

# Lista 2 - raport

Erwin Jasic

28 listopada 2020

## Cel raportu

W tym raporcie przyjrzymy się zachowaniu estymatorów parametrów pochodzących z różnych rozkładów. Zobaczymy, czy wartości jakie otrzymamy np. dla ENW zgadzają się z teorią z wykładów.

## Zadanie 1

W tabeli w zadaniu 1 opisujemy wartości statystyk dla wariancji, błędu średniokwadratowego oraz obciążenia (kolumny) dla różnych wartości parametru  $p$  (wiersze), gdzie  $p$  pochodzi z rozkładu dwumianowego. Widzimy, że dla par  $p = 0.1$  i  $p = 0.9$  oraz  $p = 0.3$  i  $p = 0.7$  wartości odpowiednich statystyk są bardzo zbliżone. Wynika to z tego, że rozkład dwumianowy jest symetryczny. Dla  $p = 0.1$  i  $p = 0.9$  wartości dla wszystkich statystyk są najmniejsze, co oznacza, że dla  $p = 0.1$  i  $p = 0.9$  estymator będzie najlepszy. Wartości dla  $p = 0.5$  cechuje największa wariancja oraz błąd średniokwadratowy, a dla  $p = 0.3$  i  $p = 0.7$  największe jest obciążenie.

Table 1: Wartość ENW

	wariancja	błąd śr kwadratowy	obciążenie
$p = 0.1$	0.0000241	0.0000247	0.0007470
$p = 0.3$	0.0014813	0.0014862	0.0022307
$p = 0.5$	0.0034939	0.0034950	-0.0010404
$p = 0.7$	0.0014903	0.0014963	-0.0024592
$p = 0.9$	0.0000244	0.0000250	-0.0007931

## Zadanie 2

W tabelach w zadaniu 2 opisujemy wartości statystyk dla wariancji, błędu średniokwadratowego oraz obciążenia (kolumny) dla różnych parametrów  $\lambda$  (wiersze), gdzie  $\lambda$  to parametr w rozkładzie Poissona. Pierwsza tabela opisuje wartości statystyk dla  $x = 0$ , druga dla  $x = 5$ , a trzecia dla  $x = 10$ . Przyjrzymy się tabeli pierwszej dla  $x = 0$ . Widzimy, że w tej tabeli wartości statystyk są dosyć spore (poza  $\lambda = 5$ ), czyli obciążenie estymatora jest duże. Dla  $x = 5$  wygląda to już znacznie lepiej, szczególnie dla  $\lambda = 0.5$  wartości są niemalże zerowe, co świadczy o tym, że estymator jest nieobciążony. Im większa wartość  $\lambda$  tym większe wartości statystyk. Dla  $x = 10$  i dla  $\lambda = 0.5$ ,  $\lambda = 1$ ,  $\lambda = 2$  wartości statystyk są zerowe lub bardzo bliskie zerowym, co oznacza, że estymator jest nieobciążony. Sytuacja wygląda nieco gorzej dla  $\lambda = 5$ , ale nadal dobrze.

Table 2: Wartość ENW, gdzie  $x = 0$

	wariancja	błąd śr kwadratowy	obciążenie
$\lambda = 0.5$	0.0037085	0.0975531	0.3063406
$\lambda = 1$	0.0028529	0.0028677	0.0038455
$\lambda = 2$	0.0007552	0.0182766	-0.1323687
$\lambda = 5$	0.0000052	0.0007139	-0.0266230

Table 3: Wartość ENW, gdzie  $x = 5$

	wariancja	błąd śr kwadratowy	obciążenie
$\lambda = 0.5$	0.0000000	0.0000001	0.0001957
$\lambda = 1$	0.0000038	0.0000122	0.0029052
$\lambda = 2$	0.0001148	0.0007222	0.0246463
$\lambda = 5$	0.0000060	0.0007625	0.0275050

Table 4: Wartość ENW, gdzie  $x = 10$

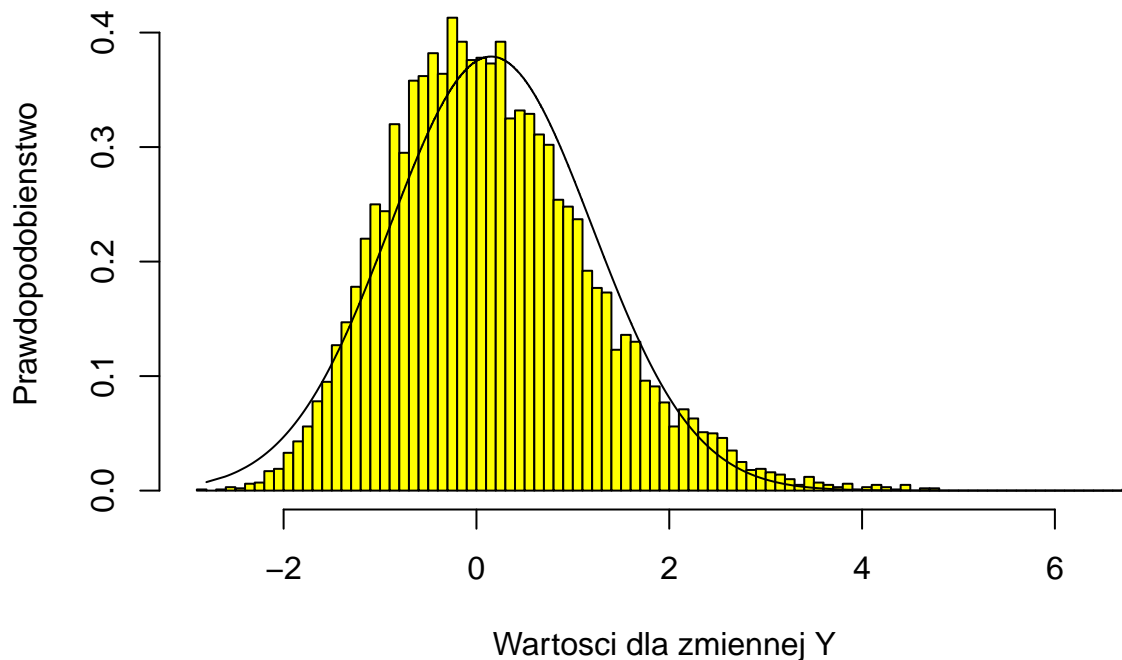
	wariancja	błąd śr kwadratowy	obciążenie
$\lambda = 0.5$	0.0000000	0.0000000	0.0000000
$\lambda = 1$	0.0000000	0.0000000	0.0000002
$\lambda = 2$	0.0000000	0.0000000	0.0000418
$\lambda = 5$	0.000034	0.0001435	0.0104679

#### Zadanie 4

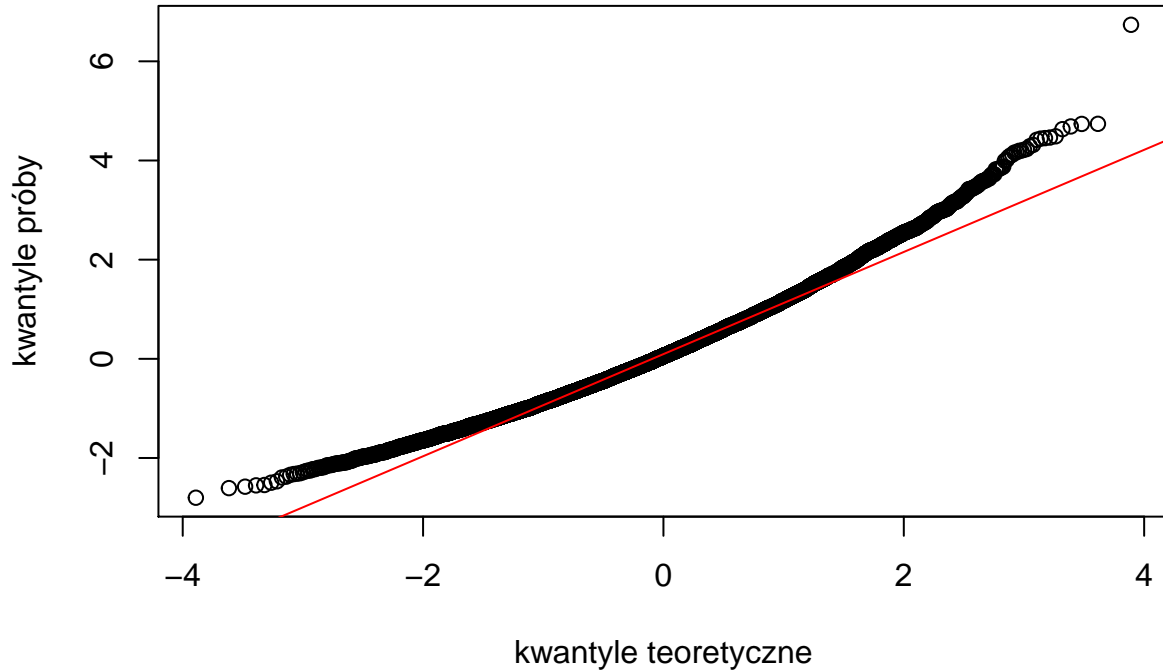
Z laboratorium wiemy, że dla rozkładu beta z parametrami  $\theta, 1$ , informacja Fishera jest wyrażona za pomocą wzoru  $I(\theta) = \frac{1}{\theta^2}$ , gdzie  $\hat{\theta}$  jest estymatorem największej wiarygodności (ENW) parametru  $\theta$ ,  $\hat{\theta} = \frac{-n}{\sum_{i=1}^n \log X_i}$

Przy tworzeniu histogramu, ważna jest liczba klas. Z PSP wiemy, że liczba klas powinna być zbliżona do pierwiastka z liczby obserwacji. U nas jest to  $\sqrt{10000} = 100$ .

### Histogram dla theta = 0.5



### wykres kwantylowo–kwantylowy dla $\theta = 0.5$



Dla każdej wartości  $\theta$  wykresy histogramów oraz wykresy kwantylowo-kwantylowe wychodzą podobnie, więc do analizy poddaliśmy przypadek, gdzie  $\theta = 0.5$ . Widzimy, że kształty histogramów układają się podobnie do rozkładu normalnego  $N(\bar{Y}, \sigma(Y))$ , ale na wykresie kwantylowo-kwantylowym widzimy, że punkty układają się wzdłuż pewnej krzywej (nie wyznaczonej prostej koloru czerwonego), co oznacza, że nie jest to rozkład normalny.

#### Zadanie 5

W tabelach w zadaniu 5, opisujemy wartości statystyk dla wariancji, błędu średniokwadratowego oraz obciążenia (kolumny) dla różnych wartości parametru  $\hat{\theta}$  (wiersze), gdzie estymatory parametru  $\theta$  są opisane wzorami jak w zadaniu. Analizując tebele, widzimy, że niezależnie od wyboru parametrów rozkładu Laplace’a  $\hat{\theta}_2$  jest najlepszym estymatorem największej wiarygodności a z kolei  $\hat{\theta}_4$  jest najgorszym. Oceniając to patrzymy oczywiście na wariancję, błędy średniokwadratowe oraz obciążenie. Tam gdzie te wartości są niskie (bliskie zeru) to jest to dobry estymator, jeśli nie to jest to słaby estymator. Na poprzedniej liście wyszło, że najlepszym estymatorem jest  $\hat{\theta}_1$  (średnia). Zatem widzimy, że estymator nie jest uniwersalny, a zależy od rozkładu.

Table 5: Laplace(50, 1, 1)

	wariancja	błąd śr kwadratowy	obciążenie
th1	0.0393437	0.0393446	-0.0009247
th2	0.0240910	0.0240911	-0.0002802
th3	0.0559837	0.0559837	0.0000870
th4	0.0402262	0.1514053	0.3334353

Table 6: Laplace(50, 4, 1)

	wariancja	błąd śr kwadratowy	obciążenie
th1	0.0397927	0.0398138	0.0045928
th2	0.0248938	0.0248989	0.0022718
th3	0.0569806	0.0569909	0.0032050
th4	0.0390296	7.1474577	-2.6661636

Table 7: Laplace(50, 1, 2)

	wariancja	błąd śr kwadratowy	obciążenie
th1	0.1609632	0.1609633	-0.0002386
th2	0.0973571	0.0973573	0.0004997
th3	0.2382997	0.2383047	-0.0022371
th4	0.1611311	2.9645932	1.6743542

**Zadanie 6**Zadanie 1 dla  $n = 20$ 

Table 8: Wartość ENW

	wariancja	błąd śr kwadratowy	obciążenie
$p = 0.1$	0.0000723	0.0000754	0.0017581
$p = 0.3$	0.0037012	0.0037335	0.0056901
$p = 0.5$	0.0082769	0.0082772	-0.0005759
$p = 0.7$	0.0036264	0.0036537	-0.0052253
$p = 0.9$	0.0000759	0.0000800	-0.0020422

Zadanie 1 dla  $n = 100$ 

Table 9: Wartość ENW

	wariancja	błąd śr kwadratowy	obciążenie
$p = 0.1$	0.0000113	0.0000115	0.0003543
$p = 0.3$	0.0007482	0.0007499	0.0013199
$p = 0.5$	0.0017442	0.0017444	0.0003977
$p = 0.7$	0.0007430	0.0007440	-0.0010138
$p = 0.9$	0.0000114	0.0000116	-0.0003922

Zwiększając rozmiar próby zmniejszamy wartości statystyk w tabeli, natomiast gdy zmniejszamy rozmiar próby dzieje się na odwrót.

Zadanie 2 dla  $n = 20$ Table 10: Wartość ENW, gdzie  $x = 0$ 

	wariancja	błąd śr kwadratowy	obciążenie
$\lambda = 0.5$	0.0090946	0.1054654	0.3104365

	wariancja	błąd śr kwadratowy	obciążenie
$\lambda = 1$	0.0070336	0.0071225	0.0094286
$\lambda = 2$	0.0019672	0.0185525	-0.1287840
$\lambda = 5$	0.0000154	0.0006942	-0.0260537

Table 11: Wartość ENW, gdzie  $x = 5$

	wariancja	błąd śr kwadratowy	obciążenie
$\lambda = 0.5$	0.0000002	0.0000003	0.0002818
$\lambda = 1$	0.0000118	0.0000234	0.0034069
$\lambda = 2$	0.0002841	0.0009490	0.0257849
$\lambda = 5$	0.0000340	0.0006607	0.0250349

Table 12: Wartość ENW, gdzie  $x = 10$

	wariancja	błąd śr kwadratowy	obciążenie
$\lambda = 0.5$	0.0000000	0.0000000	0.0000000
$\lambda = 1$	0.0000000	0.0000000	0.0000004
$\lambda = 2$	0.0000000	0.0000000	0.0000614
$\lambda = 5$	0.0000856	0.000209	0.0111095

Zadanie 2 dla  $n = 100$

Table 13: Wartość ENW, gdzie  $x = 0$

	wariancja	błąd śr kwadratowy	obciążenie
$\lambda = 0.5$	0.0018155	0.0948526	0.3050199
$\lambda = 1$	0.0014186	0.0014231	0.0021109
$\lambda = 2$	0.0003686	0.0182908	-0.1338738
$\lambda = 5$	0.0000025	0.0007195	-0.0267756

Table 14: Wartość ENW, gdzie  $x = 5$

	wariancja	błąd śr kwadratowy	obciążenie
$\lambda = 0.5$	0.0000000	0.0000000	0.0001685
$\lambda = 1$	0.0000017	0.0000091	0.0027229
$\lambda = 2$	0.0000570	0.0006504	0.0243583
$\lambda = 5$	0.0000017	0.0008051	0.0283439

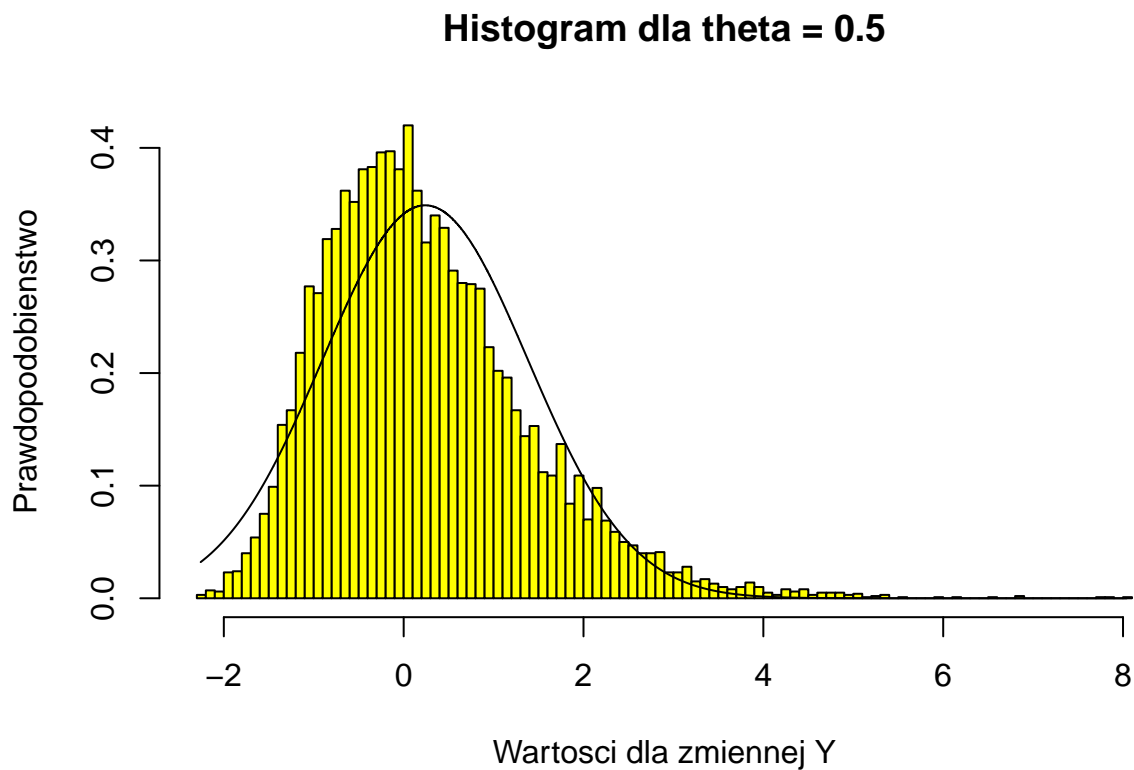
Table 15: Wartość ENW, gdzie  $x = 10$

	wariancja	błąd śr kwadratowy	obciążenie
$\lambda = 0.5$	0.0000000	0.0000000	0.0000000
$\lambda = 1$	0.0000000	0.0000000	0.0000001
$\lambda = 2$	0.0000000	0.0000000	0.0000363

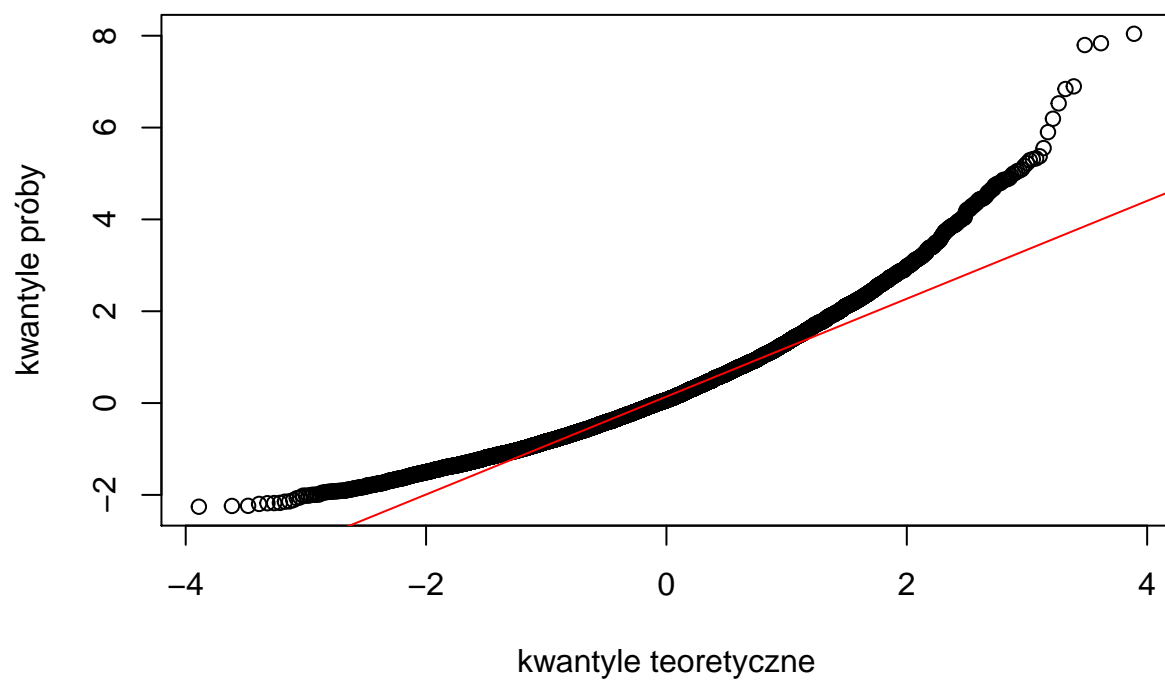
	wariancja	błąd śr kwadratowy	obciążenie
$\lambda = 5$	0.0000171	0.0001205	0.0101686

Podobnie jak dla zadania 1 zwiększając rozmiar próby zmniejszamy wartości statystyk w tabeli, natomiast gdy zmniejszamy rozmiar próby dzieje się na odwrót. Możemy zauważyć że dla próby wielkości 20, 50 oraz 100 wartości statystyk są małe co oznacza, że już dla próby wielkości 20 dostajemy dosyć dobre estymatory.

Zadanie 4 dla  $n = 20$

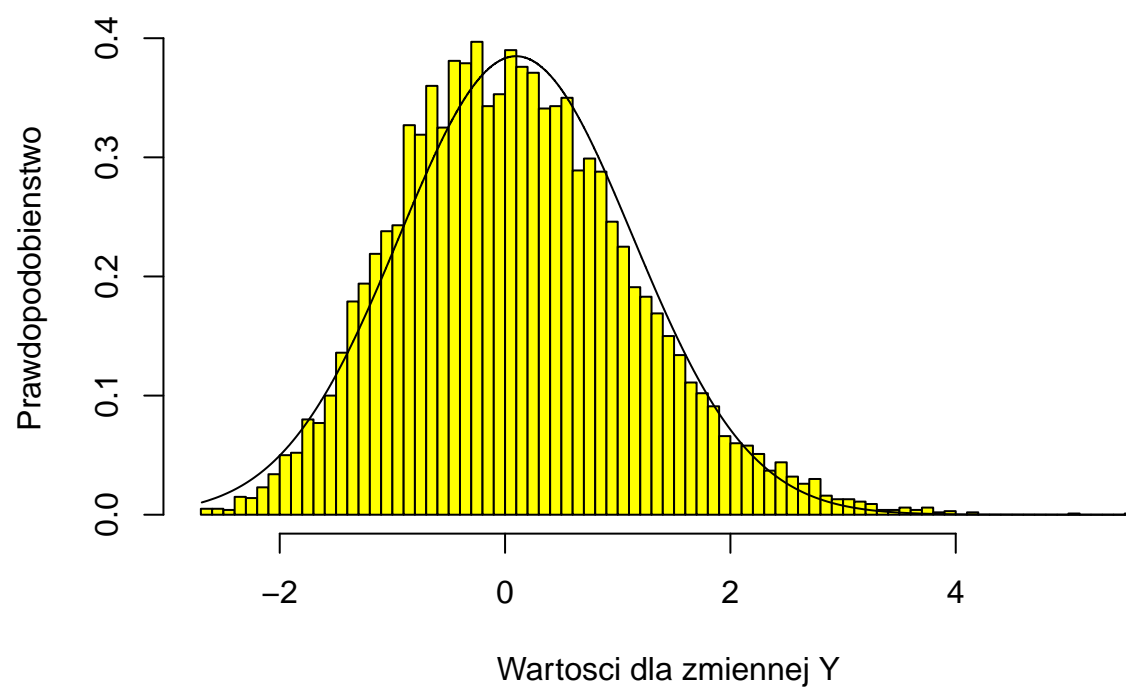


**wykres kwantylowo–kwantylowy dla  $\theta = 0.5$**



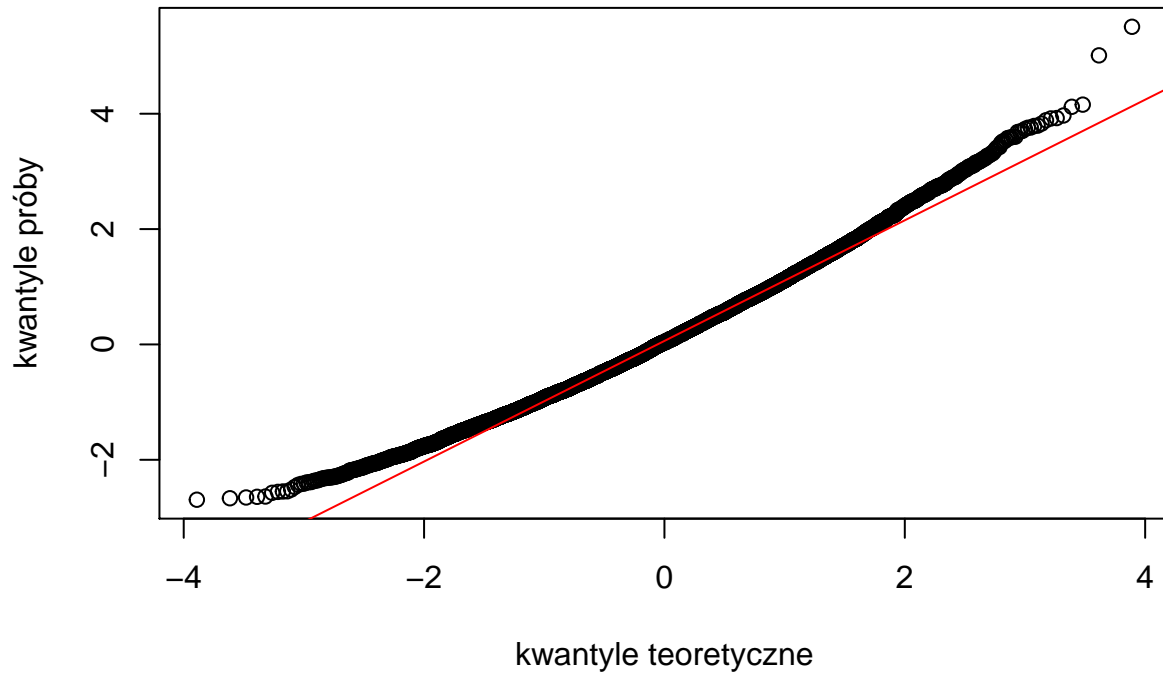
Zadanie 4 dla  $n = 100$

**Histogram dla theta = 0.5**





### wykres kwantylowo–kwantylowy dla $\theta = 0.5$



W tym zadaniu widzimy, że wraz ze wzrostem wielkości próby coraz bardziej przybliżamy się do rozkładu normalnego (poprzez analizę histogramu oraz wykresu kwantylowo-kwantylowego). Widzimy, że dla próby równej 20 histogram nie pokrywa się tak dobrze z funkcją rozkładu normalnego jak to jest w przypadku próby wielkości 100. To samo możemy zaobserwować na wykresie kwantylowo-kwantylowym. Im większa próba tym punkty układają się bliżej czerwonej prostej, co oznacza, że jesteśmy bliżej uzyskania rozkładu normalnego. Ta obserwacja zgadza się z teorią z wykładu, że zmienna  $Y$  dąży do rozkładu normalnego, gdy  $n$  dąży do nieskończoności.

Zadanie 5 dla  $n = 20$

Table 16: Laplace(50, 1, 1)

	wariancja	błąd śr kwadratowy	obciążenie
th1	0.0992121	0.0992231	-0.0033236
th2	0.0661633	0.0661638	0.0007057
th3	0.1222824	0.1222948	-0.0035120
th4	0.0891757	0.1627549	0.2712549

Table 17: Laplace(50, 4, 1)

	wariancja	błąd śr kwadratowy	obciążenie
th1	0.1003093	0.1003097	-0.0006215
th2	0.0661670	0.0661677	-0.0007949
th3	0.1222406	0.1222446	-0.0020146
th4	0.0883440	7.5350721	-2.7288694

Table 18: Laplace(50, 1, 2)

	wariancja	błąd śr kwadratowy	obciążenie
th1	0.3961970	0.3961986	-0.0012816
th2	0.2608747	0.2608757	-0.0009763
th3	0.4991300	0.4991354	-0.0023226
th4	0.3634788	2.7466566	1.5437545

Zadanie 5 dla  $n = 100$

Table 19: Laplace(50, 1, 1)

	wariancja	błąd śr kwadratowy	obciążenie
th1	0.0201059	0.0201062	-0.0005445
th2	0.0114155	0.0114165	-0.0010300
th3	0.0265389	0.0265396	-0.0008466
th4	0.0203625	0.1510849	0.3615555

Table 20: Laplace(50, 4, 1)

	wariancja	błąd śr kwadratowy	obciążenie
th1	0.0202247	0.0202251	-0.0006625
th2	0.0115853	0.0115884	-0.0017753
th3	0.0267727	0.0267729	-0.0003528
th4	0.0203608	7.0082367	-2.6434591

Table 21: Laplace(50, 1, 2)

	wariancja	błąd śr kwadratowy	obciążenie
th1	0.0800036	0.0800140	-0.0032211
th2	0.0472473	0.0472623	-0.0038667
th3	0.1053465	0.1053713	-0.0049746
th4	0.0822359	3.0192424	1.7137697

W zadaniu 5 mamy do czynienia z tą samą obserwacją. Zwiększając rozmiar próby zmniejszamy wartości statystyk w tabeli, natomiast gdy zmniejszamy rozmiar próby dzieje się na odwrót.

### Podsumowanie

Dzięki poznanej teorii na wykładzie, a następnie zastosowaniu uzyskanej wiedzy w praktyce podczas tego raportu, zobaczyliśmy, że estymatory które działają w teorii, działają również najlepiej w praktyce. Najważniejszym spostrzeżeniem wydaje się to, że im większą próbę weźmiemy tym lepsze dostaniemy rezultaty.