Projekt 6.

Mateusz Zacharecki

13 listopada 2023

Plan prezentacji

Algorytm Boxa-Müllera

Algorithm 1 Box-Müller method

1: $U_1 \leftarrow \texttt{GenerateU}$
2: $U_2 \leftarrow \texttt{GenerateU}$
3: $\phi \leftarrow 2\pi U_1$
4: $R \leftarrow \sqrt{-2\ln(U_2)}$
5: $X_1 \leftarrow R\cos(\phi)$
6: $X_2 \leftarrow R\sin(\phi)$
7: **return** (X_1, X_2)

Algorytm Boxa-Müllera z jedną zmienną

Algorithm 2 Box-Müller method with one variable

while $Y_2 \leq (1 - Y_1)^2/2$ **do**
  $Y_1 \leftarrow$ `GenerateExp(1)`
  $Y_2 \leftarrow$ `GenerateExp(1)`
end while
$U \leftarrow$ `GenerateU`
if $U \leq 0.5$ **then**
  return $-Y_1$
else
  return Y_1
end if

Metoda eliminacji

Algorithm 3 Acceptance-Rejection method

while $\sqrt{\frac{2e}{\pi}}U_1 \exp(-Y) > \sqrt{\frac{2}{\pi}} \exp(-Y^2/2)$ **do**
    $U_1 \leftarrow \texttt{GenerateU}_1$
    $Y \leftarrow \texttt{GenerateExp}(1)$
end while
$X \leftarrow Y$
$U_2 \leftarrow \texttt{GenerateU}_2$
if $U \leq 0.5$ **then**
    $X \leftarrow -X$
end if
return X

Algorytm Marsaglia-Braya

Algorithm 4 Marsaglia-Bray method

while $X > 1$ **do**
  $U_1 \leftarrow$ `GenerateU`
  $U_2 \leftarrow$ `GenerateU`
  $U_1 \leftarrow 2U_1 - 1$
  $U_2 \leftarrow 2U_2 - 1$
  $X \leftarrow U_1^2 + U_2^2$
end while
$Y \leftarrow \sqrt{\frac{-2\ln(X)}{X}}$
$Z_1 \leftarrow U_1 Y$
$Z_2 \leftarrow U_2 Y$
return (Z_1, Z_2)

Działanie metod - jedna zmienna

Rozważmy najpierw działanie metod dla jednej zmiennej. Każdym algorytmem wygenerowano jedną (bądź dwie w przypadku metod generujących parę zmiennych) zmienną. Celem jest wstępna weryfikacja algorytmów, a przede wszystkim próba odpowiedzenia na pytanie, czy zmienne te pochodzą z rozkładu normalnego. Będziemy badać prawdziwość hipotezy zerowej, że dana zmienna pochodzi z rozkładu normalnego przeciwko hipotezie alternatywnej, że nie pochodzi. Przyjmujemy poziom istotności 0.05.

Pierwsze 5 wartości każdej ze zmiennych

Algorytm Boxa-Müllera (obie zmienne):

$$-0.39776952 \quad -0.23512335 \quad -1.31026383 \quad 0.08249126 \quad -0.29654826$$

$$-0.50751393 \quad -0.08950169 \quad -1.05451559 \quad -0.43488876 \quad 0.01427994$$

Algorytm Boxa-Müllera dla jednej zmiennej:

$$0.50053482 \quad 0.96923232 \quad 0.76410219 \quad -0.32883176 \quad -1.18410218$$

Metoda akceptacji:

$$0.716475193 \quad 1.199688893 \quad -0.853392879 \quad -2.172760481 \quad 2.432400951$$

Algorytm Marsaglia-Braya (obie zmienne):

$$0.229808919 \quad 1.688293993 \quad -0.175288748 \quad 1.182702344 \quad -1.253213727$$

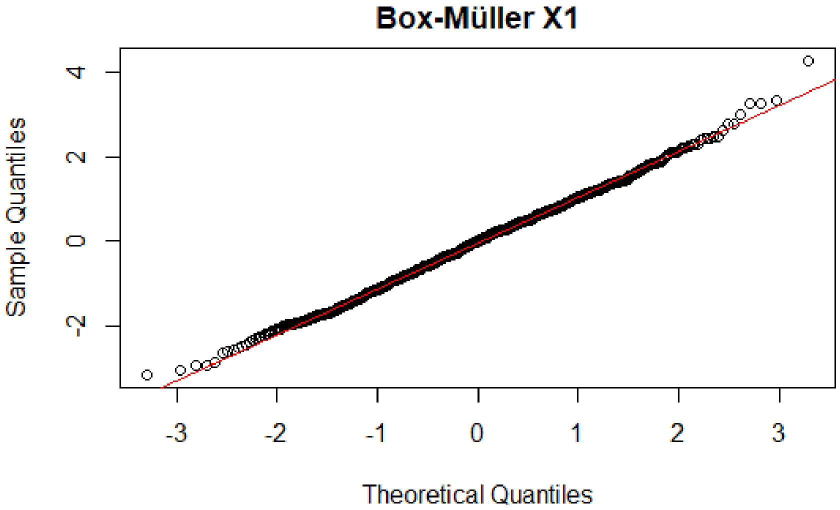
$$-0.31523887 \quad -0.39778479 \quad 0.87416236 \quad -1.05100042 \quad 1.26022079$$

Parametry dla $n = 1000$

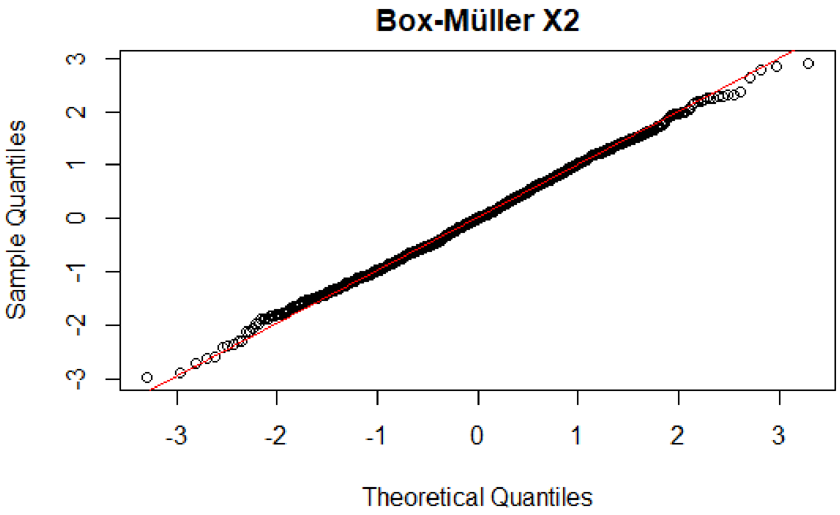
Poniższa tabela przestawia liczbę wartości wygenerowanych rozważanymi metodami zmiennych dla parametru $n = 1000$, a także ich p-wartości uzyskane testem Kołmogorowa-Smirnowa oraz testem Shapiro-Wilka:

	BM1	BM2	BMOV	AR	MB1	MB2
liczba wartości	1000	1000	767	740	797	797
p-wartości KS	0.4324	0.4005	0.8099	0.8312	0.8943	0.06723
p-wartości SW	0.2973	0.3621	0.1631	0.7994	0.5646	0.1038

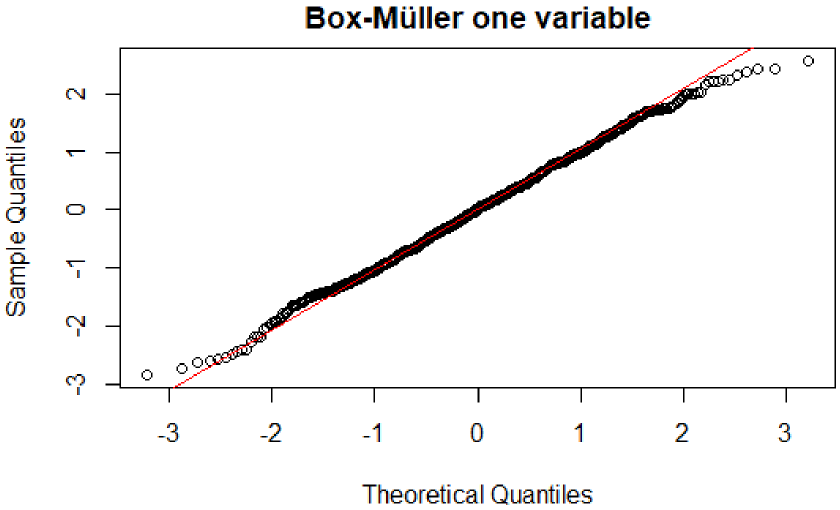
Wykresy kwantylowe



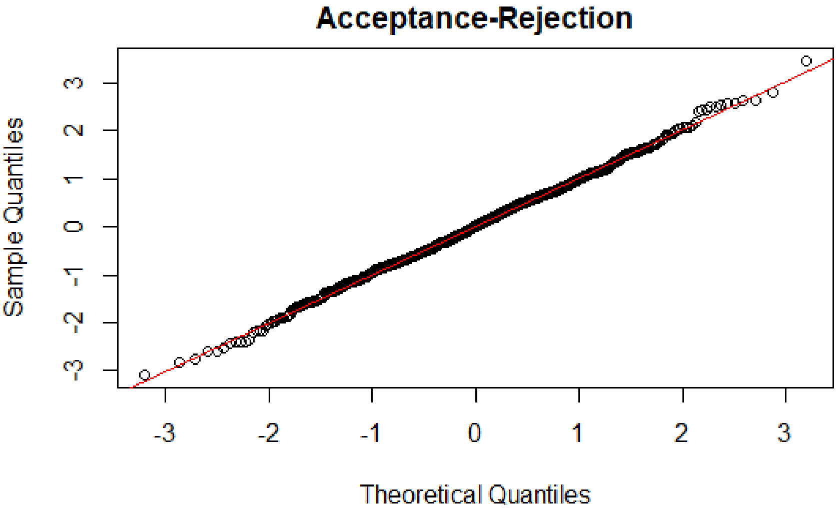
Wykresy kwantylowe



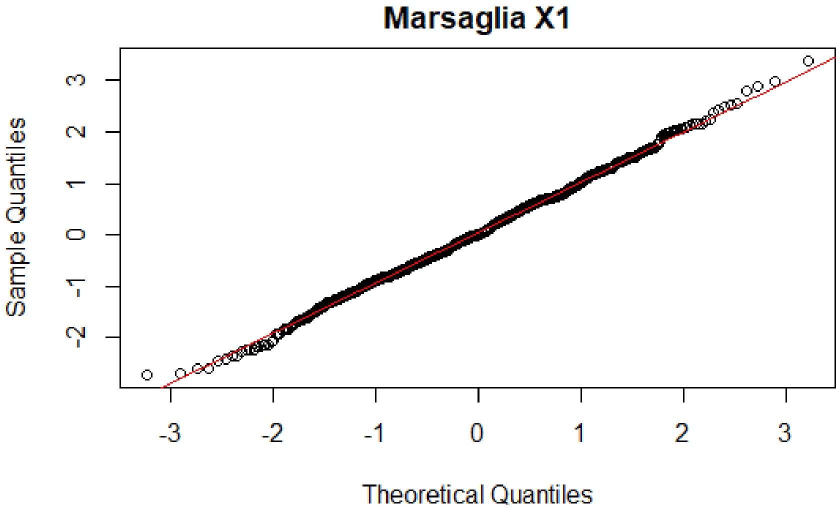
Wykresy kwantylowe



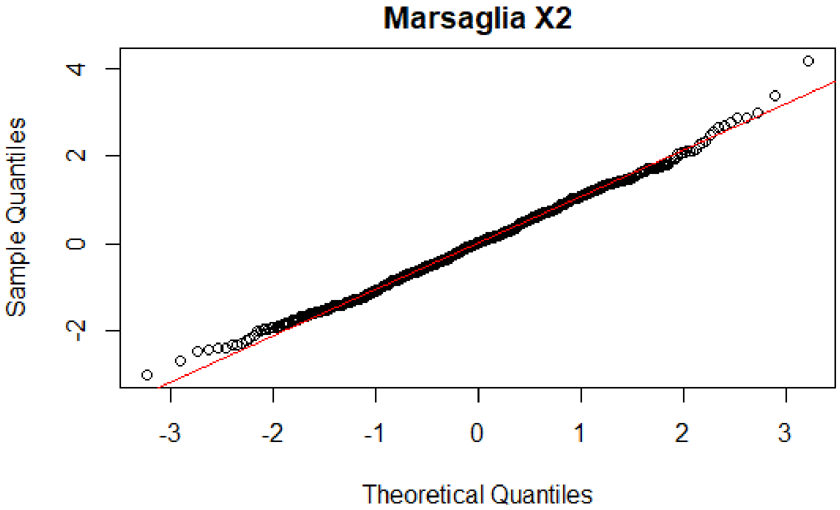
Wykresy kwantylowe



Wykresy kwantylowe



Wykresy kwantylowe



Wnioski

- Tylko pierwszym algorytmem otrzymano tyle wartości, ile zadeklarowano w funkcji.
- Żaden z dwóch testów nie dał podstaw do odrzucenia hipotezy zerowej o normalności rozkładu rozważanych zmiennych.
- Wykresy kwantylowe mimo niewielkich odchyleń dla skrajnych wartości mocno wskazują na normalność rozkładu wygenerowanych zmiennych.

Działanie metod - 100 zmiennych

W tej części rozważmy większą liczbę zmiennych. Każdym algorytmem wygenerowano 100 zmiennych (bądź 200 dla algorytmów generujących pary zmiennych). Będziemy badać działanie dla różnych wielkości prób, tj. dla $n = 100, 1000, 10000, 100000$. Zbadamy, jak duże średnio próby uzyskano, ale przede wszystkim ponownie pochylimy się nad kwestią normalności rozkładu generowanych zmiennych. Ponownie będziemy badać prawdziwość hipotezy zerowej, że dana zmienna pochodzi z rozkładu normalnego przeciwko hipotezie alternatywnej, że nie pochodzi. Przyjmujemy poziom istotności 0.05. Zweryfikujemy, jak często hipoteza zerowa jest odrzucana.

Średnie parametry dla $n = 100$

Poniższa tabela przestawia średnią liczbę wartości wygenerowanych rozważanymi metodami 100 zmiennych dla parametru $n = 100$, ich średnie p-wartości uzyskane testem Kołmogorowa-Smirnowa oraz testem Shapiro-Wilka oraz procent hipotez zerowych, które zostały odrzucone:

	BM1	BM2	BMOV	AR	MB1	MB2
liczba wartości	100	100	75.84	76.1	78.06	78.06
p-wartości KS	0.4463	0.5495	0.5079	0.5445	0.588	0.5985
p-wartości SW	0.506	0.5456	0.4866	0.5122	0.5272	0.5552
Procent odrzucanych H_0 KS	7%	3%	11.9%	3.9%	5.1%	5.1%
Procent odrzucanych H_0 SW	6%	4%	7.9%	3.9%	5.1%	6.4%

Średnie parametry dla $n = 1000$

Poniższa tabela przestawia średnią liczbę wartości wygenerowanych rozważanymi metodami 100 zmiennych dla parametru $n = 1000$, ich średnie p-wartości uzyskane testem Kołmogorowa-Smirnowa oraz testem Shapiro-Wilka oraz procent hipotez zerowych, które zostały odrzucone:

	BM1	BM2	BMOV	AR	MB1	MB2
liczba wartości	1000	1000	760.07	760.28	786.48	786.48
p-wartości KS	0.4937	0.495	0.5318	0.4674	0.5165	0.4979
p-wartości SW	0.5057	0.5029	0.4936	0.4118	0.5173	0.5304
Procent odrzucanych H_0 KS	0.3%	0.2%	0.9%	0.5%	1%	0.6%
Procent odrzucanych H_0 SW	0.4%	0.4%	0.1%	1.1%	0.8%	0.4%

Średnie parametry dla $n = 10000$

Poniższa tabela przestawia średnią liczbę wartości wygenerowanych rozważanymi metodami 100 zmiennych dla parametru $n = 10000$, ich średnie p-wartości uzyskane testem Kołmogorowa-Smirnowa oraz procent hipotez zerowych, które zostały odrzucone. Test Shapiro-Wilka nie został wzięty pod uwagę ze względu na wielkość próby:

	BM1	BM2	BMOV	AR	MB1	MB2
liczba wartości	10000	10000	7599.97	7605.15	7852.24	7852.24
p-wartości KS	0.5487	0.4816	0.5222	0.4917	0.553	0.5122
Procent odrzucanych H_0 KS	0.01%	0.04%	0.05%	0.04%	0.05%	0.04%

Średnie parametry dla $n = 100000$

Poniższa tabela przestawia średnią liczbę wartości wygenerowanych rozważanymi metodami 100 zmiennych dla parametru $n = 10000$, ich średnie p-wartości uzyskane testem Kołmogorowa-Smirnowa oraz procent hipotez zerowych, które zostały odrzucone. Test Shapiro-Wilka nie został wzięty pod uwagę ze względu na wielkość próby:

	BM1	BM2	BMOV	AR	MB1	MB2
liczba wartości	100000	100000	76025.25	76004.62	78524.54	78524.54
p-wartości KS	0.5056	0.5059	0.4889	0.4843	0.4643	0.5616
Procent odrzucanych H_0 KS	0.004%	0.006%	0.004%	0.008%	0.011%	0.004%

Wnioski

- Procent wartości, które trafiają do próbki dla każdego z algorytmów utrzymuje się na podobnym poziomie.
- Średnie p-wartości utrzymują się na podobnym poziomie między 0.4 a 0.6. Nie ma istotnych odstępstw między zmiennymi ani testami.
- Dla zdecydowanej większości zmiennych nie ma podstaw do odrzucenia hipotezy zerowej, udział hipotez zerowych, które są odrzucane jest istotnie mniejszy dla większych wartości n.

Eksperymentalna analiza prędkości algorytmów

Poniższa tabela przedstawia średnie czasy działania rozważanych algorytmów w zależności od parametru n dokonanych dla 100 iteracji:

n	BM	BMOV	AR	MB
100	$48.888\mu s$	$95.345\mu s$	$112.277\mu s$	$45.784\mu s$
1000	$314.782\mu s$	$617.19\mu s$	$738.263\mu s$	$241.981\mu s$
10000	$4183.25\mu s$	$7067.134\mu s$	$7061.956\mu s$	$2266.106\mu s$
100000	$30356.15\mu s$	$53182.88\mu s$	$61849.9\mu s$	$19602.4\mu s$

Pomimo najniższych rezultatów uzyskanych dla algorytmu Marsaglia-Braya, metoda Boxa-Müllera wydaje się być lepsza ze względu na wielkość uzyskiwanej próby. Pozostałe algorytmy istotnie od nich odstają.

Empiryczna analiza prędkości algorytmów

- Algorytm Boxa-Müllera jest jedynie przekształceniem wygenerowanych zmiennych z rozkładu jednostajnego.
- Pozostałe algorytmy wymagają potencjalnie wielokrotnego generowania zmiennych do momentu gdy nie zostaną spełnione określone warunki.
- Algorytm Boxa-Müllera wydaje się więc być najlepszy ze względu na czas działania algorytmu.

Wnioski

Każdy z rozważanych algorytmów wydaje się być dobrym do generowania zmiennych z rozkładu normalnego. Zmienne, które wydają się nie pochodzić z tego rozkładu generowane są rzadko dla każdego z algorytmów i występują tym rzadziej im większa jest próba. Na wyróżnienie zasługuje algorytm Boxa-Müllera, który bardzo dobrze wypadł w ekperymentalnej analizie prędkości algorytmów, a ponadto unika on problemu wielu prób wygenerowania zmiennej opartej na pętli.

Dziękuję za uwagę