

Sprawozdanie 1

Kacper Budnik, 262286

2023-03-31

Symulacyjna analiza własności rozkładów asymptotycznych estymatorów średniej, autokowariancji i autokorelacji

Zdefiniowanie estymatorów

Pierwsze zadanie na sprawozdaniu polega na estymacji parametrów szeregów drugiego rzędu: średniej, funkcji autokowariancji i autokorelacji. W tym celu zdefiniowaliśmy odpowiednio funkcje.

```
MEAN<-function(X){
  sum(X)/length(X)
}

COV<-function(X,h){
  x=MEAN(X)
  len=length(X)
  sum( (X[(h+1):len] - x)*(X[1:(len-h)]-x))/len
}

COR<-function(X,h){
  COV(X,h)/COV(X,0)
}
```

Generowanie szumu

W celu analizy wygenerowaliśmy szeregi czasowe typu biały szum. Do symulacji wygenerowaliśmy każdorazowo 100 realizacji szeregu. Rozpatrywaliśmy szeregi o długościach $n \in \{20, 50, 100, 1000\}$ i rozkładach: $\mathcal{U}(-1, 1)$, $\mathcal{N}(0, 1)$, $3\mathcal{B}(1, 0.5) - 1$, oznaczające kolejno: rozkład jednostajny na $[-1, 1]$, rozkład normalny standoardowy oraz rozkład dwupunktowy przyjmujący wartości w zbiorze $\{-1, 2\}$. Do każdej estymacji posłużył nowo wygenerowany szereg.

Estymowanie parametru średniej

Table 1: Wynik średnich dla jednej próby.

	n=20	n=50	n=100	n=1000
Jednostajny	0.03	-0.13	0.02	0.01
Normalny	-0.41	-0.15	-0.09	-0.01
Dwupunktowy	-0.10	0.02	0.08	-0.05

W powyższej tabeli możemy zauważyć, że wartości estymacji dla jednej próby znacząco odbiegają dla małych długości prób $n = 20$, natomiast dla długości $n = 1000$ różnicę wynoszą w przybliżeniu 0.1 od wartości teoretycznej. Możemy się przyjrzeć jeszcze rozkładowi owych estymatorów. Przyjmy jak wyglądają histogramy owych zmiennych.

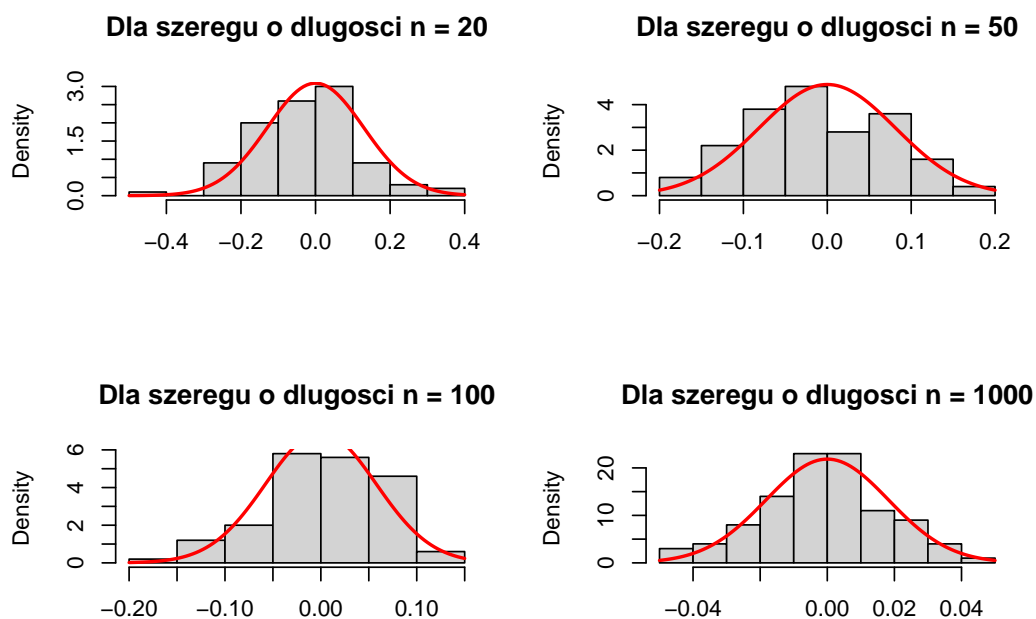


Figure 1: Wykresy średniej dla rozkładu jednostajnego

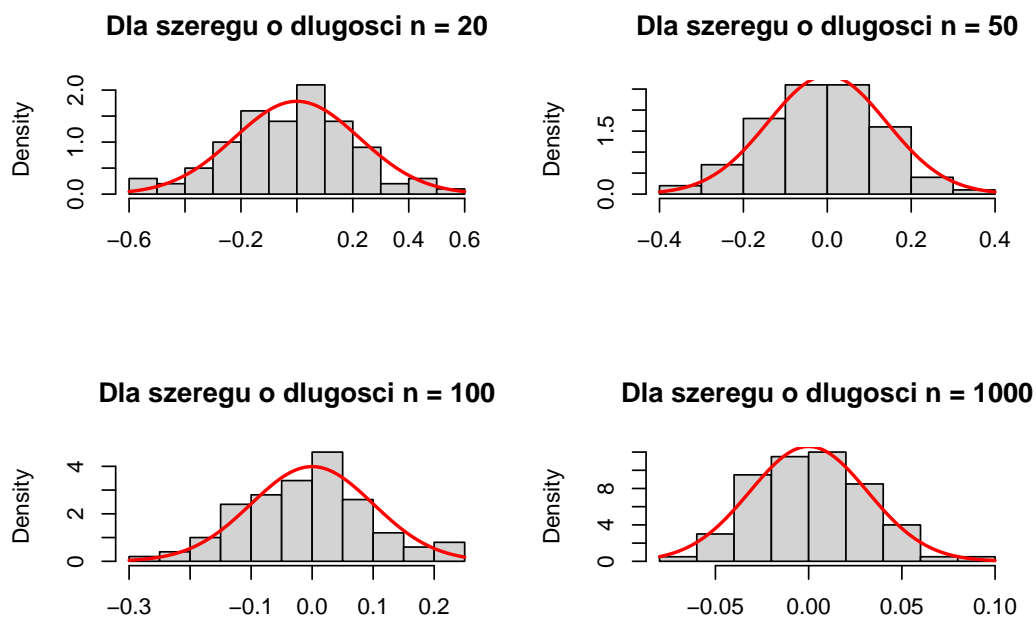


Figure 2: Wykresy średniej dla rozkładu normalnego.

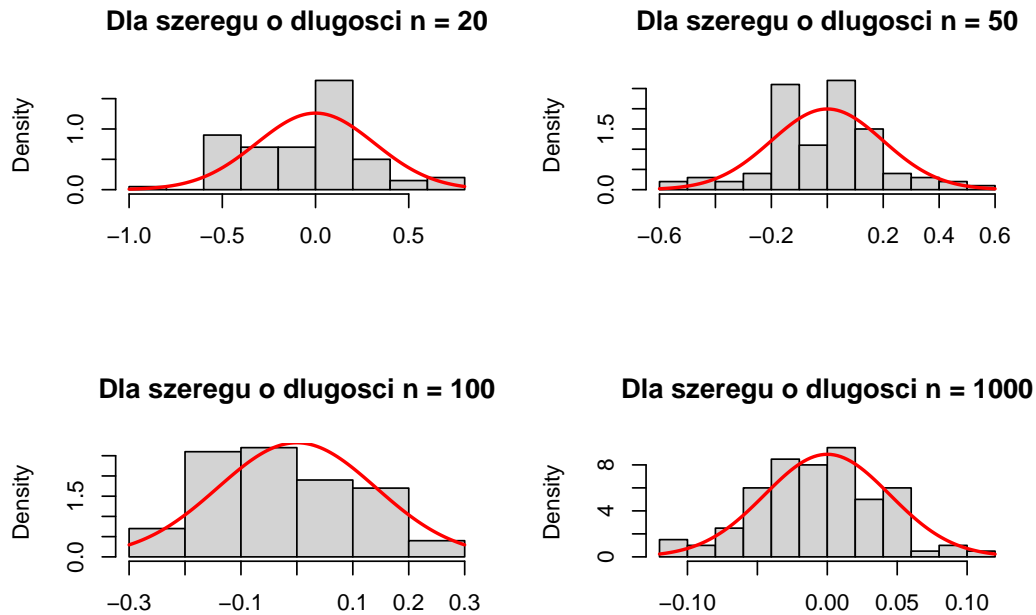


Figure 3: Wykresy średniej dla rozkładu dwupunktowego

Na powyższych wykresach możemy zauważyć, że dla szumu z rozkładu normalnego, krzywa która przybliża rozkład, rzeczywiście dobrze go porywa nie zależnie do wielkości n . Rozkład jednostajny, jak i dwupunktowy radzą sobie znacznie gorzej, jednak dla dużych wartości ponownie widać znaczne podobieństwa. Jednak z analizy graficznej, nie zawsze można wszystko wywnioskować. Sprawdźmy jak sobie poradzą nasze dane z testem Kołomomorowa-Smirnowa. W poniższej tabeli jest z jaką częstością *wartość p* osiągała wartości większe niż 0.05, która została oszacowana metodą Monte-Carlo.

Table 2: Zgodność średnich z rozkładem normalnym.

	n=20	n=50	n=100	n=1000
Jednostajny	0.95	0.95	0.96	0.95
Normalny	0.94	0.95	0.95	0.95
Dwupunktowy	0.40	0.75	0.87	0.94

Jak widzimy, dla rozkładu normalnego oraz jednostajnego test w 95% wskazywał poprawną wartość niezależnie od próby. W przypadku rozkładu dwupunktowego dla małych $n = 20$ wartość p znacząco odbiegała do pozostałych progów, natomiast dla dużego $n = 1000$ wartość ta był taka sama jaką osiągnęły pozostałe wykresy.

Estymowanie funkcji autokorelacji

Estymacja punktowa

Z estymacji dla jednej próby możemy zauważyć w tabelach, że subiektywnie dla $n \geq 20$ wartość estymowanego parametru wariancji jest zbliżona na wartości teoretycznej. Dwa pierwsze rozkłady poradziły sobie poprawnie. Natomiast rozkład dwupunktowy wykazywał znaczącą wartość covariancji nawet dla dużych wartości n . Sprawdźmy jak wygląda rozkład tego estymatora na wykresie

Table 3: Wynik estymacji funkcji autocovariancji dla opóźnienie $h = 0$ (variancji) oraz jednej próby.

	n=20	n=50	n=100	n=1000
Jednostajny	0.38	0.31	0.30	0.33
Normalny	0.65	1.19	1.18	1.03
Dwupunktowy	2.16	1.96	1.93	1.96

Table 4: Wynik estymacji funkcji autocovariancji dla opóźnienie $h = 1$ oraz jednej próby.

	n=20	n=50	n=100	n=1000
Jednostajny	0.03	0.02	0.01	0.01
Normalny	0.12	0.07	0.03	-0.02
Dwupunktowy	-0.48	0.00	-0.21	-0.01

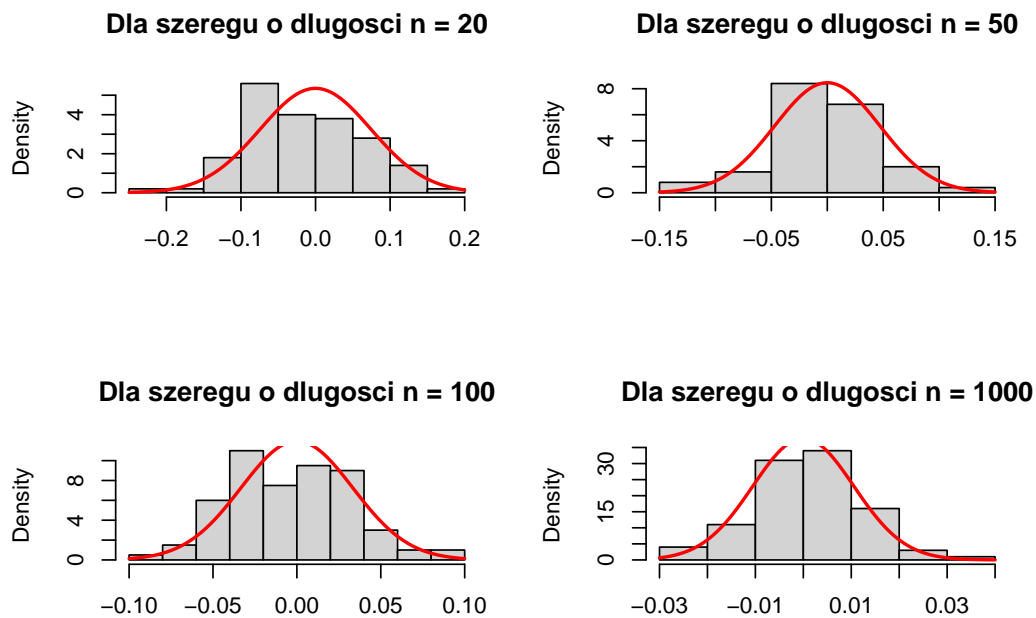


Figure 4: Wykresy funkcji autokowariancji dla rozkładu jednostajnego i opóźnienia $h=1$

Table 5: Wynik estymacji funkcji autocovariancji dla opóźnienie $h = 5$ (variancji) oraz jednej próby.

	n=20	n=50	n=100	n=1000
Jednostajny	-0.01	-0.03	-0.01	-0.02
Normalny	0.02	-0.07	0.07	-0.03
Dwupunktowy	-0.27	0.37	-0.19	-0.06

Table 6: Wynik estymacji funkcji autocovariancji dla opóźnienie $h = 15$ (variancji) oraz jednej próby.

	n=20	n=50	n=100	n=1000
Jednostajny	-0.00	0.03	-0.02	-0.00
Normalny	-0.08	0.14	-0.04	0.04
Dwupunktowy	0.34	-0.33	-0.38	0.12

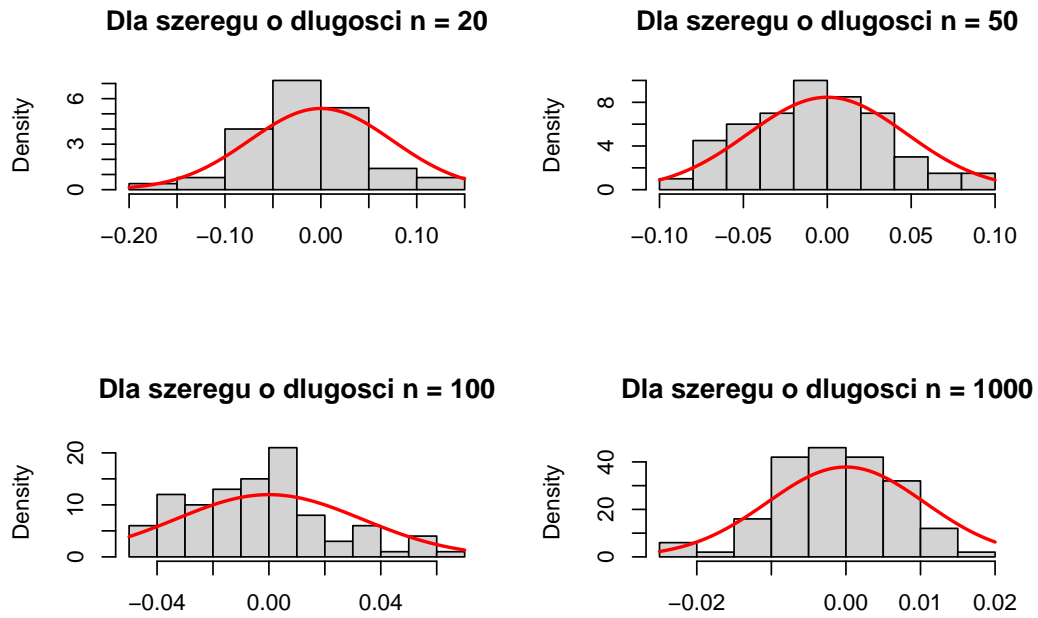


Figure 5: Wykresy funkcji autokowariancji dla rozkładu jednostajnego i opóźnienia h równe ok. $n/4$

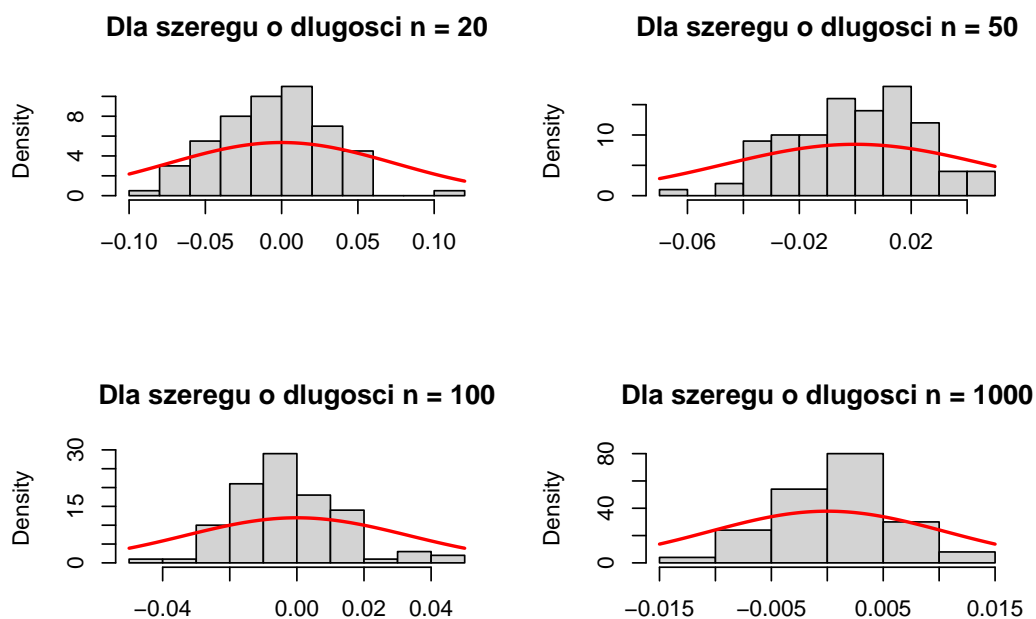


Figure 6: Wykresy funkcji autokowariancji dla rozkładu jednostajnego i opóznienia h równe ok. $3n/4$

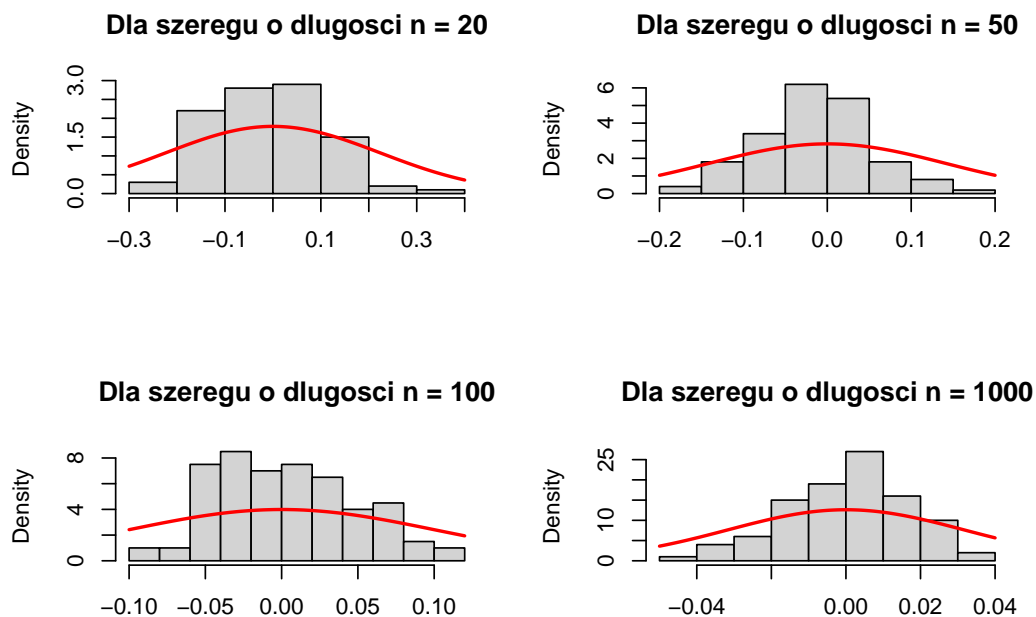


Figure 7: Wykresy funkcji autokowariancji dla rozkładu normalnego i opóznienia h równe ok. $3n/4$

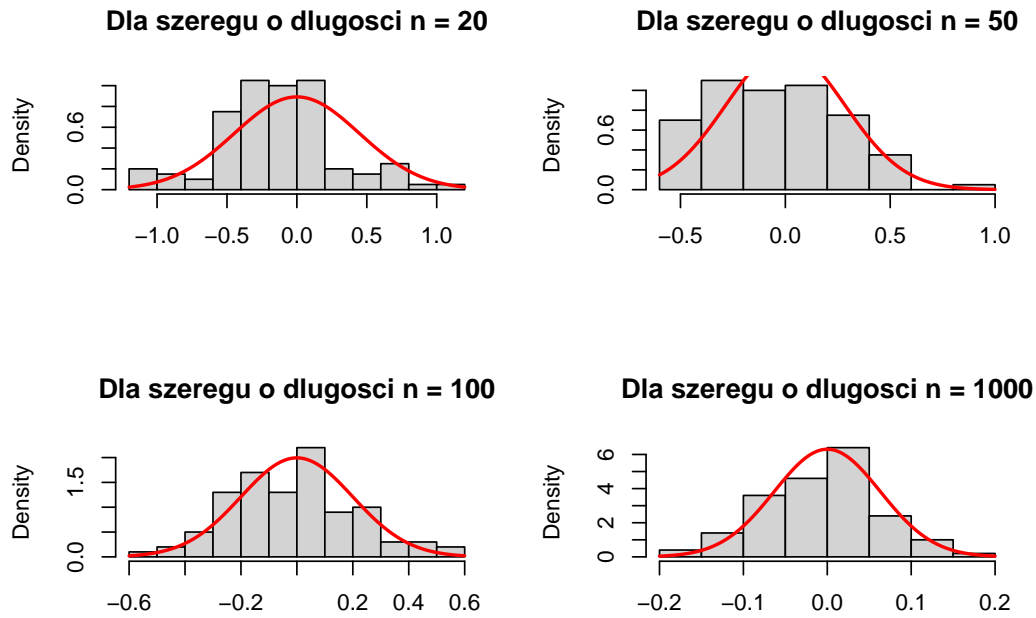


Figure 8: Wykresy funkcji autokowariancji dla rozkładu dwupunktowego i opóznienia $h = 1$

Na wykresach możemy zauważyć, że dla małych wartości h względem n ($h \leq n/4$) nałożona krzywa rozkładu normalnego dobrze pokrywa otrzymane wartości. Jedynie rozkład dwumianowy potrzebuje dużej próby by się do niej dopasować. Natomiast dla dużych wartości $h = 3n/4$ nawet rozkład normalny odstaje od asymptotycznej gęstości. Ponownie możemy sprawdzić wyniki jeszcze przy pomocy testy K-S, po zasadach podobnych jak poprzednio.

Table 7: Zgodność funkcji autokowariancji z rozkładem normalnym dla $h=1$.

	n=20	n=50	n=100	n=1000
Jednostajny	0.49	0.75	0.88	0.96
Normalny	0.41	0.78	0.87	0.94
Dwupunktowy	0.29	0.69	0.87	0.94

Table 8: Zgodność funkcji autokowariancji z rozkładem normalnym dla $h=n/4$.

	n=20	n=50	n=100	n=1000
Jednostajny	0.44	0.74	0.87	0.93
Normalny	0.35	0.74	0.83	0.92
Dwupunktowy	0.34	0.70	0.82	0.92

Table 9: Zgodność funkcji autokowariancji z rozkładem normalnym dla $h=n/2$.

	n=20	n=50	n=100	n=1000
Jednostajny	0.27	0.45	0.53	0.60
Normalny	0.16	0.40	0.46	0.60
Dwupunktowy	0.10	0.33	0.45	0.64

Na powyższych tabelach możemy zauważyć, że dla małych opóźnień ($h \leq n/4$) rozkład zbliża się do rozkładu

normalnego dla $n = 100$. Dla $n = 1000$ częstotliwość wartości p jest porównywalny z wartością p dla rozkładów normalnych. Natomiast, już przy $h = n/2$ i nawet dużych n częstość ta jest daleka od teoretycznych 95%. Ponownie potwierdziło to, że h powinno być relatywnie małe w stosunku do n .

Estymowanie funkcji autocorelacji

Estymacja punktowa

Table 10: Wynik estymacji funkcji autocorelacji dla opóźnienie $h = 1$ (variancji) oraz jednej próby.

	n=20	n=50	n=100	n=1000
Jednostajny	0.29	0.11	0.12	0.01
Normalny	-0.03	-0.07	-0.11	-0.03
Dwupunktowy	-0.19	0.03	-0.04	0.01

Table 11: Wynik estymacji funkcji autocorelacji dla opóźnienie $h = 5$ (variancji) oraz jednej próby.

	n=20	n=50	n=100	n=1000
Jednostajny	-0.14	0.02	0.04	0.05
Normalny	-0.15	0.05	0.04	-0.00
Dwupunktowy	0.15	0.03	0.01	-0.01

Table 12: Wynik estymacji funkcji autocorelacji dla opóźnienie $h = 15$ (variancji) oraz jednej próby.

	n=20	n=50	n=100	n=1000
Jednostajny	0.07	-0.20	0.10	-0.02
Normalny	-0.18	0.10	-0.00	-0.02
Dwupunktowy	0.04	0.15	0.02	-0.04

Z estymacji dla jednej próby możemy zauważyć w tabelach, że dla dużych wartości n wartość jest bliska 0. Dla mniejszych widać jednak statystycznie znaczące odstępstwa, szczególnie dla większych wartości h . Sprawdźmy jak wygląda rozkład tego estymatora na wykresie

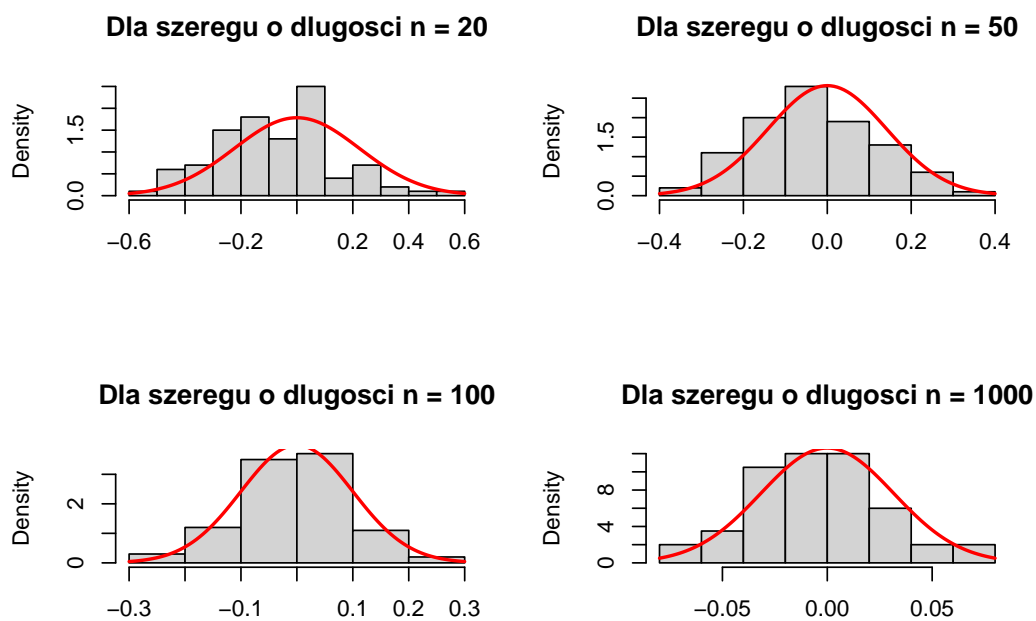


Figure 9: Wykresy funkcji autokorelacji dla rozkładu jednostajnego i opóźnienia $h=1$

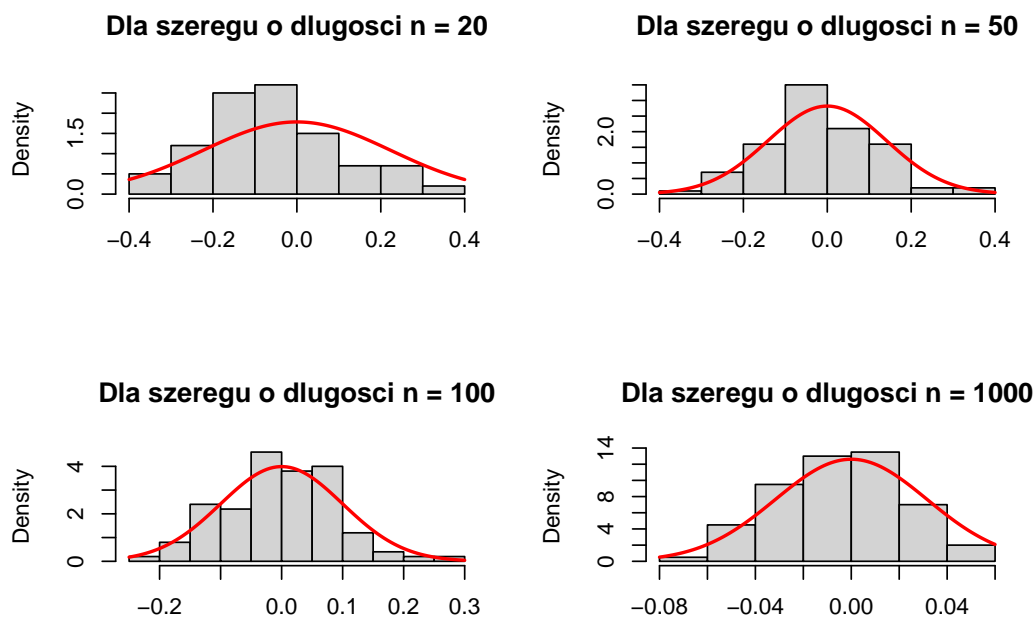


Figure 10: Wykresy funkcji autokorelacji dla rozkładu jednostajnego i opóźnienia h równe ok. $n/4$

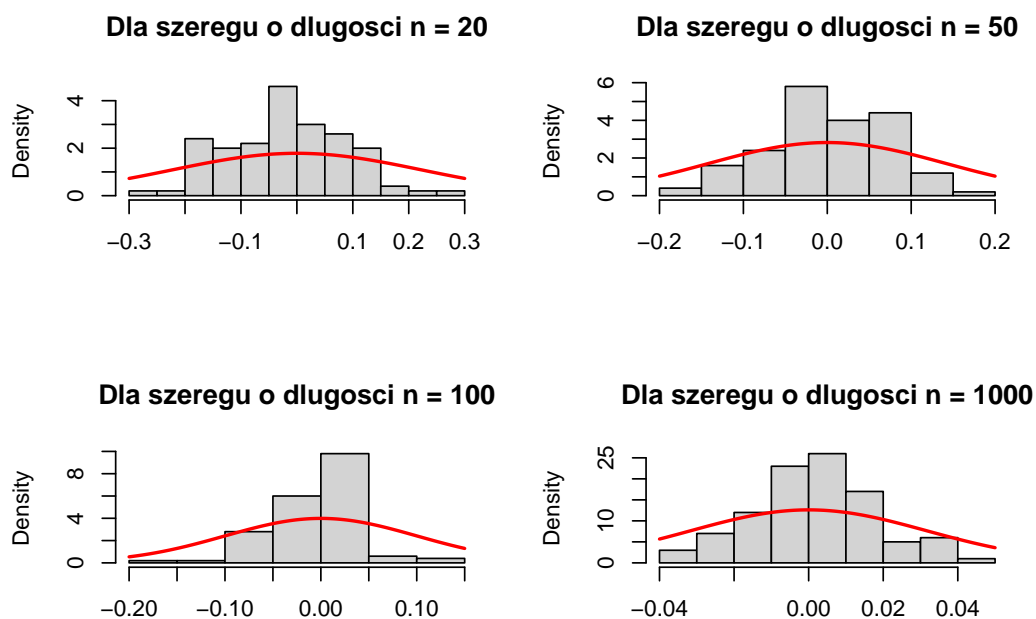


Figure 11: Wykresy funkcji autokorelacji dla rozkładu jednostajnego i opóznienia h równe ok. $3n/4$

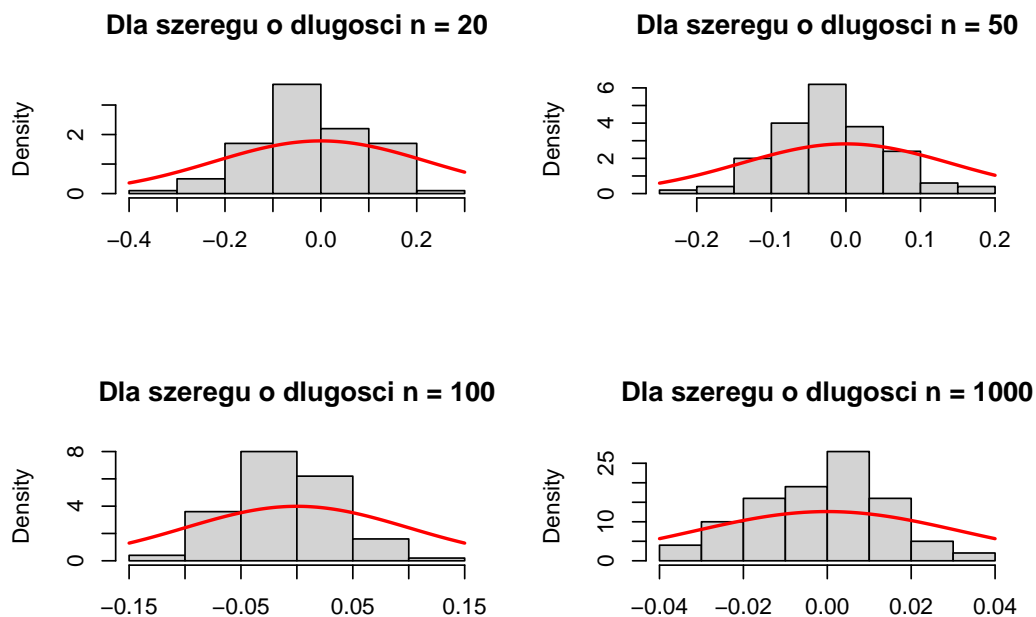


Figure 12: Wykresy funkcji autokorelacji dla rozkładu normalnego i opóznienia h równe ok. $3n/4$

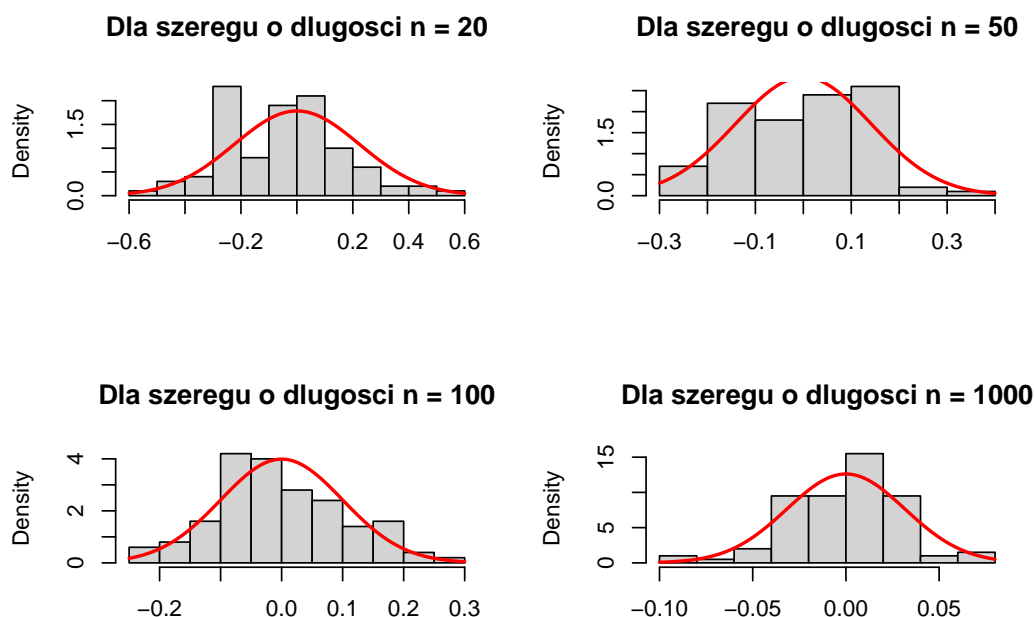


Figure 13: Wykresy funkcji autokorelacji dla rozkładu dwupunktowego i opóźnienia $h = 1$

Na wykresach, podobnie jak wcześniej, możemy zauważyć, że dla małych wartości h względem n ($h \leq n/4$) nałożona krzywa rozkładu normalnego dobrze pokrywa otrzymane wartości. Ponownie dla dużych wartości $h = 3n/4$ względem n , możemy zauważyć, niepoprawne dopasowanie krzywej gęstości rozkładu normalnego do danych. Jeśli spojrzymy jeszcze na wartości stabilizowane poniżej.

Table 13: Zgodność funkcji autokorelacji z rozkładem normalnym dla $h=1$.

	n=20	n=50	n=100	n=1000
Jednostajny	0.46	0.77	0.88	0.95
Normalny	0.47	0.80	0.87	0.95
Dwupunktowy	0.26	0.66	0.84	0.93

Table 14: Zgodność funkcji autokorelacji z rozkładem normalnym dla $h=n/4$.

	n=20	n=50	n=100	n=1000
Jednostajny	0.54	0.81	0.86	0.93
Normalny	0.52	0.77	0.85	0.91
Dwupunktowy	0.33	0.74	0.83	0.93

Table 15: Zgodność funkcji autokorelacji z rozkładem normalnym dla $h=n/2$.

	n=20	n=50	n=100	n=1000
Jednostajny	0.42	0.52	0.54	0.61
Normalny	0.38	0.52	0.52	0.62
Dwupunktowy	0.11	0.40	0.49	0.61

Również tu możemy wyciągnąć te same wnioski jakie dla funkcji autokowariancji. Dla małych opóźnień h względem $n \geq 100$ wygenerowane dane przypominają rozkład normalny z parametrami średniej $\mu = 0$ oraz

odchylenia standardowego $\sigma = 1/\sqrt{n}$. Natomiast dla $h = n/2$ i nawet dużej próby $n = 1000$ pojawiają się poważne odstępstwa od rozkładu asymptotycznego

Testowanie Białoszumowości

Test graficzny

Do zaimplementowania “graficznego” testu białoszumowości korzystam z informacji z wykładu, czyli

- Co najmniej 95% z rozpatrywanych czasów późnienia funkcji autokowariacji powinna się znajdować w przedziale $\pm 1.96/\sqrt{n}$.
- Obserwacje wystające poza przedział nie powinny przekraczać wartości $\pm 2.94/\sqrt{n}$, czyli nie powinny wykroczyć znacząco za zadany przedział. Wartość tę wybraliśmy na podstawie zasady 3 sigm.

W testach będziemy rozpatrywać maksymalne opóźnienie $h_{max} = [n/4]$, czyli część całkowitą z wartości $n/4$.

```
test<-function(X){
  n=length(X)
  h=floor(n/4)
  af=abs(acf(X,lag=h,plot=FALSE)$acf[-1])
  if(sum(af>1.96/sqrt(n))/h > 0.05) return(0)#(c(0,sum(af<1.96/sqrt(n))/h))
  if(max(af)>2.94/sqrt(n)) return(0)
  return(1)
}
```

```
set.seed(3141562)
sum(apply(matrix(runif(20*1000),nrow=1000), MARGIN = 1,test))/1000
sum(apply(matrix(rnorm(20*1000),nrow=1000), MARGIN = 1,test))/1000

X=matrix(runif(21*1000),nrow=1000, ncol = 21)
Y=X[,-21]-pi*X[,-1]
sum(apply(Y, MARGIN = 2,test))/1000
```

Test ten sprawdziliśmy początkowo dla ciągów typu właśnie białego szumu o długości 20 i rozkładzie jednostajnym lub normalnym standardowym. Symulacje powtórzyliśmy 1000 razy dla każdego rozkładu. W wyniku otrzymaliśmy, że szereg został uznany za biały szum jedynie odpowiednio 87.2% oraz 89.2%. Sprawdziliśmy jeszcze jak ocenia on szeregi MA(1) postaci

$$X_t = Z_t - aZ_{t-1}, \quad (1)$$

gdzie Z_t jest białym szumem z rozkładu jednostajnego standardowego. Wybraliśmy parametr $a = \pi$. W wyniku 1000 symulacji szereg został zaklasyfikowany jako biały szum jedynie 1.7% razy. Widać, że metoda graficzna ma skłonność do klasyfikacji szeregów jako nie białyszum

Formalny Test białego szumu

W przeciwieństwie do testów graficznych, których zasady działania często opierają się na często złudnej intuicji, stoją testy formalne. W naszym przypadku skorzystamy z dwóch testów: Testu Boxa–Pierce’a (BP) oraz Testu Ljungu–Boxa (LB). Poniżej napisaliśmy odpowiednio funkcje:

```
bp<-function(X,h=0,a=0.05){
  if(h==0) h=floor(length(X)/4)
  n=length(X)
  rho=acf(X,lag.max=h,plot=FALSE)$acf[2:(h+1)]
  a<1-pchisq(n*sum(rho^2),h)
}
```

```

lb<-function(X,h=0,a=0.05){
  if(h==0) h=floor(length(X)/4)
  rho=acf(X,lag.max=h,plot=FALSE)$acf[2:(h+1)]
  n=length(X)
  Q=0
  for(i in 1:h){
    Q=Q+rho[i]^2/(n-i)
  }
  Q=Q*n*(n+2)
  a<1-pchisq(Q,h)
}

```

Każdą z funkcji przetestowaliśmy poprzez sprawdzenie jak często testy świadczą o białoszumowości z pewnością na poziomie $\alpha = 0.05$, na podstawie 1000 symulacji MC dla każdego z rozpatrywanych rozkładów. Dla zmiennej z rozkładu MA(1), zdefiniowanej wcześniej testy BP i LB uzyskały kolejno częstość na poziomie 32.2% oraz 30%, więc dla mniejszych rozmiarów $n = 20$, testy mają jeszcze małą skuteczność. Sprawdzić powinniśmy jeszcze jak się zachowują wobec prawdziwego białego szumu. Sprawdziliśmy to dla rozkładu jednostajnego standardowego i normalnego standardowego. Test BP w 96.7% oraz 97.6% poprawnie świadczył o białoszumowości, natomiast test LB osiągnął wynik na poziomie 92.4% oraz 93.7%.

Porównanie metod

Table 16: Częstość przyjmowania hipotezy o białoszumowości dla testu graficznego.

	n=20	n=50	n=100	n=150
Jednostajny	0.87	0.65	0.77	0.63
Normalny	0.89	0.70	0.77	0.64
MA(1)	0.72	0.32	0.26	0.07
Spacer losowy	0.02	0.04	0.04	0.10

Table 17: Częstość przyjmowania hipotezy o białoszumowości dla testu BP.

	n=20	n=50	n=100	n=150
Jednostajny	0.97	0.96	0.97	0.97
Normalny	0.98	0.97	0.96	0.98
MA(1)	0.93	0.81	0.72	0.63
Spacer losowy	0.02	0.05	0.10	0.15

Table 18: Częstość przyjmowania hipotezy o białoszumowości dla testu LB.

	n=20	n=50	n=100	n=150
Jednostajny	0.92	0.93	0.92	0.92
Normalny	0.94	0.92	0.91	0.94
MA(1)	0.84	0.72	0.59	0.46
Spacer losowy	0.02	0.05	0.09	0.15

Z tablic można zauważyć, że test BP częściej świadczy o białoszumowości niż test LB, niezależnie od stanu prawdziwego. Test graficzny znacząco zaniża częstość, świadcząc o braku białego szumu. Wszystkie metody bardzo dobrze sobie radzą z identyfikowaniem spaceru losowego, natomiast mają problemy z szeregiem tymu MA(1). Dla dużych $n = 150$ wartości oba testy formalne rodzą sobie z identyfikowaniem białego szumu innego, w przypadku jeżeli nie jest to model MA. Dla większych wartości, bliskich $n = 1000$ oba testy radzą sobie również z tym modelem