

Pakiety statystyczne

Sprawozdanie 2

Paulina Muszyńska, Karolina Rakus

23 stycznia 2022

1 Wstęp

W poniższym sprawozdaniu zajmiemy się testowaniem hipotez w przypadku próbek z rozkładu dwumianowego. Rozważamy testy na poziomie istotności $\alpha = 0.05$.

Testujemy hipotezę zerową $H_0 : p = 0.5$,
przeciwko $H_1 : p > 0.5$.

2 Stosowane testy

Będziemy wykonywać testowanie hipotez stosując trzy testy:

1. test oparty o przedział Wilsona,
2. test oparty o przedział Cloppera-Pearsona,
3. test oparty o przedział Jeffreysa.

3 Standardowy przedział aproksymacji

W pierwszej kolejności zapoznajmy się ze **standardowym przedziałem aproksymacji** - to on będzie "wyjściową" stosowanych przez nas testów. W tym przypadku przybliżamy rozkład dwumianowy rozkładem normalnym.

Używając przybliżenia rozkładem normalnym mamy prawdopodobieństwo p zapisane jako:

$$\hat{p} \pm z \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

lub równoważnie

$$\frac{n_S}{n} \pm \frac{z}{n\sqrt{n}} \sqrt{n_S n_F}$$

gdzie:

n - liczba prób,

n_S - liczba sukcesów,

$\hat{p} = \frac{n_S}{n}$ - proporcja sukcesów do ogólnej liczby prób,

$n_F = n - n_S$ - liczba porażek,

$z - 1 - \frac{1}{\alpha}$ kwantyl rozkładu normalnego $N(0, 1)$.

Dla przedziału ufności wyznaczonego z prawdopodobieństwem 95% błąd α to $1 - 0.95 = 0.05$, zatem $1 - \frac{\alpha}{2} = 0.975$, więc $z = 1.96$.

4 Test oparty o przedział Wilsona

Pierwszy test, którego będziemy używać, to **test oparty o przedział Wilsona**. Przedział ten został skonstruowany po raz pierwszy przez Edwina Bidwella Wilsona w 1927 roku. W przeciwieństwie do symetrycznego normalnego przedziału aproksymacji opisanego powyżej, test oparty o przedział Wilsona jest **asymetryczny**. W odróżnieniu do poprzednio zaproponowanego przedziału, w tym przypadku nie ma problemu z "przestrzeleniem". Test ten możemy stosować także w przypadku małych próbek oraz obserwacji skośnych. Obserwowane przez nas prawdopodobieństwo pokrycia jest bliższe wartości nominalnej, czyli $1 - \alpha$.

Wilson rozpoczął od przybliżenia rozkładem normalnym:

$$z \approx \frac{(p - \hat{p})}{\sigma_n}$$

oraz użył analitycznego wzoru na odchylenie standardowe:

$$\sigma_n = \sqrt{\frac{p(1 - p)}{n}}$$

Po przyrównaniu dwóch powyższych wzorów oraz podniesieniu obu stron do kwadratu dostajemy równanie kwadratowe o zmiennej p :

$$(p - \hat{p})^2 = z^2 \cdot \frac{p(1-p)}{n}$$

Przekształcamy równanie do standardowej formy równania kwadratowego. Traktujemy \hat{p} oraz n jako znane wartości z próbki. Używamy wartości z do szacowania p z odpowiednim prawdopodobieństwem:

$$\left(1 + \frac{z^2}{n}\right)p^2 + \left(-2\hat{p} - \frac{z^2}{n}\right)p + (\hat{p}^2) = 0$$

gdzie wszystkie wartości podane w nawiasach są wartościami znanymi. Rozwiązanie dla p oszacowuje górną i dolną granicę przedziału ufności dla p i dane jest jako:

$$\frac{1}{1 + \frac{z^2}{n}} \left(\hat{p} + \frac{z^2}{2n} \right) \pm \frac{z}{1 + \frac{z^2}{n}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} \frac{z^2}{4n^2}}$$

lub równoważnie:

$$p \approx \frac{n_S + \frac{1}{2}z^2}{n + z^2} \pm \frac{z}{n + z^2} \sqrt{\frac{n_S n_F}{n} + \frac{z^2}{4}}.$$

Praktyczna korzyść z zastosowania tego przedziału jest taka, że ma on dobre właściwości nawet dla niewielkiej liczby prób oraz ekstremalnego prawdopodobieństwa (czyli takiego, które jest bliskie 0 albo 1).

Oprócz analitycznego rozwiązania powyższego równania kwadratowego, przedział Wilsona może być ustalony także numerycznie:

$$p_{k+1} = \hat{p} \pm z \cdot \sqrt{\frac{p_k \cdot (1 - p_k)}{n}}$$

gdzie $p_0 = \hat{p}$.

5 Test oparty o przedział Cloppera-Pearsona

Test oparty o przedział Cloppera-Pearsona jest często stosowaną metodą przy wyznaczaniu przedziału ufności w przypadku rozkładu dwumianowego. Zdarza się, że testowanie w oparciu o ten przedział jest nazywane

metodą "dokładną", ponieważ bazuje prawdopodobieństwie rozkładu dwumianowego, a nie na jego przybliżeniu.

Przedział Cloppera-Pearsona może być zapisany jako:

$$S_{\geq} \cap S_{\leq}$$

lub równoważnie:

$$(\inf S_{\geq}, \sup S_{\leq})$$

a S_{\leq} oraz S_{\geq} mamy dane jako:

$$S_{\leq} := \{\theta | P[Bin(n; \theta) \leq x] > \frac{\alpha}{2}\}$$

$$S_{\geq} := \{\theta | P[Bin(n; \theta) \geq x] > \frac{\alpha}{2}\}$$

gdzie:

x - liczba sukcesów,

n - liczba prób,

θ - prawdopodobieństwo sukcesu,

w próbie o rozkładzie dwumianowym zapisanym jako $Bin(n; \theta)$.

Ze względu na relację pomiędzy rozkładem dwumianowym a rozkładem beta przedział Cloppera-Pearsona może być przedstawiony w alternatywny sposób z użyciem kwantyli rozkładu beta:

$$B\left(\frac{\alpha}{2}; x, n - x + 1\right) < \theta < B\left(1 - \frac{\alpha}{2}; x + 1, n - x\right)$$

gdzie:

x - liczba sukcesów,

n - liczba prób,

$B(p; v, w)$ - p -ty kwantyl rozkładu beta z parametrami v i w .

Ponadto dla przedziału ufności $(\theta_{min}, \theta_{max})$ mamy:

$$\frac{\Gamma(n+1)}{\Gamma(x)\Gamma(n-x+1)} \int_0^{\theta_{min}} t^{x-1}(1-t)^{n-x} dt = \frac{\alpha}{2}$$

$$\frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x)} \int_0^{\theta_{max}} t^x(1-t)^{n-x-1} dt = 1 - \frac{\alpha}{2}$$

W przypadku gdy liczba sukcesów x jest równa liczbie prób n przedział można zapisać jako:

$$\left(\left(\frac{\alpha}{2} \right)^{\frac{1}{n}}, 1 \right),$$

natomiast dla $x = 0$ mamy:

$$\left(0, 1 - \left(\frac{\alpha}{2} \right)^{\frac{1}{n}} \right).$$

Przedział Cloppera-Pearsona może być również zmodyfikowany w taki sposób, by uzyskać przedział ufności dla rozkładu innego niż dwumianowy. Przykładowo, może być stosowny w przypadku, gdy próbki pobierane są bez zwracania z populacji o znanej wielkości, zamiast jak w naszych przypadkach z powtarzalnych losowań rozkładu dwumianowego. W takim przypadku jako rozkład bazowy mielibyśmy dany rozkład hipergeometryczny.

6 Test oparty o przedział Jeffreysa

Przedział Jeffreysa pochodzi z podejścia bayesowskiego, ale ma dobre właściwości częstościowe. Właściwości pokrycia ma podobne do tych, które wykazywał przedział Wilsona. Oprócz tego jest jednym z nielicznych przedziałów, których zaletą jest równostronność (przykładowo - dla przedziału ufności z prawdopodobieństwem wynoszącym 95 % prawdopodobieństwa, które nie wpadają do przedziału ufności z góry są bliskie tym z dołu i wynoszą po około 2.5 %. W przypadku przedziału Wilsona, był on zbyt wyśrodkowany w okolicy $p = 0.5$.

Założmy, że $X \sim \text{Bin}(n, p)$ oraz, że p ma początkowy rozkład $\text{Beta}(a_1, a_2)$, rozkład p wynosi zatem:

$$\text{Beta}(X + a_1, n - X + a_2)$$

Przedział Jeffreysa jest dany jako:

$$\left[B\left(\frac{\alpha}{2}; X + a_1, n - X + a_2\right), B\left(1 - \frac{\alpha}{2}; X + a_1, n - X + a_2\right) \right],$$

gdzie:

$B(\alpha; m_1, m_2)$ - α -ty kwantyl rozkładu $\text{Beta}(m_1, m_2)$.

7 Wprowadzenie do zadań

W poniższych zadaniach naszym zadaniem będzie wykonanie wykresu funkcji mocy przy pomocy symulacji Monte Carlo. Przypomnijmy sobie w pierwszej kolejności czym jest funkcja mocy, oraz zaimplementujmy odpowiednie funkcje. **Moc testu** to prawdopodobieństwo uniknięcia błędu drugiego rodzaju, czyli przyjęcia hipotezy zerowej, gdy w rzeczywistości jest ona fałszywa.

Implementujemy funkcje - zwracającą moc testu oraz rysującą wykres:

```
moc <- function(n, p, method, alpha=0.05, MC=1000, p0=0.5) {  
  mean(sapply(1:MC, function(...) {  
    S <- rbinom(1, n, p)  
    wyniki <- BinomCI(S, n, method=method, sides="left")  
    L <- wyniki[1,2]  
    R <- wyniki[1,3]  
    ile_razy <- sum( L <= p0 && p0 <= R)  
    dziala <- prop.test(ile_razy, 1, p=0.95)$p.value<alpha  
  })))  
}
```

```
draw <- function(n, p0=0.5) {  
  
  p <- seq(0,1, 0.01)  
  
  if (p0 == 0.5) {  
    pos = c(0.15, 0.85)  
  }  
  else {  
    pos = c(0.85,0.2)  
  }  
  
  moc_wilson <- sapply(p, function(x) {  
    moc(n,x, "wilson", p0=p0)  
  })  
  
  moc_clopper <- sapply(p, function(x) {
```

```

    moc(n,x, "clopper-pearson", p0=p0)
  })

moc_jeffreys <- sapply(p, function(x) {
  moc(n,x, "jeffreys", p0=p0)
})

ds <- data.frame(p, moc_wilson, moc_clopper, moc_jeffreys)

colors <- c("Wilson" = "blue",
            "Clopper-Pearson" = "red",
            "Jeffreys"="green")

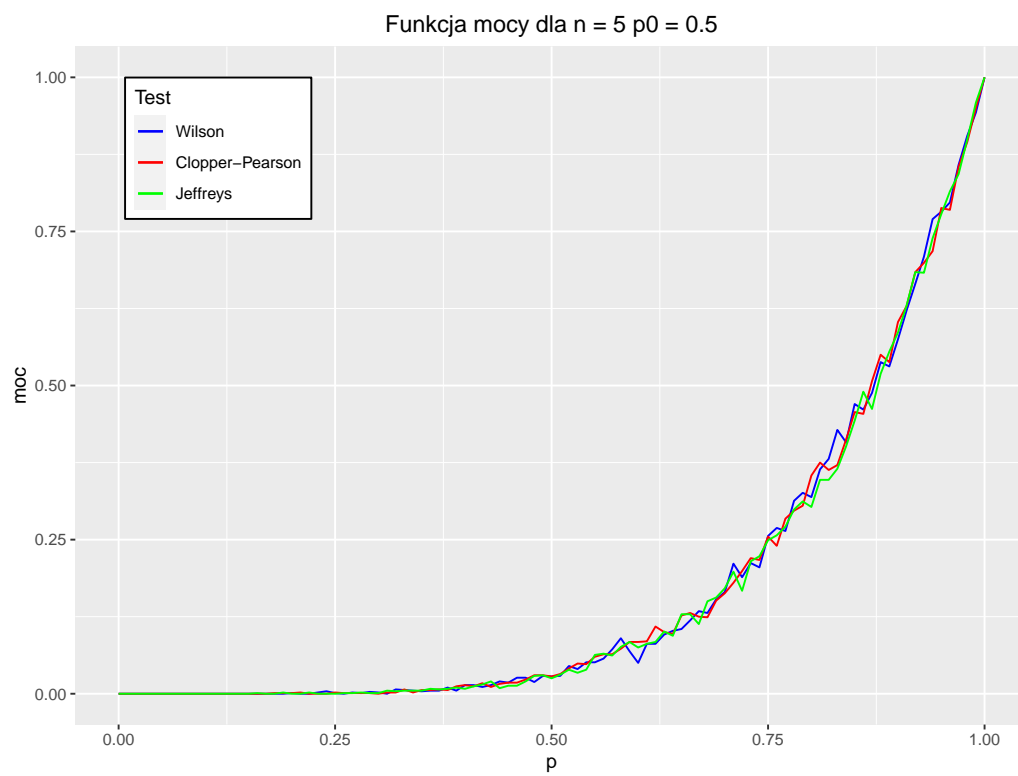
ggplot(ds, mapping = aes(x = p)) +
  geom_line(aes(y = moc_wilson, color="Wilson")) +
  geom_line(aes(y = moc_clopper, color="Clopper-Pearson")) +
  geom_line(aes(y = moc_jeffreys, color="Jeffreys")) +
  ggtitle(sprintf("Funkcja mocy dla n = %s p0 = %s", n, p0)) +
  theme(plot.title = element_text(hjust = 0.5)) +
  labs(x = "p", y = "moc") +
  theme(legend.position = pos,
        legend.background = element_rect(fill = "white",
                                           color = "black")) +
  scale_color_manual(name = "Test", values = colors)
}

```

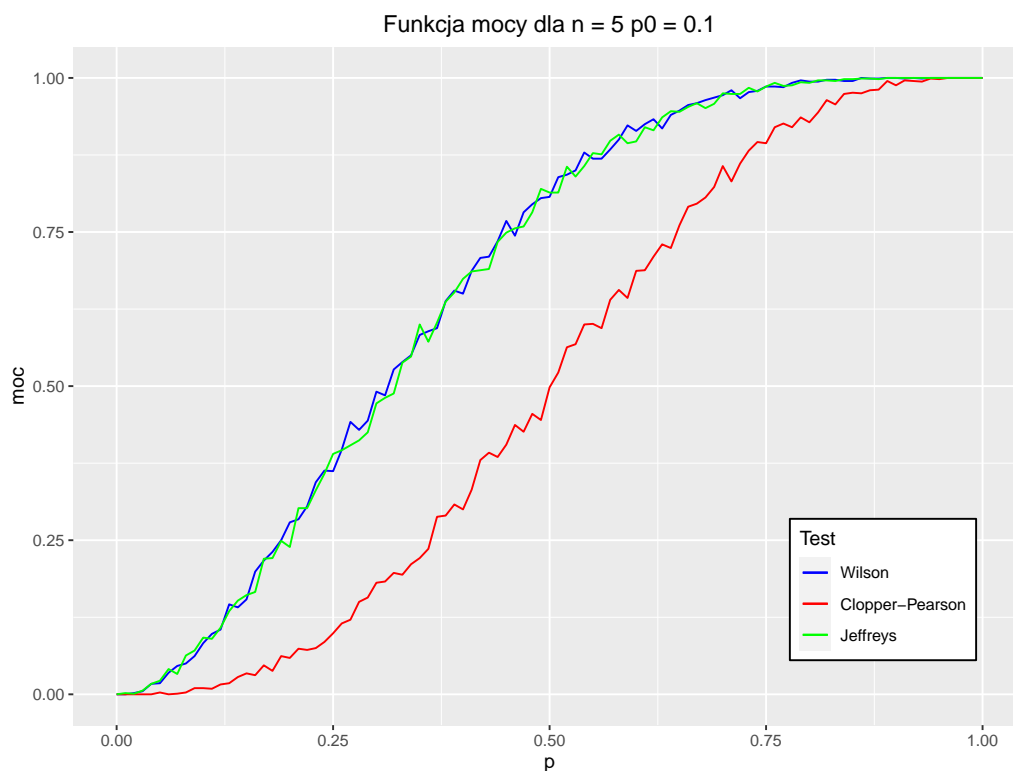
8 Zadanie 1

Zaobserwowaliśmy $S \sim B(5, p)$. Korzystając z symulacji Monte Carlo wykonaj wykres funkcji mocy w zależności od p na przedziale $(0, 1)$ dla wszystkich trzech testów. Czy istnieje test jednostajnie najmocniejszy spośród nich?

Przy użyciu wcześniej zaimplementowanych funkcji rysujemy wykres funkcji mocy:



Rysunek 1: Funkcja mocy dla $n = 5$, $p_0 = 0.5$

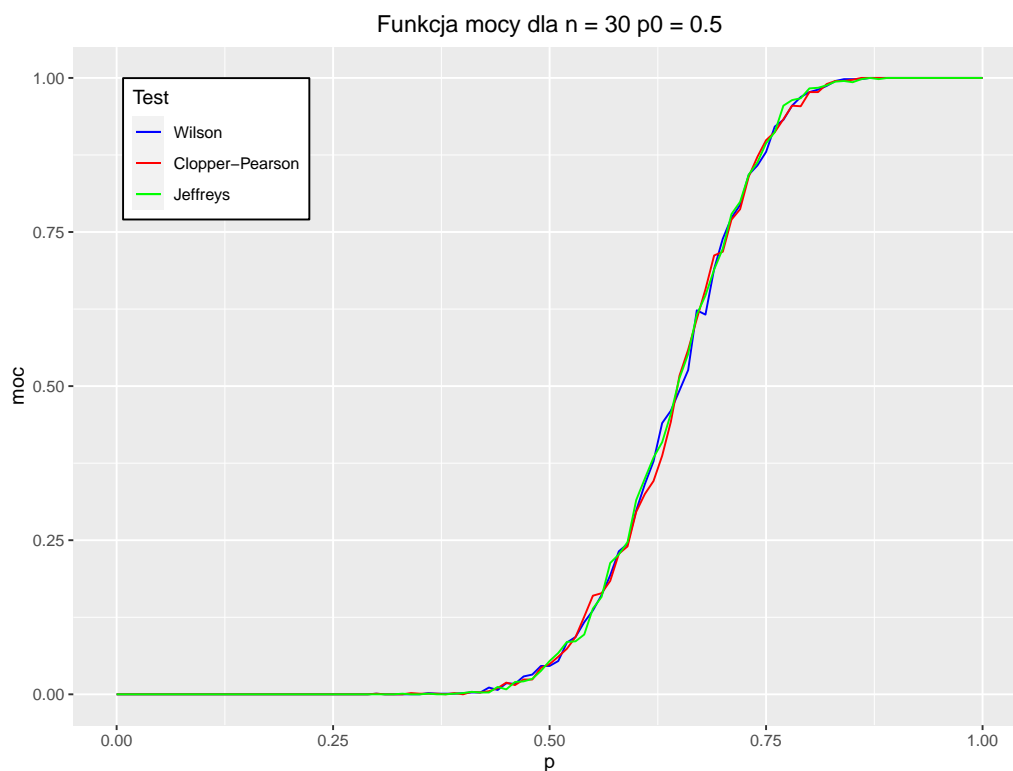


Rysunek 2: Funkcja mocy dla $n = 5$, $p_0 = 0.1$

W tym przypadku 1 mamy do czynienia z małą próbką oraz z prawdopodobieństwem, które nie jest skrajne.

Widzimy, że funkcje mocy dla przedziałów ustalonych każdą z trzech metod są do siebie bardzo podobne, mimo, że w tym przypadku badana próbka jest mała ($n = 5$). Wynika to hipotezy zerowej - środkowa wartość prawdopodobieństwa jest "łatwa w obsłudze". W każdym z tych przypadków moc testu na poziomie 80% uzyskiwana jest w okolicy $p = 0.9$.

Zmieńmy wartość p na bardziej ekstremalną, weźmy $p = 0.1$. W tym przypadku 2 widzimy, że test oparty o przedział Cloppera-Pearsona radzi sobie zdecydowanie gorzej niż testy oparte o przedziały Wilsona oraz Jeffreysa. W przypadku tego pierwszego moc na poziomie 80% uzyskujemy dopiero w okolicy $p = 0.7$ natomiast jeżeli chodzi o Jeffreysa i Wilsona ten wynik osiągany jest w okolicy $p = 0.5$. Różnica będzie tym bardziej widoczna, im wartość podana w hipotezie zerowej będzie bliższa zeru lub jedynce.



Rysunek 3: Funkcja mocy dla $n = 30$, $p_0 = 0.5$

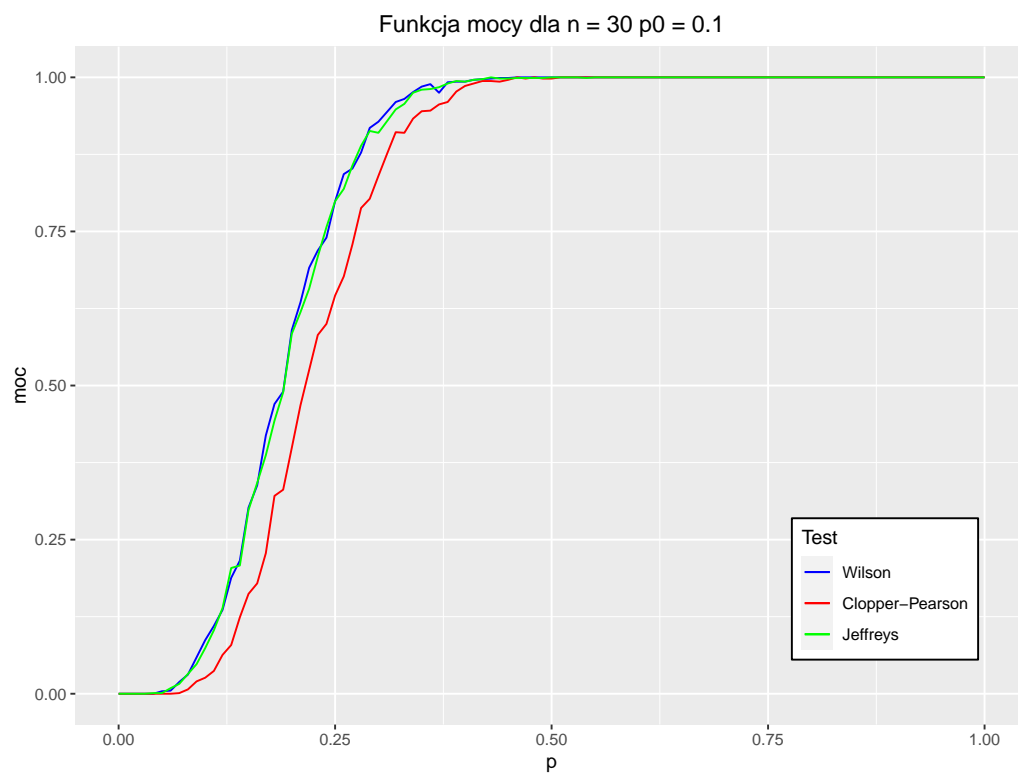
9 Zadanie 2

Zaobserwowaliśmy $S \sim B(30, p)$. Korzystając z symulacji Monte Carlo wykonaj wykres funkcji mocy w zależności od p na przedziale $(0, 1)$ dla wszystkich trzech testów. Czy istnieje test jednostajnie najmocniejszy spośród nich?

Przy użyciu wcześniej zaimplementowanych funkcji rysujemy wykres funkcji mocy:

W tym przypadku 3 badana przez nas próbka jest średniej wielkości ($n = 30$). Możemy zaobserwować, że każdy z testów radzi sobie z podobną skutecznością. Moc na poziomie 80% uzyskują w okolicy $p = 0.75$. Wykres jest niemalejący. Do $p = 0.4$ we wszystkich przypadkach funkcja jest stała i wynosi 0, potem zaczyna rosnąć.

Po zmianie 4 hipotezy zerowej na $H_0 : p = 0.1$, w dalszym ciągu widzimy,



Rysunek 4: Funkcja mocy dla $n = 30$, $p_0 = 0.1$

że test oparty o przedział Cloppera-Pearsona radzi sobie gorzej od testów bazujących na pozostałych konstruowanych przez nas przedziałach, jednak w tym przypadku nie ma między nimi już tak dużej różnicy. Moc testu na poziomie 80% uzyskiwana jest w okolicy $p = 0.25$, a dla mocniejszych w tym przypadku testów w okolicy $p = 0.2$.

10 Zadanie 3

Zaobserwowaliśmy $S \sim B(30, p)$. Korzystając z symulacji Monte Carlo wykonaj wykres funkcji mocy w zależności od p na przedziale $(0, 1)$ dla wszystkich trzech testów. Czy istnieje test jednostajnie najmocniejszy spośród nich?

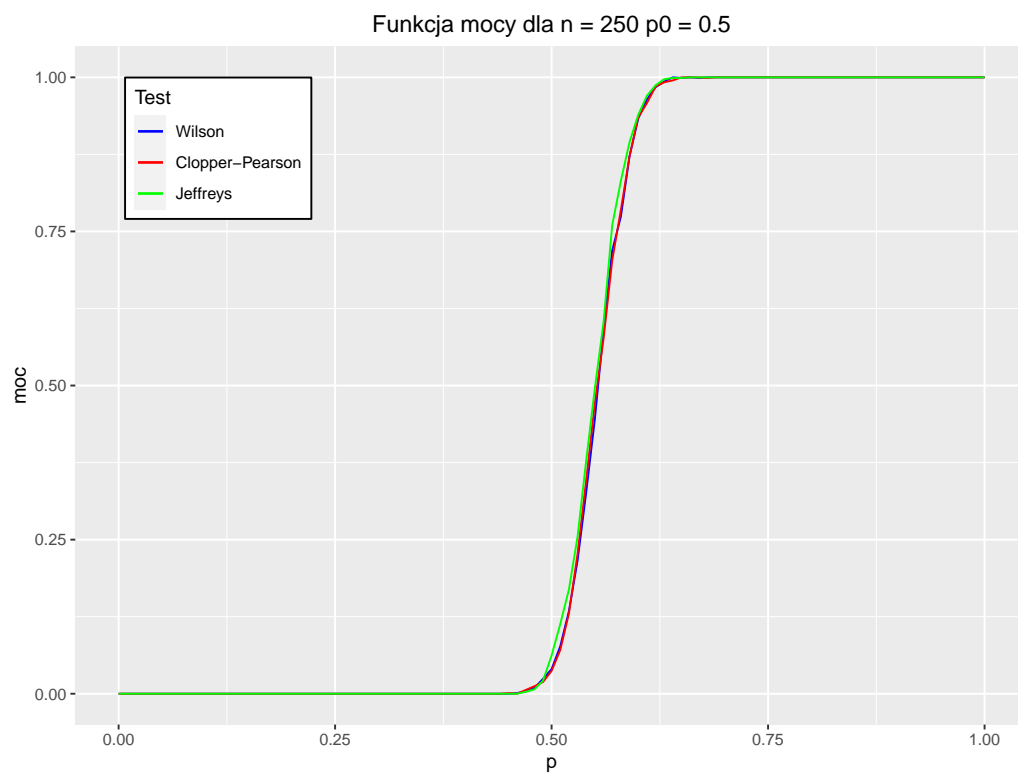
Przy użyciu wcześniej zaimplementowanych funkcji rysujemy wykres funkcji mocy:

Rozpatrywana przez nas próbka jest duża, wielkości 250. Podobnie i w tym przypadku 5 każdy z testów dla środkowego prawdopodobieństwa $p = 0.5$ wykazuje się podobną skutecznością. W tym przypadku zmiana mocy testów dokonuje się na niewielkim przedziale p - do $p \approx 0.45$ funkcje mają wartość 0 a już w okolicy $p \approx 0.6$ funkcje mocy są bliskie 1.

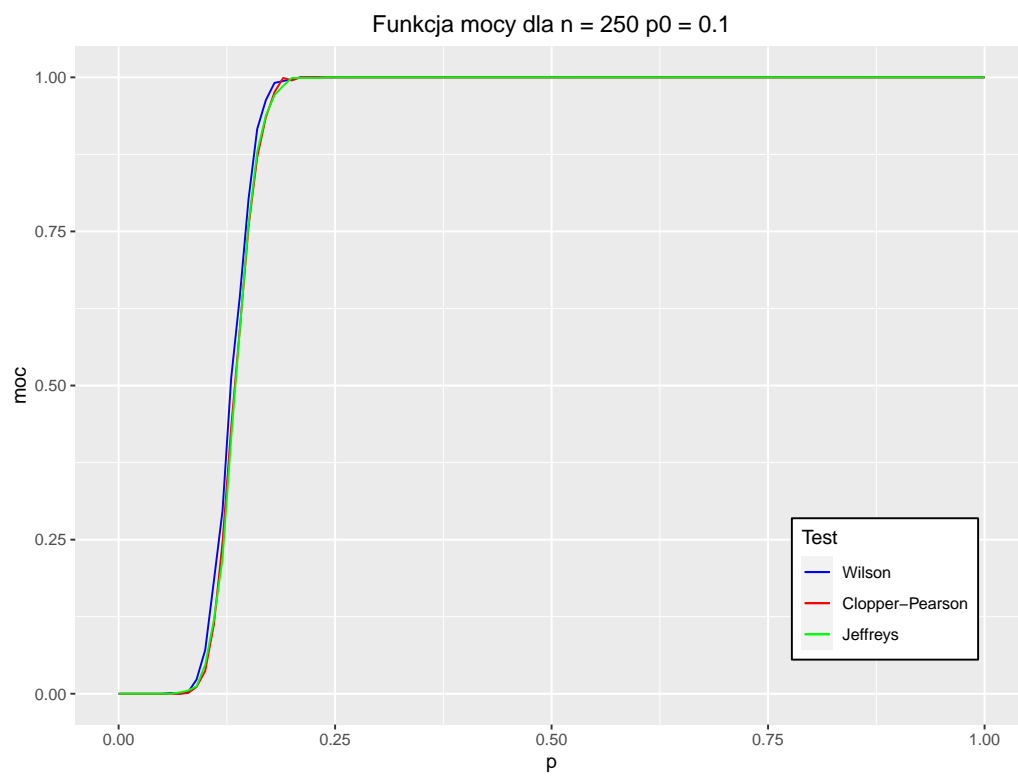
W przypadku 6 bardziej skrajnej wartości prawdopodobieństwa znajdującego się w hipotezie zerowej, weźmy $p = 0.1$, widzimy, że skuteczność testów dla próby tej wielkości jest w każdym przypadku zbliżona - nie ma takiej różnicy, jaka była widoczna w poprzednich zadaniach dla prób mniejszych rozmiarów. W okolicy $p = 0.1$ następuje gwałtowny wzrost funkcji mocy.

11 Podsumowanie

W rozpatrywanym przez nas przypadku hipoteza alternatywna jest jednostronna, dlatego otrzymywane przez nas funkcje są rosnące. Im próba jest liczniejsza, tym funkcja gwałtowniej rośnie w okolicy prawdopodobieństwa zadanego w hipotezie zerowej. Ze względu na prawdopodobieństwo $p = 0.5$ nie widać różnic w skuteczności testów, co można zauważyć natomiast w przypadkach, gdy za hipotezę alternatywną podstawiona zostaje bardziej skrajna wartość np. $p = 0.1$. Po przeprowadzeniu analizy funkcji mocy dla takiego przypadku widzimy, że najgorzej spośród stosowanych przez nas testów radził sobie ten, który był oparty o przedział Cloppera-Pearsona.



Rysunek 5: Funkcja mocy dla $n = 250$, $p_0 = 0.5$



Rysunek 6: Funkcja mocy dla $n = 250$, $p_0 = 0.1$