

Sprawozdanie 5

Jakub Markowiak
album 255705

4 czerwca 2021

Spis treści

1	Krótki opis zagadnienia	1
2	Opis eksperymentów/analiz	1
3	Wyniki	1
3.1	Porównanie funkcji centralnie asymptotycznych w rozkładzie Poissona $\mathcal{P}(\lambda)$. . .	1
4	Podsumowanie	6

1 Krótki opis zagadnienia

2 Opis eksperymentów/analiz

Przeprowadzimy następujące analizy i eksperymenty:

1. Porównanie funkcji centralnie asymptotycznych w rozkładzie Poissona $\mathcal{P}(\lambda)$.

3 Wyniki

3.1 Porównanie funkcji centralnie asymptotycznych w rozkładzie Poissona $\mathcal{P}(\lambda)$.

W tym sprawozdaniu rozpatrujemy rozkład $\mathcal{P}(\lambda)$, gdzie λ – nieznany parametr. Skupimy się na zbadaniu własności statystyk S_1, S_2 oraz S_3 , które są funkcjami centralnie asymptotycznymi dla parametru λ :

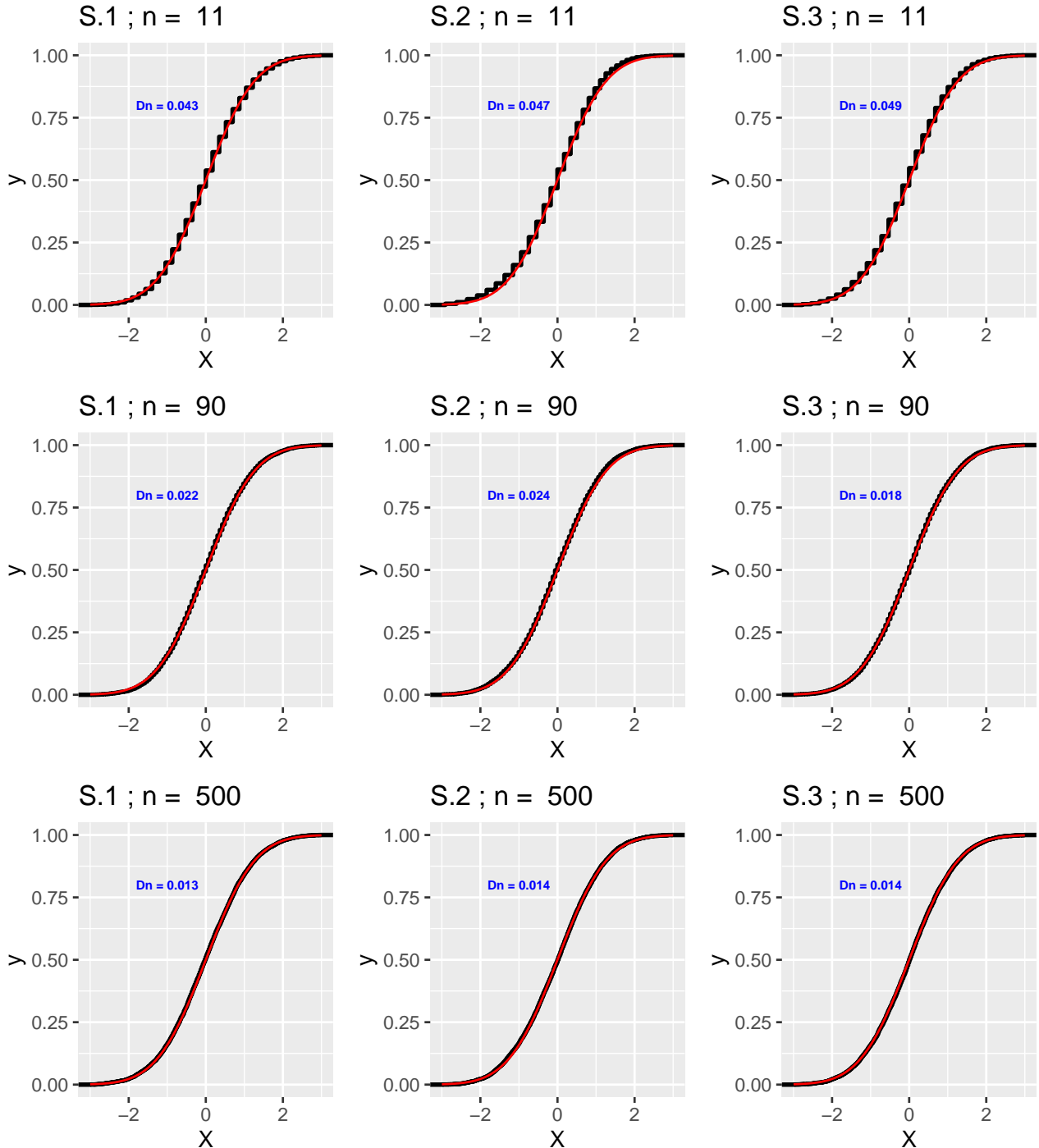
$$S_1 = \frac{\bar{X} - \lambda}{\sqrt{\lambda}} \sqrt{n}, \quad (1)$$

$$S_2 = \frac{\bar{X} - \lambda}{\sqrt{\bar{X}}} \sqrt{n}, \quad (2)$$

$$S_3 = \left(\sqrt{\bar{X}} - \sqrt{\lambda} \right) 2\sqrt{n}. \quad (3)$$

Wiemy, że statystyki S_1 , S_2 oraz S_3 zbiegają według rozkładu do $\mathcal{N}(0, 1)$. Sprawdźmy zatem, jaki powinien być minimalny rozmiar próby, aby przybliżenie rozkładem normalnym było wystarczająco dokładne.

Dystrybuanty empiryczne dla S.1, S.2, S.3



Rysunek 1: Dystrybuanty empiryczne oraz odległość Kołmogorowa dla różnych rozmiarów próby

Widzimy, że te statystyki faktycznie zbiegają według rozkładu do $\mathcal{N}(0, 1)$. Odległość Kołmogorowa D_n już dla próby rozmiaru 11 wynosi około 0.05, dla próby rozmiaru 90 – około 0.02, natomiast dla próby rozmiaru 500 – 0.01.

Możemy przyjąć, że od $n = 11$ aproksymacja rozkładem asymptotycznym jest wystarczająco dokładna ($D_n < 0.05$).

Skonstruujemy teraz asymptotyczne przedziały ufności dla S_1, S_2 oraz S_3 na poziomie ufności $1 - \alpha$. Mają one postać kolejno:

$$[L, P]_{S_1} = \left[\frac{z_{1-\frac{\alpha}{2}}^2}{2n} + \bar{X} - \frac{z_{1-\frac{\alpha}{2}}}{2\sqrt{n}} \sqrt{\frac{z_{1-\frac{\alpha}{2}}^2}{n} + 4\bar{X}}, \frac{z_{\frac{\alpha}{2}}^2}{2n} + \bar{X} - \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}} \sqrt{\frac{z_{\frac{\alpha}{2}}^2}{n} + 4\bar{X}} \right], \quad (4)$$

$$[L, P]_{S_2} = \left[\bar{X} - \frac{\sqrt{\bar{X}} z_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \bar{X} - \frac{\sqrt{\bar{X}} z_{\frac{\alpha}{2}}}{\sqrt{n}} \right], \quad (5)$$

$$[L, P]_{S_3} = \left[\left[\max \left(0, \sqrt{\bar{X}} - \frac{z_{1-\frac{\alpha}{2}}}{2\sqrt{n}} \right) \right]^2, \left(\sqrt{\bar{X}} - \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}} \right)^2 \right], \quad (6)$$

gdzie z_α – α -kwantyl rozkładu normalnego.

Zdefiniujemy teraz funkcję `przedzialy`, które dla zadanej próby i α zwrócą odpowiednie przedziały ufności na poziomie ufności $1 - \alpha$.

```
przedzialy <- function(X, alpha) {
  #wyliczenie składników
  X.sr <- mean(X)
  z.1 <- qnorm(1 - alpha / 2)
  z.2 <- qnorm(alpha / 2)
  n <- length(X)

  #lewa i prawa strona dla S.1
  L.1 <- z.1 ^ 2 / (2 * n) + X.sr - z.1 / (2 * sqrt(n)) * sqrt(z.1 ^ 2 /
                                                                    n + 4 * X.sr)
  P.1 <- z.2 ^ 2 / (2 * n) + X.sr - z.2 / (2 * sqrt(n)) * sqrt(z.2 ^ 2 /
                                                                    n + 4 * X.sr)

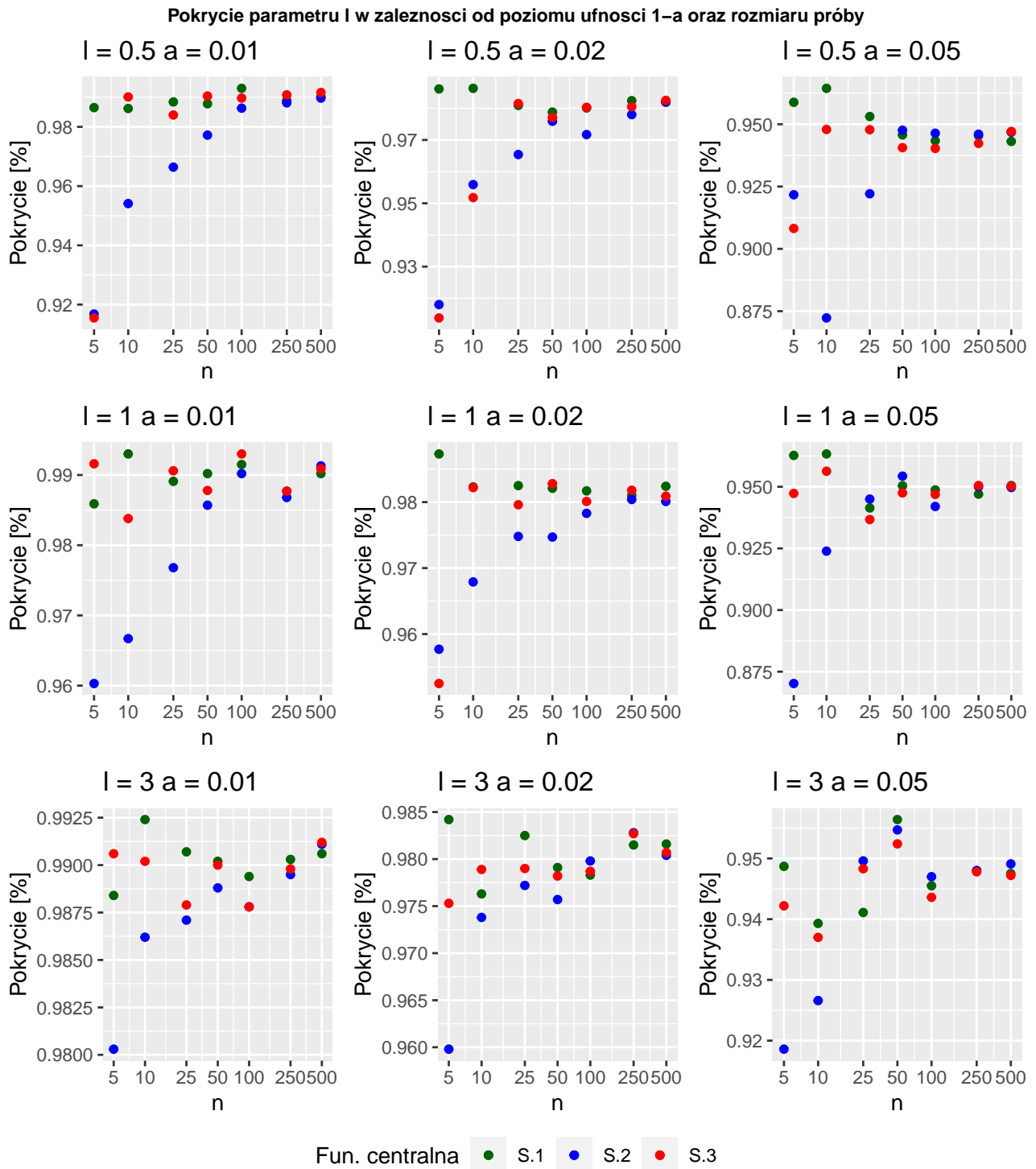
  #lewa i prawa strona dla S.2
  L.2 <- X.sr - sqrt(X.sr) * z.1 / sqrt(n)
  P.2 <- X.sr - sqrt(X.sr) * z.2 / sqrt(n)

  #lewa i prawa strona dla S.3
  L.3 <- (max(0, sqrt(X.sr) - z.1 / (2 * sqrt(n)))) ^ 2
  P.3 <- (sqrt(X.sr) - z.2 / (2 * sqrt(n))) ^ 2

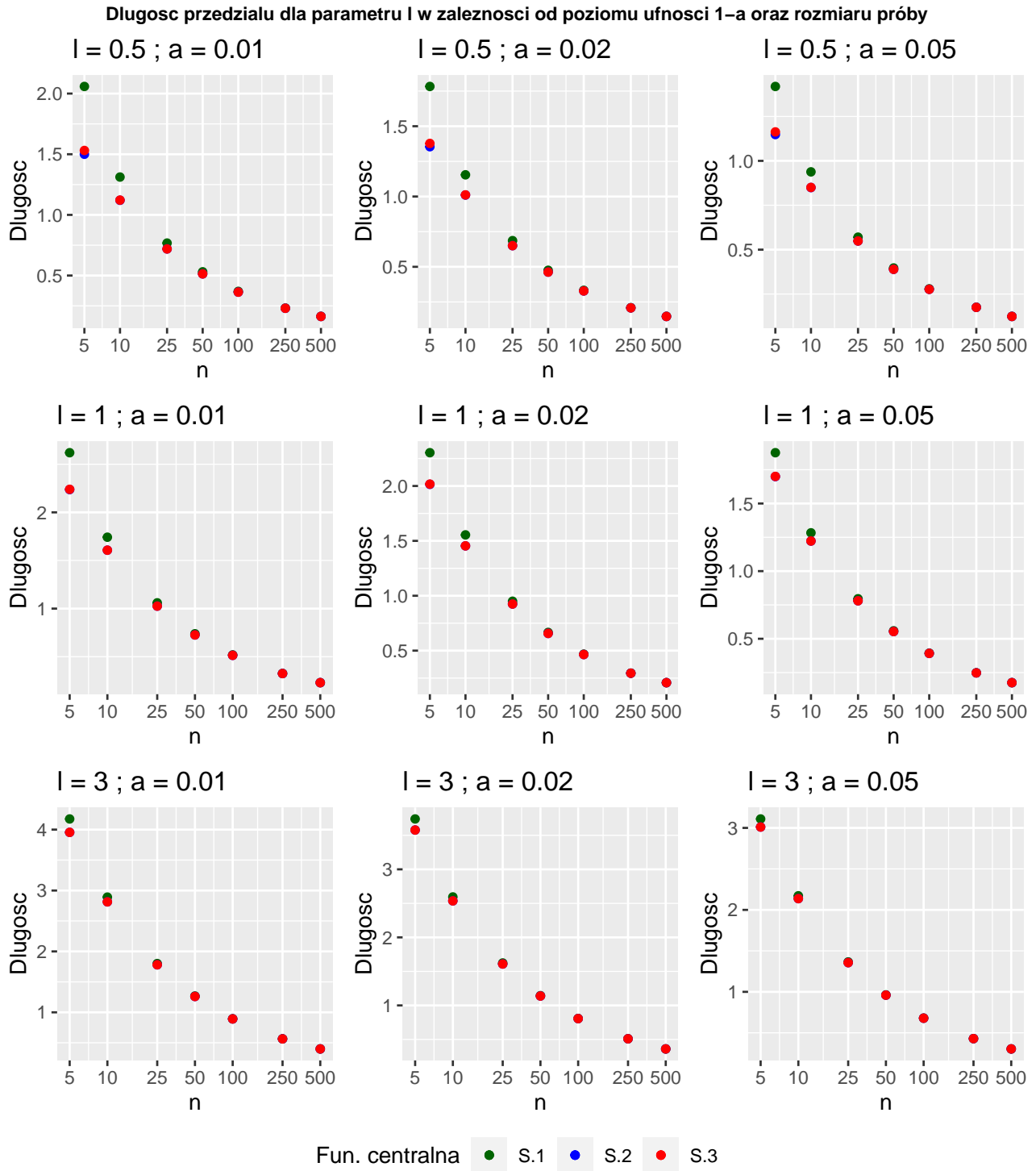
  #ramka danych zawierająca końce przedziałów ufności
  table <- data.frame(
    S.1 = c(L.1, P.1),
    S.2 = c(L.2, P.2),
    S.3 = c(L.3, P.3)
  )

  return(table)
}
```

Wykorzystując napisaną funkcję zbadamy i porównamy własności przedziałów ufności w zależności od parametru λ , poziomu ufności $1 - \alpha$ oraz rozmiaru próby n .



Rysunek 2: Własności asymptotycznych przedziałów ufności - empiryczne pokrycie



Rysunek 3: Własności asymptotycznych przedziałów ufności - długość przedziału

Widzimy, że wraz ze wzrostem parametru λ rośnie empiryczne prawdopodobieństwo pokrycia niezależnie od α oraz rozmiaru próby n . Oczywiście wraz ze wzrostem parametru α rośnie szerokość przedziału oraz prawdopodobieństwo pokrycia. Widzimy, że długości przedziałów skonstruowanych na bazie S_2 oraz S_3 są zbliżone, natomiast w przypadku S_3 przedziały ufności są zauważalnie dłuższe dla „małych” n .

Najlepszym przedziałem ufności wydaje się być ten skonstruowany z wykorzystaniem S_3 , ponieważ dla $n \geq 10$ empiryczne prawdopodobieństwo pokrycia jest zbliżone do pokrycia dla

S_1 , natomiast długość przedziału jest krótsza.

Z drugiej strony, najgorszym przedziałem ufności wydaje się być ten skonstruowany z wykorzystaniem S_2 , ponieważ najczęściej odpowiada mu najniższe empiryczne prawdopodobieństwo pokrycia, a przy tym nie wyróżnia się długością przedziału.

Dla małych rozmiarów próby najlepszy wydaje się być przedział uzyskany z wykorzystaniem S_1 . Mimo większej długości przedziału, uzyskiwane rezultaty są dość stabilne i empiryczne pokrycie jest bardzo zbliżone poziomowi ufności $1 - \alpha$ niezależnie od parametru λ . Minimalizujemy wówczas ryzyko popełnienia błędu. (dla S_2 i S_3 ryzyko popełnienia błędu może wynieść nawet 10%.)

4 Podsumowanie

Poniżej wypunktujemy najważniejsze wnioski, jakie można wyciągnąć z przeprowadzanych analiz:

- już dla $n = 11$ możemy przybliżać rozkład S_1 , S_2 oraz S_3 rozkładem normalnym $\mathcal{N}(0, 1)$,
- wraz ze wzrostem parametru λ rośnie empiryczne prawdopodobieństwo pokrycia,
- przedział ufności skonstruowany z wykorzystaniem S_3 wydaje się być najlepszy, natomiast najgorszy to ten oparty na S_2 .