# Lista_3

Analiza przeżycia

Olga Foriasz

2025-10-15

## LISTA 3

### Zadanie 1a

Dane bierzeny z zadania 3 z litsy 2. Wyznaczymy oszacowania największej wiarogodności średniego czasu do remisji choroby dla pacjentów leczonych lekiem A oraz pacjentów leczonych lekiem B.

```
T_a = sum(grupa_A[1:10]) + t0*(n-R)
T_b = sum(grupa_B[1:10]) + t0*(n-R)

theta_a <- R/T_a
theta_b <- R/T_b
```

Średni czas do remisji dla leku A wynosi: 0.7034414, natomiast dla leku B: 0.7430155

### Zadanie 1b

Wyznaczamy realizację przedziału ufności na poziomie 1-$\alpha$ dla średniego czasu do remisji choroby osobno dla pacjentów leczonych lekiem A oraz lekiem B (dane z listy 2. z zadania 3.) Zakładamy, że są to zmienne z rozkładu wykładniczego. Za $\alpha$ przyjmujemy kolejno $\alpha$=0.05 oraz $\alpha$=0.01.

```
library(binom)

## dla alfy = 0.05
alfa=0.05
T.L <- binom.confint(R,n,conf.level=1-alfa, methods="exact")$lower
T.U <- binom.confint(R,n,conf.level=1-alfa, methods = "exact")$upper

TL1= (-log(1-T.L))/t0
TU1= (-log(1-T.U))/t0

TL1 = round(TL1,2)
TU1 = round(TU1,2)
## dla alfy = 0.01
alfa=0.01
T.L <- binom.confint(R,n,conf.level=1-alfa, methods="exact")$lower
T.U <- binom.confint(R,n,conf.level=1-alfa, methods = "exact")$upper
```

```
TL2= (-log(1-T.L))/t0
TU2= (-log(1-T.U))/t0

TL2 = round(TL2,2)
TU2 = round(TU2,2)
```

Realizacją przedziału ufności dla $\alpha = 0.05$ jest przedział $[T_L, T_U]$ =[0.32, 1.3], natomiast dla $\alpha =$ 0.01 przedział $[T_L, T_U]$ =[0.25, 1.52]

### Zadanie 2

Przyjmujemy, że obserwacje czasu do remisji choroby były prowadzone do momentu, w którym u dziesięciu pacjentów zostanie ona zaobserwowana. Dla takich danych wyznaczamy najpierw oszacowania największej wiarygodności średniego czasu do remisji choroby dla pacjentów leczonych lekiem A oraz pacjentów leczonych lekiem B.

```
m=10
n=10
T2_a <- round(m/(sum(grupa_A)+(n-m)*grupa_A[m]),2)
T2_b <- round(m/(sum(grupa_B)+(n-m)*grupa_B[m]),2)
```

Zatem z wyliczeń wynika, że oszacowania największej wiarygodności średniego czasu do remisji w grupie A wynoszą: 0.73 a w grupie B: 0.96.

Następnie dla danych wyznaczamy realizację przedziału ufności na poziomie 1-$\alpha$, dla odpowiednio dwóch wartości $\alpha$: 0.05 i 0.01.

```
m=10
n=10
## dla alfy = 0.05
alfa=0.05
qm_a1 = qgamma(alfa/2,m,m)
TLa <- round((m*qm_a1)/T2_a,2)
qm_a2 = qgamma((1-alfa/2),m,m)
TUa <-round((m*qm_a2)/T2_a,2)

qm_b1 = qgamma(alfa/2,m,m)
round(TLb <- (m*qm_a1)/T2_b,2)
```

```
## [1] 5
```

```
qm_b2 = qgamma((1-alfa/2),m,m)
TUb <-round((m*qm_b2)/T2_b,2)
```

Dla $\alpha$=0.05 realizacjami przedziałów ufności dla grupy A jest przedział [6.57, 23.4] a dla grupy B: [4.9951966, 17.8]. Szerokość przedziałów wynika z niewielkiej liczebności próby danych.

Następnie wyznaczamy realizację przedziału ufności dla $\alpha$=0.01:

```
m=10
n=10
```

```r
## dla alfy = 0.01
alfa=0.01
qm_a1 = qgamma(alfa/2,m,m)
TLa <- round((m*qm_a1)/T2_a,2)
qm_a2 = qgamma((1-alfa/2),m,m)
TUa <-round((m*qm_a2)/T2_a,2)

qm_b1 = qgamma(alfa/2,m,m)
TLb <- round((m*qm_a1)/T2_b,2)
qm_b2 = qgamma((1-alfa/2),m,m)
TUb <- round((m*qm_b2)/T2_b,2)
```

Stąd otrzymujemy, że realizacjami przedziałów ufności dla grupy A jest przedział [5.09, 27.4] a dla grupy B: [3.87, 20.83]

Realizacją przedziału ufności dla $\alpha = 0.05$ jest przedział $[T_L, T_U]$ =[**??**,**??**], natomiast dla $\alpha = 0.01$ przedział $[T_L, T_U]$ =[**??**,**??**]

### Zadanie 3

Przeprowadzono symulację w celu porównania dwóch estymatorów parametru $\vartheta$ rozkładu wykładniczego przy danych cenzurowanych I-go typu:

$$\hat{\vartheta} = \frac{R}{T_1} \tag{1}$$

$$\tilde{\vartheta} = -\frac{\log(1 - R/n)}{t_0} \tag{2}$$

gdzie $R = \sum_{i=1}^{n} 1(X_i \le t_0)$ oraz $T_1 = \sum_{i=1}^{R} X_{(i)} + t_0(n - R)$.

Symulację wykonano dla $\vartheta = 1$, liczności prób $n = 10, 30$ oraz czasów obserwacji $t_0 = 0.5, 1, 2$, powtarzając eksperyment $M = 10{,}000$ razy. Dla każdego przypadku obliczono obciążenie (Bias) oraz błąd średniokwadratowy (MSE) obu estymatorów.

```r
library(knitr)

typ_I <- function(t0,theta,n){
  X<-numeric(n)
  delta <-numeric(n)
  check <- 0

  while (check == 0){
    X<-numeric(n)
    delta <-numeric(n)
    X=rexp(n, 1/theta)
    X <- sort(X)
    for(i in 1:n){
      if(X[i]>t0){
```

```r
        X[i] <- t0
        delta[i] <- 0
        check <- 1
      }else{delta[i]<-1}}
  }
  return(data.frame(X = X, delta = delta))
  X <- sort(X)
}
symulacja <- function(theta,n,t0,M){
  est1 <- numeric(M)
  est2 <- numeric(M)
  for (i in (1:M)){
    dane <- typ_I(t0,theta,n)
    R <- sum(dane$delta)
    T1 <- sum(dane$X)
    est1[i] <- R/T1
    est2[i] <- -log(1-R/n)/t0
  }
  bias1 <- est1-theta
  bias2 <- est2-theta
  mse1 <- bias1^2
  mse2 <- bias2^2
  bias_c1 <- mean(bias1)
  bias_c2 <- mean(bias2)
  mse_c1 <- mean(mse1)
  mse_c2 <- mean(mse2)
  return(c(bias_c1, bias_c2, mse_c1, mse_c2))
}
#symulacja(1,10,1,10000)


n_list <- c(10,30)
t0_list <- c(0.5,1,2)
wyniki <- list(6)
i <- 1
for (n in n_list){
  for (t0 in t0_list){
    wyniki[[i]] <- symulacja(1,n,t0,10000)
    i = i+1
  }
}

wyniki_df <- data.frame(
  n = rep(n_list, each = length(t0_list)),
  t0 = rep(t0_list, times = length(n_list)),
  Bias_Est1 = round(sapply(wyniki, function(x) x[1]), 4),
  Bias_Est2 = round(sapply(wyniki, function(x) x[2]), 4),
```

```
  MSE_Est1 = round(sapply(wyniki, function(x) x[3]), 4),
  MSE_Est2 = round(sapply(wyniki, function(x) x[4]), 4)
)

kable(wyniki_df,
      booktabs = TRUE,
      col.names = c("n", "t0", "Bias Est1", "Bias Est2", "MSE Est1", "MSE Est2"),
      caption = "Wyniki symulacji dla różnych wartości n i t0")
```

Tabela 1: Wyniki symulacji dla różnych wartości n i t0

| n | t0 | Bias Est1 | Bias Est2 | MSE Est1 | MSE Est2 |
|---|---|---|---|---|---|
| 10 | 0.5 | 0.0584 | 0.0680 | 0.2987 | 0.3233 |
| 10 | 1.0 | 0.0606 | 0.0847 | 0.1890 | 0.2273 |
| 10 | 2.0 | -0.0465 | -0.0765 | 0.0793 | 0.0604 |
| 30 | 0.5 | 0.0178 | 0.0203 | 0.0892 | 0.0928 |
| 30 | 1.0 | 0.0226 | 0.0317 | 0.0571 | 0.0675 |
| 30 | 2.0 | 0.0181 | 0.0478 | 0.0392 | 0.0670 |

Z przeprowadzonej analizy wynika, że wraz ze wzrostem liczności próby $n$ obciążenie i MSE obu estymatorów maleją, co potwierdza ich zgodność. Większe wartości $t_0$ (czyli mniejsze cenzurowanie) prowadzą do dokładniejszych oszacowań. W większości przypadków estymator $\hat{\vartheta} = R/T_1$ charakteryzuje się mniejszym błędem średniokwadratowym, co wskazuje na jego większą efektywność w porównaniu z estymatorem $\tilde{\vartheta} = -\log(1 - R/n)/t_0$.