

Raport 2

Magdalena Potok

2023-11-06

Zadanie 1

Wygenerowałam 50 obserwacji z rozkładu dwumianowego $b(5, p)$, gdzie $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

Na podstawie wygenerowanych obserwacji wyznaczyłam wartość estymatora największej wiarygodności wielkości $P(X \geq 3)$, gdzie $X \sim b(5, p)$. Doświadczenie powtórzyłam 10 000 razy, oszacowałam wariancję, błąd średniokwadratowy oraz obciążenie analizowanego estymatora.

Aby policzyć wartość estymatora największej wiarygodności $P(X \geq 3)$ trzeba wiedzieć, że ENW dla rozkładu $b(5, p)$ jest $\hat{p} = \frac{\bar{x}}{5}$ - łatwo można się o tym przekonać licząc pierwszą i drugą pochodną funkcji prawdopodobieństwa tego rozkładu. Jak już znamy nasz estymator, zauważmy, że

$$P(\widehat{X} \geq 3) = 1 - (P(\widehat{X} = 0) + P(\widehat{X} = 1) + P(\widehat{X} = 2)) = 1 - \sum_{i=0}^2 \binom{5}{i} \hat{p}^5 (1 - \hat{p})^{5-i} = 1 - \sum_{i=0}^2 \binom{5}{i} \left(\frac{\bar{x}}{5}\right)^5 \left(1 - \frac{\bar{x}}{5}\right)^{5-i}$$

| p | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|-----|------------------|-----------------|---------|---------|----------|
| 0.1 | 0.011795 | 0.00856 | 0.00002 | 0.00002 | 0.00075 |
| 0.3 | 0.105636 | 0.16308 | 0.00149 | 0.00148 | 0.00223 |
| 0.5 | 0.485003 | 0.50000 | 0.00349 | 0.00349 | -0.00104 |
| 0.7 | 0.847341 | 0.83692 | 0.00150 | 0.00149 | -0.00246 |
| 0.9 | 0.990433 | 0.99144 | 0.00002 | 0.00002 | -0.00079 |

Z tabeli możemy zaobserwować, że MSE i Var mają rozkład symetryczny, czyli im bardziej skrajny przypadek parametru p, tym wymienione wartości są niższe, czyli rozrzut ENW $P(X \geq 3)$ od prawdziwej wartości jest najmniejszy. Obciążenie estymatora jest najmniejsze dla parametru $p = 0.1$.

Zadanie 2

Wygenerowałam 50 obserwacji z rozkładu Poissona z parametrem λ , gdzie $\lambda \in \{0.5, 1, 2, 5\}$.

Na podstawie wylosowanych obserwacji wyznaczyłam wartość estymatora największej wiarygodności wielkości $P(X = x)$, $x = 0, 1, \dots, 10$, gdzie $X \sim \pi(\lambda)$. Doświadczenie zostało powtórzone 10 000 razy i na tej podstawie oszacowałam wariancję, błąd średniokwadratowy oraz obciążenie analizowanego estymatora.

Aby policzyć estymator największej wiarygodności wielkości $P(X = x)$, $x = 0, 1, \dots, 10$, przypomnijmy, że estymatorem największej wiarygodności jest $\hat{\lambda} = \hat{X}$ oraz funkcja prawdopodobieństwa tego rozkładu dyskretnego to: $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$, zatem

$$P(\widehat{X} = x_i) = \frac{\hat{\lambda}^{x_i} e^{-\hat{\lambda}}}{x_i!} = \frac{\hat{X}^{x_i} e^{-\hat{X}}}{x_i!}$$

- $P(X = 1)$

| lambda | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|--------|------------------|-----------------|---------|---------|----------|
| 0.5 | 0.232368 | 0.303265 | 0.00099 | 0.00097 | -0.00459 |
| 1.0 | 0.367879 | 0.367879 | 0.00004 | 0.00003 | -0.00384 |
| 2.0 | 0.230595 | 0.270671 | 0.00070 | 0.00070 | 0.00024 |
| 5.0 | 0.029627 | 0.033690 | 0.00008 | 0.00008 | 0.00096 |

Ponownie wyniki MSE i Var są do siebie zbliżone, warto zauważyć, że najmniejszą wartość Var otrzymujemy dla największej $\lambda = 1$. Oznacza to, że dla tej wielkości parametru w naszym rozkładzie, badany estymator jest bardziej stabilny i mniej podatny na sytuacje odstające, co jest ważną cechą estymatora. Dla tej λ jest również niskie MSE, co oznacza, że ten estymator jest dokładny. Najmniejsze obciążenie estymatora jest dla $\lambda = 2$.

- $P(X = 2)$

| lambda | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|--------|------------------|-----------------|---------|---------|----------|
| 0.5 | 0.094175 | 0.075816 | 0.00051 | 0.00051 | 0.00035 |
| 1.0 | 0.156485 | 0.183940 | 0.00066 | 0.00066 | -0.00199 |
| 2.0 | 0.270671 | 0.270671 | 0.00002 | 0.00001 | -0.00267 |
| 5.0 | 0.073671 | 0.084224 | 0.00026 | 0.00026 | 0.00105 |

Tym razem najmniejsze MSE i Var badanego estymatora jest dla $\lambda = 2$. Najmniejsze obciążenie estymatora jest dla $\lambda = 0.5$.

- $P(X = 5)$

| lambda | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|--------|------------------|-----------------|---------|---------|----------|
| 0.5 | 0.000472 | 0.000158 | 0.00000 | 0.00000 | 0.00005 |
| 1.0 | 0.004792 | 0.003066 | 0.00000 | 0.00000 | 0.00032 |
| 2.0 | 0.032908 | 0.036089 | 0.00012 | 0.00012 | 0.00059 |
| 5.0 | 0.175405 | 0.175467 | 0.00001 | 0.00001 | -0.00173 |

Dla tego prawdopodobieństwa wyniki są zdecydowanie bardziej zbliżone, dla $\lambda = 0.5$, 1 wariancja oraz błąd średniokwadratowy w przybliżeniu wynoszą 0, co świadczy o precyzyjności badanego estymatora. Dla $\lambda = 0.5$ obciążenie estymatora jest również bardzo niskie, bo wynosi jedynie 0.00005.

- $P(X = 7)$

| lambda | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|--------|------------------|-----------------|---------|---------|----------|
| 0.5 | 0.000003 | 0.000001 | 0.00000 | 0.00000 | 0.00000 |
| 1.0 | 0.000029 | 0.000073 | 0.00000 | 0.00000 | 0.00002 |
| 2.0 | 0.001682 | 0.003437 | 0.00000 | 0.00000 | 0.00029 |
| 5.0 | 0.114074 | 0.104445 | 0.00017 | 0.00017 | -0.00053 |

Dla tego prawdopodobieństwa można zobaczyć już bardzo niskie wyniki MSE, Var oraz Bias dla każdej λ . Można pomyśleć, że dla większego x w funkcji prawdopodobieństwa $P(X = x)$, tym nasz estymator jest bardziej dokładny, sprawdźmy jeszcze dla ostatniej wartości.

- $P(X = 10)$

| lambda | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|--------|------------------|-----------------|-------|-------|---------|
| 0.5 | 0.000000 | 0.000000 | 0e+00 | 0e+00 | 0.00000 |
| 1.0 | 0.000000 | 0.000000 | 0e+00 | 0e+00 | 0.00000 |
| 2.0 | 0.000122 | 0.000038 | 0e+00 | 0e+00 | 0.00001 |
| 5.0 | 0.014407 | 0.018133 | 3e-05 | 3e-05 | 0.00052 |

Wyniki dla tego prawdopodobieństwa potwierdzają powyższe spekulacje o tym, że im większy x , tym estymator daje nam lepszy wynik. Wszystkie wyniki są bardzo bliskie 0. Można jeszcze powiedzieć, że mniejsza lambda daje nam lepszą estymację, ale te wyniki są na tyle bliskie 0, że dla $\lambda = 5$ również estymacja jest poprawna.

W tym zadaniu ominęłam badanie wartości estymatora $P(\widehat{X} = x_i)$ $x_i \in \{3, 4, 6, 8\}$, ponieważ wyniki były bardzo zbliżone do wyżej przeanalizowanych sytuacji i dochodziłam do tych samych wniosków.

Zadanie 4

Wygenerowałam 50 obserwacji z rozkładu beta z parametrami θ i 1, gdzie $\theta \in \{0.5, 1, 2, 5\}$.

Doświadczenie powtórzyłam 10 000 razy, na podstawie tych danych wyznaczyłam wartość estymatora $\widehat{I}(\hat{\theta})$ informacji Fishera parametru θ . Estymator największej wiarygodności $\hat{\theta}$ dla rozkładu beta z gęstością $\theta x^{\theta-1} \mathbb{I}_{(0,1)}(x)$, $\theta > 0$ wynosi: $-\frac{\sum_{i=1}^n \log x_i}{n}$. Informację Fishera liczymy w następujący sposób: $I(\theta) = -\mathbb{E}\left(\frac{\partial \log f(X, \theta)}{\partial \theta}\right)^2 = -\mathbb{E}\left(\frac{\partial^2 \log f(X, \theta)}{\partial \theta^2}\right)$, po wyliczeniu drugiej pochodnej z funkcji logwiarygodności wychodzi, że $I(\theta) = -\mathbb{E}\left(-\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}$. A więc estymator informacji Fishera, to:

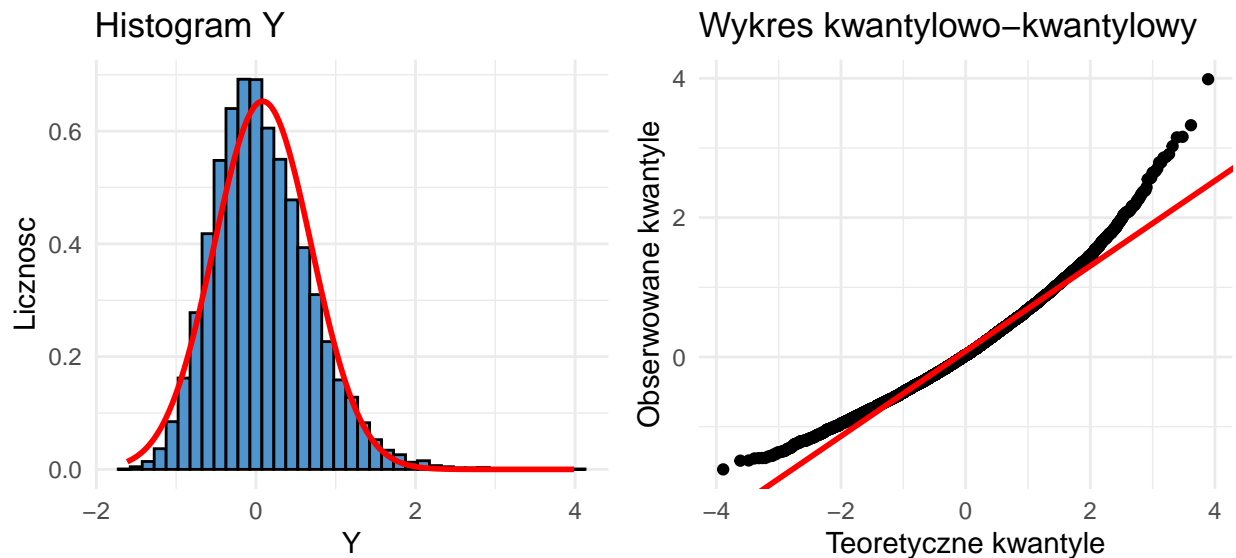
$$\widehat{I}(\hat{\theta}) = \left(\frac{\sum_{i=1}^n \log x_i}{n}\right)^2$$

| theta | Wynik_estymowany | Wynik_prawdziwy |
|-------|------------------|-----------------|
| 0.5 | 3.988325 | 4.00 |
| 1.0 | 0.899634 | 1.00 |
| 2.0 | 0.243197 | 0.25 |
| 5.0 | 0.053519 | 0.04 |

Każdy z wyników estymacji $\widehat{I}(\hat{\theta})$ jest bardzo bliski prawdziwej wartości informacji Fishera.

Następnie wygenerowałam niezależnie 50 obserwacji z rozkładu beta z parametrami j.w. Wyznaczyłam wartość estymatora największej wiarygodności parametru θ i wyliczę dla nowej zmiennej zdefiniowanej jako $Y = \sqrt{n\widehat{I}(\hat{\theta})}(\hat{\theta} - \theta)$ jej wartość na podstawie zaobserwowanej próby oraz wcześniejszego wyniku. Doświadczenie powtórzyłam 10 000 razy i narysowałam histogram oraz wykres kwantylowo-kwantylowy dla uzyskanych wyników.

- $\theta = 0.5$



Widzimy, że histogram przypomina przesunięty wykres funkcji gęstości rozkładu normalnego. Na histogram została nałożona (czerwonym kolorem) teoretyczny wykres funkcji gęstości i widać, że znacznie większa część histogramu pokrywa się z tą krzywą. Możemy przypuszczać więc, że rozkład zachowuje się w przybliżeniu jak rozkład normalny.

Jeśli jednak spojrzymy na wykres kwantylowo-kwantylowy zobaczymy, że rozkład empiryczny danych odstaje od rozkładu teoretycznego rozkładu normalnego. Wskazuje to na obecność obserwacji odstających, rozkład więc nie jest dokładnie normalny, ale zachowuje się w podobny sposób.

- $\theta \in \{1, 2, 5\}$

Dla wyżej wymienionych θ wykresy nie odbiegają znacząco od wykresów, jak wyżej. Zatem wnioski są takie same jak powyższe.

Rozkład zmiennej losowej $Y = \sqrt{nI(\theta)}(\hat{\theta} - \theta)$ jest bliski rozkładu normalnego.

Zadanie 5

Wygenerowałam 50 obserwacji z rozkładu Laplace'a z parametrem θ i skali σ , gdzie

- $\theta = 1, \sigma = 1,$
- $\theta = 4, \sigma = 1,$
- $\theta = 1, \sigma = 2.$

Na podstawie tych rozkładów obliczyłam wartość estymatora θ postaci:

- $\hat{\theta}_1 = \bar{X},$
- $\hat{\theta}_2 = \text{Me}\{X_1, \dots, X_n\},$
- $\hat{\theta}_3 = \sum_{i=1}^n w_i X_i, \sum_{i=1}^n w_i = 1, 0 \leq w_i \leq 1,$ gdzie w_i to ciąg liczb zaczynających się od 0.02 i zwiększających się o 0.02. W ten sposób cały wektor sumuje się do 1 i każda waga jest z przedziału $(0, 1),$
- $\hat{\theta}_4 = \sum_{i=1}^n w_i X_{i:n},$ gdzie $X_{i:n}$ to uporządkowane obserwacje X_i oraz $w_i = \varphi(\Phi^{-1}(\frac{i-1}{n})) - \varphi(\Phi^{-1}(\frac{i}{n})),$ φ i Φ to kolejno gęstość oraz dystrybuanta standardowego rozkładu normalnego.

Doświadczenie powtórzyłam 10 000 razy i na tej podstawie oszacowałam wariancję, błąd średniokwadratowy oraz obciążenie każdego z estymatorów.

- (a) $\theta = 1, \sigma = 1$

| | MSE | Var | Bias |
|------------------|---------|---------|---------|
| $\hat{\theta}_1$ | 0.04020 | 0.04020 | 0.00130 |
| $\hat{\theta}_2$ | 0.02464 | 0.02464 | 0.00098 |

| | MSE | Var | Bias |
|------------------|---------|---------|---------|
| $\hat{\theta}_3$ | 0.05281 | 0.05281 | 0.00041 |
| $\hat{\theta}_4$ | 0.15243 | 0.03987 | 0.33549 |

Najmniejsze obciążenie estymatora jest dla $\hat{\theta}_3$, natomiast najniższy MSE i Var jest dla $\hat{\theta}_2$, dla tego estymatora Bias nie jest też zbyt wysokie, więc dla tych parametrów rozkładu Laplace'a $\hat{\theta}_2$ byłby najlepszy wyborem estymatora.

- (b) $\theta = 4, \sigma = 1$

| | MSE | Var | Bias |
|------------------|---------|---------|----------|
| $\hat{\theta}_1$ | 0.04061 | 0.04061 | 0.00022 |
| $\hat{\theta}_2$ | 0.02461 | 0.02461 | -0.00039 |
| $\hat{\theta}_3$ | 0.05338 | 0.05338 | -0.00120 |
| $\hat{\theta}_4$ | 7.12749 | 0.03983 | -2.66227 |

Można zauważyć, że wzrost parametru θ w naszym rozkładzie spowodował wzrost MSE i Var dla $\hat{\theta}_1, \hat{\theta}_3$ i $\hat{\theta}_4$, a $\hat{\theta}_2$ minimalnie zmalało, a więc ten estymator nadal wydaje się najlepszym wyborem.

- (c) $\theta = 1, \sigma = 2$

| | MSE | Var | Bias |
|------------------|---------|---------|----------|
| $\hat{\theta}_1$ | 0.15959 | 0.15959 | 0.00201 |
| $\hat{\theta}_2$ | 0.09759 | 0.09759 | -0.00102 |
| $\hat{\theta}_3$ | 0.21379 | 0.21375 | 0.00696 |
| $\hat{\theta}_4$ | 2.94832 | 0.15908 | 1.67010 |

Ponownie najniższe wyniki (co do modułu) wychodzą dla estymatora $\hat{\theta}_2$, a więc próbkowa mediana okazała się najlepszym estymatorem dla każdego z przypadków, wzzrost θ oraz σ nie wpływa na wybór estymatora.

Zadanie 6

W tym zadaniu powótrzyłam eksperymenty z zadań 1, 2, 4, 5 dla $n = 20$ i $n = 100$.

Zadanie 1'

Estymator $P(X \geq 3)$, gdzie $X \sim b(5, p)$, $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$

- $n = 20$

| p | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|-----|------------------|-----------------|---------|---------|----------|
| 0.1 | 0.004525 | 0.00856 | 0.00008 | 0.00008 | 0.00198 |
| 0.3 | 0.074434 | 0.16308 | 0.00374 | 0.00372 | 0.00483 |
| 0.5 | 0.682560 | 0.50000 | 0.00847 | 0.00846 | -0.00132 |
| 0.7 | 0.809474 | 0.83692 | 0.00384 | 0.00382 | -0.00517 |
| 0.9 | 0.993659 | 0.99144 | 0.00008 | 0.00007 | -0.00189 |

Ponownie można zaobserwować symetryczne rozłożenie MSE i Var. Zmniejszenie próby wywołało wzrost wszystkich wartości, co mówi nam, że estymator jest mniej dokładny, podatniejszy na sytuacje odstające i bardziej obciążony dla każdego p .

- $n = 100$

| p | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|-----|------------------|-----------------|---------|---------|----------|
| 0.1 | 0.026612 | 0.00856 | 0.00001 | 0.00001 | 0.00042 |
| 0.3 | 0.160444 | 0.16308 | 0.00074 | 0.00074 | 0.00132 |
| 0.5 | 0.470020 | 0.50000 | 0.00174 | 0.00174 | 0.00000 |
| 0.7 | 0.834264 | 0.83692 | 0.00074 | 0.00074 | -0.00063 |
| 0.9 | 0.983586 | 0.99144 | 0.00001 | 0.00001 | -0.00035 |

Wszystkie wartości MSE, Var i Bias zmalały przy zwiększeniu próby. Oznacza to, że zwiększenie próby wpłynęło na dokładność naszego estymatora, czego moglibyśmy się oczywiście spodziewać. Najlepszy wynik (tj. najniższy błąd średniokwadratowy, wariancja oraz niskie obciążenie) wyszło dla $p = 0.9$.

Zadanie 2'

Estymator $P(X = x)$, $x = 0, 1, \dots, 10$, gdzie $X \sim \pi(\lambda)$, $\lambda \in \{0.5, 1, 2, 5\}$

- $n = 20$
- $P(X = 1)$

| lambda | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|--------|------------------|-----------------|---------|---------|----------|
| 0.5 | 0.246641 | 0.303265 | 0.00263 | 0.00250 | -0.01139 |
| 1.0 | 0.354275 | 0.367879 | 0.00028 | 0.00019 | -0.00950 |
| 2.0 | 0.310562 | 0.270671 | 0.00165 | 0.00165 | 0.00014 |
| 5.0 | 0.031093 | 0.033690 | 0.00022 | 0.00021 | 0.00259 |

- $P(X = 5)$

| lambda | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|--------|------------------|-----------------|---------|---------|----------|
| 0.5 | 0.000242 | 0.000158 | 0.00000 | 0.00000 | 0.00013 |
| 1.0 | 0.007286 | 0.003066 | 0.00001 | 0.00001 | 0.00086 |
| 2.0 | 0.060196 | 0.036089 | 0.00029 | 0.00028 | 0.00172 |
| 5.0 | 0.174785 | 0.175467 | 0.00005 | 0.00003 | -0.00422 |

- $P(X = 7)$

| lambda | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|--------|------------------|-----------------|-------|-------|----------|
| 0.5 | 0.000000 | 0.000001 | 0e+00 | 0e+00 | 0.00000 |
| 1.0 | 0.000214 | 0.000073 | 0e+00 | 0e+00 | 0.00006 |
| 2.0 | 0.003886 | 0.003437 | 1e-05 | 1e-05 | 0.00080 |
| 5.0 | 0.119987 | 0.104445 | 4e-04 | 4e-04 | -0.00169 |

Wszystkie wyniki badające dokładność estymatora dla każdego z funkcji prawdopodobieństwa wyszły o drobinę gorsze (tj. większe) niż dla próby $n = 50$. Jednak nadal są to bardzo niskie wyniki i już dla próby $n = 20$ można uznać ten estymator za odpowiedni. Wnioski, że im większe x_i dla funkcji $P(X = x_i)$, tym wynik estymatora będzie bardziej satysfakcjonujący.

- $n = 100$
- $P(X = 1)$

| lambda | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|--------|------------------|-----------------|---------|---------|----------|
| 0.5 | 0.297016 | 0.303265 | 0.00047 | 0.00047 | -0.00191 |
| 1.0 | 0.367879 | 0.367879 | 0.00001 | 0.00001 | -0.00191 |
| 2.0 | 0.276083 | 0.270671 | 0.00036 | 0.00036 | -0.00031 |
| 5.0 | 0.022294 | 0.033690 | 0.00004 | 0.00004 | 0.00066 |

- $P(X = 5)$

| lambda | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|--------|------------------|-----------------|-------|-------|----------|
| 0.5 | 0.000205 | 0.000158 | 0e+00 | 0e+00 | 0.00002 |
| 1.0 | 0.002287 | 0.003066 | 0e+00 | 0e+00 | 0.00020 |
| 2.0 | 0.047587 | 0.036089 | 6e-05 | 6e-05 | 0.00024 |
| 5.0 | 0.175065 | 0.175467 | 0e+00 | 0e+00 | -0.00086 |

- $P(X = 7)$

| lambda | Wynik_estymowany | Wynik_prawdziwy | MSE | Var | Bias |
|--------|------------------|-----------------|-------|-------|----------|
| 0.5 | 0.000002 | 0.000001 | 0e+00 | 0e+00 | 0.00000 |
| 1.0 | 0.000109 | 0.000073 | 0e+00 | 0e+00 | 0.00001 |
| 2.0 | 0.002513 | 0.003437 | 0e+00 | 0e+00 | 0.00015 |
| 5.0 | 0.100636 | 0.104445 | 8e-05 | 8e-05 | -0.00011 |

Ponownie dla większej próbki estymator jest lepszy, ma każdy wynik, dla każdej lambdy niższy niż w przypadku gry $n = 50$ lub $n = 20$. Dużo szybciej również wyniki zbiegają do 0, tzn. już dla $P(X = 5)$ możemy zauważyć zerowe MSE i Var (z dokładnością do 4 miejsc po przecinku) dla każdej z lambd, czego nie mogliśmy zauważyć we wcześniejszych przypadkach.

Zadanie 4'

Estymator $\widehat{I}(\theta)$, gdzie $X \sim b(\theta, 1)$, $\theta \in \{0.5, 1, 2, 5\}$

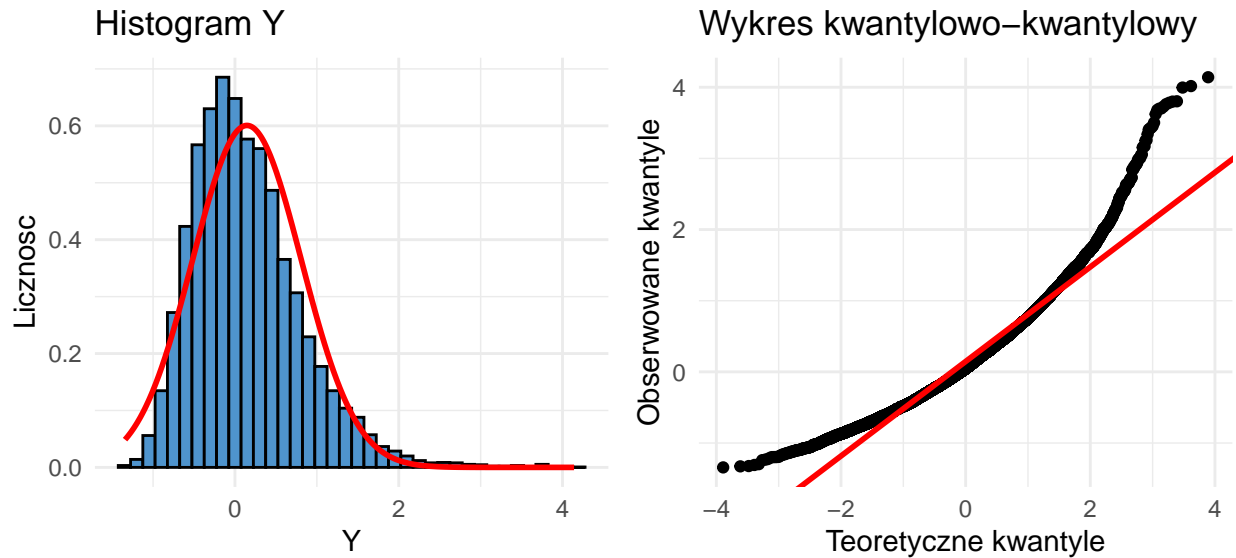
- $n = 20$

| theta | Wynik_estymowany | Wynik_prawdziwy |
|-------|------------------|-----------------|
| 0.5 | 6.844510 | 4.00 |
| 1.0 | 0.852164 | 1.00 |
| 2.0 | 0.103305 | 0.25 |
| 5.0 | 0.036012 | 0.04 |

Dla takiej wielkości próby nasz estymator informacji Fishera znacznie bardziej odbiega od rzeczywistej wartości, estymator jest mniej dokładny niż dla $n = 50$.

Wykres $Y = \sqrt{n\widehat{I}(\theta)}(\hat{\theta} - \theta)$

- $\theta = 0.5$



Histogram dalej w większej części znajduje się pod teoretycznie wyznaczoną krzywą gęstości, ale wystaje większa jego część niż w przypadku, gdy $n = 50$. Na wykresie kwantylowo-kwantylowym można zaobserwować tkzw. ciężkie ogony, które odstają od teoretycznej prostej.

- $\theta \in \{0.5, 1, 2, 5\}$

Ponownie jak w przypadku $n = 50$. Wykresy nie różnią się znacząco od siebie.

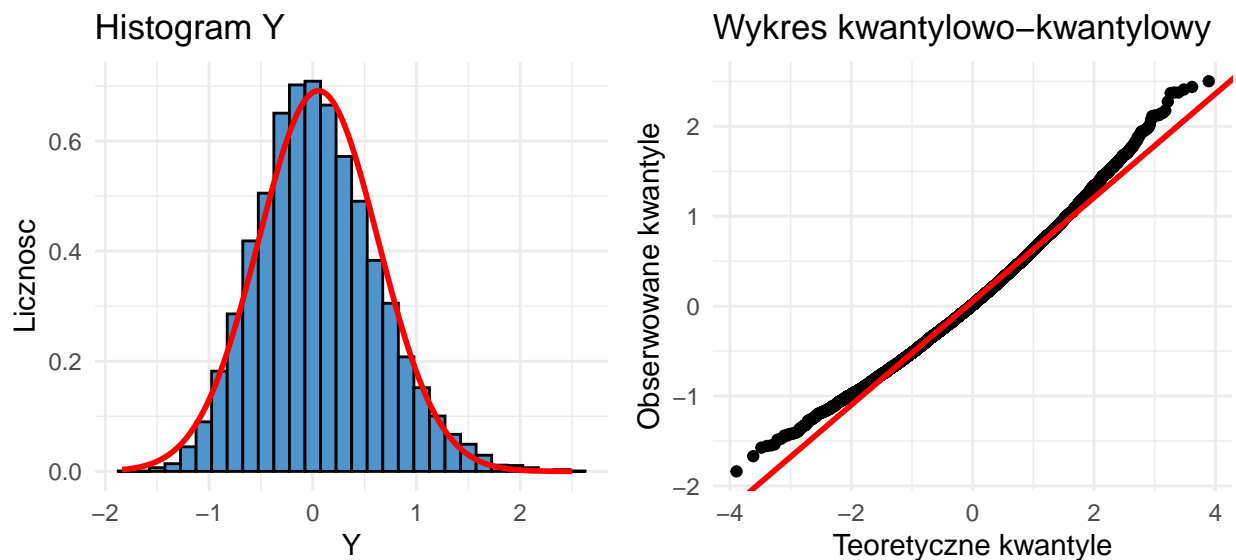
- $n = 100$

| theta | Wynik_estymowany | Wynik_prawdziwy |
|-------|------------------|-----------------|
| 0.5 | 5.075118 | 4.00 |
| 1.0 | 1.066902 | 1.00 |
| 2.0 | 0.236015 | 0.25 |
| 5.0 | 0.064077 | 0.04 |

Wyniki estymowane są lepsze niż dla $n = 20$, ale są gorsze niż dla $n = 50$, jest to spowodowane losowością wybranej próby, spodziewalibyśmy się, żeby wyniki były lepsze dla większego n .

Wykres $Y = \sqrt{n\widehat{I}(\hat{\theta})}(\hat{\theta} - \theta)$

- $\theta = 0.5$



Tym razem histogram prawie pokrywa się z krzywą gęstości, co wskazywałoby na zachowanie podobne dla rozkładu normalnego. Wykres kwantylowo-kwantylowy również coraz bardziej zbliża się do wyznaczonej teoretycznej prostej. Wynik nie powinien nas dziwić, zgodnie z Centralnym Twierdzeniem Granicznym.

- $\theta \in \{0.5, 1, 2, 5\}$

Podobnie jak dla pozostałych n , wykresy dla podanych wartości parametru θ nie odbiega znacząco od powyższego.

Zadanie 5'

Estymatory $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$ oraz $\hat{\theta}_4$ (zdefiniowane w zadaniu 5.), gdzie $X \sim L(\theta, \sigma)$

- $n = 20$

(a) $\theta = 1, \sigma = 1$

| | MSE | Var | Bias |
|------------------|---------|---------|---------|
| $\hat{\theta}_1$ | 0.09981 | 0.09981 | 0.00113 |
| $\hat{\theta}_2$ | 0.06750 | 0.06749 | 0.00166 |
| $\hat{\theta}_3$ | 0.12983 | 0.12983 | 0.00269 |
| $\hat{\theta}_4$ | 0.16765 | 0.09173 | 0.27554 |

Każdy z wyników dla każdego estymatora $\hat{\theta}_i$ jest większy niż dla $n = 50$. Najlepszym estymatorem, który ma najmniejszy błąd średniokwadratowy, najmniejszą wariancję i prawie najniższe obciążenie jest $\hat{\theta}_2$ dla naszej próby $n = 20$.

(b) i (c)

Dla tych podpunktów również wyniki były gorsze niż dla większej próby. Najlepszym estymatorem również okazał się $\hat{\theta}_2$.

- $n = 100$

(a) $\theta = 1, \sigma = 1$

| | MSE | Var | Bias |
|------------------|---------|---------|----------|
| $\hat{\theta}_1$ | 0.01980 | 0.01978 | 0.00336 |
| $\hat{\theta}_2$ | 0.01133 | 0.01132 | 0.00137 |
| $\hat{\theta}_3$ | 0.02659 | 0.02658 | 0.00377 |
| $\hat{\theta}_4$ | 6.99362 | 0.02079 | -2.64061 |

Można zaobserwować, że MSE i Var są niższe dla każdego $\hat{\theta}_i$, jednak Bias prawie w każdym z przypadków jest wyższe, co oznaczałoby, że estymator jest bardziej obciążony dla większej próby. Ponownie najlepszym wyborem estymatora jest $\hat{\theta}_2$.

(b) i (c)

Takie same wnioski jak powyżej dla obu podpunktów. Estymator $\hat{\theta}_2$ jest najlepszym wyborem estymatora dla tego rozkładu niezależnie od wyboru wielkości próbki.