

Lista 1

Analiza szeregów czasowych

Karol Pustelnik 249828 i Jan Solarz

2021-03-02

Spis treści

1	Zadanie 1	3
1.1	Wstęp	3
1.2	Opis eksperymentów	3
1.3	Wyniki dla średniej próbkowej	3
1.4	Pierwsza symulacja	3
1.5	Druga symulacja	6
1.6	Trzecia Symulacja	7
1.7	Czwarta symulacja	7
1.8	Piąta symulacja	10
1.9	Szosta symulacja	10
1.10	Podsumowanie	13
1.11	Wyniki dla estymatora autokowariancji	13
1.12	Symulacja pierwsza	14
1.13	Symulacja druga	15
1.14	Symulacja trzecia	15
1.15	Symulacja czwarta	19
1.16	Symulacja piąta	20
1.17	Wyniki dla estymatora autokorelacji	20
1.18	Pierwsza symulacja	20
1.19	Druga symulacja	22
1.20	Symulacja trzecia	23
2	Zadanie 3	24
2.1	Formalne testy białoszumowości	28
2.2	Porównanie testu graficznego i formalnego	29
2.3	Symulacja Pierwsza	29
2.4	Druga symulacja	30
2.5	Trzecia symulacja	31
2.6	Symulacja czwarta	33

2.7	Piąta symulacja	34
2.8	Szósta symulacja	35
2.9	Podsumowanie porównania testów	36
3	Bibliografia	37

1 Zadanie 1

1.1 Wstęp

Zajmiemy się badaniem asymptotycznych własności estymatorów parametrów rozkładu stacjonarnego szeregu czasowego drugiego rzędu. Rozpatrzmy próby z różnych rozkładów i sprawdzimy czy zachowują się jak szereg czasowy typu biały szum. Przypomijmy, że wspomniane wyżej estymatory wyrażają się wzorem:

1. estymator próbkowy wartości oczekiwanej μ – średnia próbkowa:

$$\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t$$

2. estymator funkcji autokowariancji $\gamma(h)$:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X}_n)(X_t - \bar{X}_n), \text{ dla } h = 0, 1, 2, 3, \dots, n-1,$$

3. estymator funkcji autokorelacji $\rho(h)$:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \text{ dla } h = 0, 1, 2, 3, \dots, n-1$$

Biały szum definiujemy jako ciąg nieskorelowanych zmiennych losowych ze średnią 0 i wariancją σ^2

1.2 Opis eksperymentów

Będziemy losować 100 szeregów o zadanej długości n . Wykorzystamy rozkład normalny oraz wykładniczy. Wyniki przedstawimy za pomocą wykresów:

- histogram oraz gęstość teoretyczną,
- estymator jądrowy gęstości oraz gęstość teoretyczną,
- dystrybuante empiryczną oraz teoretyczną,
- wykresy kwantylowe.

Ponadto zastosujemy test statystyczny normalności Shapiro-Wilka do sprawdzenia własności estymatora średniej.

1.3 Wyniki dla średniej próbkowej

1.4 Pierwsza symulacja

Ustawiamy parametry na:

- długości szeregu $n = 30$
- próbie z rozkładu normalnego o średniej 0 i wariancji 1.

Poniżej prezentujemy kod z programu R:

```
set.seed(123)
n <- 30 # długość szeregu
```

```

k <- 100 # liczba realizacji
# realizacje białego szumu
realizacje <- matrix(rnorm(n*k),n,k)
# średnie próbkowe dla poszczególnych realizacji
srednie <- apply(realizacje, MARGIN=2, FUN=mean)
par(mfrow=c(2,2))
# histogram
hist(srednie, probability=T, col="lightblue", breaks="FD", main="Histogram
i gęstość teoretyczna", ylab="gęstość / częstość", xlab="x",cex.lab=1.5,
      cex.axis=1.5, cex.main=1.5, cex.sub=1.5)
curve(dnorm(x,mean=0,sd=1/sqrt(n)), col="red", add=T, lwd=2)
# estymator jądrowy
dens <- density(srednie)
plot(dens, ylab="gęstość", xlab="x", main="Estymator jądrowy i gęstość
teoretyczna",cex.lab=1.5, cex.axis=1.5, cex.main=1.5, cex.sub=1.5)
curve(dnorm(x,mean=0,sd=1/sqrt(n)), col="red", add=T, lwd=2)
# dystrybuanta empiryczna

plot(ecdf(srednie), cex=0.5, main="Dystrybuanta empiryczna i teoretyczna", xlab="x",
      cex.lab=1.5, cex.axis=1.5, cex.main=1.5, cex.sub=1.5)
curve(pnorm(x,mean=0,sd=1/sqrt(n)), col="red", add=T, lwd=2)
# wykresy kwantylowe
qqnorm(srednie, main="Wykresy kwantylowe", xlab= "Kwantyle Teoretyczne",
        ylab="Kwantyle Próbkowe",cex.lab=1.5, cex.axis=1.5,
        cex.main=1.5, cex.sub=1.5)
qqline(srednie)

```

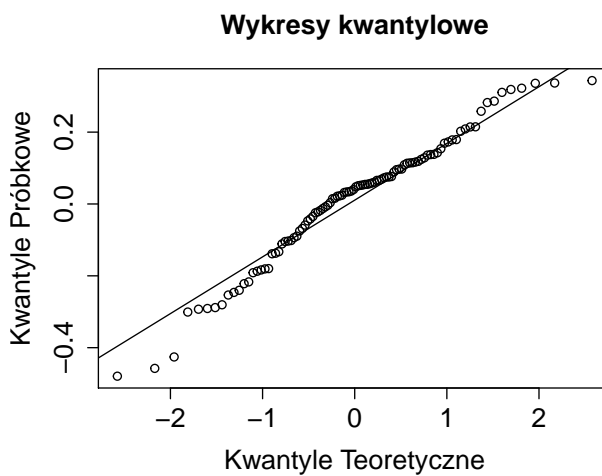
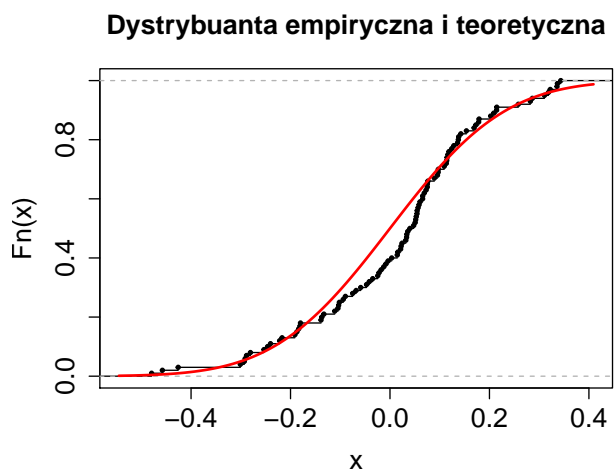
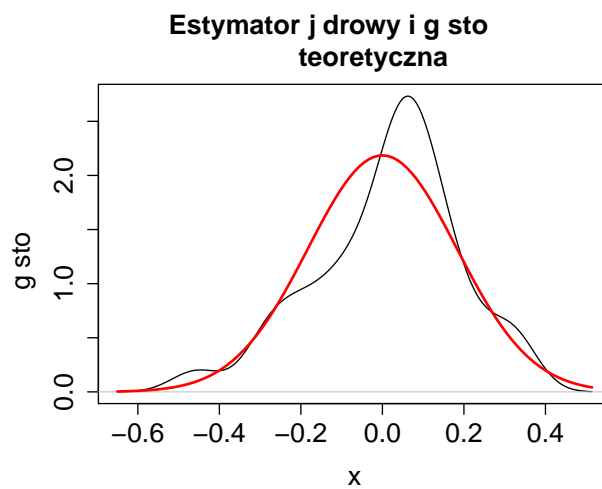
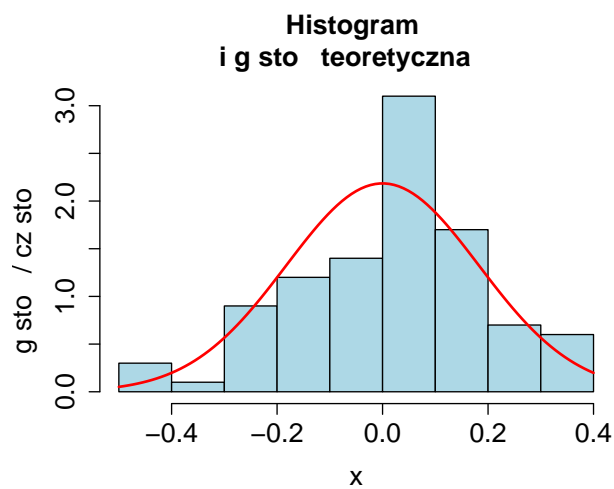
Na podstawie wykresów wyciągamy następujące wnioski:

- Histogram w przybliżeniu odzwierciedla wykres gęstości. Różni się jednak skośnością lewostronną.
- Podobnie na wykresie estymatora gęstości jądrowej zauważamy lewostronną skośność.
- Dystrybuanta empiryczna oraz teoretyczna zachowują się podobnie choć widać, że na pewnym przedziale dystrybuanta teoretyczna leży powyżej dystrybuanty empirycznej.
- Na krańcach wykresów kwantylowych widać, że nie przypomina linii prostej.

Tablica 1: Wynik Testu Shapiro-Wilka

	Statystyka testowa	P-value
Wartość	0.9702051	0.0229191

Test normalności Shapiro-Wilka sugeruje, aby odrzucić hipotezę zerową, że próba pochodzi

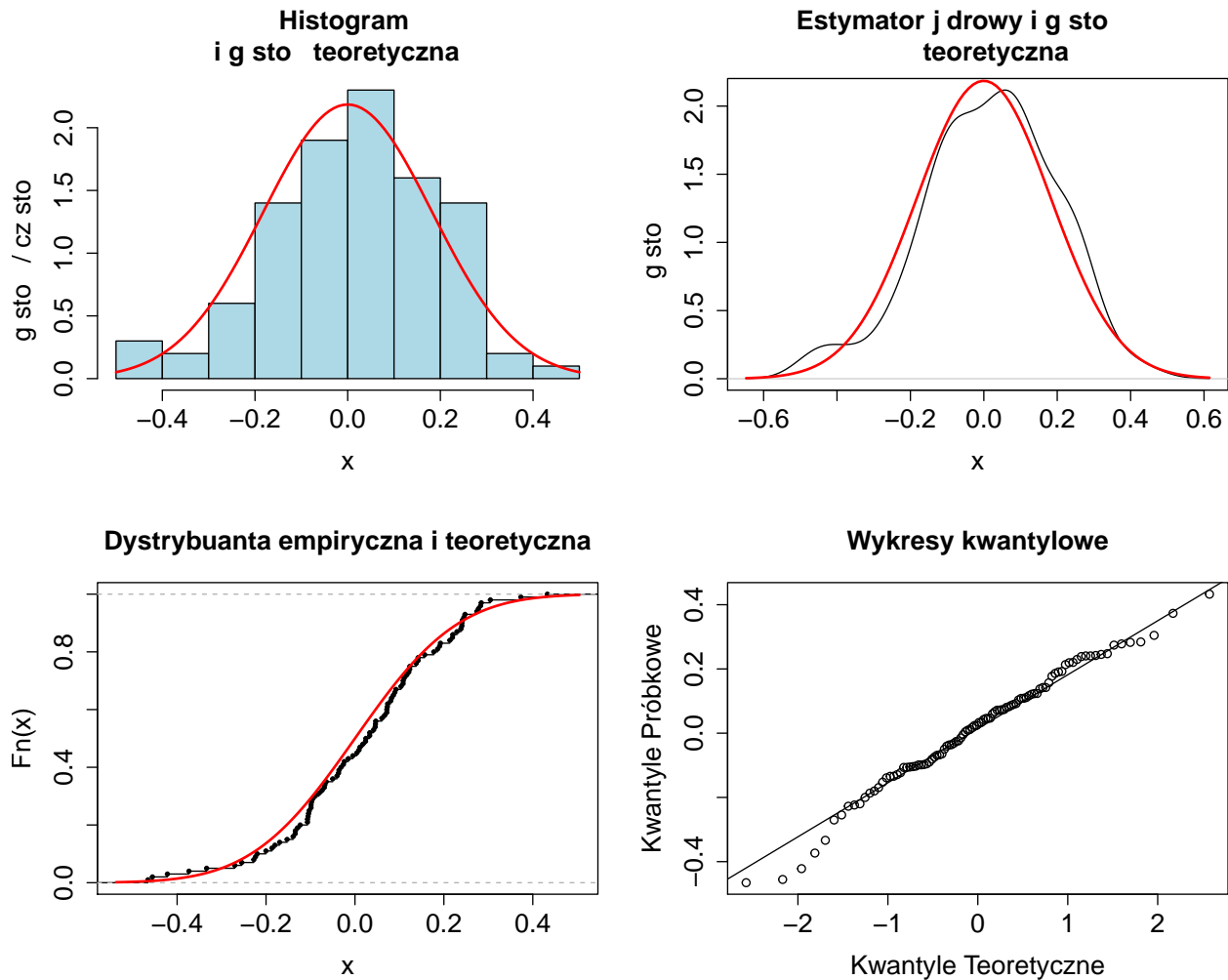


Rysunek 1: Wykresy dla pierwszej symulacji

z rozkładu normalnego i przyjąć alternatywną - próba nie pochodzi z rozkładu normalnego. ($p < 0.05$)

1.5 Druga symulacja

Symulujemy dane ponownie nie zmieniając parametrów. Uwzględniamy jednak losowość - za pomocą funkcji `set.seed()`.



Rysunek 2: Wykresy dla drugiej symulacji

- Tym razem, wszystkie wykresy wyglądają na zdecydowanie lepiej dopasowane.
- Nie widać skośności ani lewo ani prawo stronnej.
- Jedynie na początku wykresu kwantylowego widać, że nie układa się w linię prostą.
- Ogólnie dla tej próby dane o wiele bardziej przypominają rozkład normalny o średniej zero i zadanej wariancji.

Tablica 2: Wynik Testu Shapiro-Wilka

	Statystyka testowa	P-value
Wartość	0.985017	0.3184898

Na podstawie wyników testu Shapiro-Wilka nie ma podstaw do odrzucenia hipotezy zerowej mówiącej o tym, że próba pochodzi z rozkładu normalnego. ($p > 0.05$)

1.6 Trzecia Symulacja

Sprawdzimy teraz, czy długość szeregu istotnie wpłynie na wyniki testu oraz wygląd wykresów. Ustawiamy parametry na:

- długości szeregu $n = 200$,
- rozkład normalny o średniej 0 i wariancji 1
- Pierwszy wykres, sugeruje, że występuje skośność prawostronna.
- Wykresy gęstości nie wydają się potwierdzać powyższego wniosku. Zauważalny jest charakterystyczny “dołek” na wykresie gęstości estymatora, który nie występuje w typowym rozkładzie normalnym.
- Natomiast wykresy dystrybuant oraz kwantylowe wyglądają na lepiej dopasowane.
- Naszym zdaniem, próba faktycznie przypomina rozkład normalny. Sprawdźmy to formalnym testem.

Tablica 3: Wynik Testu Shapiro-Wilka

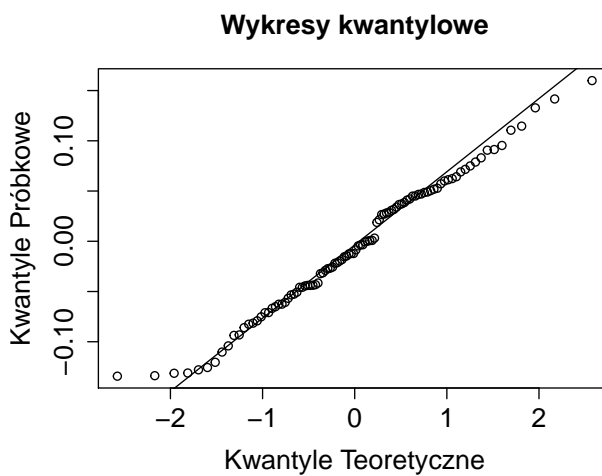
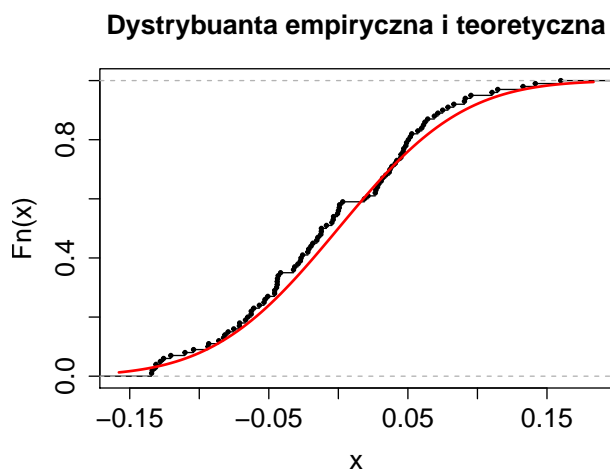
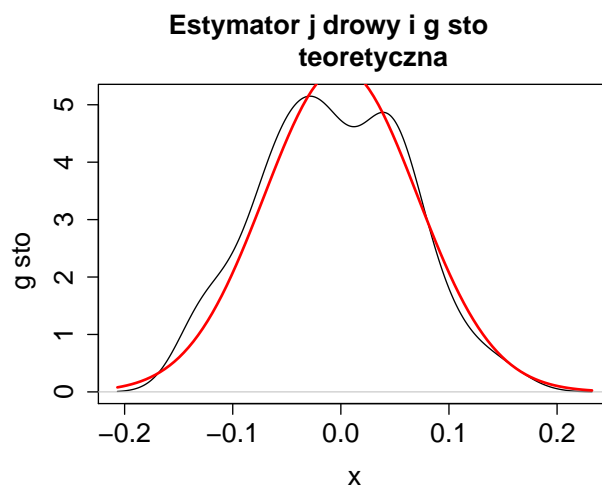
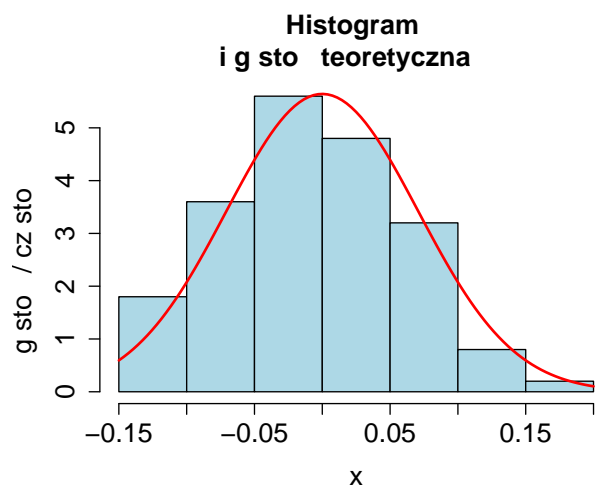
	Statystyka testowa	P-value
Wartość	0.9855751	0.3491148

Podobnie jak wcześniej, test Shapiro-Wilka nie pozwala odrzucić hipotezy zerowej. To oznacza, że dane mogą pochodzić z rozkładu normalnego.

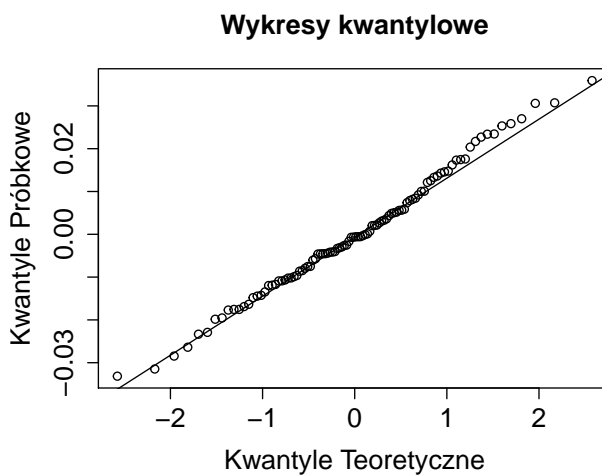
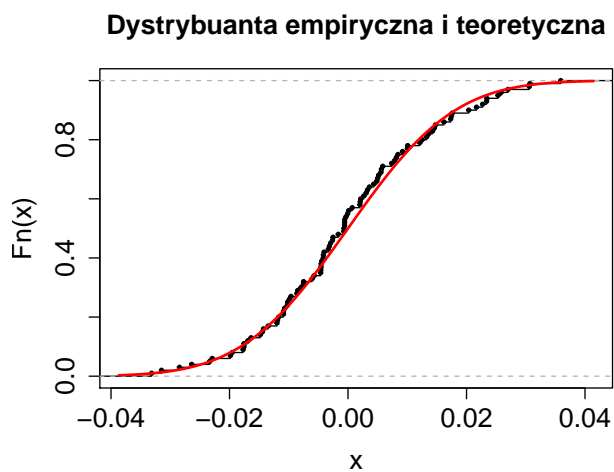
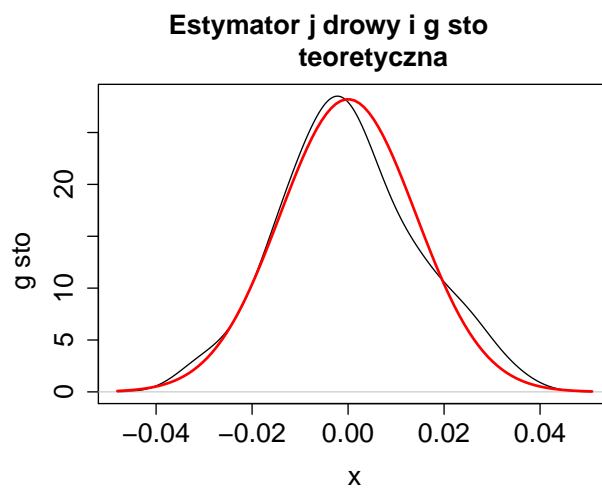
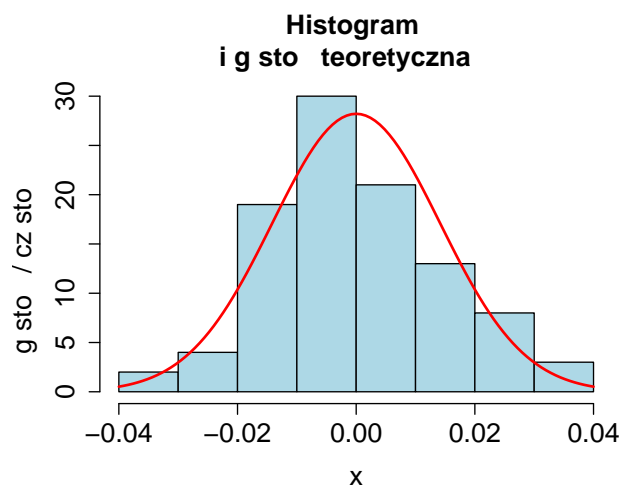
1.7 Czwarta symulacja

Powtórzymy symulację dla jeszcze dłuższych szeregów - spodziewamy się, że wykresy będą jeszcze lepiej dopasowane, a p-value dla testu normalności znacznie większe. Ustawiamy parametry na:

- długość szeregu $n = 5000$,
- rozkład normalny o średniej 0 i wariancji 1



Rysunek 3: Wykresy dla trzeciej symulacji



Rysunek 4: Wykresy dla czwartej symulacji

- Z wykresu pierwszego i drugiego widać, że dane charakteryzują się lekką skośnością prawostronną.
- Natomiast wykres dystrybuanty i kwantylowy sugerują bardzo dobre odzwierciedlenie zachowania rozkładu normalnego.
- Według nas są to najlepsze wyniki do tej pory.

Tablica 4: Wynik Testu Shapiro-Wilka

	Statystyka testowa	P-value
Wartość	0.9914494	0.7800561

1.8 Piąta symulacja

Sprawdźmy czy inna wariancja cokolwiek zmieni. Ustawiamy parametry na:

- długość szeregu $n = 5000$,
- rozkład normalny o średniej zero i wariancji 7

Zmiana wariancji nie wpłynęła na dopasowanie wykresów.

Tablica 5: Wynik Testu Shapiro-Wilka

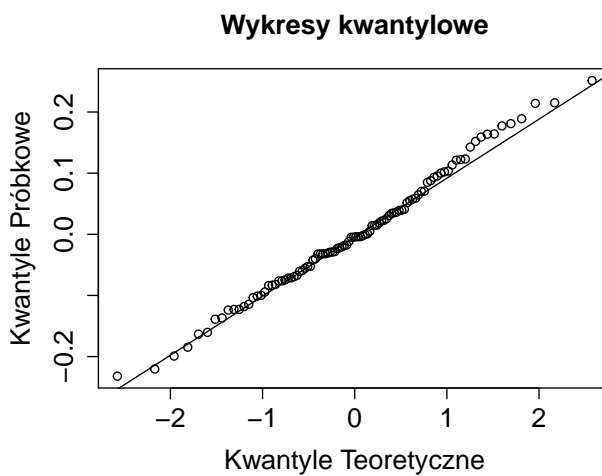
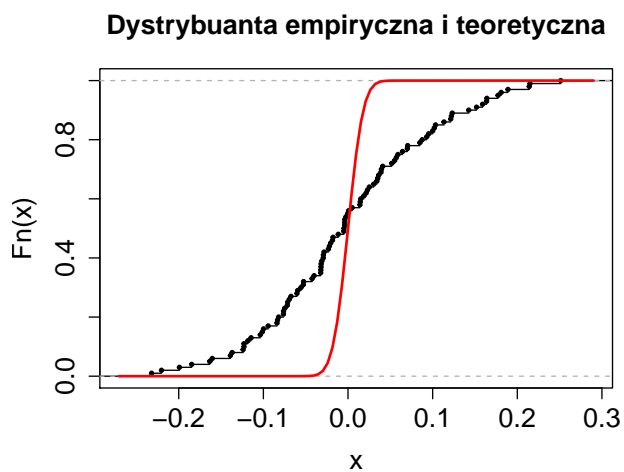
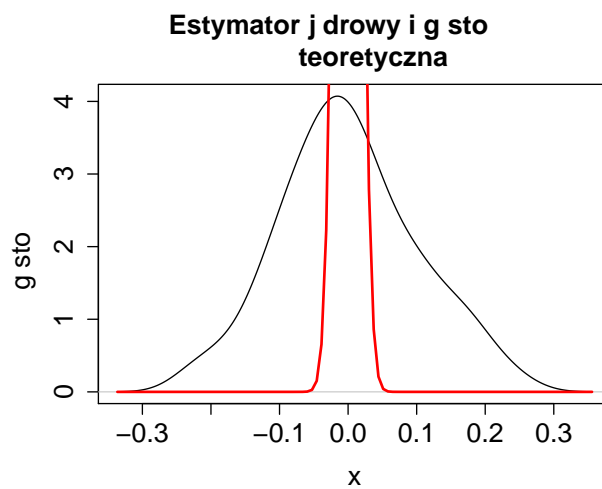
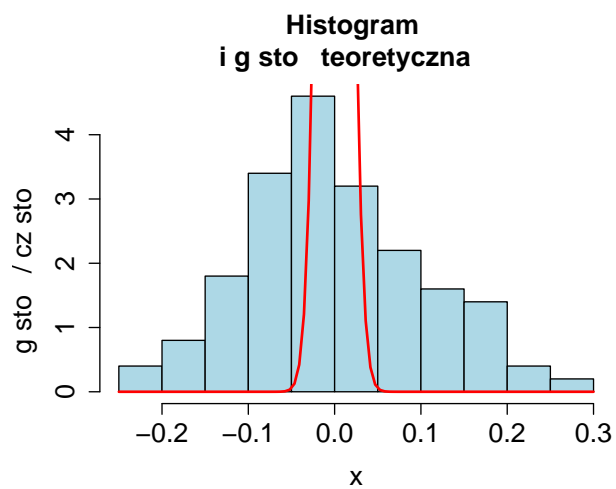
	Statystyka testowa	P-value
Wartość	0.9914494	0.7800561

Wyniki testu Shapiro-Wilka nie zmieniły się dla większej wariancji.

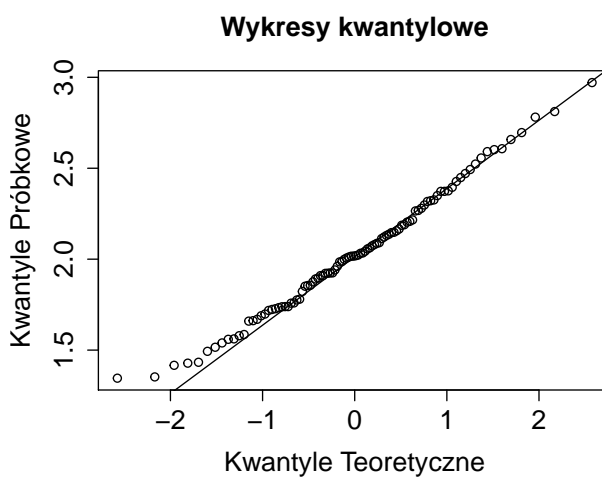
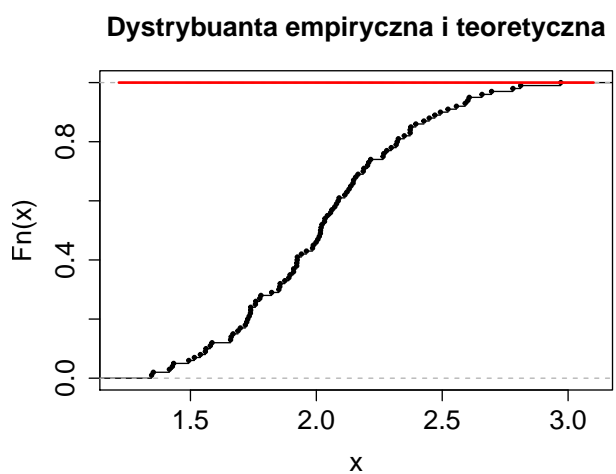
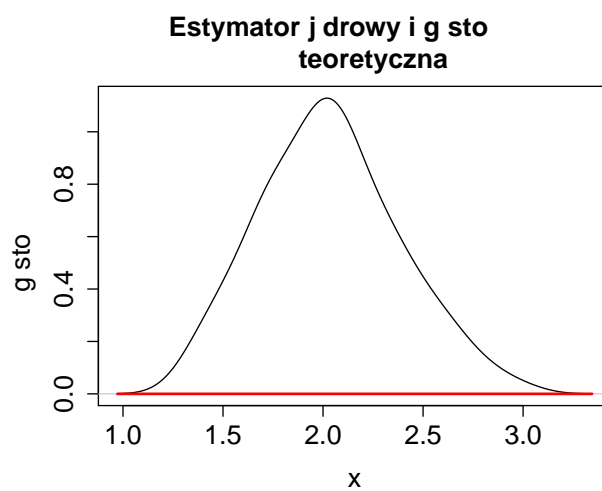
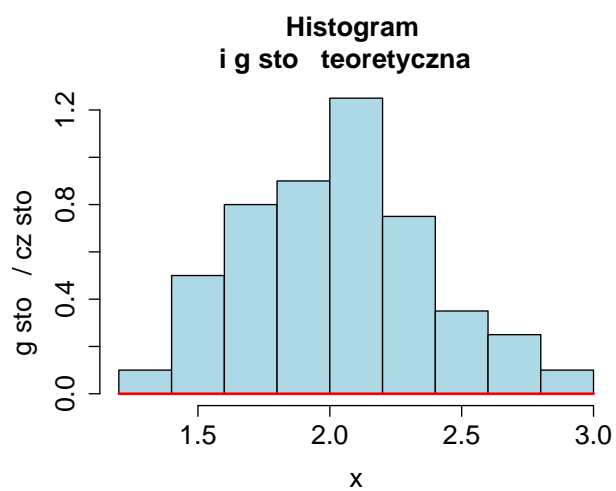
1.9 Szósta symulacja

Zbadajmy teraz, co się dzieje dla innych rozkładów. Ustawiamy parametry na:

- długość szeregu $n = 30$,
- rozkład wykładniczy z parametrem $\lambda = 0.5$
- Wykres kwantylowy wzbudza wątpliwości. Jest podobny do wykresów, gdzie dane faktycznie pochodziły z rozkładu normalnego - tutaj pozostałe wykresy jednoznacznie na to nie wskazują.



Rysunek 5: Wykresy dla piątej symulacji



Rysunek 6: Wykresy dla szóstej symulacji

Tablica 6: Wynik Testu Shapiro-Wilka

	Statystyka testowa	P-value
Wartość	0.9891022	0.5926263

Test Shapiro-Wilka nie sugeruje odrzucenia hipotezy zerowej i przyjęcia alternatywnej (tzn. przyjęcia, że próba nie pochodzi z rozkładu normalnego). Jest to dosyć interesujące, ponieważ interpretacja wykresów sugeruje, że próba nie pochodzi z rozkładu normalnego.

1.10 Podsumowanie

Na podstawie przeprowadzonych symulacji można wyciągnąć następujące wnioski:

- wraz ze wzrostem długości szeregu, dane wydają się dokładniej odzwierciedlać zachowanie rozkładu normalnego,
- wykres kwantylowy jest ciężki do interpretacji - dla próby z rozkładu wykładniczego dostaliśmy bardzo podobny wykres jak dla próby z rozkładów normalnych,
- wariancja nie wpływa na wyniki testu Shapiro-Wilka i na dopasowanie wykresów,
- Skośność danych niekoniecznie jest powiązana z odrzuceniem hipotezy zerowej (że próba nie pochodzi z rozkładu normalnego),
- Samo sprawdzenie wyniku testu Shapiro-Wilka nie jest wystarczające aby wnioskować o normalności próby. Warto też spojrzeć jak dane prezentują się na wykresach.

Tablica 7: Podsumowanie symulacji

Nr. Symulacji	Typ rozkładu	Parametry rozkładu	Długość szeregu	P-value testu Shapiro-Wilka
Pierwsza symulacja	Normalny	$\mu = 0, \sigma=1$	$n = 30$	0.23
Druga symulacja	Normalny	$\mu = 0, \sigma=1$	$n = 30$	0.32
Trzecia Symulacja	Normalny	$\mu = 0, \sigma=1$	$n = 200$	0.35
Czwarta symulacja	Normalny	$\mu = 0, \sigma=1$	$n = 5000$	0.78
Symulacja piąta	Normalny	$\mu = 0, \sigma=7$	$n = 5000$	0.78
[Symulacja szósta]	Wykładniczy	$\lambda = 0.5$	$n = 30$	0.59

1.11 Wyniki dla estymatora autokowariancji

Na podstawie oceny wykresów gęstości teoretycznej z parametrami rozkładu takimi jak przy losowaniu danych oraz wykresu gęstości teoretycznej z parametrami średniej i wariancji ustalonej odpowiednio na średnią i wariancję macierzy autokorelacji próbkowej. Ocenimy też jak opóźnienie wpływa na wyniki testu normalności.

1.12 Symulacja pierwsza

Ustawiamy parametry na:

- długość szeregu $n = 30$,
- rozkład normalny o średniej 0 i wariancji 1,
- opóźnienia : 1, 10, 20, 29

```
## Symulacja dla n=30, k=100, set.seed=123, wariancja=1
set.seed(123)
n <- 30 # długość szeregu
k <- 100 # liczba realizacji
# realizacje białego szumu
realizacje1 <- matrix(rnorm(n*k),n,k)
# wyznaczamy macierz ACF (w wierszach mamy realizacje acf(h)
# dla h=1,2,3,...,
# h.max)
h.max <- n # maksymalne opóźnienie (można uzależnić od n)
acf.matrix <- apply(realizacje1, 2, function(x) acf(x, lag.max=h.max,
type="covariance", plot=FALSE)$acf)
# usuwamy ACF(0)
acf.matrix <- acf.matrix[-1,]
h.wybrane <- c(1, 10, 20, 29) # można uzależnić od n
# histogramy dla ACF(h)
par(mfrow=c(2,2))
for (h in h.wybrane)
{
  tytul <- paste0("Wykresy dla acf(",h,")")
  acf.h <- acf.matrix[h,]
  hist( acf.h, freq=FALSE, col="lightblue", main=tytul, xlab="",
        ylab="gęstość / częstość",cex.lab=1.5, cex.axis=1.5, cex.main=1.5, cex.sub=1.5)
  curve(dnorm(x,mean=mean(acf.h), sd=sd(acf.h)), add=T, col="blue", lwd=2,cex.lab=1.5, cex.axis=1.5)
  curve(dnorm(x,mean=0, sd=1/sqrt(n)), add=T, col="red", lwd=2,cex.lab=1.5, cex.axis=1.5)
}
```

Można zaobserwować, że dla coraz większych opóźnień estymator funkcji autokowariancji ma coraz gorsze własności.

Tablica 8: Wyniki testów Shapiro-Wilka dla różnych opóźnień - symulacja 1

Opóźnienie	Statystyka Testowa	P-value
$h = 1$	$W = 0.98853$	0.5483
$h = 10$	$W = 0.98122$	0.1651
$h = 20$	$W = 0.98169$	0.1797

Opóźnienie	Statystyka Testowa	P-value
$h = 29$	$W = 0.78316$	$7.773e-11$

Wyniki testów sugerują, że p-value maleje wraz ze wzrostem opóźnienia h . Jedynie dla $h = 29$ odrzucamy hipotezę, że próba pochodzi z rozkładu normalnego. Dla pozostałych opóźnień, nie ma podstaw do odrzucenia hipotezy zerowej - jest duża szansa, że próba faktycznie pochodzi z rozkładu normalnego.

1.13 Symulacja druga

Parametry nie ulegają zmianie. Uwzględniamy losowość funkcją `set.seed()`

Nasuwa się podobny wniosek jak wcześniej - dla większych opóźnień dostajemy gorsze własności estymatora.

Tablica 9: Wyniki testów Shapiro-Wilka dla różnych opóźnień - symulacja 2

Opóźnienie	Statystyka Testowa	P-value
$h = 1$	$W = 0.98488$	0.3115
$h = 10$	$W = 0.99268$	0.868
$h = 20$	$W = 0.99169$	0.7987
$h = 29$	$W = 0.90597$	$2.729e-06$

P-value dla coraz większych opóźnień zmniejszają się (pomijając $h = 1$) - rośnie prawdopodobieństwo, że próba nie pochodzi z rozkładu normalnego. Dla $h = 29$ przyjmujemy hipotezę, że próba nie pochodzi z rozkładu normalnego.

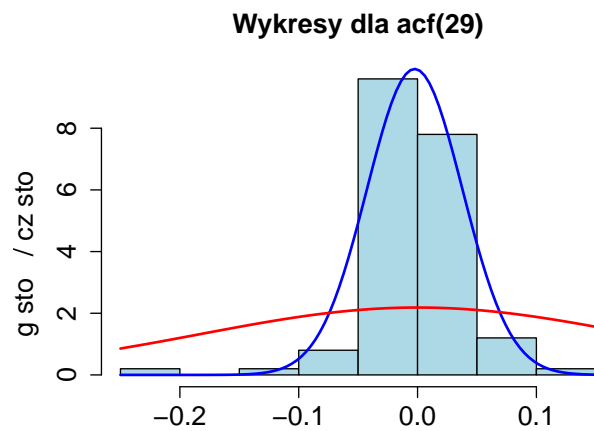
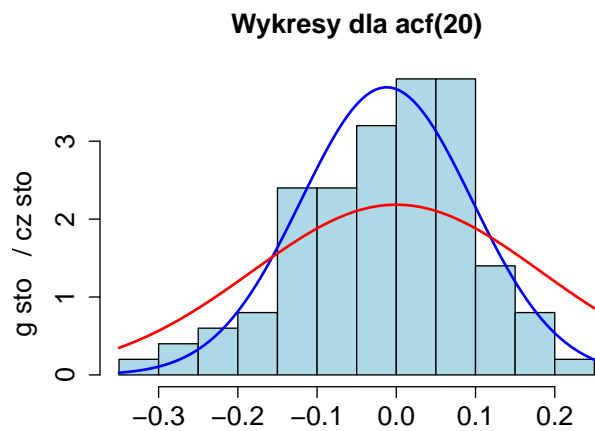
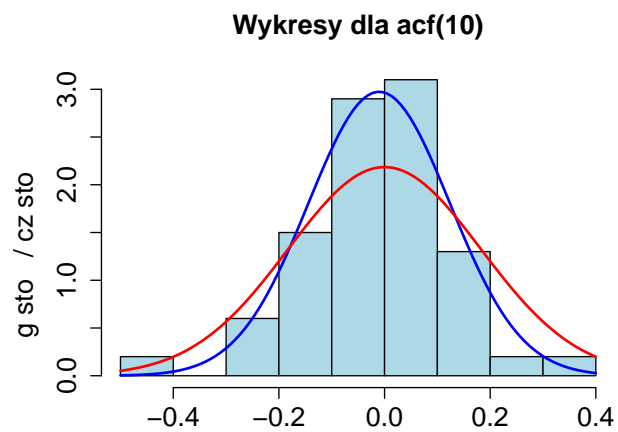
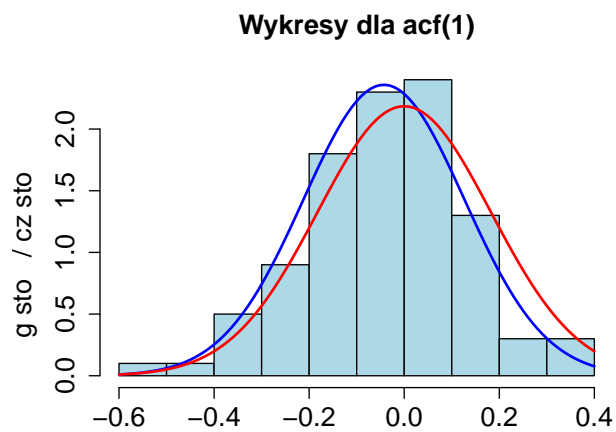
1.14 Symulacja trzecia

Teraz zwiększymy długość szeregu oraz zmienimy opóźnienia h .

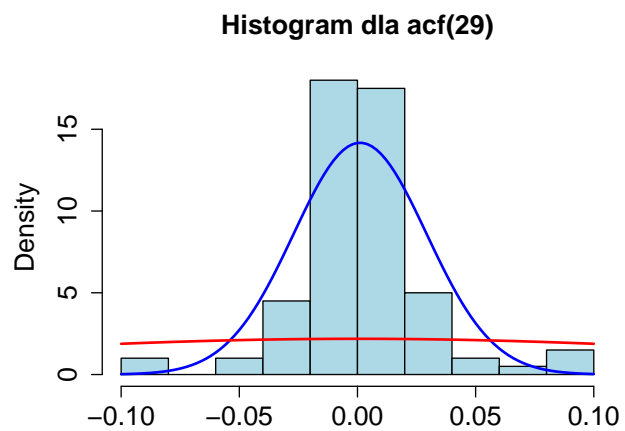
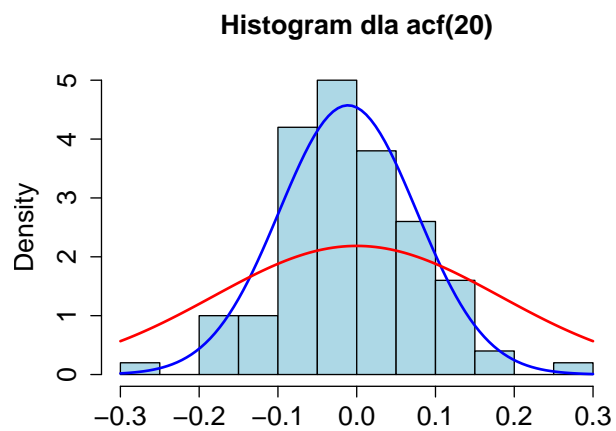
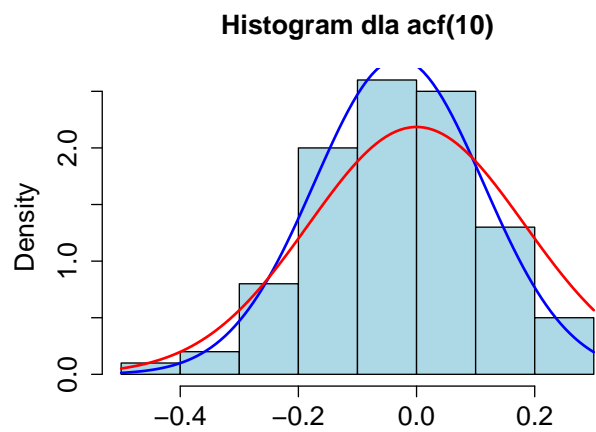
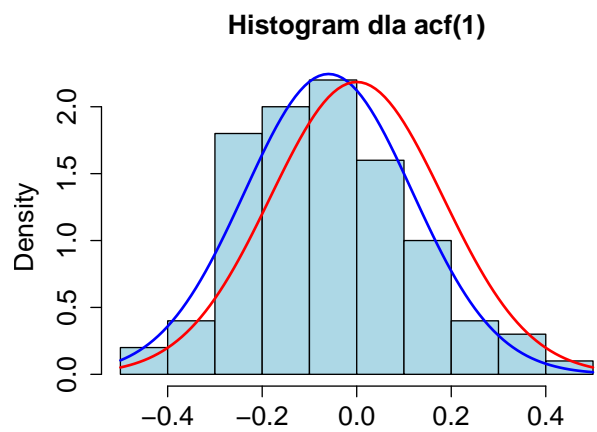
Ustawiamy parametry:

- długość szeregu $n = 200$,
- rozkład normalny o średniej 0 i wariancji 1
- $h = 1, 20, 100, 199$

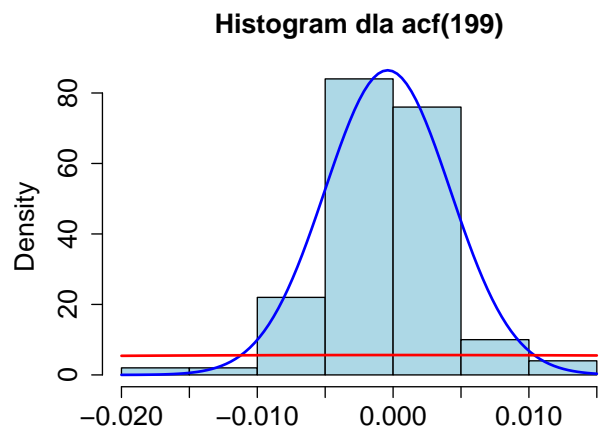
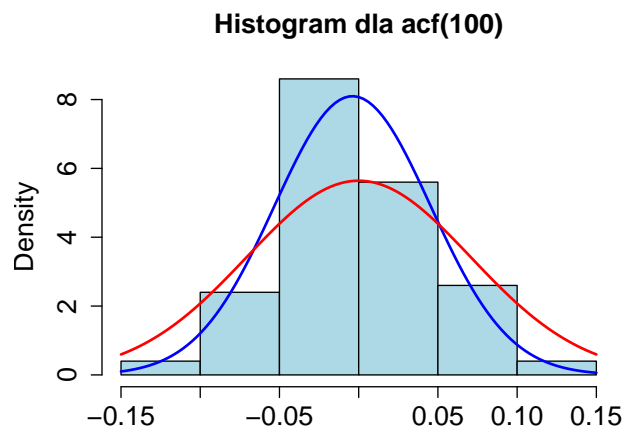
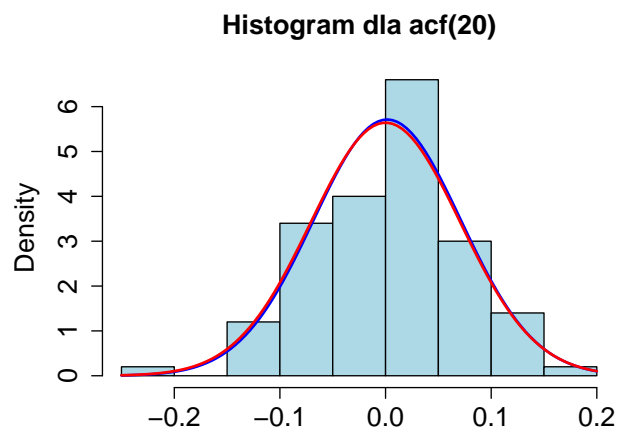
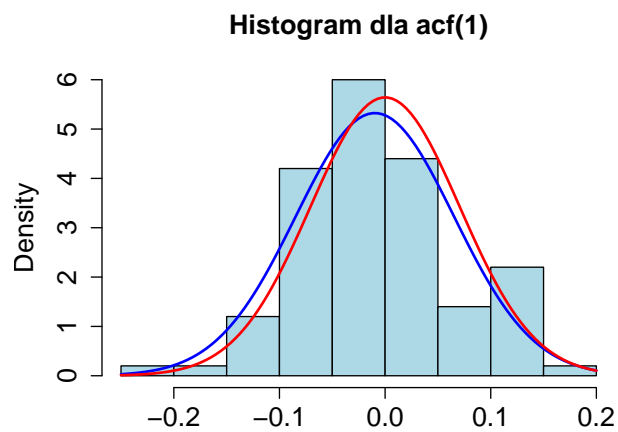
Wnioski są podobne jak dla poprzednich symulacji. Dla większych opóźnień próba coraz mniej przypomina rozkład normalny. Rzuci się w oczy, że dla $h = 20$ gęstość z parametrami średniej i wariancji będącymi estymatorami z próby pokrywa się prawie w całości z gęstością rozkładu normalnego o średniej 0 i wariancji 1. Oznacza to, że dla opóźnienia $h = 20$, wyznaczony estymator ma bardzo dobre własności.



Rysunek 7: Wykresy dla pierwszej symulacji



Rysunek 8: Wykresy dla drugiej symulacji



Rysunek 9: Wykresy dla trzeciej symulacji

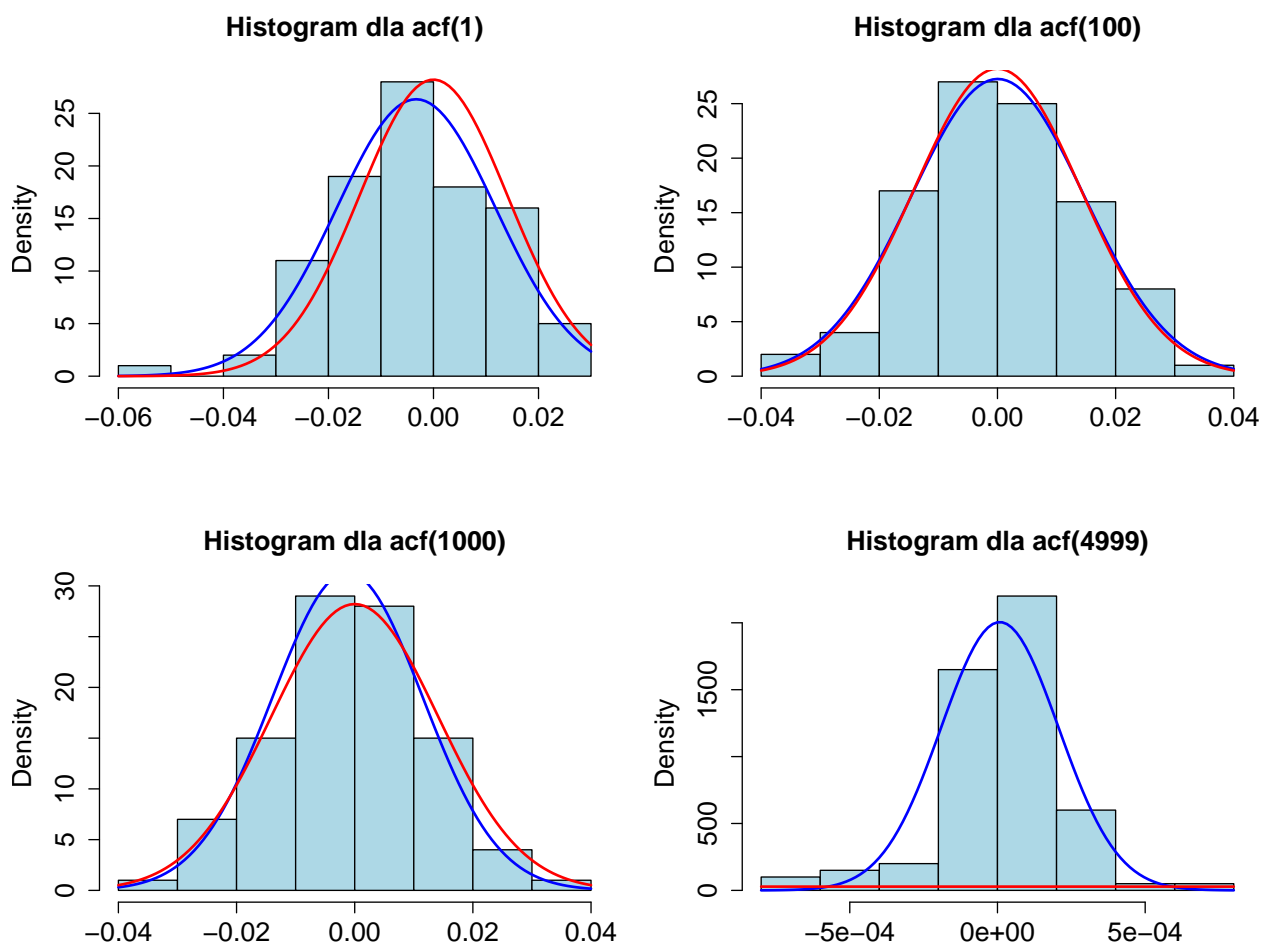
Opóźnienie	Statystyka Testowa	P-value
$h = 1$	$W = 0.98281$	0.2188
$h = 20$	$W = 0.98859$	0.5524
$h = 100$	$W = 0.98886$	0.5737
$h = 199$	$W = 0.93687$	0.0001249

Utrzymuje się tendencja, że dla dużych opóźnień odrzucamy hipotezę o normalności rozkładu, z którego pochodzą dane.

1.15 Symulacja czwarta

Parametry ustawiliśmy na:

- długość szeregu $n = 5000$,
- rozkład $N(0,1)$,



Rysunek 10: Wykresy dla czwartej symulacji

Dla opóźnienia $h = 100$ własności estymatora są bardzo dobre. Dla 10-krotnie większego opóźnienia również otrzymujemy dobre własności estymatora autokowariancji. Natomiast, gdy opóźnienie jest podobnej wartości co długość szeregu estymator nie jest wiarygodny.

Opóźnienie	Statystyka Testowa	P-value
$h = 1$	$W = 0.98176$	0.1817
$h = 100$	$W = 0.99374$	0.9285
$h = 1000$	$W = 0.99264$	0.8657
$h = 4999$	$W = 0.88681$	3.564e-07

Wyniki testów dla h małych w stosunku do długości szeregu nie wskazują na odrzucenie hipotezy zerowej. Natomiast dla opóźnienia bliskiego n odrzucamy hipotezę, że dane pochodzą z rozkładu normalnego.

1.16 Symulacja piąta

Ustawiamy parametry na:

- długość szeregu $n = 5000$,
- rozkład wykładniczy,
- opóźnienia $h = 1, 100, 1000, 4999$

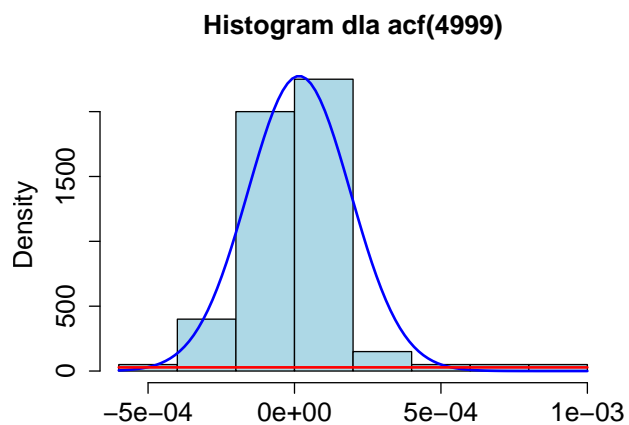
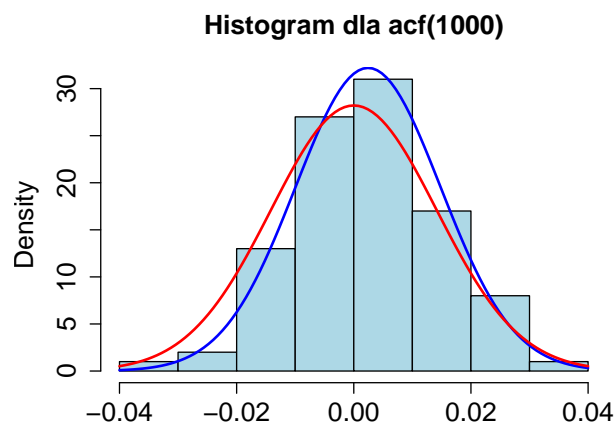
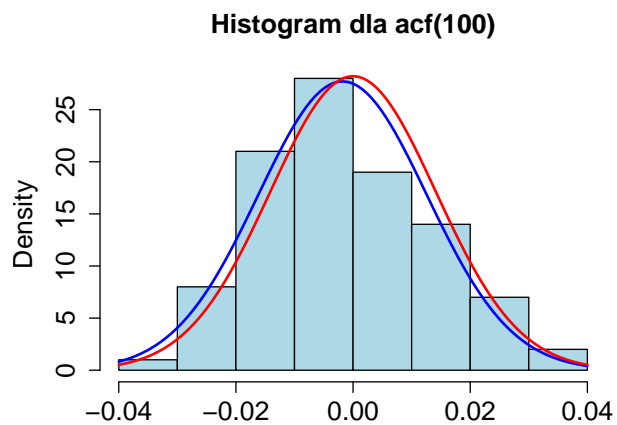
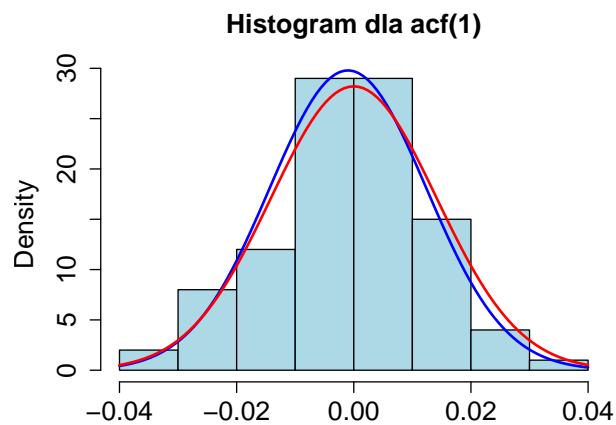
1.17 Wyniki dla estymatora autokorelacji

W analizie estymatorów autokorelacji uwzględnimy regułę Boxa-Jenkinsa, która mówi, że użyteczny estymator autokorelacji możemy otrzymać jedynie wówczas, gdy:

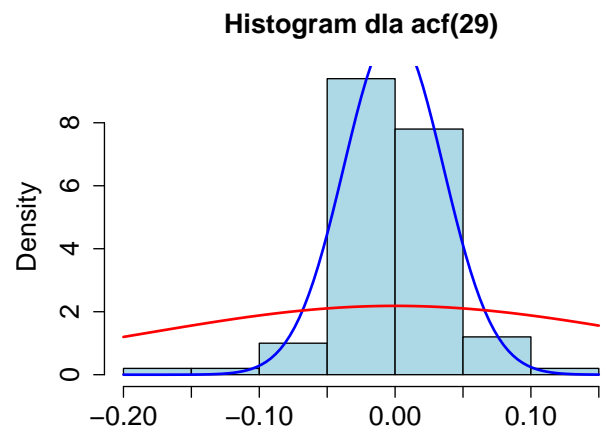
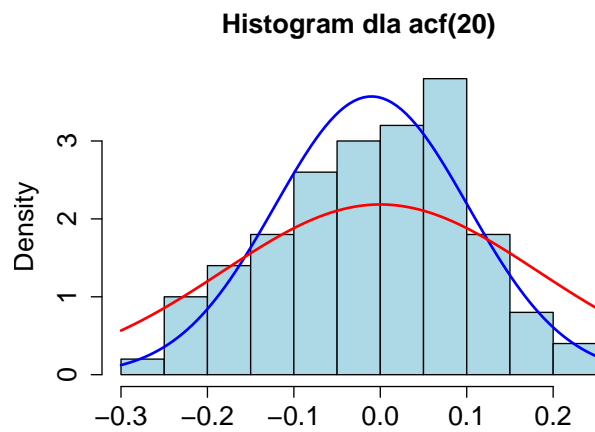
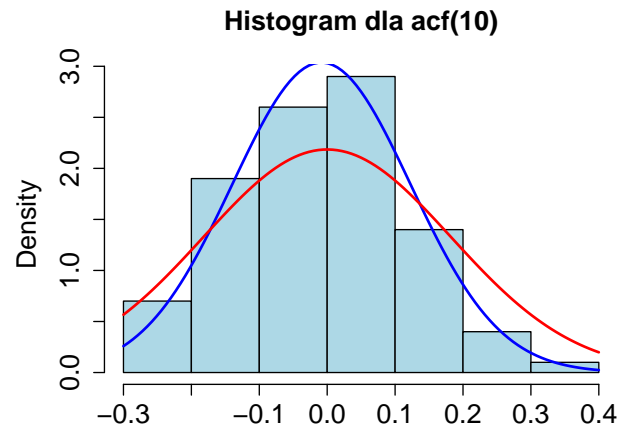
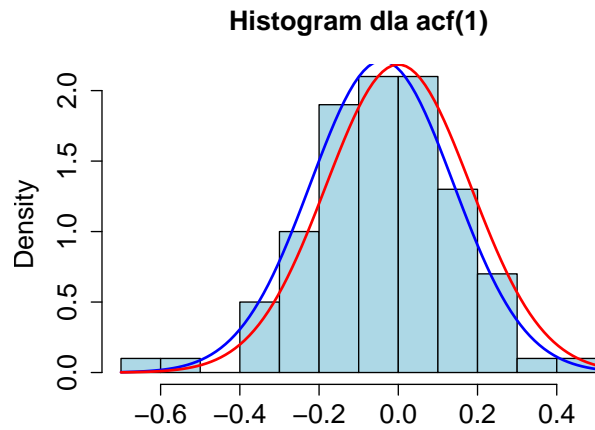
1. długość szeregu: $n \geq 50$,
2. maksymalne opóźnienie: $h \leq n/4$.

1.18 Pierwsza symulacja

W pierwszej symulacji będziemy tworzyli szeregi o długości $n = 30$, jako próby z rozkładu normalnego o średniej 0 i wariancji 1.



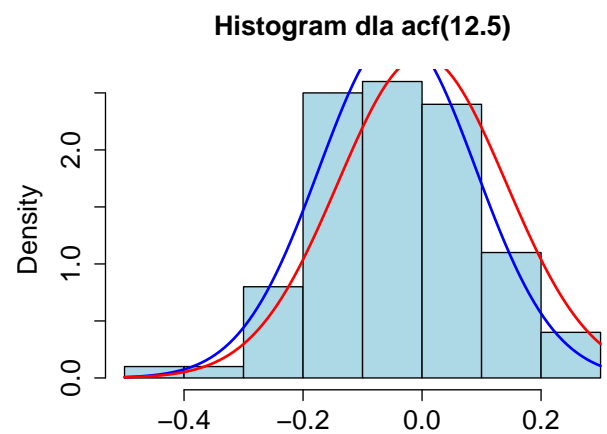
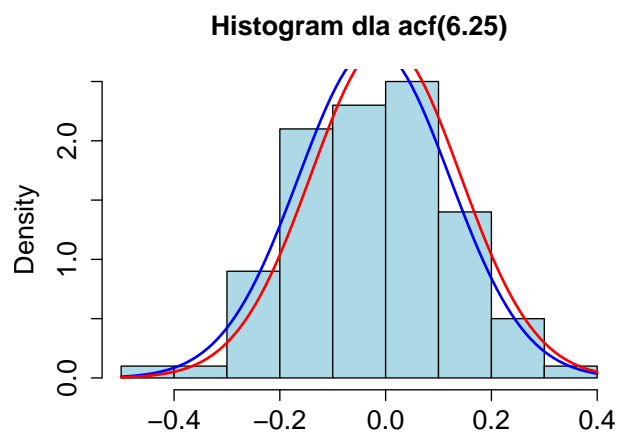
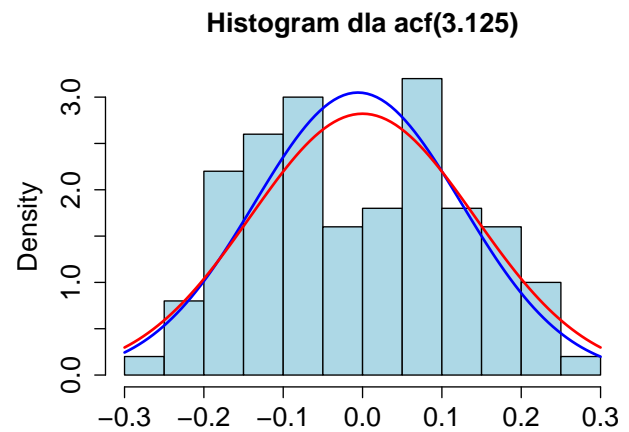
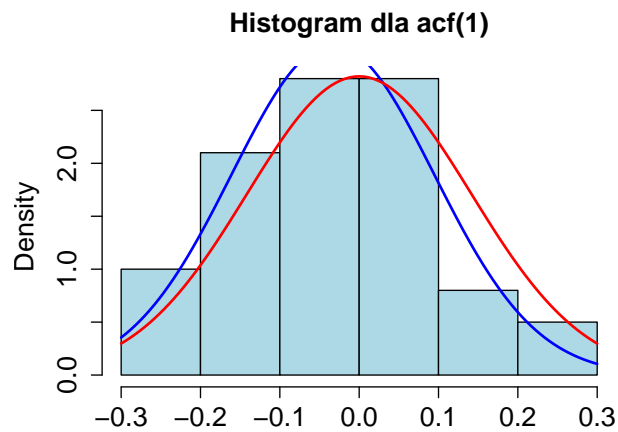
Rysunek 11: Wykresy dla piątej symulacji



- Słabe własności estymatora widoczne są już dla małych opóźnień
- Dla większych opóźnień dostajemy co raz gorsze własności estymatora autokorelacji.
- Przypadek wydaje się potwierdzać regułę Boxa-Jenkinsa - u nas jej założenia nie są spełnione i mamy kiepski estymator.

1.19 Druga symulacja

Teraz dobierzemy długość szeregu n i opóźnienie h według reguły Box'a-Jenkins'a.

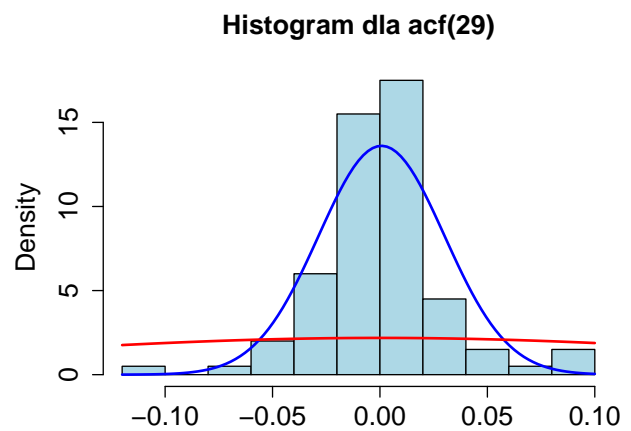
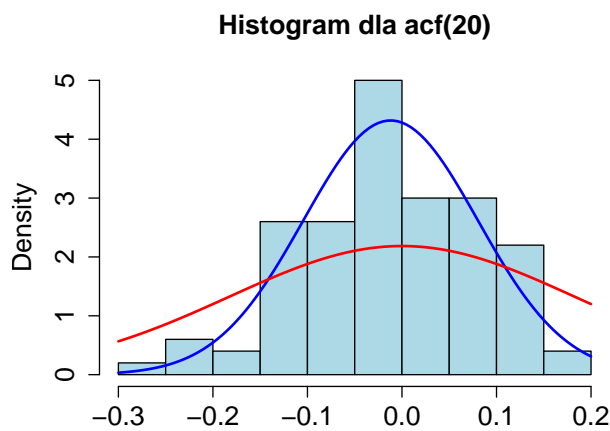
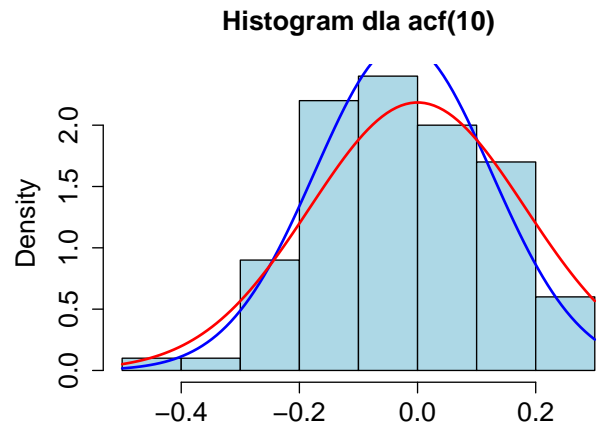
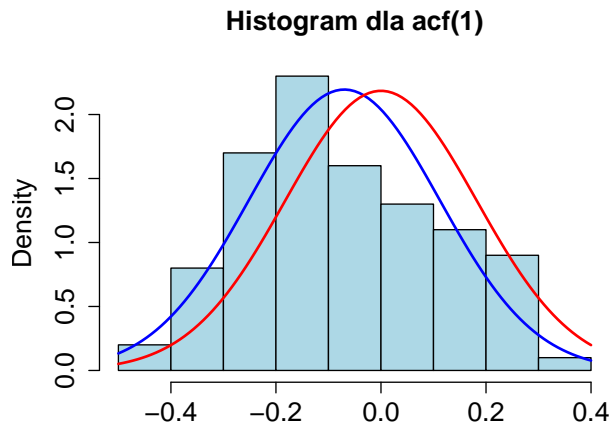


- Z wykresów wnioskujemy, że estymator autokorelacji ma dosyć dobre własności.
- W tym przypadku reguła Box'a-Jenkins'a sprawdziła się.

1.20 Symulacja trzecia

Sprawdźmy co się stanie, gdy znacznie zwiększymy długość szeregu.

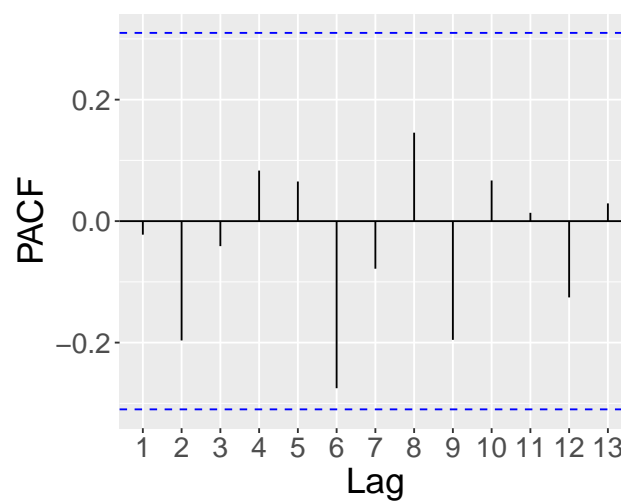
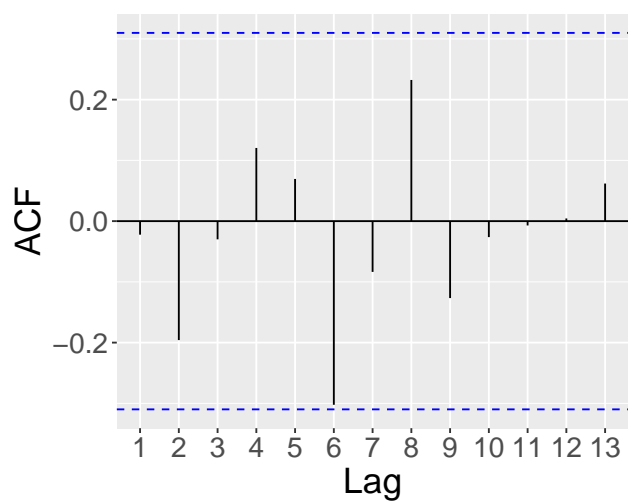
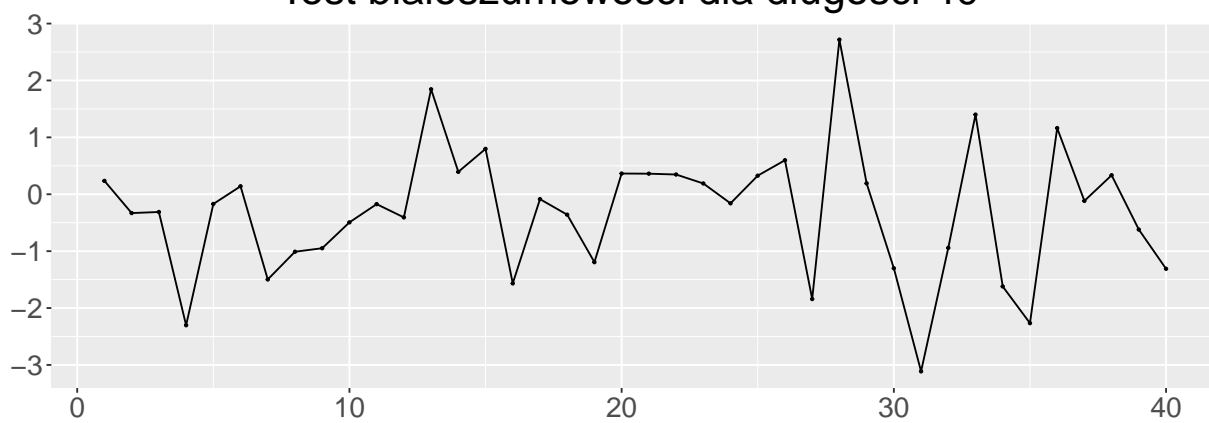
```
## [1] "h=1"
## [1] "h=10"
## [1] "h=20"
## [1] "h=29"
```



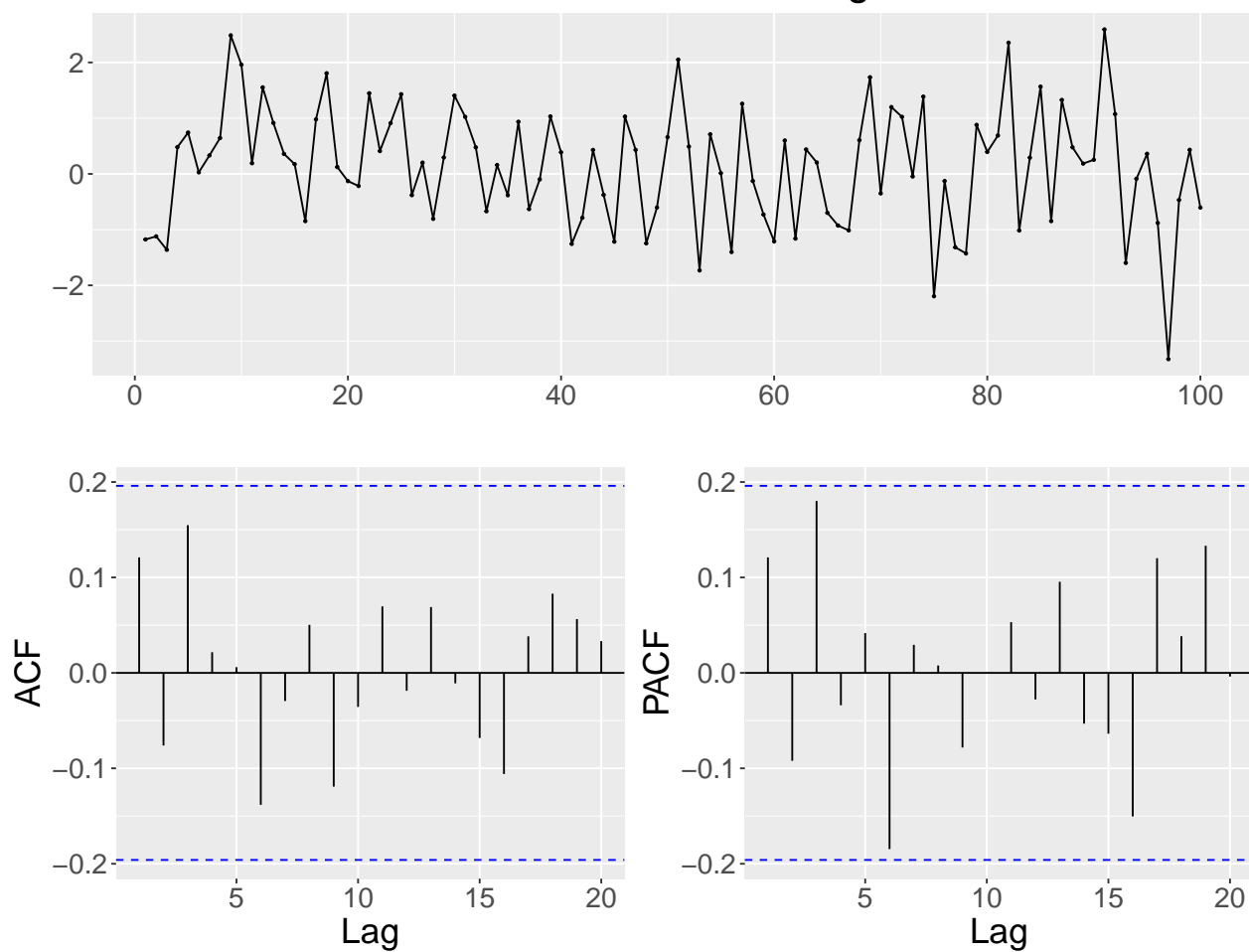
2 Zadanie 3

W zadaniu trzecim wykorzystamy własności funkcji autokorelacji białego szumu do zaimplementowania testu graficznego weryfikującego hipotezę, że obserwowany szereg czasowy jest szeregiem czasowym typu biały szum. Skonstruujemy też formalny test statystyczny weryfikujący hipotezę, że obserwowany szereg czasowy jest szeregiem typu biały szum na asymptotycznym poziomie istotności równym alfa. Przeprowadzimy badania symulacyjne aby ocenić, która metod jest lepsza.

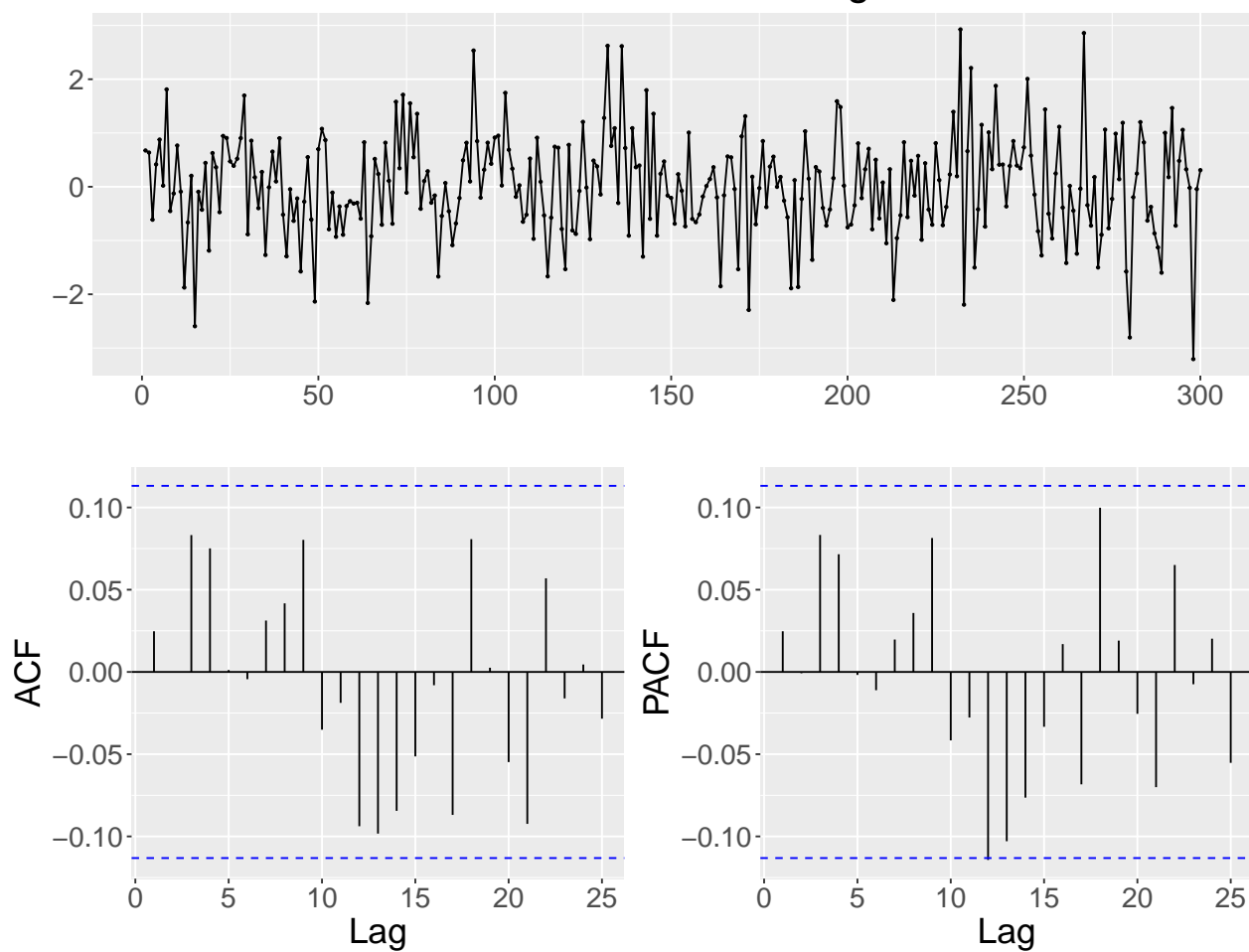
Test białoszumowości dla długości 40



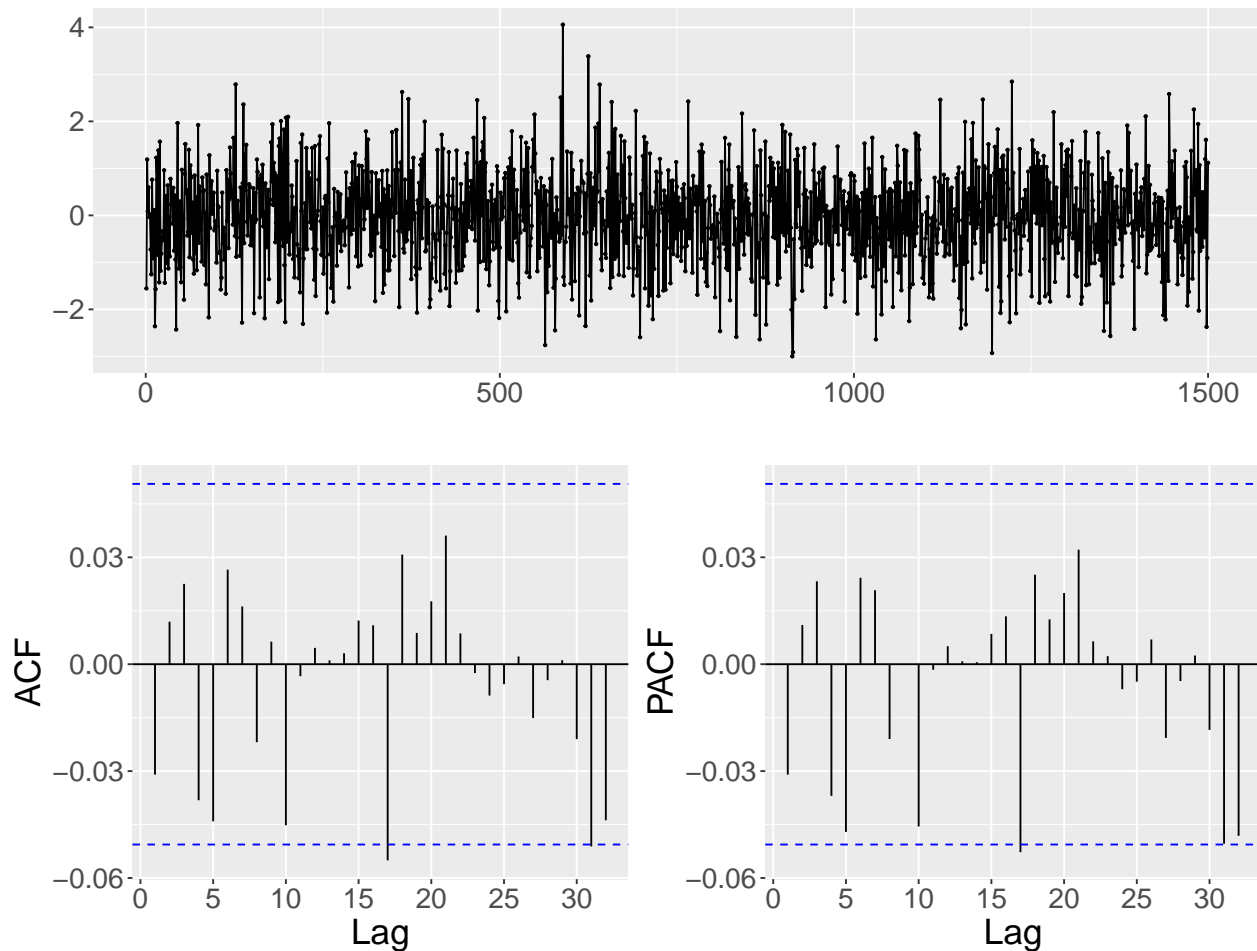
Test białoszumowosci dla dlugosci 100



Test białoszumowości dla długości 300



Test białoszumowosci dla dlugosci 1500



- Na wykresach dla długości szeregu 40, 100 i 300 funkcje ACF i PACF znajdują się pomiędzy krańcami realizacji przedziałów istotności, co potwierdza, że mamy do czynienia z szeregiem typu biały szum.
- Dla szeregu o długości 1500 wartość autokorelacji dla opóźnienia 17 przekracza dolną granicę realizacji przedziału istotności co przeczy tezie, że szereg jest typu biały szum.
- Żadne z wykresów szeregów nie prezentują trendów ani sezonowości. Jest natomiast zauważalna stacjonarność. Są to kolejne argumenty za tym, aby nazwać szereg białym szumem.

2.1 Formalne testy białoszumowości

Tablica 12: Porównanie testów Box'a-Pierce'a i Ljung'i-Box'a pod względem częstości odrzucania hipotezy zerowej.

Długość szeregu (n)	40	100	300	1500
Odsetek odrzuconych hipotez o białoszumowości wg. testu Box'a-Pierce'a	4.36%	4.8%	4.95%	5.02%
Odsetek odrzuconych hipotez o białoszumowości wg. testu Ljung'i-Box'a	5.11%	5.04%	5.1%	5.04%

Test Box'a-Pierce'a rzadziej karze odrzucić hipotezę o białoszumowości szeregu - korzystając z tego testu mamy mniejsze szanse na popełnienie błędu pierwszego rodzaju (wiemy, że hipoteza zeroowa jest prawdziwa, bo losujemy dane z rozkładu normalnego o średniej 0). Nie możemy stąd jednak wnioskować o mocy tych testów.

2.2 Porównanie testu graficznego i formalnego

Stworzymy trzy różne grupy symulacji. Pierwsza z nich dotyczy realizacji szeregu z rozkładu normalnego przy długości szeregu 100 i 500. Następnie realizujemy szereg z rozkładu Cauciego również dla długości 100 i 500. Na koniec przedstawiamy grupę realizacji innych typów rozkładów. Skorzystamy z modelu autoregresji rzędu 1 (AR(1)) i modelu ruchomej średniej rzędu 1 (MA(1)).

2.3 Symulacja Pierwsza

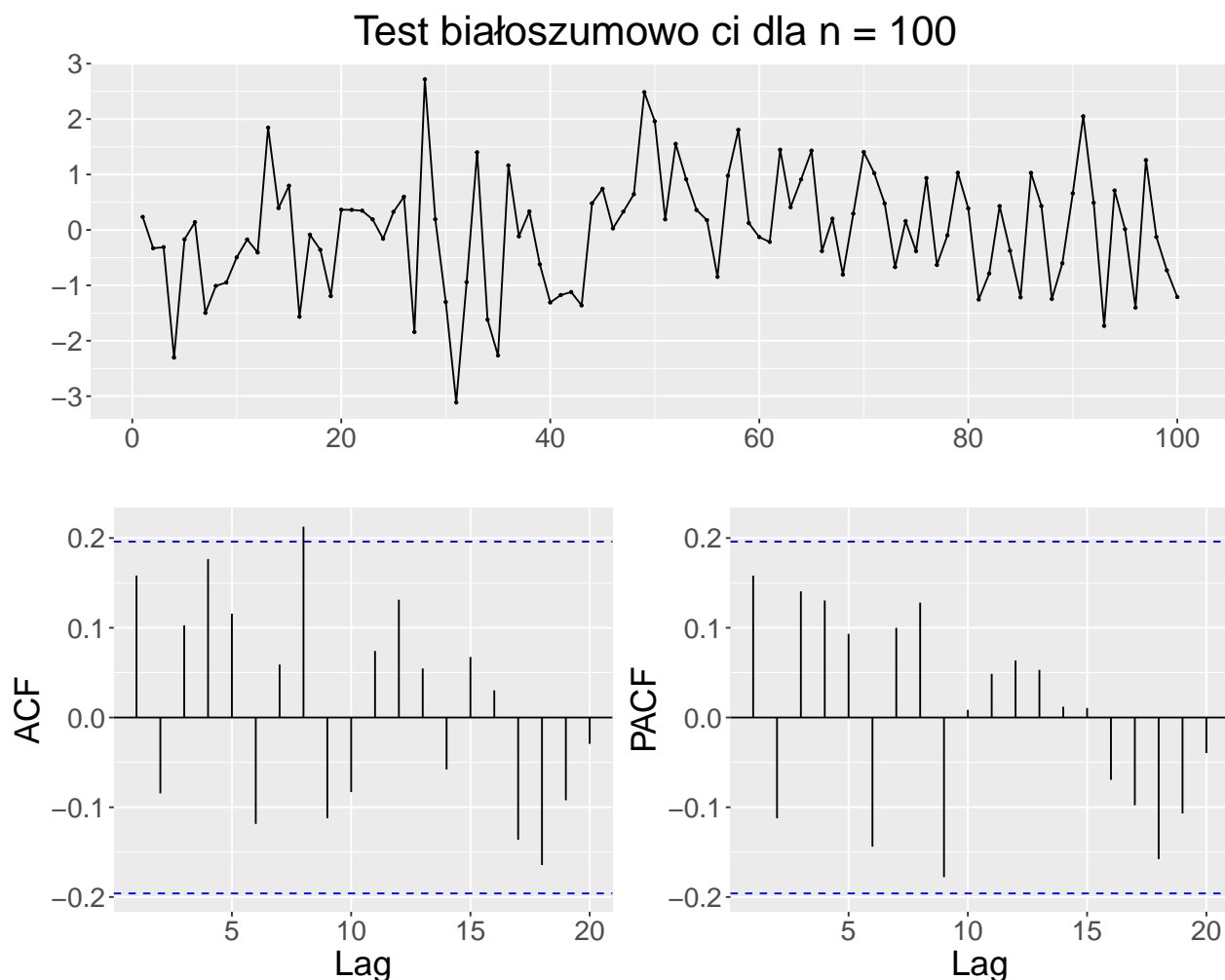
Losujemy 100 liczb z rozkładu normalnego o średniej 0 i wariancji 1.

```

alfa=0.05
#n=100, rozklad normalny
set.seed(111)
par(mfrow=c(1,1))
n=100
X <- rnorm(n,0,1)

ggtsdisplay(X,main="Test białoszumowości dla n = 100",
            theme=theme(text = element_text(size = 20),
                        plot.title = element_text(hjust = 0.5)))

```



```
h.max <- floor(log10(n/4))
# generujemy realizacje białego szumu o długości n
```

Na podstawie graficznego testu odrzucamy hipotezę zerową, że szereg jest białym szumem - obserwacja dla opóźnienia $h = 9$ wystaje poza górną granicę realizacji przedziału istotności dla autokorelacji próbkowej. Wyniki testów sugerują jednak, żeby nie odrzucać hipotezy zerowej. W przypadku testu graficznego popełniamy błąd pierwszego rodzaju.

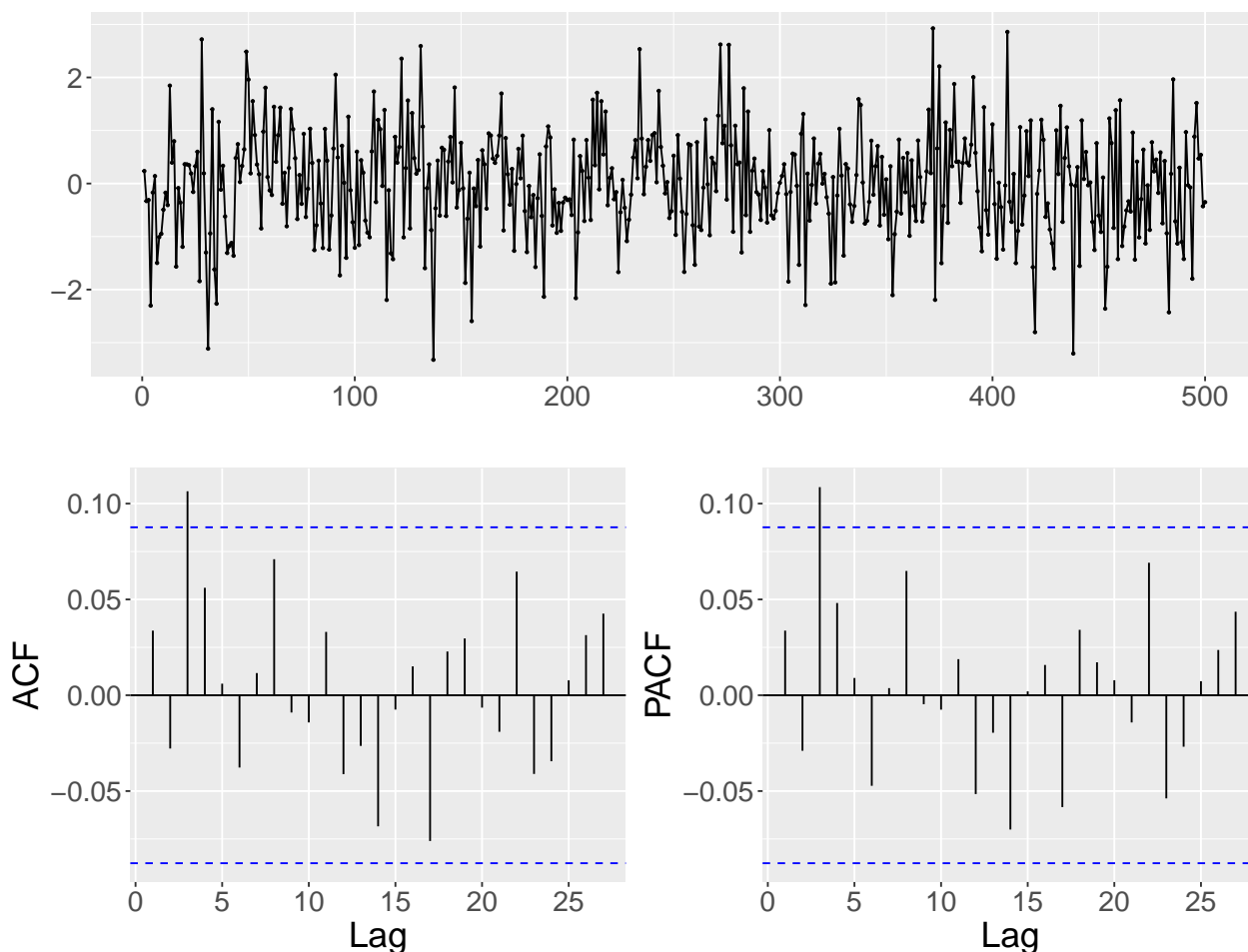
Tablica 13: Porównanie wyników testów formalnych dla pierwszej symulacji

	Statystyka Testowa	P-value
Test Box'a-Pierce'a	2.501	0.1138
Test Ljung'i-Box'a	2.5768	0.1084

2.4 Druga symulacja

Dane losujemy z tego samego rozkładu, ale zwiększamy długość szeregu do 500.

Test białoszumowości dla $n = 500$



Dla opóźnienia $h = 3$ wartość autokorelacji próbkowej wystaje poza realizację przedziału istotności. Zatem test graficzny sugeruje odrzucenie hipotezy zerowej. Wyniki testów zalecają przeciwnie. Tutaj również w przypadku testu graficznego popełniamy błąd pierwszego rodzaju.

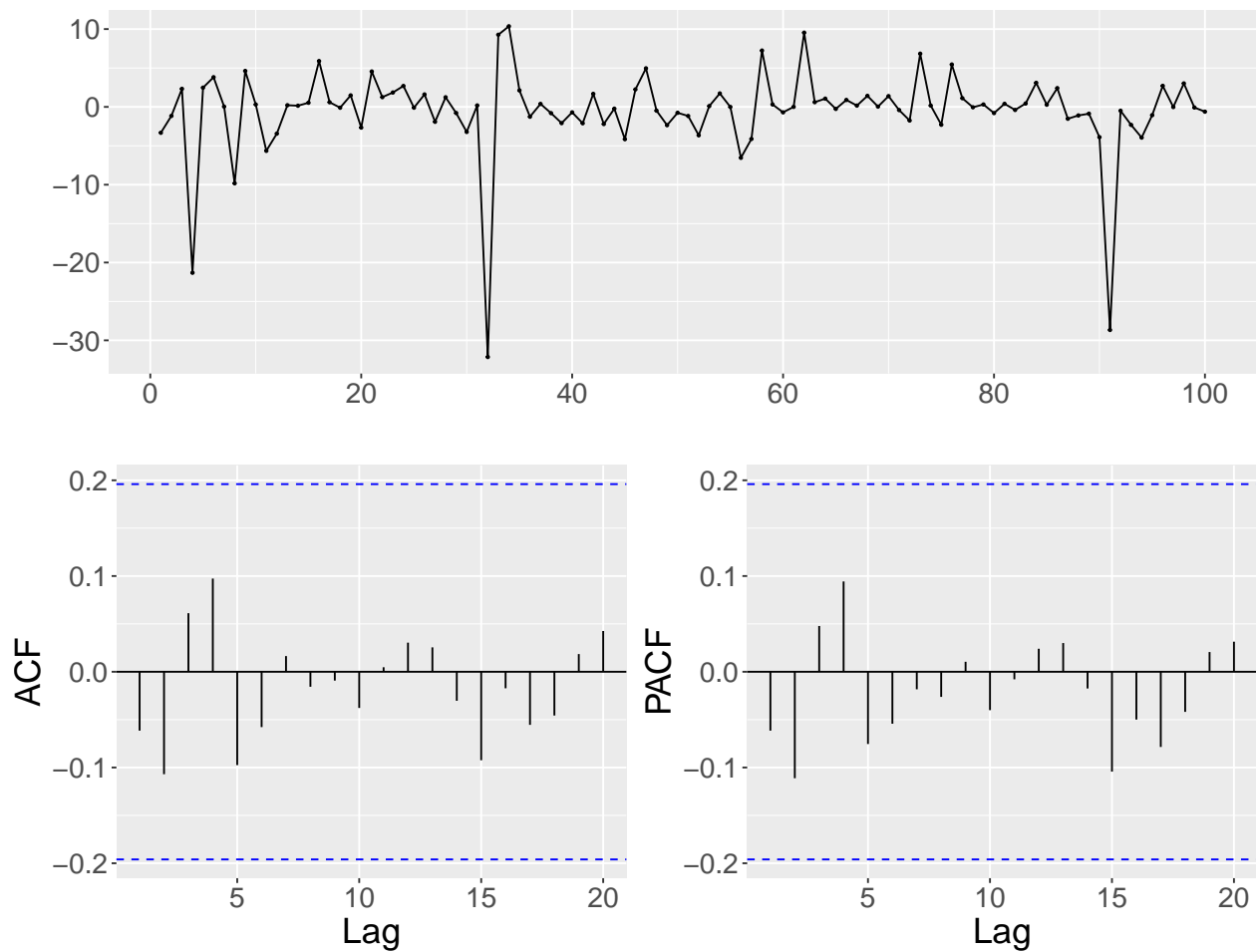
Tablica 14: Wyniki testów formalnych dla drugiej symulacji

	Statystyka Testowa	P-value
Test Box'a-Pierce'a	0.95497	0.6203
Test Ljung'i-Box'a	0.96149	0.6183

2.5 Trzecia symulacja

Teraz dane będziemy losować z rozkładu Cauchy'ego z parametrem skali = 0.0001. Tak jak wcześniej, uwzględnimy długości szeregu 100 i 500.

Test białoszumowości dla $n = 100$



```
##
## Box-Pierce test
##
## data: X
## X-squared = 0.37782, df = 1, p-value = 0.5388
##
## Box-Ljung test
##
## data: X
## X-squared = 0.38927, df = 1, p-value = 0.5327
```

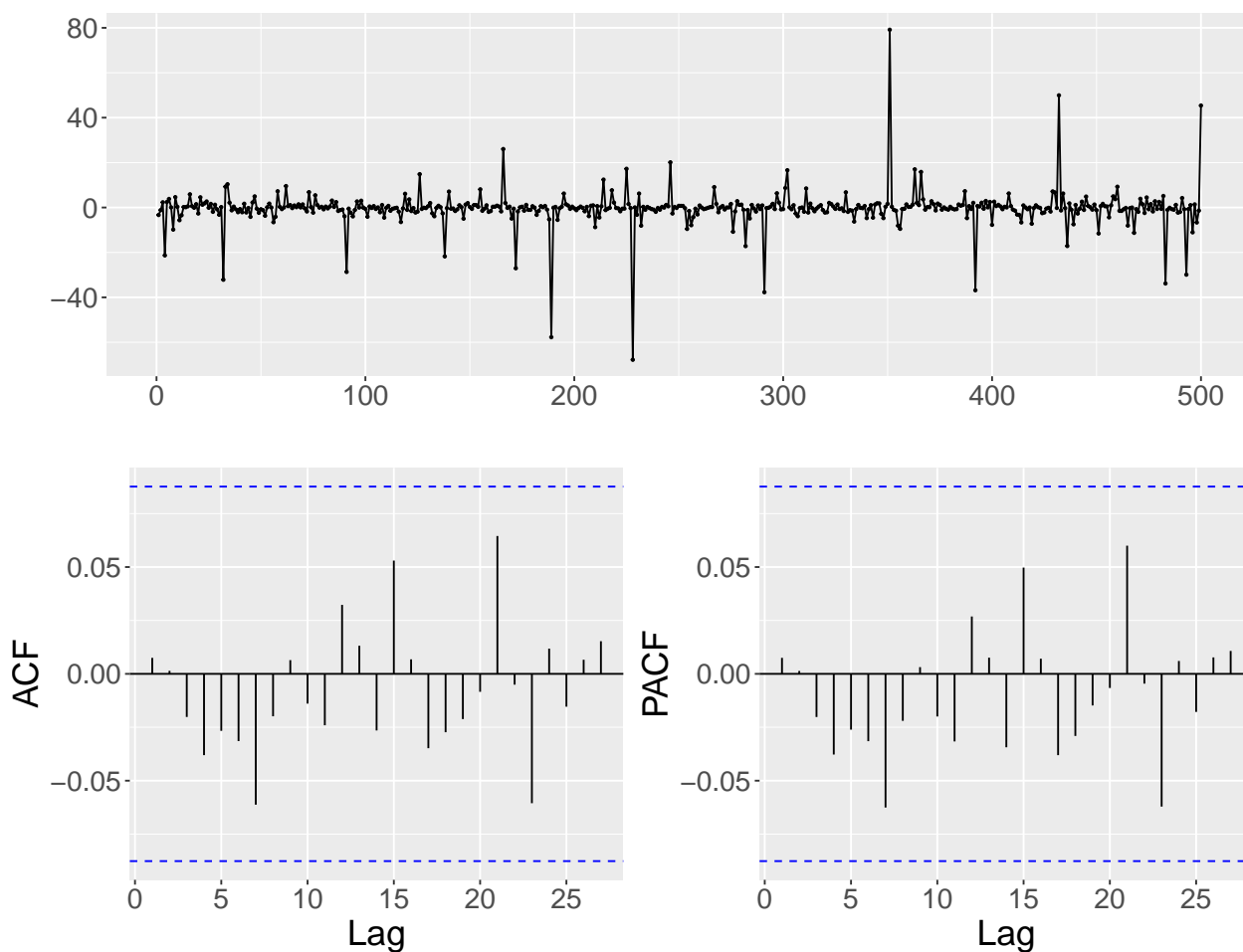
Zarówno test graficzny i jak i testy formalne, nie sugerują odrzucenia hipotezy zerowej.

Tablica 15: Porównanie wyników testów dla symulacji trzeciej

	Statystyka Testowa	P-value
Test Box'a-Pierce'a	0.37782	0.5388
Test Ljung'i-Box'a	0.38927	0.5327

2.6 Symulacja czwarta

Test białoszumowo ci dla $n = 500$



Tutaj również w obu przypadkach nie odrzucamy hipotezy zerowej. Dane losowane z rozkładu Cauchy'ego wydają się bardziej przypominać biały szum, niż te losowane z rozkładu normalnego.

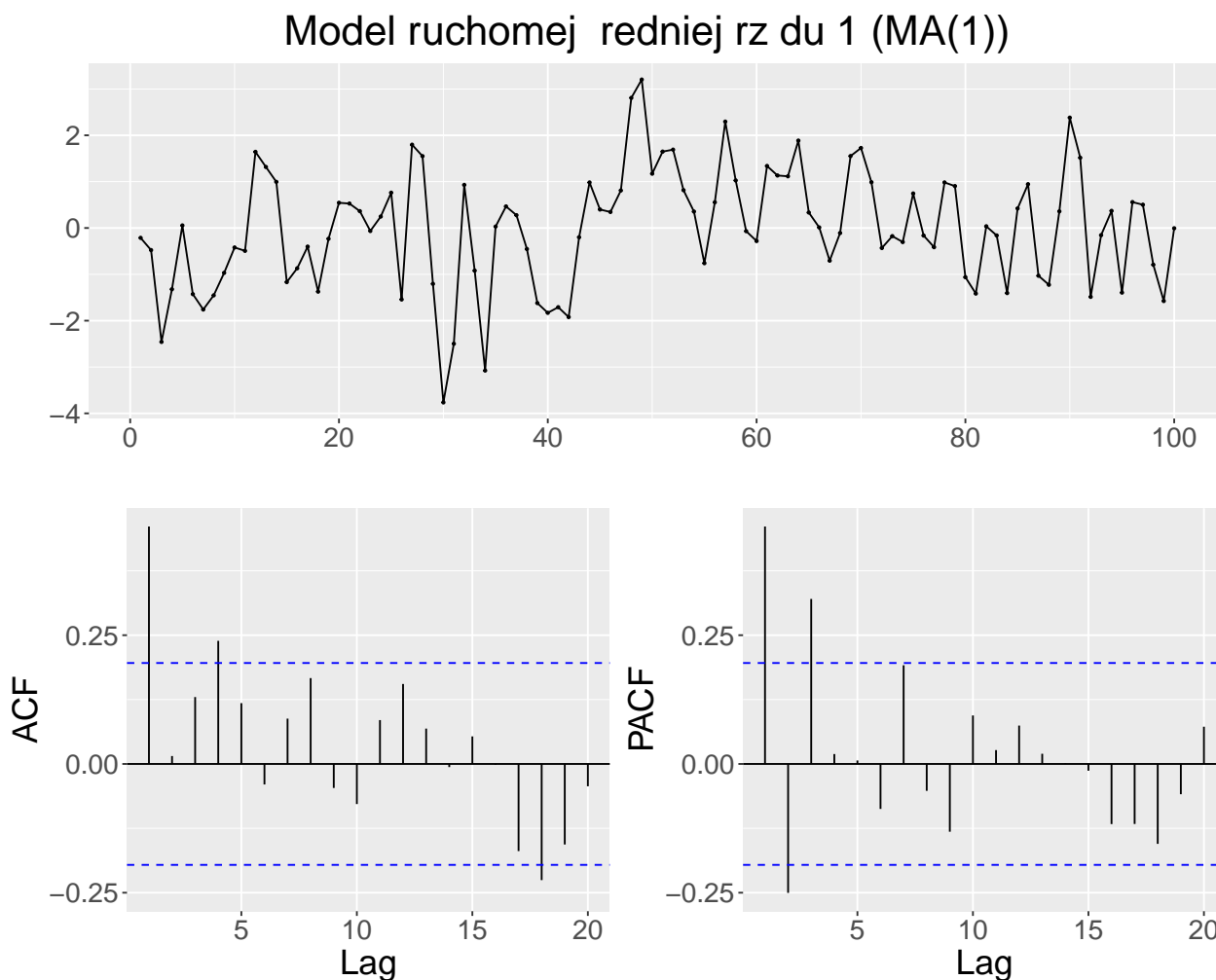
Tablica 16: Porównanie wyników testów dla symulacji czwartej

	Statystyka Testowa	P-value
Test Box'a-Pierce'a	0.029055	0.9856
Test Ljung'i-Box'a	0.029232	0.9855

2.7 Piąta symulacja

Uwzględnimy teraz model ruchomej średniej rzędu 1 (MA(1)) . Z założenia wiemy, że szeregi tego typu nie są białym szumem.

```
set.seed(111)
theta <- 0.5
MA1 <- arima.sim(n=100, model = list(ma=theta))
ggtsdisplay(MA1, main="Model ruchomej średniej rzędu 1 (MA(1))",
  theme=theme(text = element_text(size = 20),
    plot.title = element_text(hjust = 0.5)))
```



```
realizacje <- matrix(MA1,n, 100)
```

Oba testy formalne sugerują odrzucenia hipotezy zerowej. Również test graficzny tak zaleca. Nie mamy do czynienia z błędem żadnego rodzaju.

Tablica 17: Porównanie wyników testów dla symulacji piątej

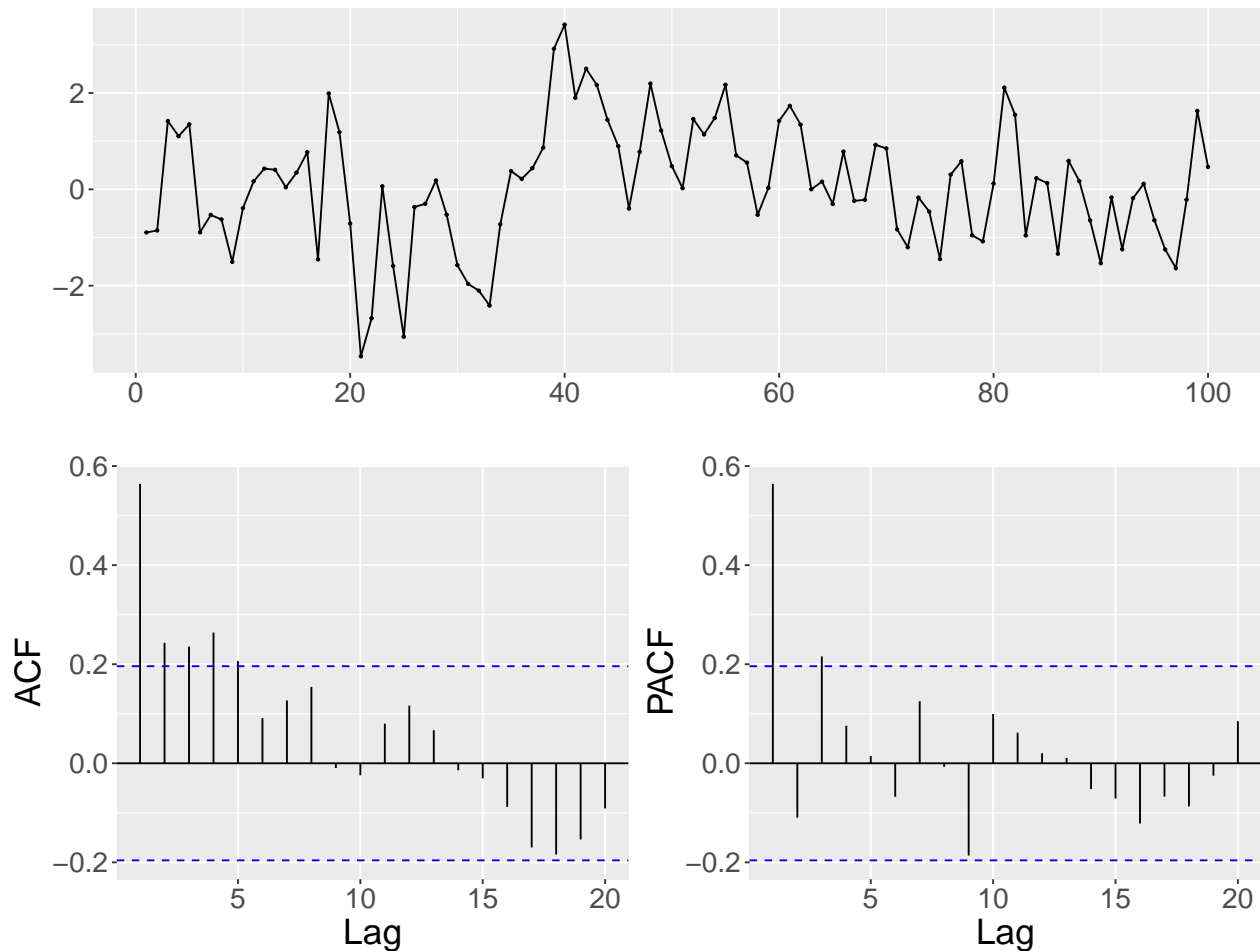
	Statystyka Testowa	P-value
Test Box'a-Pierce'a	21.252	4.027e-06
Test Ljung'i-Box'a	21.896	2.878e-06

2.8 Szósta symulacja

W kolejnej symulacji przeanalizujemy model autoregresji rzędu 1 (AR(1)).

```
set.seed(111)
phi <- 0.5
AR1 <- arima.sim(n=100, model = list(ar=phi))
ggtsdisplay(AR1, theme=theme(text = element_text(size = 20),
  plot.title = element_text(hjust = 0.5)),
  main="Model autoregresji rzędu 1 (AR(1))")
```

Model autoregresji rzędu 1 (AR(1))



```
realizacje <- matrix(AR1,n, 100)
```

Na podstawie wyników testów formalnych jak i testu graficznego odrzucamy hipotezę o tym, że populacja z której pochodzą dane ma korelację 0 i przyjmujemy hipotezę alternatywną - istnieje korelacja w naszych danych. Nie popełniamy błędu żadnego rodzaju.

Tablica 18: Porównanie wyników testów dla symulacji szóstej

	Statystyka Testowa	P-value
Test Box'a-Pierce'a	84.702	4.077e-07
Test Ljung'i-Box'a	95.513	9.363e-09

2.9 Podsumowanie porównania testów

- Gdybyśmy rozstrzygali hipotezy tylko o testy graficzne, częściej popełnialibyśmy błąd pierwszego rodzaju.

- Zwiększenie długości szeregu skutkuje większym p-value dla testów formalnych.
- Przeprowadzając test graficzny, możemy również ocenić czy istnieje sezonowość lub trend w danych. Testy formalne białoszumowości nie mają takiej zalety.
- Żadna metoda nie skutkowała popełnieniem błędu drugiego rodzaju (groźniejszym).
- Ciekawą symulacją byłoby porównanie mocy testów Box'a-Pearce'a i Ljung'i-Boxa

3 Bibliografia

Literatura

- [1] Robert H. Shumway, David S. Stoffer, *Time Series Analysis and Its Applications*, Springer 3rd Edition.
- [2] Dr. Adam Zagdański, *Analiza szeregów czasowych, semestr letni 2021*