H_0$  jest prawdziwa. Korzystając z tego, zapisujemy

$$P\left(-z < \frac{S - np_0}{\sqrt{np_0(1 - p_0)}} < z\right) \approx 1 - \alpha,$$

gdzie  $z$  jest kwantylem rzędu  $1 - \frac{\alpha}{2}$  rozkładu  $\mathcal{N}(0, 1)$ . Podnosząc strony nierówności wewnątrz funkcji prawdopodobieństwa do kwadratu dostajemy

$$P\left(\frac{(S - np_0)^2}{np_0(1 - p_0)} < z^2\right) \approx 1 - \alpha,$$

a następnie rozwiązując ukrytą wewnątrz nierówność kwadratową względem  $p_0$  otrzymamy

$$P(p_0 \in W) \approx 1 - \alpha,$$

gdzie

$$W = \left[ \frac{S + \frac{1}{2}z^2}{n + z^2} - \frac{z}{n + z^2} \sqrt{\frac{S(n - S)}{n} + \frac{z^2}{4}}, \quad \frac{S + \frac{1}{2}z^2}{n + z^2} + \frac{z}{n + z^2} \sqrt{\frac{S(n - S)}{n} + \frac{z^2}{4}} \right]$$

nazywamy przedziałem Wilsona. Widzimy zatem, że przedział ten będzie zmieniać się w zależności od wartości  $S$  i jeśli hipoteza zerowa jest prawdziwa,  $p_0$  będzie wpadać do niego z częstością zbliżoną do  $1 - \alpha$ .

### 3. Test oparty o przedział Cloppera-Pearsona

Oznaczmy kwantyl rzędu  $\gamma$  rozkładu  $\mathcal{Beta}(\alpha, \beta)$  jako  $b_\gamma(\alpha, \beta)$ . Dla pewnej realizacji zmiennej  $S$ , przedział Cloppera-Pearsona przedstawia się w postaci

$$CP = \left[ b_{\frac{\alpha}{2}}(S, n - S + 1), \quad b_{1 - \frac{\alpha}{2}}(S + 1, n - S) \right],$$

gdzie

$$P(p_0 \in CP) \approx 1 - \alpha$$

pod warunkiem prawdziwości hipotezy zerowej.

### 4. Test oparty o przedział Jeffreysa

Przedział ten definiuje się podobnie jak przedział Cloppera-Pearsona, z tą różnicą, że tutaj korzystamy wyłącznie z kwantyli rozkładu  $\mathcal{B}(S + 0.5, n - S + 0.5)$ . Dla pewnej realizacji zmiennej  $S$  ma on zatem następującą postać:

$$J = \left[ b_{\frac{\alpha}{2}}(S + 0.5, n - S + 0.5), b_{1-\frac{\alpha}{2}}(S + 0.5, n - S + 0.5) \right],$$

gdzie

$$P(p_0 \in J) \approx 1 - \alpha$$

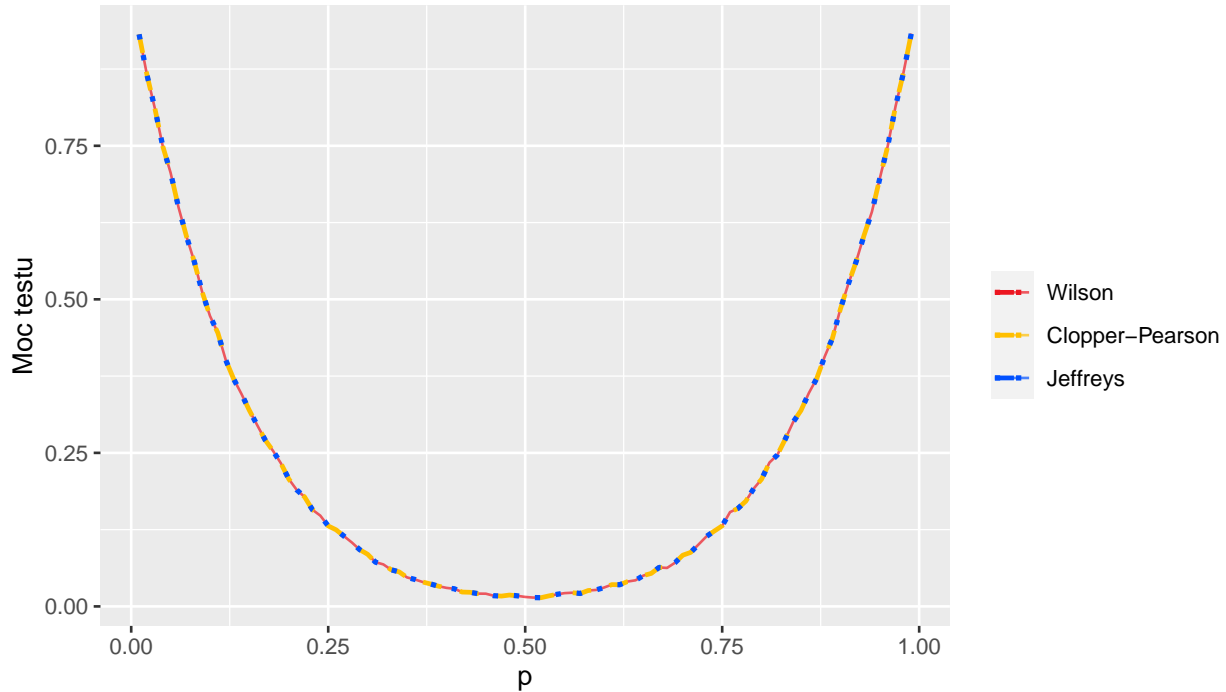
pod warunkiem, że hipoteza  $H_0$  jest prawdziwa.

## 5. Porównanie testów

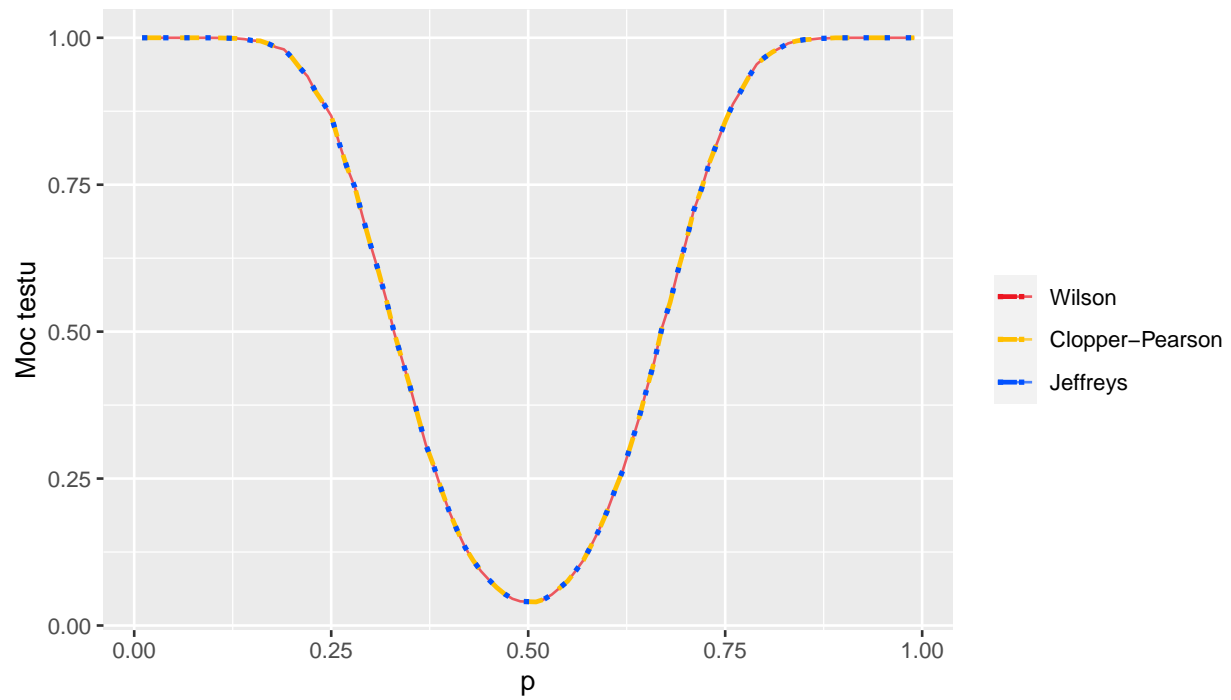
W powyższych sekcjach przedstawione zostały trzy różne sposoby na testowanie parametru  $p$  w rozkładzie dwumianowym. Odpowiemy teraz na pytanie, który z nich oferuje największą moc. W tym celu skorzystamy z metody Monte Carlo. Dla kolejnych wartości  $p \in (0, 1)$  wykonujemy następujące kroki:

1. generujemy  $N = 10\,000$  realizacji zmiennej losowej  $S \sim \mathcal{B}(n, p)$ ,
2. dla każdego  $S$  z próbki testujemy hipotezę  $H_0 : p = p_0 = 0.5$  na poziomie istotności  $\alpha = 0.05$ , przy pomocy wszystkich trzech testów,
3. dla każdego z testów zliczamy ile razy hipoteza zerowa została zaakceptowana i dzielimy tę liczbę przez  $N$ , co daje nam przybliżoną wartość mocy dla danego  $p$ .

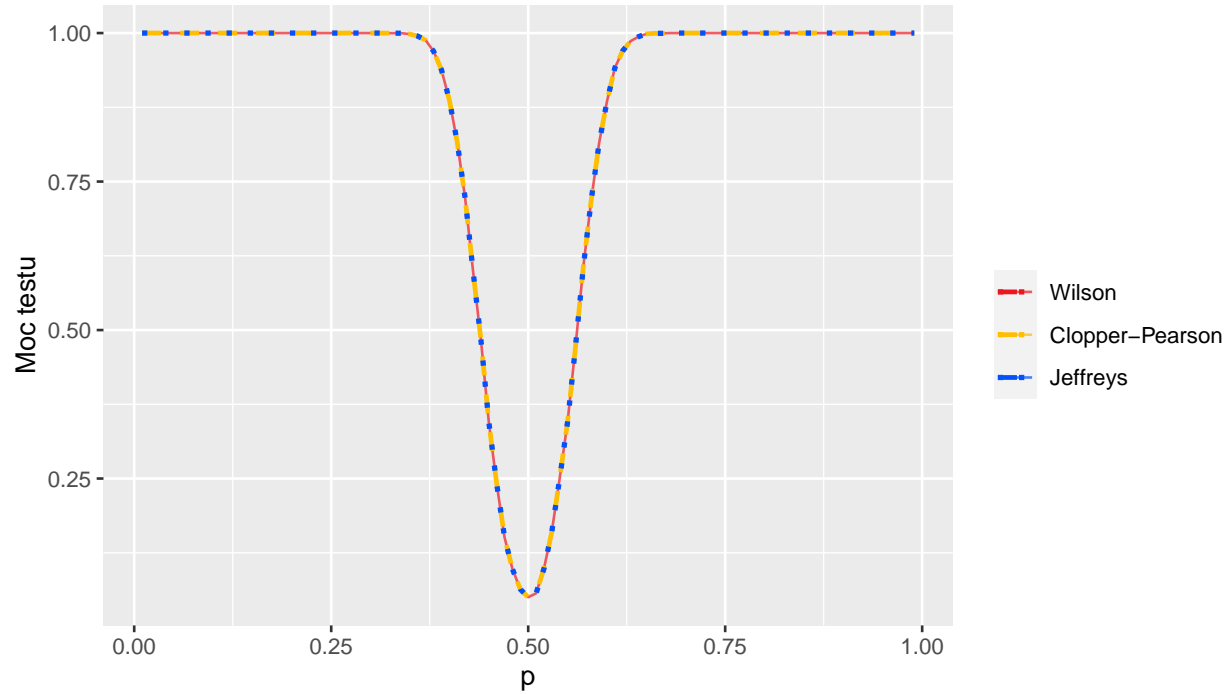
Porównanie mocy wykonamy dla  $n \in \{7, 35, 250\}$ .



Wykres 1: Wykres zależności mocy testów od parametru  $p$  dla  $n = 7$ .

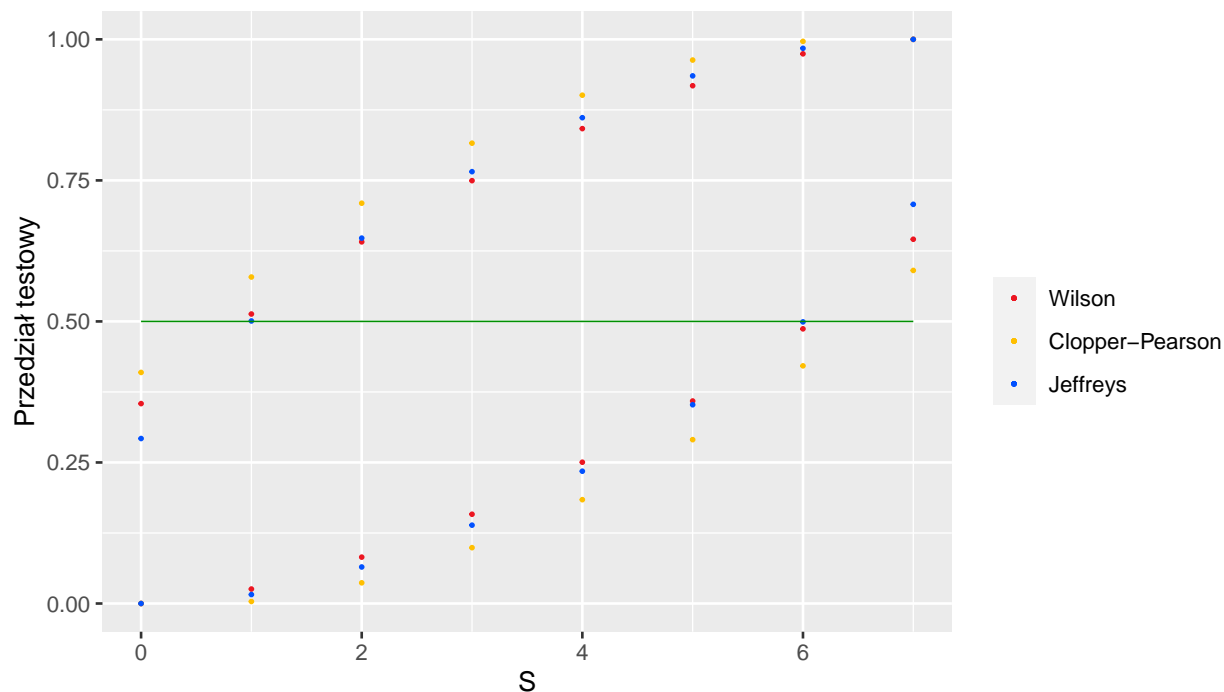


Wykres 2: Wykres zależności mocy testów od parametru  $p$  dla  $n = 35$ .



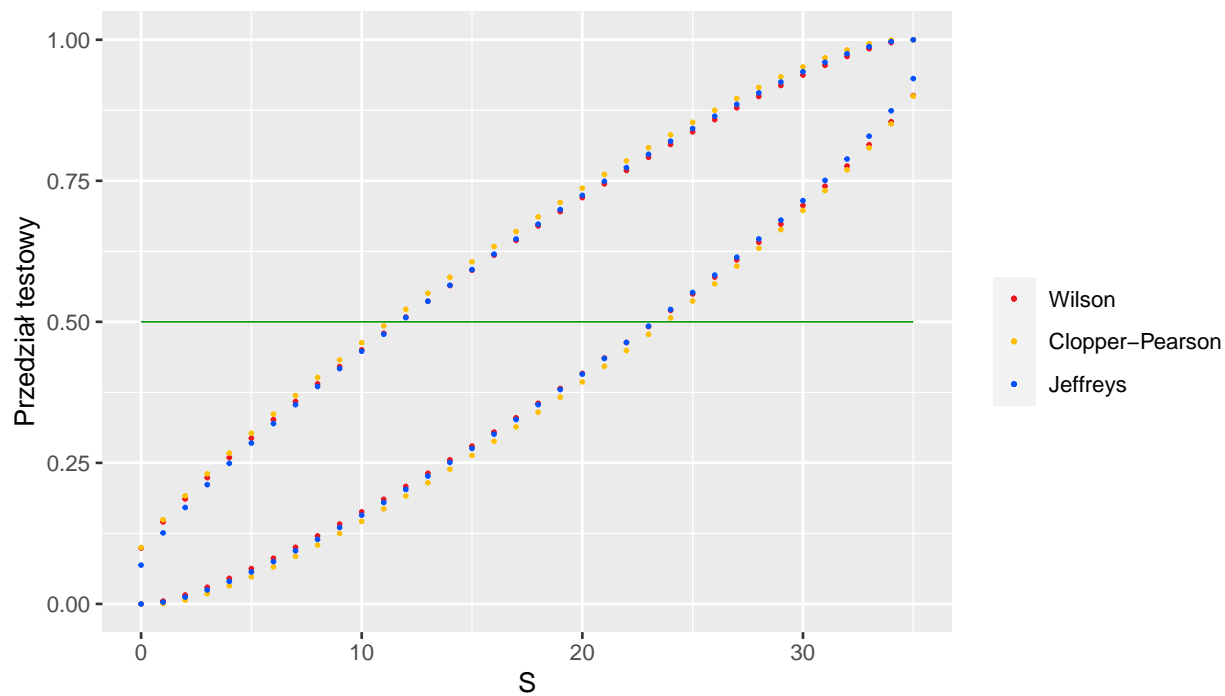
Wykres 3: Wykres zależności mocy testów od parametru  $p$  dla  $n = 250$ .

Na wykresach 1, 2 i 3 dostrzec można, że krzywe mocy testów dla każdego  $n$  wyraźnie się pokrywają. Jak się okazuje, wyliczone wartości mocy są identyczne dla wszystkich trzech testów. Wynik ten, na pierwszy rzut oka, może budzić pewne wątpliwości, jednakże jest on w pełni poprawny. Główną przyczyną takich samych wartości mocy jest to, że przeprowadzane testy dotyczą rozkładu dwumianowego, który jest rozkładem dyskretnym. Oznacza to, że zmienna  $S \sim \mathcal{B}(n, p)$  przyjmuje ograniczoną liczbę wartości. Przykładowo dla  $n = 7$ , zmienna  $S$  przyjmuje jedynie wartości  $0, 1, 2, \dots, 7$ . Zatem w tym przypadku mamy 8 możliwych przedziałów.

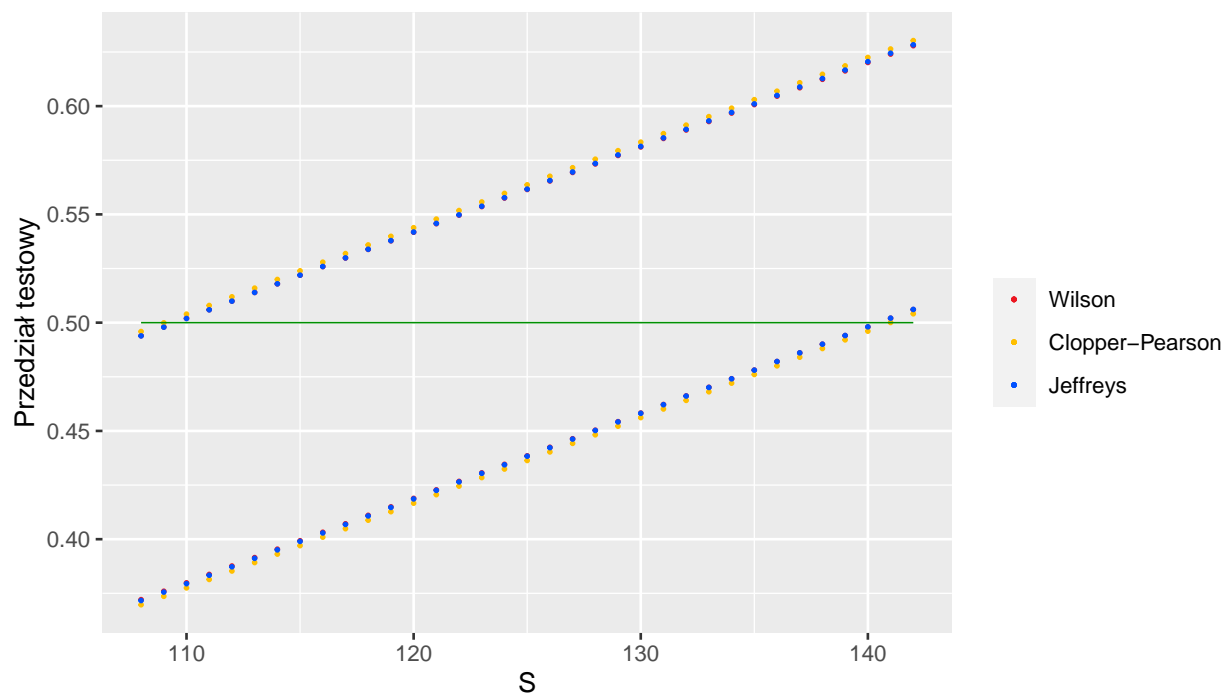


Wykres 4: Przedziały testowe w zależności od wartości  $S$  dla  $n = 7$  z wartością  $p_0 = 0.5$  oznaczoną zieloną linią.

Okazuje się, że gdy  $S$  przyjmuje wartości od 1 do 6, wszystkie z trzech testowych przedziałów zawierają wartość  $p_0 = 0.5$ , co prowadzi do akceptacji  $H_0$ . Z kolei dla  $S = 0$  i  $S = 7$  wartość 0.5 nie wpada do żadnego z przedziałów (wykres 4). Stąd każdy z testów zachowuje się dokładnie tak samo.



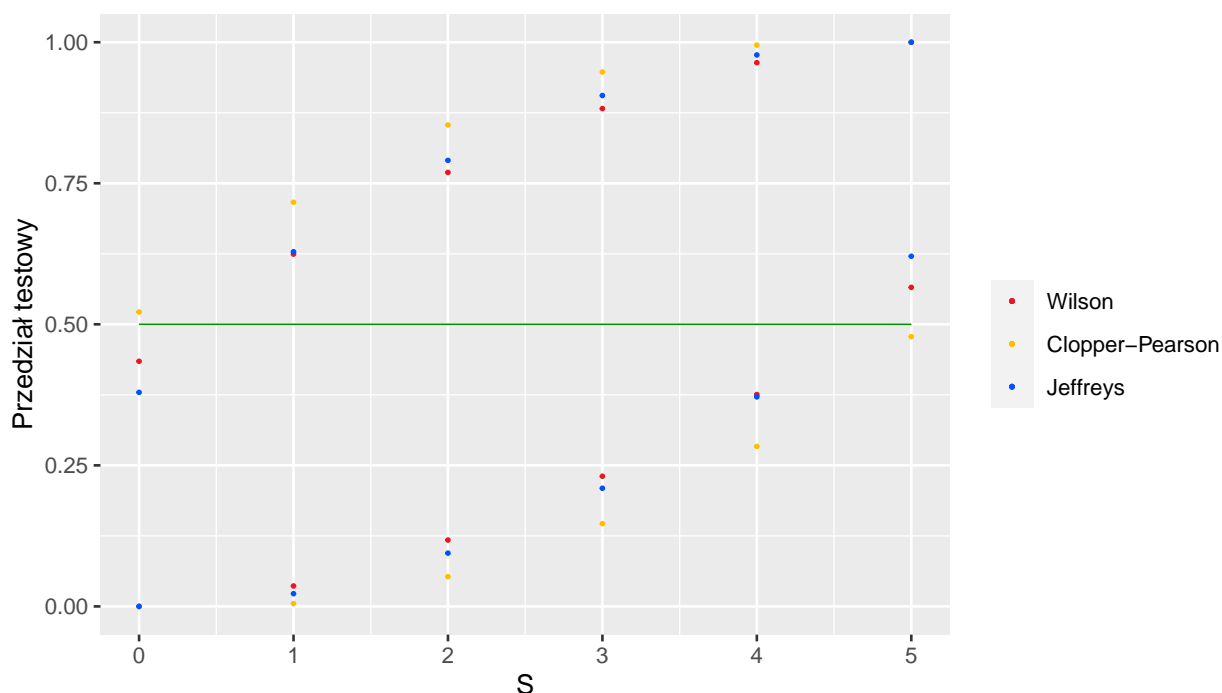
Wykres 5: Przedziały testowe w zależności od wartości  $S$  dla  $n = 35$  z wartością  $p_0 = 0.5$  oznaczoną zieloną linią.



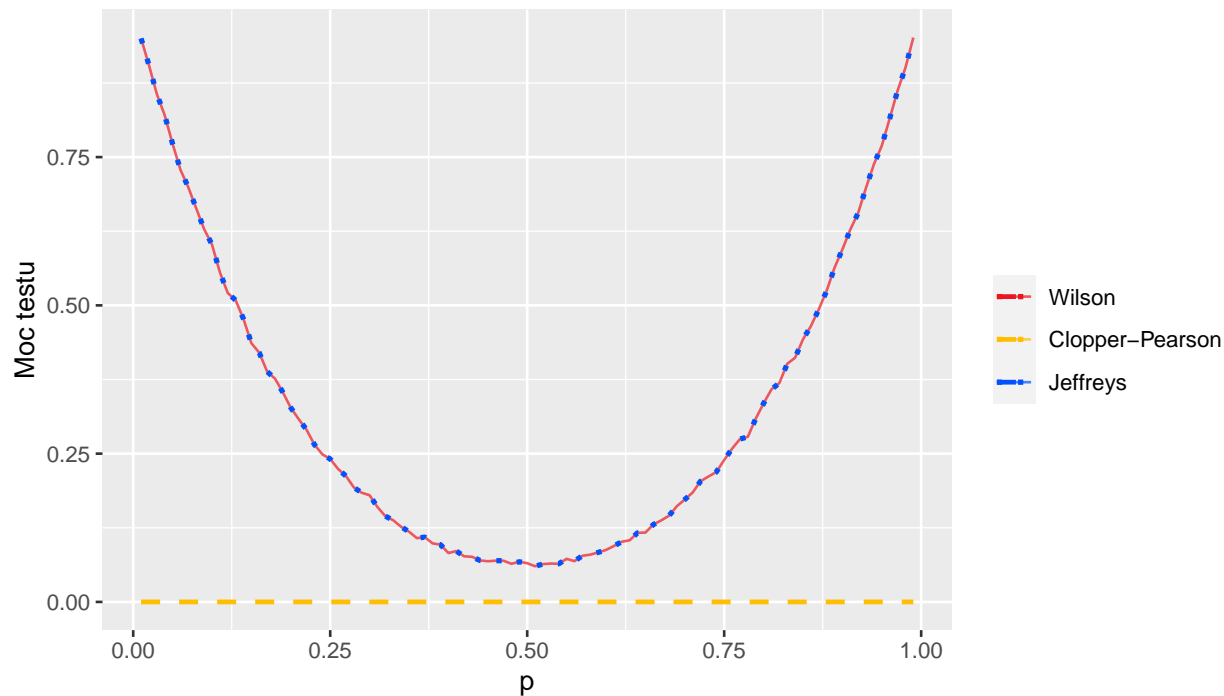
Wykres 6: Przedziały testowe w zależności od wartości  $S$  dla  $n = 250$  z wartością  $p_0 = 0.5$  oznaczoną zieloną linią.

Na wykresach 5 i 6 zobaczyć można jak wyglądają przedziały trzech testów w zależności od wartości  $S$  w kolejnych dwóch przypadkach, czyli  $n = 35$  oraz  $n = 250$ . Jak się okazuje, mamy do czynienia z tą samą sytuacją. Nie występuje ani jedna wartość  $S$ , dla której któryś z przedziałów zawierałby 0.5, a któryś nie (taka sytuacja spowodowałaby różnicę w mocy). Zatem dla rozpatrywanych przez nas wartości  $n$  oraz  $\alpha = 0.05$  nie da się wyłonić testu jednostajnie najmocniejszego, ponieważ wszystkie testy są równe. Nie oznacza to jednak, że jest tak dla każdego  $n$  i dowolnego poziomu istotności  $\alpha$ . Zwiększenie wartości  $\alpha$  spowodowałoby zwiększenie długości przedziałów, co mogłoby skutkować tym, że któryś z przedziałów “zjadłby” 0.5. Na podstawie wykresów możemy wywnioskować, że tym przedziałem byłby ten pochodzący z testu Cloppera-Pearsona, ponieważ jest on największy.

Rozpatrzmy jeszcze przypadek  $n = 5$ . Na wykresie 7 dostrzec można coś ciekawego. Mianowicie przedział Cloppera-Pearsona dla każdego  $S$  zawiera  $p_0$ , czego nie można powiedzieć o dwóch pozostałych przedziałach. Takie zachowanie będzie prowadziło do sytuacji, w której niezależnie jaką wartość przyjmie  $S$ , test Cloppera-Pearsona będzie akceptował fałszywą hipotezę zerową (czyli będzie popełniał błąd II rodzaju), skutkiem czego test ten będzie miał moc równą zero dla każdego  $p \in (0, 1)$ , co zobaczyć możemy na wykresie 8.



Wykres 7: Przedziały testowe w zależności od wartości  $S$  dla  $n = 5$  z wartością  $p_0 = 0.5$  oznaczoną zieloną linią.



Wykres 8: Wykres zależności mocy testów od parametru  $p$  dla  $n = 5$ .

## 6. Podsumowanie

Rozważaliśmy trzy testy dla parametru  $p$  rozkładu dwumianowego  $\mathcal{B}(n, p)$  oparte o następujące przedziały:

- przedział Wilsona,
- przedział Cloppera-Pearsona,
- przedział Jeffreysa.

Aby stwierdzić, który z tych testów jest jednostajnie najmocniejszy, przeprowadziliśmy symulację mocy dla  $n \in \{7, 35, 250\}$ . Jak się okazało, wszystkie testy były tak samo mocne. Nie oznacza to jednak, że dla innych wartości  $n$  jest tak samo - dla  $n = 5$  przeprowadziliśmy analogiczną symulację. Okazało się, że w tym przypadku można wyłonić test najsłabszy - test Cloppera-Pearsona.