Lista 5

Analiza szeregów czasowych zadanie 3

Kacper Szmigielski (282255)

Spis treści

1 Zadanie 3 **2**

1.1 a) 2

1.1.1 Definicje teoretyczne potrzebne do tego zadania 2

1.1.2 Reguła graficznej identyfikacji białego szumu 5

1.1.3 Formalne testy białoszumowości oparte na ACF 5

1.2 b) 9

1.2.1 Porównywanie testów graficznych oraz formalnych dla testowania białoszumowości 9

Spis rysunków

Spis tabel

1 Wyniki procentowe kwalifikacji danych jako biały szum 11

2 Wyniki procentowe kwalifikacji danych jako biały szum 11

3 Wyniki procentowe kwalifikacji danych jako biały szum 11

1 Zadanie 3

1.1 a)

1.1.1 Definicje teoretyczne potrzebne do tego zadania

Definicja 1 (Biały szum)**.** *Proces losowy* $\{X_t\}_{t\in\mathbb{Z}}$ *nazywamy* białym szumem*, jeśli spełnia następujące warunki:*

1. $E[X_t] = \mu$ *(stała średnia, zwykle przyjmuje się 0, bo zawsze można dane wycentrować),*
2. $\mathrm{Var}(X_t) = \sigma^2 < \infty$ *(stała wariancja),*
3. X_t *i* X_s *są nieskorelowane dla* $t \neq s$*, tzn.*

$$\mathrm{Cov}(X_t, X_s) = 0 \quad \text{dla wszystkich } t \neq s.$$

Jeśli dodatkowo zmienne X_t *są wzajemnie niezależne i mają rozkład normalny* $\mathcal{N}(0, \sigma^2)$*, to proces nazywamy* gaussowskim białym szumem.

Definicja 2 (Szereg stacjonarny)**.** *Szereg czasowy* $\{X_t : t \in \mathbb{Z}\}$ *nazywamy (słabo) stacjonarnym, jeśli dla dowolnych* $t, r, s \in \mathbb{Z}$ *spełnione są następujące warunki:*

1. $E[X_t^2] < \infty$,
2. *średnia jest stała:*
$$\mu_X(t) = E[X_t] = \mu = const,$$
3. *kowariancja zależy wyłącznie od przesunięcia czasowego:*
$$\gamma_X(r, s) = \gamma_X(r + t, s + t),$$
to znaczy zależy tylko od różnicy $r - s$.

Definicja 3 (Estymator wartości średniej)**.** *Niech $x_1, x_2, \ldots, x_n$ będzie realizacją szeregu czasowego $\{X_t\}$. Estymator wartości średniej μ_X definiujemy jako*

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

```
mn <- function(data){

  return (mean(data))

}
```

Definicja 4 (Próbkowa funkcja autokowariancji)**.** *Niech $x_1, x_2, \ldots, x_n$ będzie realizacją szeregu czasowego. Próbkową funkcję autokowariancji dla opóźnienia h, gdzie $-n < h < n$, definiujemy jako*

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \overline{x})(x_t - \overline{x}),$$

gdzie $\overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$ jest średnią próbkową.

```
ACVF <- function(data, h){

  x_bar = mn(data)

  result = 0

  n = length(data)

  if( abs(h) > n){
    stop("h must be greater than -n and lower than n")
  }

  for (t in 1:(n-abs(h))){
    result  = result + (data[t + abs(h)]-x_bar) * (data[t] - x_bar)
  }
  result = result / n
  return (result)
}
```

Definicja 5 (Próbkowa funkcja autokorelacji)**.** *Niech* $x_1, x_2, \ldots, x_n$ *będzie realizacją szeregu czasowego, a* $\hat{\gamma}(h)$ *niech oznacza prób-kową funkcję autokowariancji. Próbkową funkcję autokorelacji dla opóźnienia* h*, gdzie* $-n < h < n$*, definiujemy jako*

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

```
ACF <- function(data,h){

  n = length(data)

  if( abs(h) > n){
    stop("h must be greater than -n and lower than n")
  }

  part1 = ACVF(data,h)

  part2 = ACVF(data,0)

  result = part1/part2

  return (result)

}
```

Definicja 6 (Własności asymptotyczne estymatorów średniej, autokowariancji i autokorelacji)**.** *Niech* $\{X_t\}$ *będzie procesem stacjonarnym. Własności drugiego rzędu tego procesu są całkowicie określone przez średnią* μ *oraz funkcję autokowariancji* $\gamma(\cdot)$*.*

Estymacja parametrów μ*,* $\gamma(\cdot)$ *oraz funkcji autokorelacji*

$$\theta(h) = \frac{\gamma(h)}{\gamma(0)},$$

na podstawie obserwacji $X_1, X_2, \ldots, X_n$ *odgrywa zasadniczą rolę w metodach wnioskowania statystycznego, w szczególności przy doborze modelu oraz konstrukcji optymalnych prognoz.*

Dlatego też istotne jest, aby stosowane estymatory $\hat{\mu}$*,* $\hat{\gamma}(h)$ *oraz* $\hat{\theta}(h)$ *posiadały odpowiednie własności asymptotyczne, takie jak zgodność, asymptotyczna nieobciążoność oraz asymptotyczna normalność, zapewniające wiarygodność wnioskowania przy rosnącej liczbie obserwacji.*

1.1.2 Reguła graficznej identyfikacji białego szumu

Szereg możemy uznać za realizację białego szumu jeżeli:

1. Co najmniej 95% autokorelacji próbkowych ($ACF(h), h = 1, 2, ..., h_{max}$) znajduje się w przedziale ufności : $\frac{\pm 1{,}96}{\sqrt{n}}$
2. Nie powinno być autokorelacji "istotnie" wychodzących pzoa przedziały ufności $\frac{\pm 1{,}96}{\sqrt{n}}$

1.1.3 Formalne testy białoszumowości oparte na ACF

1. Podstawowe testy oparte na ACF
   a) text Boxa-Pierce'a
   b) test Ljungi-Boxa
2. Idea obu testów jest następująca: zamiast sprawdzać (tak jak w graficznym teście białoszumowości) dla każdego h = 1,2,3,... czy autokorelacja próbkowa ACF(h) znajduje się pomiędzy przedziałami istotności $\frac{\pm 1{,}96}{\sqrt{n}}$, analizujemy pojedyczną wartość, tzn. statystykę testową opartą na ACF dla kilku/ kilkudziesięciu początkowych opóźnień

1.1.3.1 Test Boxa-Pierce'a

W jego przypadku postać statystyki testowej można wyrazić jako : $Q_{BP} = n \sum_{j=1}^{h} \hat{p}^2(j)$ gdzie h ozn. pewne maksymalne opóźnienie, a $\hat{p}(j)$ to próbkowa autokorelacja dla opóźnienia j.

Duża wartość Q_{BP} oznacza, że wartości ACF są zbyt duże, aby uznać dane za realizację ciągu nieskorelowanego (białego szumu)

Wiadomo, że jeżeli $X_1, X_2, ..., X_n$ jest realizacją ciągu iid o skończonej wariancji, wówczas Q_{BP} ma w przybliżeniu rozkład chi-kwadrat z h stopniami swobody

Uzasadnienie : przy tych założeniach $\sqrt{n}\hat{\theta j}$, j = 1,...,h są w przybliżeniu niezaleznymi zm.losowymi o rozkładzie N(0,1). Statystyka Q_{BP} jest więc sumą kwadratów h niezależnych zm. losowych o rozkładzie N(0,1)

W praktyce aby stwierdzić, czy wartość statystyki Q_{BP} jest zbyt duża wykorzystujemy wskaźnik p-value dla ustalonego poziomu istotności (zazwyczaj: $\alpha = 0.05$)

Zbyt mała p-wartość przemawia, przeciwko przypuszczeniu o losowości(jeżeli p-value <0.05 to odrzucamy hipotezę o niezależności)

1.1.3.2 Test Ljungi-Boxa

W przypadku Kjungi-Boxa (L-B) statystykę testową Q_{BP} zastępujemy przez

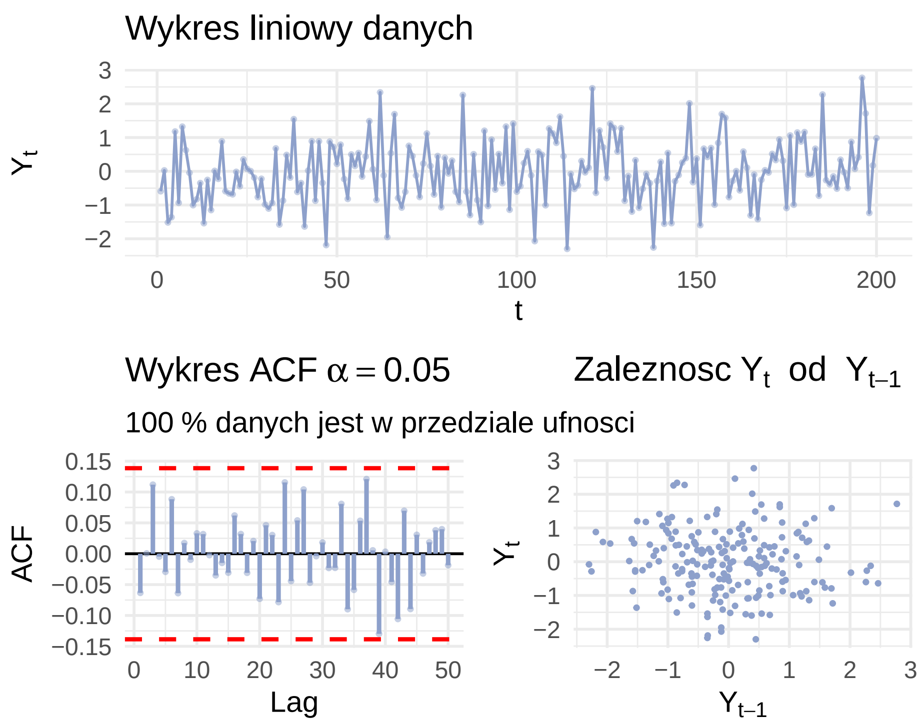
$Q_{LB} = n(n+2)\sum_{j=1}^{h}\frac{\hat{p}^2(j)}{n-j}$

Rozkład tego testu, jest lepiej przybliżony rozkładem granicznym od testu Boxa-Pierce'a i dlatego jest preferowany w praktyce

Oba powyższe testy zaimplementowane są w bibliotece Box.test(), będziemy z niej korzystać w tym sprawozdaniu

1.1.3.3 Zilustrowanie działania testów dla danych (n = 200) typu biały szum (rozkład $WN(0, \sigma^2)$)

Prezentacja metody graficznej :



Test graficzny opiera się na wygenerowaniu wykresu funkcji ACF, oraz sprawdzenia, czy 95% autokorelacji próbkowych znajduje się w przedziale ufności dla $\alpha = 0.05$.

Oraz nie ma obserwacji istotnie wychodzących poza przedziały ufności

Na wykresie przedział ufności zaznaczony jest czerwonymi przerywanymi liniami

98% autokorelacji leży w przedziale ufności, i nie ma żadnych przypadków znacznie odstających

Teraz zastosujemy testy formalne dla wygenerowanych danych

Test Boxa-Pierce'a dla przedstawionych danych zwraca wartość

```
## [1] TRUE
```

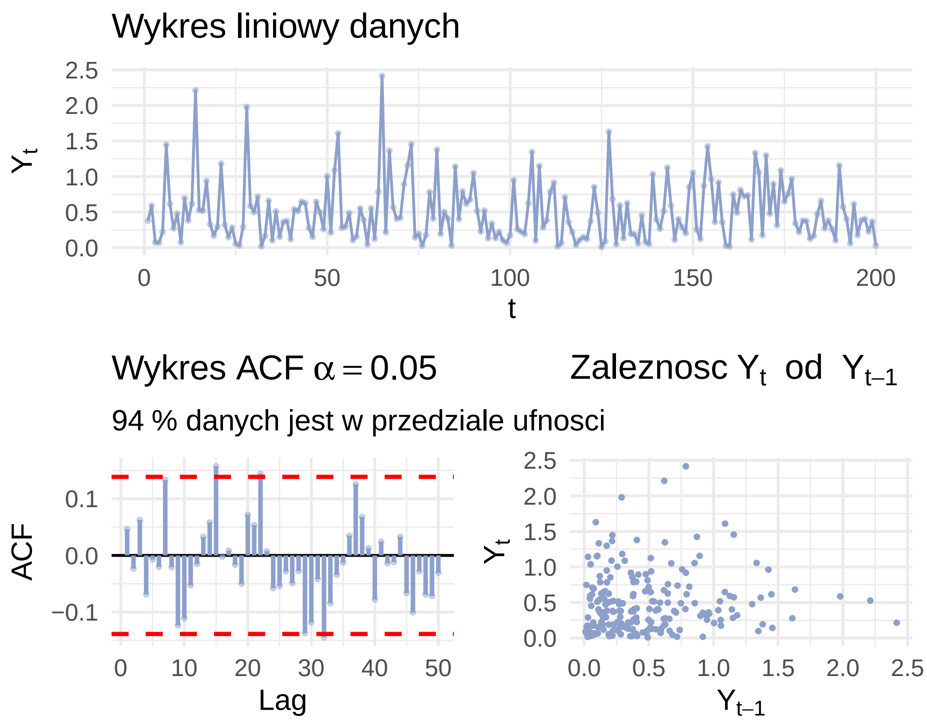
Według testu przedstawione dane są białym szumem

Test Ljunga-Boxa dla danych testowych zwraca wartość

```
## [1] TRUE
```

Według testu Ljunga-Boxa przedstawione dane są białym szumem

1.1.3.4 Zilustrowanie działania testów dla danych (n = 200) typu (rozkład $Exp(2)$)



94% autokorelacji leży w przedziale ufności, plus pojawiają się informacje odstające, według testu graficznego nie jest to biały szum

Pomimo faktu, że z definicji jest to biały szum o średniej $\frac{1}{2}$ i wariancji $\frac{1}{4}$

Teraz zastosujemy testy formalne dla wygenerowanych danych

Test Boxa-Pierce'a dla przedstawionych danych zwraca wartość

```
## [1] TRUE
```

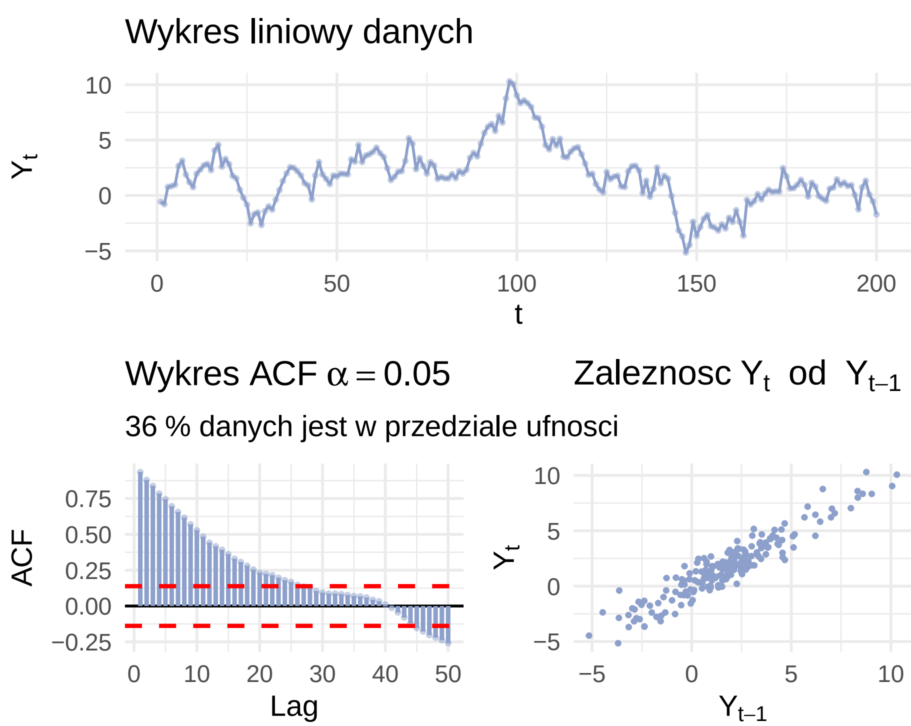
Według testu przedstawione dane są białym szumem

Test Ljunga-Boxa dla danych testowych zwraca wartość

```
## [1] TRUE
```

Według testu przedstawione dane są białym szumem

1.1.3.5 Zilustrowanie działania testów dla danych (n = 200) z autokorelacją (nie biały szum)



36% autokorelacji leży w przedziale ufności, więc według testu graficznego to nie jest biały szum

Test Boxa-Pierce'a dla przedstawionych danych zwraca wartość

```
## [1] FALSE
```

Według testu przedstawione dane nie są białym szumem

Test Ljunga-Boxa dla danych testowych zwraca wartość

```
## [1] FALSE
```

Według testu przedstawione dane nie są białym szumem

1.2 b)

1.2.1 Porównywanie testów graficznych oraz formalnych dla testowania biało-szumowości

Definicja 7 (Zautomatyzowany test graficzny ACF dla białego szumu)**.** *Niech* $\{X_t\}_{t=1}^n$ *będzie szeregiem czasowym, a* $\hat{\rho}(k)$ *jego empiryczną funkcją autokorelacji*

$$\hat{\rho}(k), \quad k = 0, 1, \ldots, L,$$

gdzie L *oznacza maksymalny analizowany lag (bez zera).*

Chcemy zweryfikować hipotezę

$$H_0 : \ \{X_t\} \text{ jest białym szumem.}$$

(1) Przybliżony rozkład ACF i przedział ufności. *Dla białego szumu przyjmujemy przybliżenie*

$$\hat{\rho}(k) \approx \mathcal{N}\Big(0, \frac{1}{n}\Big),$$

co prowadzi do 95% przedziału ufności

$$CI = \Big[-\frac{1.96}{\sqrt{n}}, \ \frac{1.96}{\sqrt{n}}\Big].$$

(2) Liczba lagów wychodzących poza przedział ufności. *Dla każdego laga* $k = 1, \ldots, L$ *definiujemy zmienną wskaźnikową*

$$I_k = \begin{cases} 1, & \text{gdy } |\hat{\rho}(k)| > \dfrac{1.96}{\sqrt{n}}, \\ 0, & \text{w przeciwnym razie.} \end{cases}$$

Liczbę lagów wychodzących poza przedział ufności opisujemy zmienną

$$X = \sum_{k=1}^{L} I_k.$$

Przyjmujemy przybliżony model

$$X \sim \mathrm{Bin}(L, 0.05),$$

co odpowiada założeniu, że pojedyncza autokorelacja ma prawdopodobieństwo ok. 5% *wyjścia poza 95% przedział ufności.*

Niech $x_{0.95}$ *będzie 95-tym percentylem rozkładu* $\mathrm{Bin}(L, 0.05)$*, tzn.*

$$P\Big(X \leq x_{0.95}\Big) = 0.95.$$

(3) Znacznie odstające autokorelacje. *Wprowadzamy dodatkowo pojęcie znacznie odstającej autokorelacji: lag k uznajemy za znacznie odstający, jeśli*

$$|\hat{\rho}(k)| > c \cdot \frac{1.96}{\sqrt{n}},$$

gdzie $c > 1$ jest stałą ustaloną heurystycznie; w praktyce przyjmujemy $c = 1.7$.

(4) Reguła decyzyjna zautomatyzowanego testu graficznego ACF.

Na podstawie powyższych elementów definiujemy zautomatyzowany test graficzny ACF*:*

- ***Akceptujemy*** *hipotezę H_0 („szereg jest zgodny z białym szumem”), jeśli spełnione są oba warunki:*
  1. *$X \leq x_{0.95}$, tzn. co najmniej 95% autokorelacji leży w przedziale ufności,*
  2. *nie występuje żadna znacznie odstająca autokorelacja, tzn.*

     $$|\hat{\rho}(k)| \leq 1.7 \cdot \frac{1.96}{\sqrt{n}} \quad \text{dla każdego } k = 1, \ldots, L.$$

- ***Odrzucamy*** *hipotezę H_0, jeśli:*
  1. *$X > x_{0.95}$ (zbyt wiele lagów wychodzi poza przedział ufności)* lub
  2. *istnieje lag k taki, że*

     $$|\hat{\rho}(k)| > 1.7 \cdot \frac{1.96}{\sqrt{n}},$$

     co interpretujemy jako obecność znacznie odstającej autokorelacji.

Test ten stanowi formalizację klasycznego „testu graficznego” ACF: wymaga, aby zdecydowana większość słupków autokorelacji mieściła się w przedziale ufności oraz aby nie występowały pojedyncze, bardzo duże piki, które wskazywałyby na istotną autokorelację.

W testach przyjujemy $\alpha = 0.05$

1.2.1.1 Porównanie pod względem liczby powtórzeń

1.2.1.1.1 Dla $WN(0, \sigma^2)$

```
##   graficzny_test test_pearsona test_ljunga
## 1          0.924         0.954        0.95
## 2          0.925         0.945        0.94
## 3          0.860         0.960        0.96
## 4          0.920         1.000        1.00
## 5          1.000         1.000        1.00
```

Tabela 1: Wyniki procentowe kwalifikacji danych jako biały szum

	500	200	50	25	5
graficzny_test	0.92	0.92	0.86	0.92	1
test_pearsona	0.95	0.94	0.96	1.00	1
test_ljunga	0.95	0.94	0.96	1.00	1

1.2.1.1.2 Dla #Exp(4)

```
##   graficzny_test test_pearsona test_ljunga
## 1          0.908          0.96       0.958
## 2          0.910          0.95       0.950
## 3          0.840          0.96       0.960
## 4          0.880          1.00       1.000
## 5          1.000          1.00       1.000
```

Tabela 2: Wyniki procentowe kwalifikacji danych jako biały szum

	500	200	50	25	5
graficzny_test	0.91	0.91	0.84	0.88	1
test_pearsona	0.96	0.95	0.96	1.00	1
test_ljunga	0.96	0.95	0.96	1.00	1

1.2.1.1.3 Dla danych skorelowanych

```
##   graficzny_test test_pearsona test_ljunga
## 1              0             0           0
## 2              0             0           0
## 3              0             0           0
## 4              0             0           0
## 5              0             0           0
```

Tabela 3: Wyniki procentowe kwalifikacji danych jako biały szum

	500	200	50	25	5
graficzny_test	0	0	0	0	0
test_pearsona	0	0	0	0	0

	500	200	50	25	5
test_ljunga	0	0	0	0	0