Raport drugiej listy z laboratorium Statystyki

Kamil Zdancewicz

November 5, 2025

Zadanie 1

Cel

Celem zadania było oszacowanie wariancji, błędu średniokwadratowego, obciążenia oraz zmienności wyników w zależności od parametru p estymatora największej wiarogodności wielkości $P(X \geq 3)$ dla rozkładu dwumianowego $b(5, p)$.

Stosowane metody

Dla każdego podpunktu (a) - (e) wygenerowałem 10 000 razy próbę wielkości $n = 50$. Dla każdej z tych prób obliczyłem estymator największej wiarygodności wielkości $P(X \geq 3)$, który został wyznaczony w następujący sposób:

Na początku chcemy wyznaczyć mle parametru p, ponieważ PMF rozkładu dwumianowego to:

$$f(k, n, p) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Logarytm funkcji wiarogodności dla próby wielkości n to:

$$l(p) = \sum_{i=1}^{n} \log L(x_i; p) = \text{const} + \sum_{i=1}^{n} X_i \log p + (5n - \sum_{i=1}^{n} X_i) \log{(1-p)}$$

Bierzemy pochodną, zakładamy, że $0 < p < 1$:

$$l'(p) = \frac{\sum_{i=1}^{n} X_i}{p} - \frac{5n - \sum_{i=1}^{n} X_i}{1-p}$$

Równanie $l'(p) = 0$ daje:

$$\frac{\sum_{i=1}^{n} X_i}{p} = \frac{5n - \sum_{i=1}^{n} X_i}{1-p}$$

Stąd otrzymujemy:

$$\sum_{i=1}^{n} X_i = 5np$$

Więc nasz estymator to:

$$\hat{p} = \frac{1}{5n} \sum_{i=1}^{n} X_i = \frac{\overline{X}}{5}$$

Sprawdźmy, czy to na pewno maksimum biorąc drugą pochodną:

$$l''(p) = -\frac{\sum_{i=1}^{n} X_i}{p^2} - \frac{5n - \sum_{i=1}^{n} X_i}{(1-p)^2}$$

Co widocznie jest ujemne, więc $L(\hat{p})$ jest maksimum.

Teraz rozważymy estymator $P(X \geq 3)$, samo prawdobieństwo wynosi (u nas $n = 5$):

$$P(X \geq 3) = \sum_{k=3}^{5} \binom{5}{k} p^k (1-p)^{5-k} = 10p^3(1-p)^2 + 5p^4(1-p) + p^5$$

Ponieważ własność estymatora największej wiarogodności jest zachowana przez nałożenie funkcji na ten estymator, przez podstawienie $\hat{p}$ do tego wzoru otrzymujemy mle $P(\widehat{X \geq 3})$.

W przypadku, gdy $\sum_{i=1}^{n} X_i = 0$ lub $\sum_{i=1}^{n} X_i = 5n$. $\hat{p}$ wynosi odpowiednio 0 lub 1.

Wyniki

(a) $p = 0.1$

Estymator	Wariancja	MSE	Obciążenie
$\hat{p}$	0.000025	0.000025	0.000845

(b) $p = 0.3$

Estymator	Wariancja	MSE	Obciążenie
$\hat{p}$	0.001464	0.001467	0.001933

(c) $p = 0.5$

Estymator	Wariancja	MSE	Obciążenie
$\hat{p}$	0.003452	0.003452	-0.000328

(d) $p = 0.7$

Estymator	Wariancja	MSE	Obciążenie
$\hat{p}$	0.001473	0.001478	-0.002206

(e) $p = 0.9$

Estymator	Wariancja	MSE	Obciążenie
$\hat{p}$	0.000024	0.000025	-0.000756

Wnioski

- Największe wariancje i błędy średniokwadratowe występują w wartościach pośrednich. Wynika to z faktu, że największa wariancja pojedynczej obserwacji X_i występuje w okolicach $p = 0.9$. Ponadto wyniki dla p i $1-p$ są bardzo podobne, co jest zgodne z funkcją gęstości.
- We wszytkich podpunktach obciążenie jest minimalne, co wskazuje na to, że estymator nie jest obciążony.

Załączniki

Generacja danych, obliczanie estymatora i dalsza logika znajduje się w załączonym pliku `main.cpp`.

Zadanie 2

Cel

Celem zadania było oszacowanie wariancji, błędu średniokwadratowego, obciążenia oraz zmienności wyników w zależności od parametru λ estymatora największej wiarogodności wielkości $P(X = x)$ dla $x = 0 \ldots 10$ dla rozkładu Poissona $\mathrm{Pois}(\lambda)$.

Stosowane metody

Dla każdego podpunktu (a) - (d) wygenerowałem 10 000 razy próbę wielkości $n = 50$. Dla każdej z tych prób obliczyłem estymator największej wiarygodności wielkości $P(X = x)$ dla $x = 0 \ldots 10$, który został wyznaczony w następujący sposób:

Na początku chcemy wyznaczyć mle parametru λ, ponieważ PMF rozkładu dwumianowego to:

$$f(k; \lambda) = P(X = k) = \frac{\lambda^k e^{-k}}{k!}$$

Logarytm funkcji wiarogodności dla próby wielkości n to:

$$l(p) = \sum_{i=1}^{n} \log L(x_i; \lambda) = \mathrm{const} - n\lambda + \ln \lambda \sum_{i=1}^{n} x_i$$

Biorą pochodną:

$$l'(p) = -n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i$$

Przyrównując do zera otrzymujemy estymator:

$$\widehat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Sprawdźmy czy $\widehat{\lambda}$ rzeczywiście maksymalizuje funkcję wiarogodności, druga pochodna:

$$l''(p) = -\frac{1}{\lambda^2} \sum_{i=1}^{n} x_i$$

Co w oczywisty sposób jest zawsze ujemne, więc $L(\widehat{\lambda})$ jest maksimum.

Przechodzimy do rozważania estymatora $P(X = x)$, ponieważ własność estymatora największej wiarogodności jest zachowana przez nałożenie funkcji na ten estymator, przez podstawienie $\widehat{\lambda}$ do tego wzoru otrzymujemy mle $P(\widehat{X = x})$.

Wyniki

(a) $\lambda = 0.5$

x	Wariancja	MSE	Obciążenie
0	0.003700	0.003714	0.003717

x	Wariancja	MSE	Obciążenie
1	0.000968	0.000992	-0.004906
2	0.000502	0.000502	0.000136
3	0.000043	0.000043	0.000745
4	0.000002	0.000002	0.000251
5	0.000000	0.000000	0.000049
6	0.000000	0.000000	0.000007
7	0.000000	0.000000	0.000001
8	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000

(b) $\lambda = 1.0$

x	Wariancja	MSE	Obciążenie
0	0.002703	0.002716	0.003771
1	0.000028	0.000041	-0.003644
2	0.000649	0.000652	-0.001856
3	0.000290	0.000290	0.000547
4	0.000044	0.000045	0.000729
5	0.000004	0.000004	0.000329
6	0.000000	0.000000	0.000098
7	0.000000	0.000000	0.000022
8	0.000000	0.000000	0.000004
9	0.000000	0.000000	0.000001
10	0.000000	0.000000	0.000000

(c) $\lambda = 2.0$

x	Wariancja	MSE	Obciążenie
0	0.000761	0.000768	0.002533
1	0.000707	0.000707	-0.000219
2	0.000014	0.000021	-0.002694
3	0.000319	0.000322	-0.001658
4	0.000314	0.000314	0.000119
5	0.000117	0.000118	0.000783
6	0.000025	0.000025	0.000630
7	0.000003	0.000004	0.000323
8	0.000000	0.000000	0.000127
9	0.000000	0.000000	0.000041
10	0.000000	0.000000	0.000011

(d) $\lambda = 5.0$

x	Wariancja	MSE	Obciążenie
0	0.000005	0.000005	0.000340
1	0.000076	0.000077	0.001006
2	0.000254	0.000256	0.001158
3	0.000303	0.000303	0.000262
4	0.000118	0.000119	-0.001045
5	0.000006	0.000009	-0.001723

x	Wariancja	MSE	Obciążenie
6	0.000085	0.000087	-0.001433
7	0.000166	0.000167	-0.000618
8	0.000147	0.000147	0.000121
9	0.000083	0.000083	0.000493
10	0.000034	0.000034	0.000533

Wnioski

- Ponieważ rozkład Poissona ma największą masę w okolicy $x \approx \lambda$, to obciążenie zachowuje się w następujący sposób:
  - gdy $x \ll \lambda$, to $P(X = x)$ jest bardzo małe i stabilne (małe zmiany w $\widehat{\lambda} \to$ małe zmiany w $P(X = x)$)
  - gdy $x \approx \lambda$, to $P(X = x)$ jest największe, więc najbardziej czułe na zmiany.
  - gdy $x \gg \lambda$, to $P(X = x)$ jest ponownie małe, więc nieczułe na zmiany $\widehat{\lambda}$
- Przez małe wartości obciążenia, pomimo fluktuacji jego wartości możemu wnioskować, że estymator $P(\widehat{X = x})$ nie jest obciążony
- Trochę inna sytuacja jest przy wariancji oraz błędzie średniokwadratowym. Oba rosną gdy x zbliża się do λ, lecz po przekroczeniu jego wartości, pomimo zmniejszania się $P(X = x)$ czynnik $(x - \lambda)^2$ wariancji rośnie szybciej niż tłumienie wykładnicze, więc wariancja rośnie przez parę wartości. Lecz później eksponens znowu zmniejsza te wartości.
- Z książki wiemy dodatkowo, że ten estymator jest efektywny

Załączniki

Generacja danych, obliczanie estymatora i dalsza logika znajduje się w załączonym pliku `main.cpp`.

Zadanie 3

Cel

Celem zadania było przedyskutowanie wybrania terminu **liczby pseudolosowe** zamiast losowe przez autorów książki `Statystyka dla studentów kierunków technicznych i przyrodniczych Koronacki; J., Mielniczuk, J. (2009)`

Omówienie

Autorzy wybrali termin **liczby pseudolosowe**, ponieważ w rozdziale `8.2.1` definiują ciąg z $u_1 = G(u_0), u_2 = G(u_1), \dots, u_i = G(u_{i-1}), i = 1, 2, \dots$ dla wybranej funkcji G określonej na odcinku $[0, 1]$ i ustalonego u_0.

Tak zdefiniowany ciąg umożliwia ponowne wygenerowanie próby losowej z rozkładu $U[0, 1]$. Imituje on zachowanie realizacji prawdziwie losowej próby $U_1, U_2, \dots, U_n$ rozkładu $U[0, 1]$. Pomimo pseudolosowości, powinna ona spełniać standardowe testy zgodności i niezależności elementów.

Czyli nie są to liczby faktycznie losowe, ale zachowuję się jak one.

Zadanie 4

Cel

Celem zadania było przeanalizowanie zachowania zmiennej losowej $Y = \sqrt{n\widehat{I(\theta)}}(\hat{\theta} - \theta)$, gdzie $\widehat{I(\theta)}$ i $\hat{\theta}$ są obliczane na podstawie niezależnie generowanych wartości w zależności od parametru θ rozkładu beta $\mathrm{B}(\theta, 1)$.

Stosowane metody

Dla każdego podpunktu (a) - (d) wygenerowałem 10 000 razy próbę wielkości $n = 50$. Dla każdej z tych prób obliczyłem estymator największej wiarygodności informacji Fishera parametru θ. Następnie niezależnie ponownie wygenerowałem 10 000 razy próbę wielkości $n = 50$, tym razem obliczając estymator największej wiarygodności parametru θ. Na koniec obliczyłem dla każdej z prób wartość zmiennej Y_i i wygenerowałem histogramy oraz wykresy kwantylowo-kwantylowe.

Mle informacji Fishera parametru θ oraz θ zostały obliczone w następujący sposób, zaczynamy od gęstości, u nas:

$$f(x;\theta) = \theta x^{\theta-1}$$

Log-wiarygodność:

$$l(\theta) = \sum_{i=1}^{n} (\ln\theta + (\theta-1)\ln X_i) = n\ln\theta + (\theta-1)\sum_{i=1}^{n}\ln X_i$$

Bierzemy pochodną:

$$l'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n}\ln X_i$$

Po przyrównaniu do zera otrzymujemy estymator:

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} X_i}$$

Aby upewnić się, że $\hat{\theta}$ maksymalizuje funkcję wiarogodności sprawdzamy drugą pochodną:

$$l''(\theta) = -\frac{n}{\theta^2}$$

Wiec $\hat{\theta}$ jest maksimum. Teraz informacja fishera dla jednej obserwacji:

$$I(\theta) = -\mathbb{E}[l''(\theta)] = \frac{1}{\theta^2}$$

Ponieważ własność estymatora największej wiarogodności jest zachowana przez nałożenie funkcji na ten estymator, $I(\hat{\theta}) = \widehat{I(\theta)}$.

W histogramie wybrałem $\sqrt{\text{repeats}} = 100$ jako liczbę klas, ponieważ daje ona dobry kompromis między przedstawieniem kształtu rozkładu liczb Y_i a klarownością wykresu.

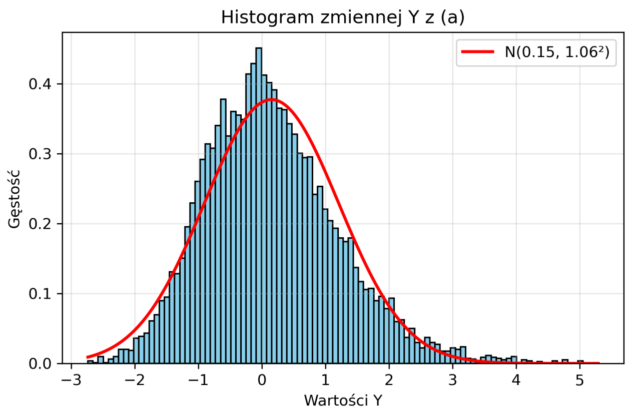
Na wykresie kwantylowo-kwantylowym kwantyle teoretyczne są wyznaczane na podstawie założonego rozkładu odniesienia - normalnego $N(0,1)$. Dla uporządkowanych wartości liczb Y_i. Odpowiadające im kwantyle teoretyczne są obliczane przez wzór:

$$q_i = F^{-1}\left(\frac{i-0.5}{n}\right)$$

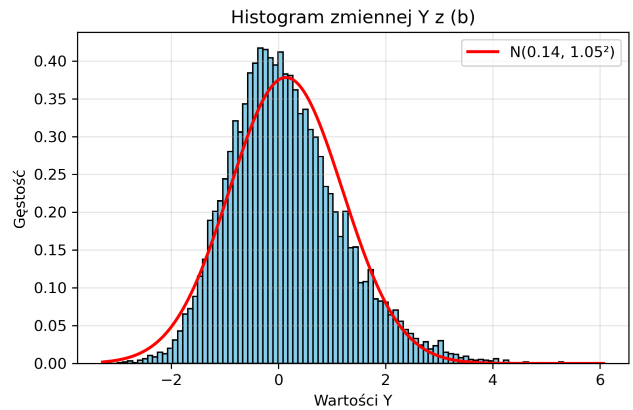
Gdzie F^{-1} jest odwrotnością dystrybuanty rozkładu normalnego.

Wyniki

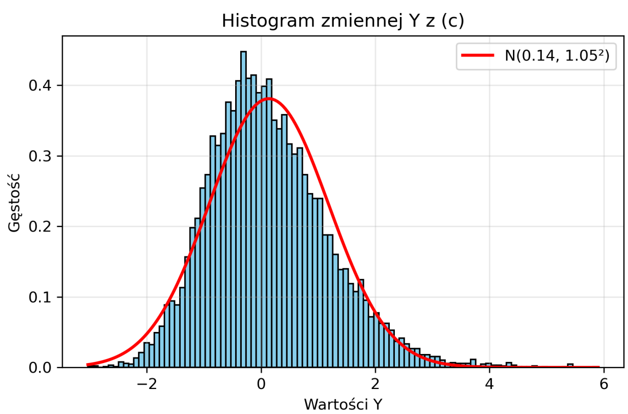
Histogramy



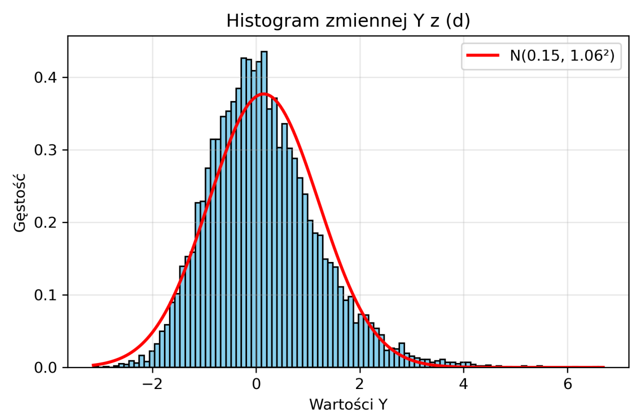
(a) $\theta = 0.5$



(b) $\theta = 1$

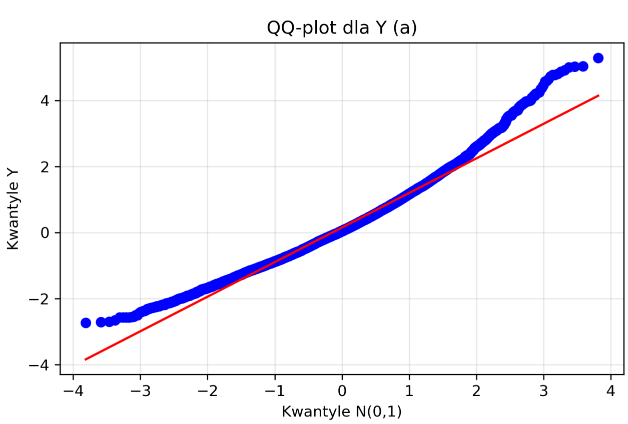


(c) $\theta = 2$

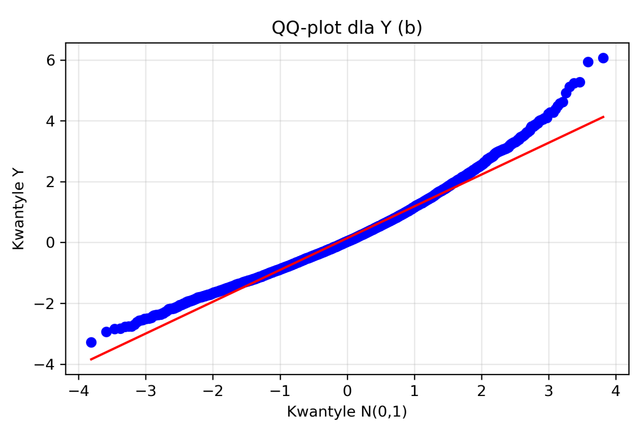


(d) $\theta = 5$

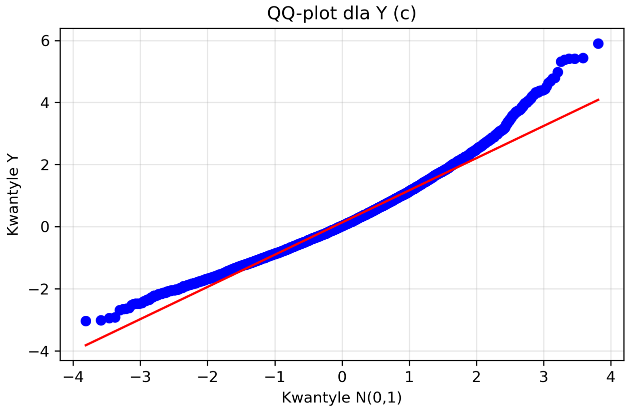
Wykresy kwantylowo-kwantylowe



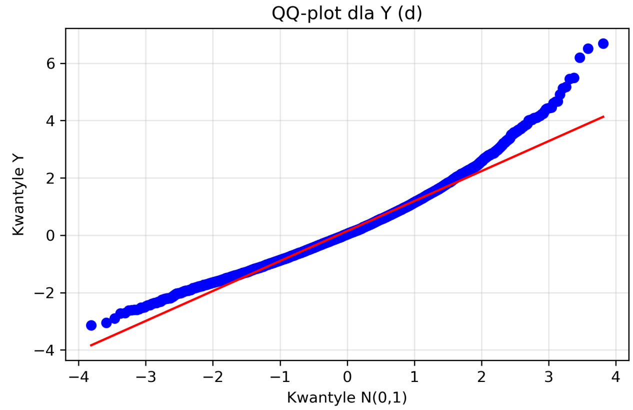
(a) $\theta = 0.5$



(b) $\theta = 1$



(c) $\theta = 2$



(d) $\theta = 5$

Wnioski

- Z histogramów widać, że rozkład zmiennych losowych Y_i bardzo przypomina rozkład normalny $N(0, 1)$. Kształt histogramu przypomina krzywą dzwonową.
- Na wykresach kwantylowo-kwantylowych widać, że uporządkowane obserwacje leżą blisko prostej $y = x$, co sugeruje, że mają rozkład podobny do rozkładu normalnego.
- Zgadza się to z teorią:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \theta^2)$$

Załączniki

Generacja danych, obliczanie estymatorów i dalsza logika znajduje się w załączonym pliku `main.cpp`. Natomiast generowanie wykresów znajduje się w załączonym pliku `graphE4.py`

Zadanie 5

Cel

Celem zadania było oszacowanie wariancji, błędu średniokwadratowego oraz obciążenia za pomocą czterech estymatorów przesunięcia rozkładu Laplace'a Laplace(θ, σ^2) dla różnych wartości θ i σ. Następnie wyznaczenie estymatora optymalnego i porównanie wyników z zadaniem 1 z pierwszej listy laboratoryjnej.

Stosowane metody

Dla każdego podpunktu (a) - (c) wygenerowałem 10 000 razy próbę wielkości $n = 50$. Dla każdej z tych prób obliczyłem cztery estymatory $\hat{\theta}$:

(i) $\hat{\theta}_1$ - średnia arytmetyczna próby

(ii) $\hat{\theta}_2$ - mediana próby

(iii) $\hat{\theta}_3$ - eksperymentalny nieobciążony estymator liniowy z losowymi wagami

(iv) $\hat{\theta}_4$ - potencjalnie obciążony estymator ważony sumy elementów próby, gdzie wagi są obliczane z pomocą funkcji gęstości i kwantyli

Następnie oszacowałem wariancję, błąd średniokwadratowy oraz obciążenie dla każdego z estymatorów we wszystkich podpunktach.

Wyniki

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

Estymator	Wariancja	MSE	Obciążenie
$\hat{\theta}_1$	0.038989	0.038988	0.001524
$\hat{\theta}_2$	0.023779	0.023777	-0.000401
$\hat{\theta}_3$	0.052072	0.052067	0.000610
$\hat{\theta}_4$	0.040071	0.153441	0.336711

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

Estymator	Wariancja	MSE	Obciążenie
$\hat{\theta}_1$	0.040309	0.040305	-0.000650
$\hat{\theta}_2$	0.024151	0.024149	-0.000858
$\hat{\theta}_3$	0.053519	0.053514	0.000176
$\hat{\theta}_4$	0.039588	7.135785	-2.663870

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

Estymator	Wariancja	MSE	Obciążenie
$\hat{\theta}_1$	0.157628	0.157634	-0.004689
$\hat{\theta}_2$	0.098103	0.098099	-0.002309
$\hat{\theta}_3$	0.210296	0.210281	-0.002632
$\hat{\theta}_4$	0.156752	2.960178	1.674348

Wnioski

- We wszystkich podpunktach estymatory (i) - (iii) mają bardzo małe obciążenie (tak samo jak w zadaniu 1 z listy 1).
- Estymator (ii) ma konsekwentnie najmniejszą wariancję oraz błąd średniokwadratowy (kiedy w zadaniu 1 z listy 1 najlepszy był estymator (i)).
- We wszystkich przypadkach oprócz (iv) wariancje i błędy średniokwadratowe wzrastają kwadratowo do wzrostu σ, co jest zgodne z teorią (tak samo jak w zadaniu 1 z listy 1).
- Po wszystkich podpunktach widać, że estymator (iv) jest obciążony (bo bazuje na $N(0,1)$, któremu dalego do rozkładu Laplace'a), co skutkuje dużymi błędami średniokwadratowymi. (Jednak wariancja (iv) nie jest już najmniejsza spośród tych estymatorów, przeciwnie do zadania 1 z listy 1)

Załączniki

Generacja danych, obliczanie estymatorów i dalsza logika znajduje się w załączonym pliku `main.cpp`.

Zadanie 6

Cel

Przez zmianę wielkości próby na $n = 20$ albo $n = 100$ dla zadań 1,2,4 i 5 należy zauważyć różnicę między wynikami dla oryginalnej wielkości próby a zmodyfikowanej. (Tu będzie **BARDZO** dużo danych)

Wyniki dla $n = 20$

Zadanie 1

(a) $p = 0.1$

Estymator	Wariancja	MSE	Obciążenie
$\hat{p}$	0.000079	0.000083	0.002075

(b) $p = 0.3$

Estymator	Wariancja	MSE	Obciążenie
$\hat{p}$	0.003693	0.003719	0.005142

(c) $p = 0.5$

Estymator	Wariancja	MSE	Obciążenie
$\hat{p}$	0.008256	0.008256	0.000806

(d) $p = 0.7$

Estymator	Wariancja	MSE	Obciążenie
$\hat{p}$	0.003751	0.003782	-0.005594

(e) $p = 0.9$

Estymator	Wariancja	MSE	Obciążenie
$\hat{p}$	0.000075	0.000079	-0.001903

Zadanie 2

(a) $\lambda = 0.5$

x	Wariancja	MSE	Obciążenie
0	0.008922	0.008971	0.007030
1	0.002398	0.002516	-0.010871
2	0.001178	0.001179	0.001039
3	0.000115	0.000119	0.001985
4	0.000006	0.000006	0.000660
5	0.000000	0.000000	0.000134
6	0.000000	0.000000	0.000020
7	0.000000	0.000000	0.000003
8	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000

(b) $\lambda = 1.0$

x	Wariancja	MSE	Obciążenie
0	0.006927	0.006991	0.008020
1	0.000172	0.000257	-0.009215
2	0.001570	0.001585	-0.003835
3	0.000704	0.000707	0.001824
4	0.000118	0.000122	0.001983
5	0.000011	0.000012	0.000881
6	0.000001	0.000001	0.000266
7	0.000000	0.000000	0.000063
8	0.000000	0.000000	0.000012
9	0.000000	0.000000	0.000002
10	0.000000	0.000000	0.000000

(c) $\lambda = 2.0$

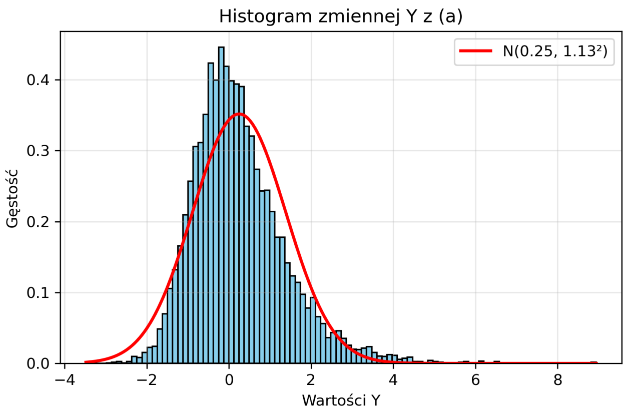
x	Wariancja	MSE	Obciążenie
0	0.001995	0.002057	0.007902
1	0.001621	0.001621	0.000915
2	0.000085	0.000128	-0.006556
3	0.000766	0.000791	-0.005010
4	0.000723	0.000724	-0.000774
5	0.000280	0.000282	0.001215
6	0.000064	0.000066	0.001215
7	0.000010	0.000011	0.000678
8	0.000001	0.000001	0.000282
9	0.000000	0.000000	0.000096
10	0.000000	0.000000	0.000028

(d) $\lambda = 5.0$

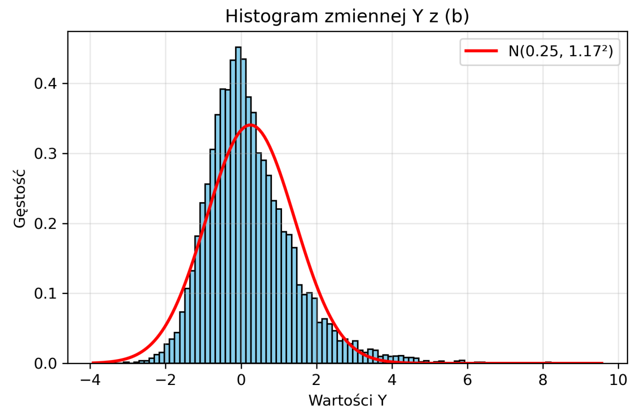
x	Wariancja	MSE	Obciążenie
0	0.000015	0.000016	0.000833
1	0.000206	0.000211	0.002381
2	0.000633	0.000640	0.002594
3	0.000710	0.000710	0.000342
4	0.000272	0.000279	-0.002721
5	0.000033	0.000050	-0.004168
6	0.000212	0.000223	-0.003325
7	0.000387	0.000389	-0.001319
8	0.000344	0.000344	0.000436
9	0.000199	0.000200	0.001289
10	0.000085	0.000086	0.001352

Zadanie 4

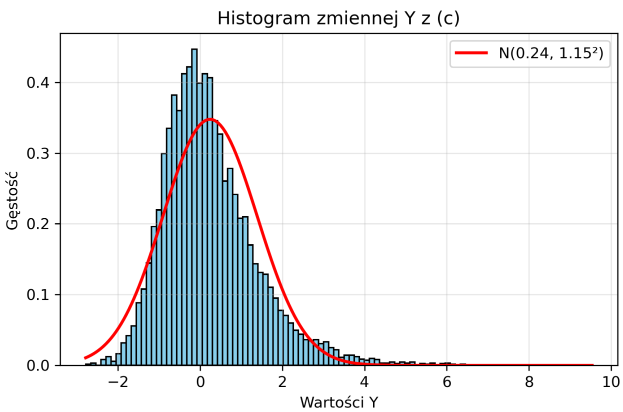
Histogramy



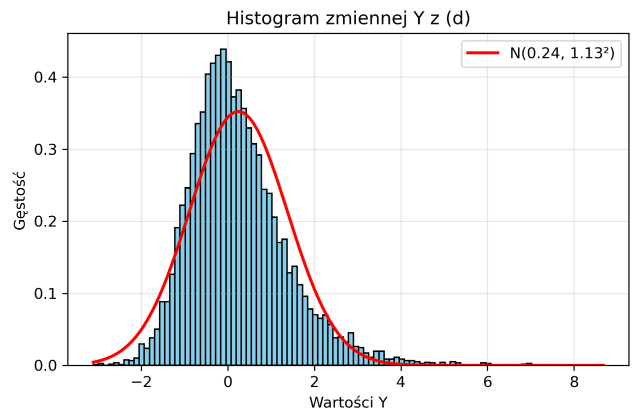
(a) $\theta = 0.5$



(b) $\theta = 1$

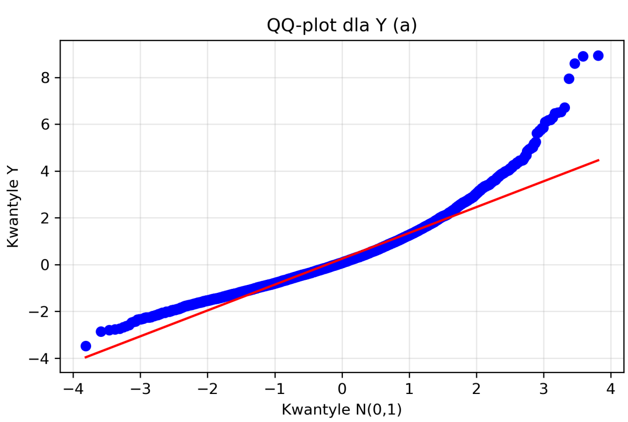


(c) $\theta = 2$

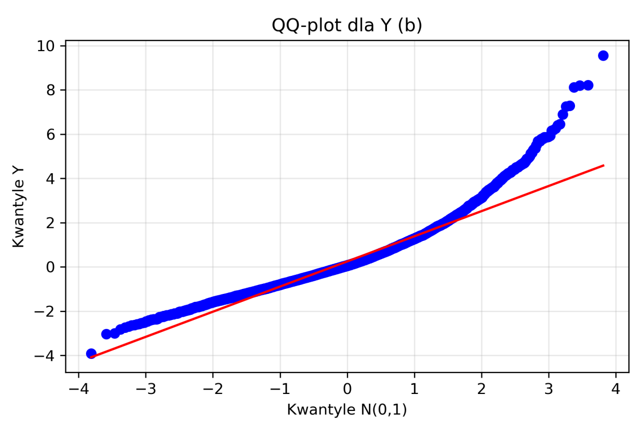


(d) $\theta = 5$

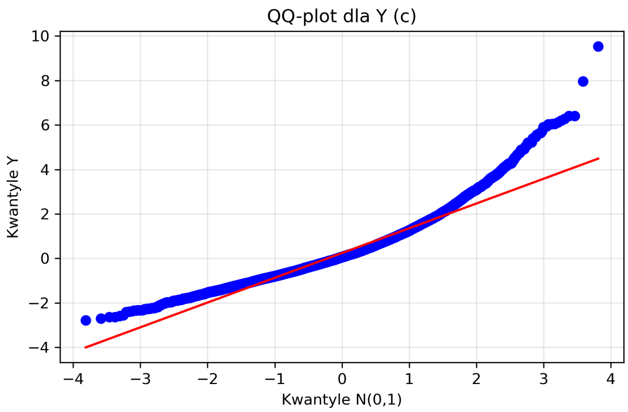
Wykresy kwantylowo-kwantylowe

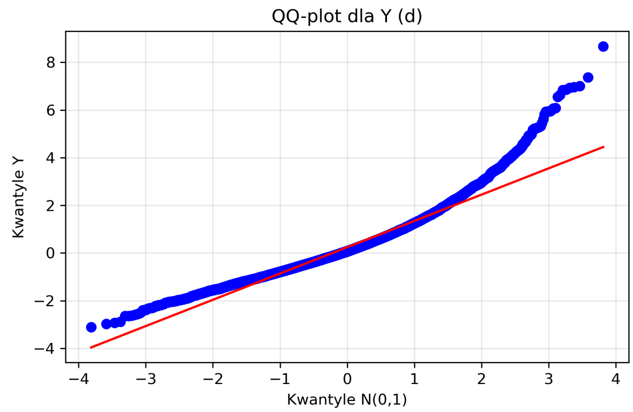


(a) $\theta = 0.5$



(b) $\theta = 1$





(c) $\theta = 2$ (d) $\theta = 5$

Zadanie 5

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

Estymator	Wariancja	MSE	Obciążenie
$\widehat{\theta}_1$	0.099425	0.099422	-0.002504
$\widehat{\theta}_2$	0.065934	0.065932	0.002104
$\widehat{\theta}_3$	0.129775	0.129773	-0.003342
$\widehat{\theta}_4$	0.090486	0.162603	0.268563

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

Estymator	Wariancja	MSE	Obciążenie
$\widehat{\theta}_1$	0.099862	0.099860	-0.002722
$\widehat{\theta}_2$	0.066490	0.066488	-0.002053
$\widehat{\theta}_3$	0.129644	0.129634	-0.001786
$\widehat{\theta}_4$	0.088795	7.536974	-2.729137

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

Estymator	Wariancja	MSE	Obciążenie
$\widehat{\theta}_1$	0.397263	0.397226	0.001569
$\widehat{\theta}_2$	0.261114	0.261088	-0.000763
$\widehat{\theta}_3$	0.521866	0.521815	0.001041
$\widehat{\theta}_4$	0.364839	2.757211	1.546741

Wyniki dla $n = 100$

Zadanie 1

(a) $p = 0.1$

Estymator	Wariancja	MSE	Obciążenie
$\widehat{p}$	0.000012	0.000012	0.000422

(b) $p = 0.3$

Estymator	Wariancja	MSE	Obciążenie
$\widehat{p}$	0.000723	0.000724	0.001010

(c) $p = 0.5$

Estymator	Wariancja	MSE	Obciążenie
$\widehat{p}$	0.001711	0.001711	0.000521

(d) $p = 0.7$

Estymator	Wariancja	MSE	Obciążenie
$\widehat{p}$	0.000727	0.000728	-0.000900

(e) $p = 0.9$

Estymator	Wariancja	MSE	Obciążenie
$\widehat{p}$	0.000012	0.000012	-0.000412

Zadanie 2

(a) $\lambda = 0.5$

x	Wariancja	MSE	Obciążenie
0	0.001858	0.001863	0.002258
1	0.000476	0.000483	-0.002661
2	0.000256	0.000256	-0.000079
3	0.000021	0.000021	0.000336
4	0.000001	0.000001	0.000118
5	0.000000	0.000000	0.000023
6	0.000000	0.000000	0.000003
7	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000

(b) $\lambda = 1.0$

x	Wariancja	MSE	Obciążenie
0	0.001381	0.001385	0.002203
1	0.000007	0.000010	-0.001872
2	0.000338	0.000340	-0.001095
3	0.000151	0.000151	0.000197
4	0.000022	0.000022	0.000346
5	0.000002	0.000002	0.000161
6	0.000000	0.000000	0.000048
7	0.000000	0.000000	0.000011
8	0.000000	0.000000	0.000002
9	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000

(c) $\lambda = 2.0$

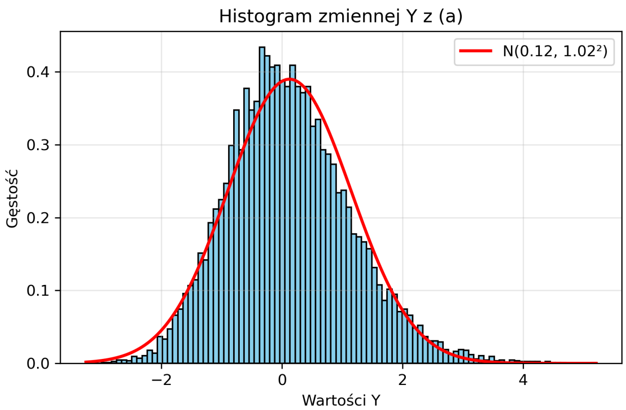
x	Wariancja	MSE	Obciążenie
0	0.000374	0.000376	0.001314
1	0.000360	0.000360	-0.000052
2	0.000004	0.000005	-0.001350
3	0.000161	0.000162	-0.000867
4	0.000160	0.000160	0.000028
5	0.000059	0.000059	0.000375
6	0.000012	0.000012	0.000307
7	0.000002	0.000002	0.000158
8	0.000000	0.000000	0.000062
9	0.000000	0.000000	0.000020
10	0.000000	0.000000	0.000005

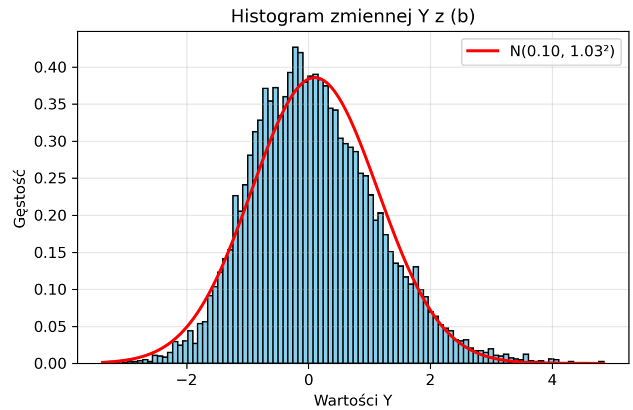
(d) $\lambda = 5.0$

x	Wariancja	MSE	Obciążenie
0	0.000002	0.000002	0.000160
1	0.000038	0.000038	0.000465
2	0.000129	0.000129	0.000507
3	0.000156	0.000156	0.000044
4	0.000061	0.000061	-0.000587
5	0.000002	0.000002	-0.000877
6	0.000043	0.000044	-0.000682
7	0.000086	0.000086	-0.000245
8	0.000076	0.000076	0.000128
9	0.000042	0.000042	0.000300
10	0.000017	0.000017	0.000301

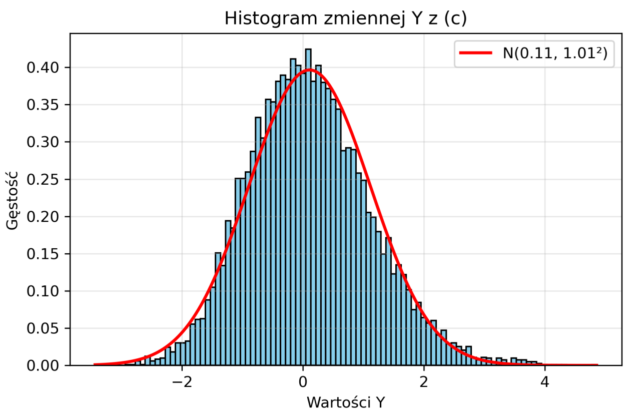
Zadanie 4

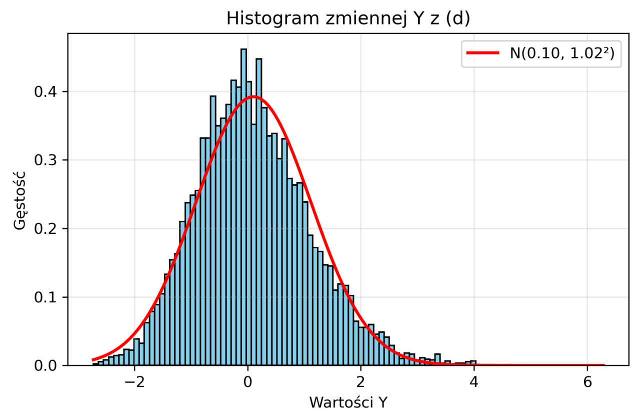
Histogramy





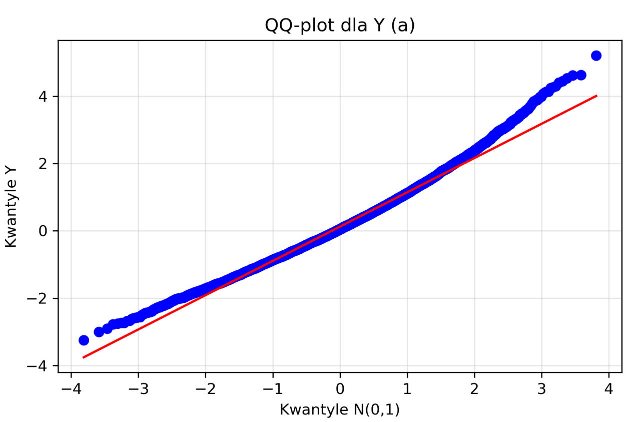
(a) $\theta = 0.5$ (b) $\theta = 1$

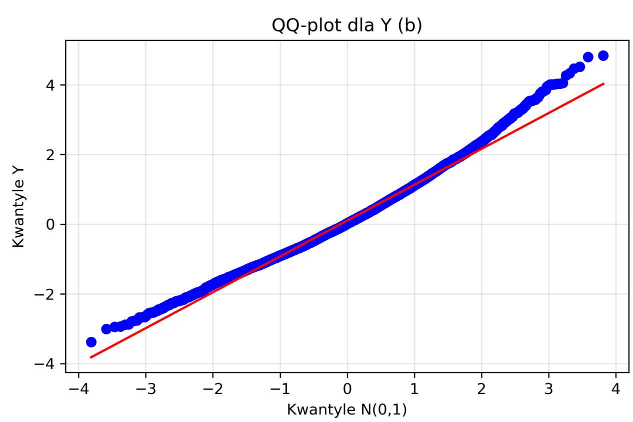




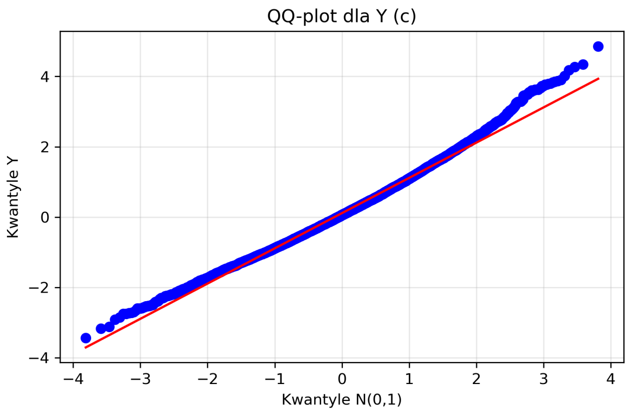
(c) $\theta = 2$ (d) $\theta = 5$

Wykresy kwantylowo-kwantylowe

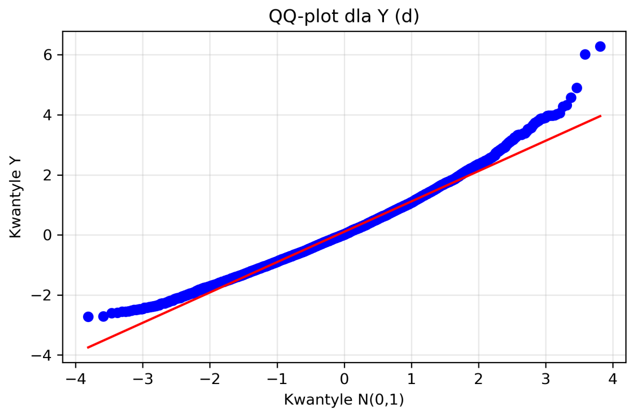




(a) $\theta = 0.5$ (b) $\theta = 1$



(c) $\theta = 2$



(d) $\theta = 5$

Zadanie 5

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

Estymator	Wariancja	MSE	Obciążenie
$\hat{\theta}_1$	0.019839	0.019837	0.000074
$\hat{\theta}_2$	0.011544	0.011549	0.002443
$\hat{\theta}_3$	0.026689	0.026687	0.000340
$\hat{\theta}_4$	0.020916	0.149759	0.358950

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

Estymator	Wariancja	MSE	Obciążenie
$\hat{\theta}_1$	0.020558	0.020556	-0.000016
$\hat{\theta}_2$	0.011835	0.011834	-0.000587
$\hat{\theta}_3$	0.026870	0.026868	-0.000604
$\hat{\theta}_4$	0.020559	7.002676	-2.642370

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

Estymator	Wariancja	MSE	Obciążenie
$\hat{\theta}_1$	0.081894	0.081888	0.001399
$\hat{\theta}_2$	0.047051	0.047050	0.002101
$\hat{\theta}_3$	0.106095	0.106094	0.003090
$\hat{\theta}_4$	0.081484	3.025137	1.715710

Wnioski

Zadanie 1 Wariancje i błędy średniokwadratowe zdają się zachowywać odwrotnie proporcjonalnie względem wielkości n, co dobrze wskazuje na zachowanie MLE w rozkładzie dwumianowym.

Ponownie n zdaje się nie wpływać na obciążenie, więc estymator prawdopodobnie nie jest obciążony.

Zadanie 2 Wariancje, błędy średniokwadratowe i **obciążenie** zmieniają się 2-3 krotnie, przy 2 krotnym powiększeniu lub zmniejszeniu n (ciągle odwrotnie proporcjonalnie względem n).

Obciążenie zmniejszające się z wzrostem n jest wartościową własnością.

Zadanie 4 Wraz ze zwiększeniem n zarówno histogramy, jak i wykresy kwantylowo-kwantylowe wskazują, że rozkład zmiennych Y_i dąży względem rozkładu do rozkładu normalnego $N(0, 1)$

Przy mniejszym n więcej masy jest skoncentrowane przy kwantylu 0, oraz przesunięcie jest większe.

Zadanie 5 We wszystkich estymatorach zauważalny jest spadek wariancji i błędu średniokwadratowego wzgędem n.

Niezależnie od n, estymatory (i) - (iii) mają obciążenie bliskie 0, co jeszcze bardziej wskazuje na ich nieobciążoność. Natomiast obciążenie estymatora (iv) nie zmienia się względem n.

Przy większych n estymatory (i) oraz (iv) mają podobną wariancję.

Estymator (iv) nawet w próbie $n = 100$ dla podpunktów (b) i (c) utrzymuje duży błąd średniokwadratowy i obciążenie wskazując na to, że nie jest odporny na przesunięcia ani zmianę skali.

Źródła

- *Introduction to Mathematical Statistics (6th edition) - Hogg, McKean, Craig*
- *Statystyka dla studentów kierunków technicznych i przyrodniczych Koronacki; J., Mielniczuk, J. (2009)*