

Sprawozdanie 4

Jakub Markowiak
album 255705

14 czerwca 2021

Spis treści

1	Krótki opis zagadnienia	1
2	Opis eksperymentów/analiz	1
3	Wyniki	1
3.1	Porównanie estymatorów MW i MNW dla rozkładu Gamma	1
3.2	Badanie stabilności estymatora NW	6
4	Podsumowanie	7

1 Krótki opis zagadnienia

W tym sprawozdaniu zbadamy własności rozkładu $G(a, s)$ oraz, korzystając z metody momentów oraz metody największej wiarygodności, wyznaczymy estymatory parametrów a oraz s a następnie porównamy je i spróbujemy rozstrzygnąć, który jest lepszy. Przeprowadzimy również doświadczalną analizę stabilności estymatora NW.

2 Opis eksperymentów/analiz

Przeprowadzimy następujące analizy i eksperymenty:

1. Porównanie estymatorów MM i NW dla rozkładu Gamma,
2. badanie stabilności estymatora NW.

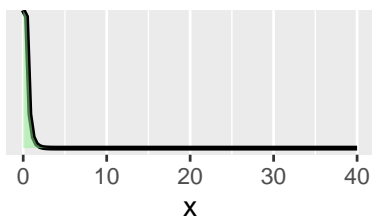
3 Wyniki

3.1 Porównanie estymatorów MW i MNW dla rozkładu Gamma

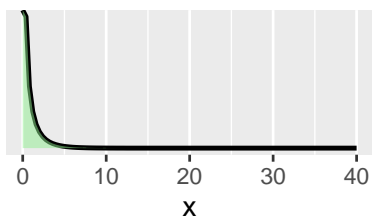
W tym sprawozdaniu rozpatrujemy rozkład $\mathcal{G}(a, s)$, gdzie a – parametr kształtu, a s – parametr skali. Sprawdźmy, jak wyglądają wykresy gęstości w zależności od dobrania tych parametrów.

Gęstość rozkładu Gamma dla różnych a i s

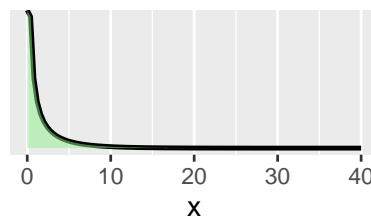
$a = 0.2, s = 0.5$



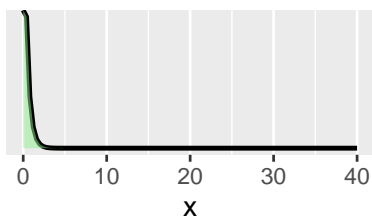
$a = 0.2, s = 2$



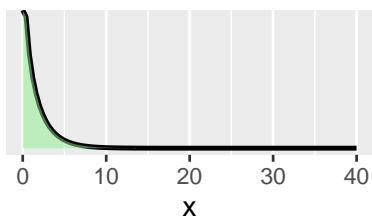
$a = 0.2, s = 5$



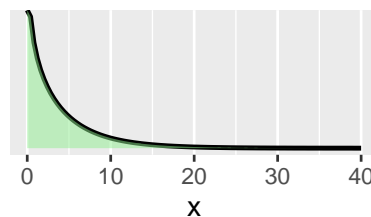
$a = 0.8, s = 0.5$



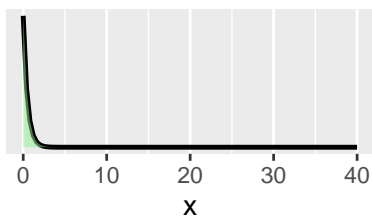
$a = 0.8, s = 2$



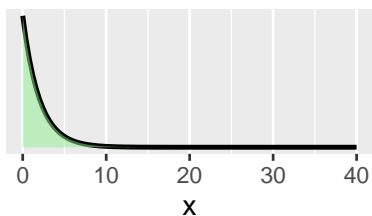
$a = 0.8, s = 5$



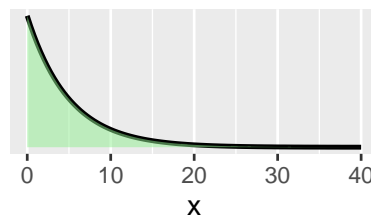
$a = 1, s = 0.5$



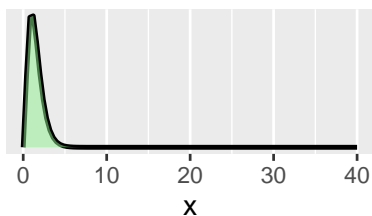
$a = 1, s = 2$



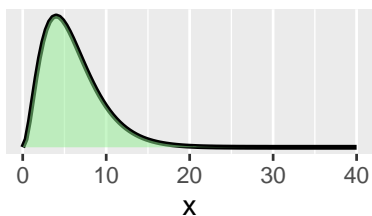
$a = 1, s = 5$



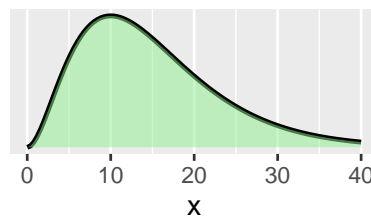
$a = 3, s = 0.5$



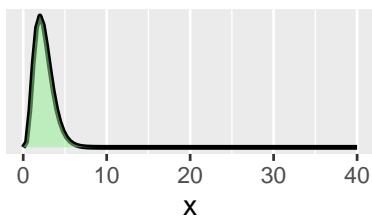
$a = 3, s = 2$



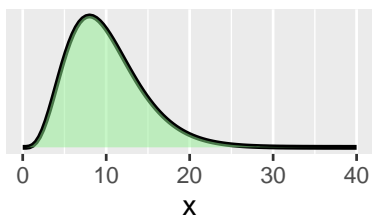
$a = 3, s = 5$



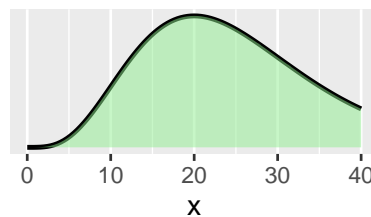
$a = 5, s = 0.5$



$a = 5, s = 2$



$a = 5, s = 5$



Rysunek 1: Rozkład Gamma dla różnych wartości parametrów a i s

Możemy zaobserwować, że im wyższa wartość parametru s , tym bardziej spłaszczony jest wykres. Dla parametru a widzimy natomiast, że wraz z jego wzrostem maleje skośność wykresu i staje się on bardziej wygładzony. Dla $a \leq 1$ obserwujemy, że wykres gęstości osiąga maksimum w zerze, natomiast dla $a > 1$ maksimum jest poza zerem.

Wyznamy teraz analityczną postać estymatorów uzyskanych metodą momentów. Wiemy, że jeżeli \mathbf{X} ma rozkład $\mathcal{G}(a, s)$ to

$$\begin{aligned} E(\mathbf{X}) &= sa, \\ E(\mathbf{X}^2) &= s^a(1+a). \end{aligned}$$

Przyrównujemy zatem momenty zwykłe do momentów teoretycznych i otrzymujemy układ równań:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i &= as, \\ \frac{1}{n^2} \sum_{i=1}^n X_i^2 &= as^2 + a^2 s^2. \end{aligned}$$

Rozwiązując ten układ otrzymujemy estymatory MM w postaci:

$$\begin{aligned} \hat{a}_{MM} &= \frac{\bar{X}^2}{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2}, \\ \hat{s}_{MM} &= \frac{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2}{\bar{X}}. \end{aligned}$$

Napišemy teraz R-funkcję, która dla zadanej próby zwróci wartości tych estymatorów.

```
gamma_MM <- function(X) {
  n <- length(X)
  X_hat <- mean(X)
  X_sq <- sum(X ^ 2)
  a_hat <- X_hat ^ 2 / (1 / n * X_sq - X_hat ^ 2)
  s_hat <- (1 / n * X_sq - X_hat ^ 2) / X_hat
  return(c(a_hat, s_hat))
}
```

Analogicznie Wyznamy analityczną postać estymatorów uzyskanych metodą największej wiarygodności:

$$\begin{aligned} p_{\theta}(\mathbf{X}) &= \prod_{i=1}^n \frac{X_i^{(a-1)}}{s^a \Gamma(a)} \exp\left(\frac{-X_i}{s}\right) 1_{[0,\infty)}(X_i) = \\ &= 1_{[0,\infty)}(X_{n:n}) \left(\frac{1}{s^a \Gamma(a)}\right)^n \prod_{i=1}^n \exp\left(\frac{-X_i}{s}\right). \end{aligned}$$

Badamy logarytm gęstości

$$l_{\theta}(\mathbf{X}) = \ln 1_{[0,\infty)}(X_{n:n}) + n(-a \ln s - \ln \Gamma(a)) + (a-1) \sum_{i=1}^n \ln X_i - \frac{\sum_{i=1}^n X_i}{s}.$$

Szukamy maksimum funkcji l_{θ} , zatem

$$\begin{cases} \frac{\partial l_\theta}{\partial a} = -n \ln s - n \frac{\Gamma'(a)}{\Gamma(a)} + \sum_{i=1}^n \ln X_i = 0, \\ \frac{\partial l_\theta}{\partial s} = -\frac{a}{s} + \frac{1}{s^2} \sum_{i=1}^n X_i = 0 \end{cases}$$

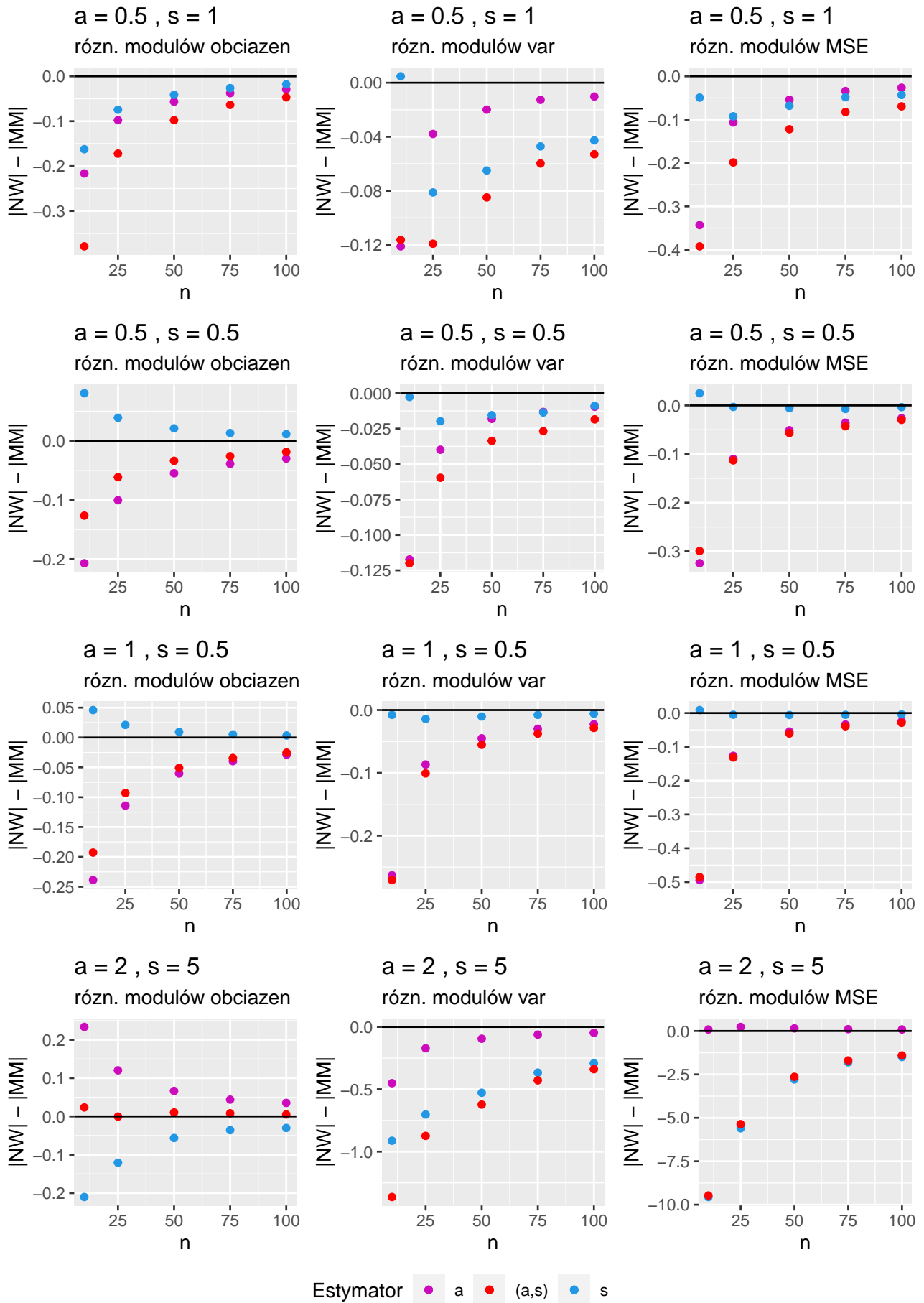
Nasze estymatory NW są rozwiązaniem układu równań

$$\begin{aligned} -n \ln \frac{\sum_{i=1}^n X_i}{\hat{a}_{NW}} - n\psi(\hat{a}_{NW}) + \sum_{i=1}^n \ln X_i &= 0, \\ \hat{s}_{NW} - \hat{a}_{NW} \sum_{i=1}^n X_i &= 0. \end{aligned}$$

Ponieważ pojawia się tutaj funkcja specjalna **digamma**, estymator ten można wyznaczyć tylko przy użyciu metod numerycznych. Napišemy zatem funkcję, która rozwiąże powyższy układ równań dla zadanej próby i zwróci estymatory NW.

```
gamma_NW <- function(X) {
  n <- length(X)
  equation <- function (a, X)
  {
    y <-
      -n * log(sum(X) / a) + n * log(n) - n * digamma(a) + sum(log(X))
  }
  a_hat <- uniroot(f = equation,
                  interval = c(0.01, 100),
                  X = X)$root
  s_hat <- mean(X) / a_hat
  return(c(a_hat, s_hat))
}
```

Porównamy teraz estymatory uzyskane obiema metodami. Przedstawimy na wykresie różnice modułów obciążeń, wariancji oraz błędów średniokwadratowych dla obu estymatorów. Odejmujemy moduły kolejnych wartości dla estymatora MM od ich odpowiedników dla NW.

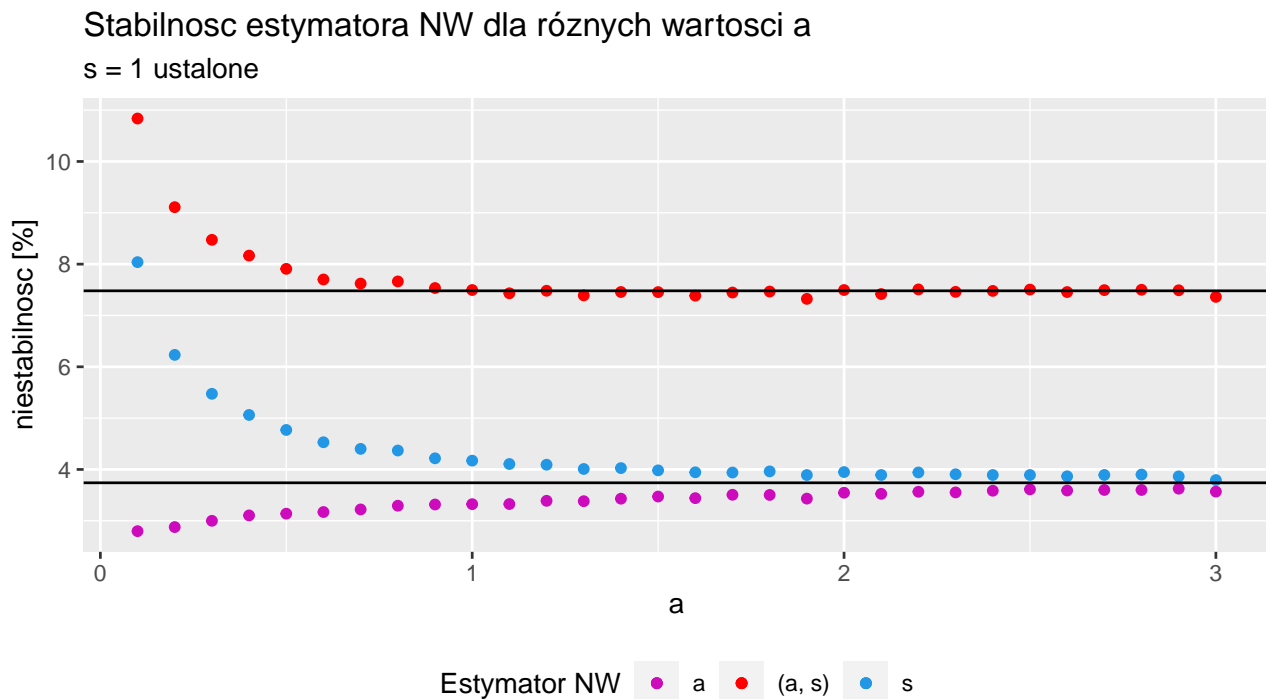


Rysunek 2: Porównanie estymatorów MM i NW

Widzimy, że niemal dla wszystkich rozpatrywanych kombinacji parametrów a i s oraz rozmiaru próby n estymator NW ma mniejszą wariancję oraz mniejszy bądź podobny błąd średniokwadratowy. Obciążenia natomiast różnią się w zależności od dobranych parametrów. Widzimy, że dla np. $a = 2, s = 5$ estymator NW ma większe obciążenie, natomiast w innych przypadkach to estymator MM jest bardziej obciążony. Możemy zatem wywnioskować, że dla rozkładu $\mathcal{G}(a, s)$ estymator NW jest lepszym estymatorem niż estymator MM (ze względu na wariancję oraz błąd średniokwadratowy).

3.2 Badanie stabilności estymatora NW

Sprawdźmy teraz, czy estymatory NW są stabilne. Eksperyment, który przeprowadzimy, polega na wylosowaniu 100 wartości z rozkładu $\mathcal{G}(a, s)$, podzieleniu próby na 10 losowych podzbiorów (za każdym razem usuwamy 10 innych wartości) i wyznaczeniu estymatorów NW dla tych podzbiorów, a następnie zmierzenia zmiany wartości tego estymatora względem estymatora NW dla całej próby. Eksperyment powtórzymy 1000 razy dla każdej badanej wartości parametru a , a wyniki przedstawimy na wykresie.



Rysunek 3: Badanie stabilności estymatora NW

Widzimy, że dla $a \leq 1$ zdecydowanie bardziej stabilny jest estymator \hat{a}_{NW} , występuje natomiast duża niestabilność estymatora \hat{s}_{NW} . Dla $a > 1$ różnice pomiędzy niestabilnością estymatorów stają się coraz mniejsze i oscylują wokół prostej $y = 3,74$. Przeprowadzając tę analizę dla różnych wartości parametru s otrzymalibyśmy bardzo zbliżone wyniki. Spoglądając natomiast na „całkowitą” niestabilność estymatora NW widzimy, że jest ona wyraźnie większa dla $a \leq 1$. Dla małych wartości parametru a , przy losowym usunięciu 10% elementów próby, możemy otrzymać wartości estymatorów NW różne nawet o 10%. Dla $a > 1$ średnia niestabilność utrzymuje się na poziomie około 7,48%.

4 Podsumowanie

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

- w rozkładzie $\mathcal{G}(a, s)$ parametr skali a odpowiada za pozycję maksimum funkcji gęstości, natomiast parametr kształtu s za skośność wykresu,
- nie zawsze można analitycznie wyznaczyć estymator NW,
- w rozkładzie $\mathcal{G}(a, s)$ estymator NW jest lepszym estymatorem niż estymator MM,
- poziom niestabilności estymatora NW zależy od parametru a – im bliżej zera, tym większa niestabilność, natomiast dla $a > 1$ średnia niestabilność utrzymuje się na poziomie około 7,48%.