

# Sprawozdanie I

## Analiza przeżycia

Olga Foriasz, Tomasz Warzecha

2025-10-08

### Spis treści

<b>1 LISTA 1</b>	<b>1</b>
1.1 Zadanie 1 - Deklaracje . . . . .	1
1.2 Zadanie 2 - Wykresy dla wybranych parametrów rozkładu . . . . .	2
1.3 Zadanie 3 - Program do generowania zmiennych . . . . .	3
1.4 Zadanie 4 - Wykresy dla ustalonych parametrów i n=50, n=100 . . . . .	3
1.5 Zadanie 5 - Podstawowe statystyki opisowe . . . . .	4
<b>2 LISTA 2</b>	<b>6</b>
2.1 Zadanie 1 - Generowanie n zmiennych cenzurowanych . . . . .	6
2.2 Zadanie 2 - Generowanie danych cenzurowanych . . . . .	8
2.3 zadanie 3 - Zadanie praktyczne . . . . .	9
<b>3 LISTA 3</b>	<b>9</b>
3.1 Zadanie 1a - Oszacowania największej wiarygodności . . . . .	9
3.2 Zadanie 1b - Realizacje przedziału ufności . . . . .	10
3.3 Zadanie 2 - Oszacowania największej wiarygodności i realizacja przedziału ufności dla danych z listy 2. zad.3 . . . . .	10
3.4 Zadanie 3 - Porównanie dokładności estymatorów . . . . .	11
<b>4 LISTA 4</b>	<b>12</b>
4.1 Zadanie 1 - Funkcja ilorazu wiarygodności . . . . .	12
4.2 Zadanie 2 - Oszacowanie testu mocy za pomocą symulacji . . . . .	13
4.3 Zadanie 3 - Zadanie praktyczne . . . . .	14

### 1 LISTA 1

#### 1.1 Zadanie 1 - Deklaracje

Poniżej kolejno deklarowane są funkcje gęstości, dystrybuanty, dystrybuanty odwrotnej (funkcji kwantylowej oraz funkcji hazardu rozkładu  $EW(\alpha, \beta, \gamma)$ )

- Do wyznaczenia gęstości funkcji użyto wzoru:

$$f(x; \alpha, \beta, \gamma) = \frac{\alpha\gamma}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \left[1 - \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right)\right]^{\gamma-1} \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right). \quad (1)$$

```
gestosc <- function(x, alfa, beta, gamma) {
  (alfa * gamma / beta) *
  (x / beta)^(alfa - 1) *
  (1 - exp(-(x / beta)^alfa))^(gamma - 1) *
  exp(-(x / beta)^alfa) * (x > 0)
}
```

- Do wyznaczenia dystrybuanty funkcji zastosowano wzór:

$$F(x, \alpha, \beta, \gamma) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right] \quad (2)$$

```
dystrybuanta <- function(x, alfa, beta){
  return((1-exp(-(x/beta)^alfa))^gamma)
}
```

- Do wyznaczenia funkcji kwantylowej wzięto wzór:

$$Q(p) = \beta * (-\ln(1 - p^{\frac{1}{\gamma}}))^{\frac{1}{\alpha}} \quad (3)$$

```
funkcja_kwantylowa <- function(p, alfa, beta, gamma){
  return(beta*(-log((1-p^(1/gamma))^(1/alfa))))
}
```

- Do wyznaczenia funkcji hazardu wykorzystano wzór:

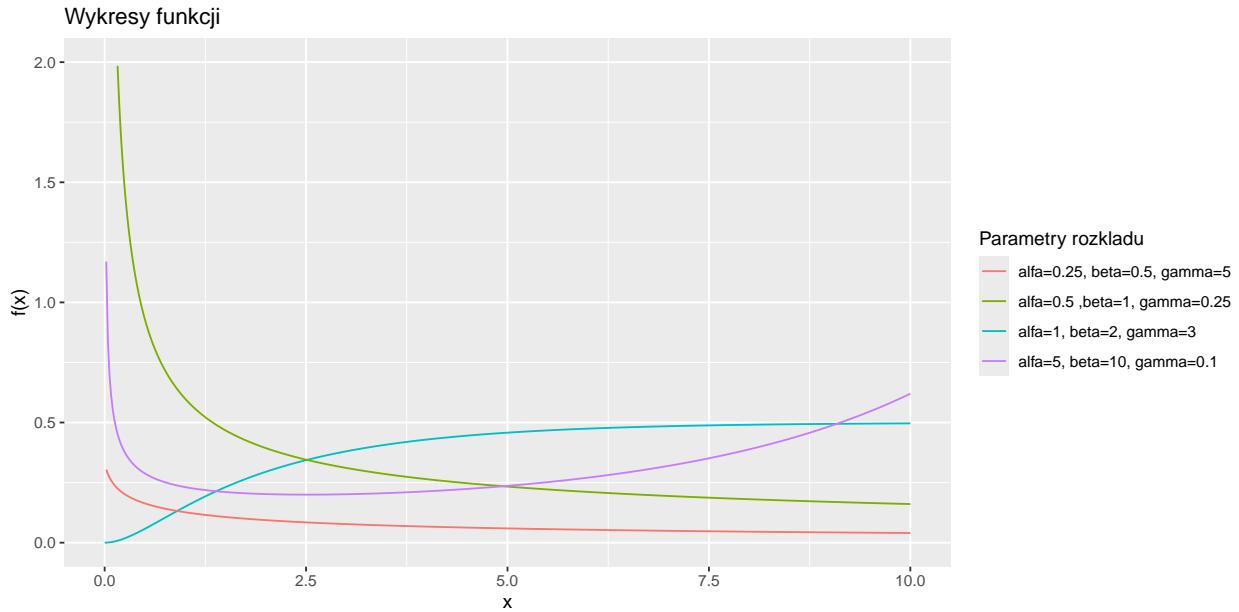
$$h(x) = \frac{f(x)}{1 - F(x)} \quad (4)$$

gdzie  $f(x)$  jest funkcją gęstości rozkładu  $WE(\alpha, \beta, \gamma)$  oraz  $F(x)$  jest dystrybuantą tego rozkładu.

```
funkcja_hazardu <- function(x, alfa, beta, gamma){
  return(gestosc(x, alfa, beta, gamma)/(1-dystrybuanta(x, alfa, beta, gamma)))
```

## 1.2 Zadanie 2 - Wykresy dla wybranych parametrów rozkładu

Poniżej przedstawiono na wykresie funkcję hazardu rozkładu  $WE(\alpha, \beta, \gamma)$  dla różnych parametrów:



Jak możemy zauważyć na wykresie, funkcja hazardu zmienia swoje zachowanie w zależności od przyjętych parametrów. W niektórych przypadkach jest funkcją rosnącą, w innych malejącą. Odpowiedni dobór parametrów pozwala kontrolować tempo wzrostu lub spadku wartości funkcji. Dzięki temu możemy modelować różne typy rozkładów hazardu, od szybko malejących po powoli rosnące, co jest istotne w analizie ryzyka i modelowaniu procesów stochastycznych.

### 1.3 Zadanie 3 - Program do generowania zmiennych

Do generowania zmiennych z rozkładu  $WE(\alpha, \beta, \gamma)$  na początku wykorzystujemy funkcje runif generującą liczby z rozkładu jednostajnego na odcinku  $(0,1)$ . Następnie używając wcześniej zdefiniowanej funkcji kwantylowej (wzór (3)) do wyznaczenia zmiennych z rozkładu  $WE(\alpha, \beta, \gamma)$ .

```
# ile zmiennych chcemy wylosować
u=10

# losujemy z U(0,1)
p <- runif(u)

#funkcja kwantylowa
Q <- function(p, alfa, beta, gamma){
  return(beta*(-log((1-p^(1/gamma))^(1/alfa))))
}

#wywołanie funkcji z parametrami alfa=1, beta=1, gamma=1
#Q(p, 1, 1, 1)
```

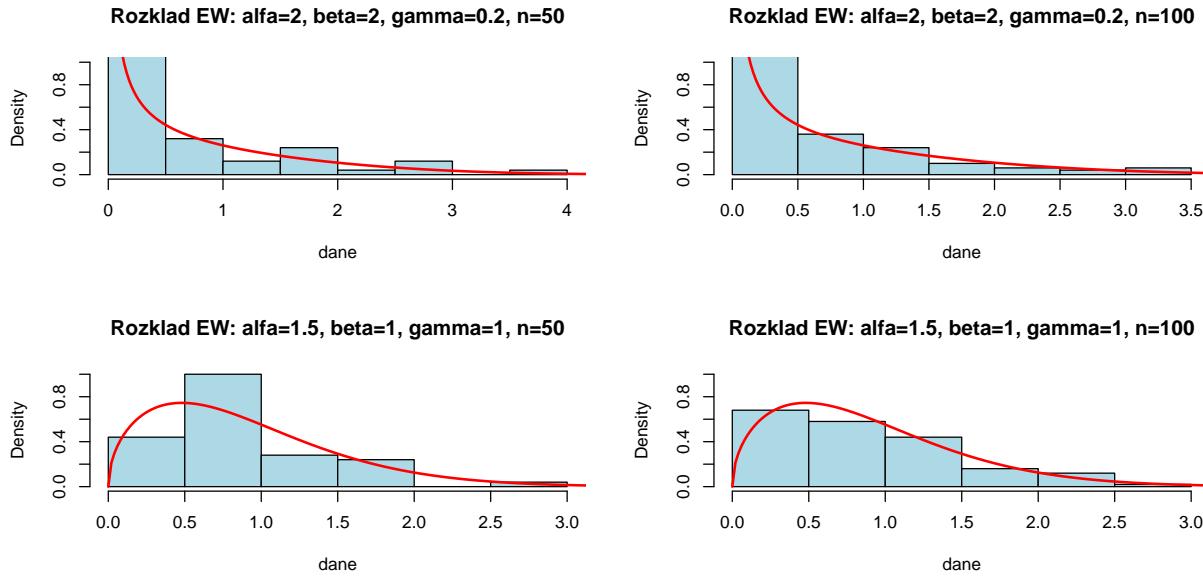
### 1.4 Zadanie 4 - Wykresy dla ustalonych parametrów i n=50, n=100

Wygenerowane wykresy zostały dla parametrów:

- W pierwszym przypadku:  $\alpha = 2$ ,  $\beta = 2$  oraz  $\gamma = 0.2$ .

- W drugim przypadku:  $\alpha = 1.5$ ,  $\beta = 1$  oraz  $\gamma = 1$ .

Dla każdej trójki parametrów wykresy wygenerowano dla n= 50 oraz n=100.



Widać, że dla pierwszej trójki parametrów ( $\alpha = 2, \beta = 2, \gamma = 0.2$ ) funkcja gęstości ma charakter „wannowy” ( $\alpha > 1, \alpha\gamma < 1$ ), natomiast dla drugiej trójki parametrów ( $\alpha = 0.5, \beta = 1, \gamma = 1$ ) funkcja gęstości wykazuje rosnącą intensywność ( $\alpha > 1, \alpha\gamma > 1$ ).

Można również zauważać wpływ liczności próby na dopasowanie histogramu do funkcji gęstości. Dla większego n, np. n=100, histogram w przybliżeniu odwzorowuje kształt funkcji gęstości, choć nie pokrywa się z nią idealnie. Przy mniejszym n, np. n=50, rozbieżności między histogramem a wykresem gęstości stają się bardziej wyraźne.

## 1.5 Zadanie 5 - Podstawowe statystyki opisowe

Wyznaczamy kolejne statystyki takie jak średnia, mediana, odchylenie standardowe, minimum i maksimum, rozstęp, kwartyl dolny i kwartyl górny.

```
x = c(4,5,5,6,7,7)

statystyki <- function(x){
  ## liczenie sredniej
  suma <- 0
  for (el in x){
    suma <- suma + el
  }
  srednia <- suma/length(x)

  ##liczenie mediany
  if (length(x)%2 == 0) {
    mediana <- (x[length(x)/2]+x[(length(x)/2)+1])/2
  } else { mediana <- x[floor(length(x)/2) + 1]}
```

```

##liczenie odchylenia standardowego
suma <- 0
for (el in x){
  suma <- suma + (el-srednia)^2
}
odch <- (suma/(length(x)-1))^(1/2)

##liczenie minimum
minimum <- x[1]
for(el in x){
  if(el < minimum){
    minimum <- el
  }
}

##liczenie maksimum
maksimum <- x[1]
for(el in x){
  if(el > maksimum){
    maksimum <- el
  }
}

##liczenie rozstępu
rozstep = maksimum - minimum

##liczenie kwartylu dolnego
kwartyl_d <- quantile(x, 0.25)

##liczenie kwartylu górnego
kwartyl_g <- quantile(x, 0.75)

## zwracanie odpowiedzi
wyn <- c(srednia, mediana, odch, minimum, maksimum, rozstep, kwartyl_d, kwartyl_g)
return(wyn)
}

```

% latex table generated in R 4.4.1 by xtable 1.8-4 package % Mon Nov 3 21:31:16 2025

Tabela 1: Podstawowe statystyki opisowe

	średnia	mediana	odch_stand	min	maks	rozstęp	kw_dolny	kw_górny
dane_1.1	0.69027	0.00368	0.94123	0.00001	4.82407	4.82406	0.08964	0.94719
dane_1.2	0.70931	3.41392	0.94386	0.00008	4.04296	4.04288	0.05552	0.99856
dane_2.1	0.87369	0.87913	0.63480	0.00601	2.90601	2.90000	0.39583	1.21352
dane_2.2	0.93366	0.57929	0.60475	0.06427	2.73254	2.66827	0.50623	1.38712

Gdzie poprzez:

**dane\_1.1** oznaczono statystyki opisowe dla danych wygenerowanych z rozkładu  $WE(\alpha = 2, \beta =$

$2, \gamma=0.2)$  dla  $n=50$ .

**dane\_1.2** oznaczono statystyki opisowe dla danych wygenerowanych z rozkładu  $WE(\alpha = 2, \beta = 2, \gamma=0.2)$  dla  $n=100$ .

**dane\_2.1** oznaczono statystyki opisowe dla danych wygenerowanych z rozkładu  $WE(\alpha = 0.5, \beta = 1, \gamma=1)$  dla  $n=50$ .

**dane\_2.2** oznaczono statystyki opisowe dla danych wygenerowanych z rozkładu  $WE(\alpha = 0.5, \beta = 1, \gamma=1)$  dla  $n=100$ .

% latex table generated in R 4.4.1 by xtable 1.8-4 package % Mon Nov 3 21:31:16 2025

Tabela 2: Teoretyczne podstawowe statystyki opisowe

	medianą	kwartyl dolny	kwartyl górny
dane_1	0.35636	0.06252	1.04095
dane_2	0.78322	0.43579	1.24328

Gdzie:

**dane\_1** oznaczają teoretyczne wartości statystyk opisowych dla danych wygenerowanych z rozkładu  $WE(\alpha = 2, \beta = 2, \gamma=0.2)$ .

**dane\_2** oznaczają teoretyczne wartości statystyk opisowych dla danych wygenerowanych z rozkładu  $WE(\alpha = 0.5, \beta = 1, \gamma=1)$ .

Jak możemy zauważyć, wartości teoretyczne mediany oraz kwartyle dolnego i górnego odbiegają od wyników przedstawionych w Tabeli 1. Są one zbliżone, jednak nie pokrywają się dokładnie. Dla parametrów  $\alpha = 2, \beta = 2, \gamma = 0.2$ , wartości kwartyle dolnego i górnego wraz ze wzrostem liczebności próby ( $n$ ) zbliżają się do wartości teoretycznych, jednak mediana znacząco od nich odbiega. Możemy to rozumieć, jako fakt, że mediana może być bardziej wrażliwa na zmienność próby niż kwartyle w przypadku tych parametrów.

Dla parametrów  $\alpha = 0.5, \beta = 1, \gamma = 1$ , wartości statystyk obliczonych z próby są bliskie wartościom teoretycznym dla  $n = 50$ . Przy większych liczebnościach próby różnice te się zwiększały. Dla tych parametrów powiększanie próby nie gwarantuje lepszego dopasowania mediany do wartości teoretycznej.

## 2 LISTA 2

### 2.1 Zadanie 1 - Generowanie n zmiennych cenzurowanych

Do wygenerowania  $n$  zmiennych cenzurowanych z uogólnionego rozkładu wykładniczego  $GE(\lambda, \alpha)$  I-go typu wykorzystaliśmy generowanie zmiennych losowych z rozkładu jednostajnego  $U(0,1)$ . Następnie dla każdej zmiennej zastosowano wzór odwróconej dystrybuanty, a kolejno dla  $X[i]$  większych od  $t_0$  zamieniono je na wartość  $t_0$ , i zaznaczono cenzurowanie poprzez  $\delta = 0$ .

```
## wzor : F(t)=(1-e)^(-lambda*t)
## t= -(1/lambda)*ln(1-p^(1/alfa))

typ_I <- function(t0, alfa, lambda, n){
  p <- runif(n)
  X<-numeric(n)
  delta <- numeric(n)
```

```

for(i in 1:n){
  X[i]=(-(1/lambda))*log(1-(p[i])^(1/alfa))
  if(X[i]>t0){
    X[i] <- t0
    delta[i] <- 0
  }else{delta[i]<-1}
  return(data.frame(X = X, delta = delta))
}

```

Aby wygenerować n zmiennych cenzurowanych z uogólnionego rozkładu wykładniczego  $\text{GE}(\lambda, \alpha)$  typu II, zastosowaliśmy metodę dystrybuanty odwrotnej. Następnie posortowaliśmy wektor X, a od wartości  $X_m$  każdą kolejną obserwację zastąpiliśmy właśnie tą wartością. Wektorowi cenzurowania  $\delta$  przypisaliśmy odpowiednio: 1 dla obserwacji niecenzurowanych oraz 0 dla cenzurowanych.

```

typ_II <- function(m, alfa, lambda, n){
  p <- runif(n)
  X<-numeric(n)
  delta <-numeric(n)
  for(i in 1:n){
    X[i]=(-(1/lambda))*log(1-(p[i])^(1/alfa))}
  sort(X)
  for (i in (n-m):n){
    X[i] <- X[m]
    delta[i] <- 0
  }
  for(i in 1:m){delta[i]=1}
  return(data.frame(X = X, delta = delta))
}

```

Do wygenerowania n zmiennych z rozkładu  $\text{GE}(\lambda, \alpha)$  przy cenzurowaniu losowym (niezależnym) wygenerowano n liczb losowych z rozkładu  $U(0,1)$  i zastosowano metodę odwróconej dystrybuanty. Następnie wygenerowano wektor C z rozkładu wykładniczego  $E(\eta)$  oraz przypisano wskaźnik cenzurowania  $\delta=1$  dla obserwacji niecenzurowanych i 0 dla cenzurowanych.

```

cenz_losowo <- function(eta, alfa, lambda, n){
  p <- runif(n)
  X<-numeric(n)
  C<-numeric(n)
  delta <-numeric(n)
  for(i in 1:n){
    X[i]=(-(1/lambda))*log(1-(p[i])^(1/alfa))
    C[i]=rexp(1, rate=1/eta)
    X[i]=min(X[i],C[i])
    if(X[i]== C[i]){
      delta[i]=0
    }else{delta[i]=1}
  }
  return(data.frame(X = X, delta = delta))
}

```

```
}
```

## 2.2 Zadanie 2 - Generowanie danych cenzurowanych

Korzystając z napisanych funkcji wygenerowaliśmy zestaw 40 danych dla cenzurowania:

- typu I-go:

```
dane_1 <- typ_I(1,0.5,1,40)
head(dane_1,5)
```

```
##          X delta
## 1 0.05792357    1
## 2 0.63727549    1
## 3 0.05234012    1
## 4 0.10696046    1
## 5 0.03073798    1
```

- typu II-go:

```
dane_2 <- typ_II(10,0.5,1,40)
head(dane_2,5)
```

```
##          X delta
## 1 0.434859145    1
## 2 1.016440062    1
## 3 1.557177254    1
## 4 0.160606129    1
## 5 0.008191538    1
```

- losowego:

```
dane_3 <- cenz_losowo(0.23,0.5,1,40)
head(dane_3,5)
```

```
##          X delta
## 1 0.31459948    0
## 2 0.10022689    0
## 3 0.11156800    0
## 4 0.09690865    0
## 5 0.09769169    0
```

Następnie dla danych wyznaczyliśmy statystyki opisowe przedstawione w Tabeli 3. Poprzez **dane\_1** oznaczono dane zawierające zmienne w przypadku cenzurowania I-go typu. **Dane\_2** - zawierają zmienne w przypadku cenzurowania II-go typu, a **dane\_3** - w przypadku cenzurowania losowego(niezależnego).

Tabela 3: Statystyki opisowe

dane	min	dolny_kwartyl	gorny_kwartyl	mediana	maks	rozstep	liczba cenz	odsetek
dane_1	0.00048	0.07290	1.00000	0.32617	1.00000	0.99952	11.00000	0.27500
dane_2	0.00318	0.23262	0.78963	0.47431	2.08292	2.07973	30.00000	0.75000
dane_3	0.00001	0.02477	0.20876	0.09331	0.46374	0.46372	23.00000	0.57500

Przy statystykach opisowych dla danych cenzurowanych wzięliśmy pod uwagę wartości takich statystyk jak minimum, maksimum, rozstęp, kwartyl górny i dolny, medianę, liczbę danych cenzurowanych oraz ich odsetek. Dzięki temu możliwa jest szybka ocena stopnia cenzurowania oraz porównanie podstawowych charakterystyk rozkładu między różnymi zestawami danych. Należy jednak pamiętać, że uzyskane statystyki klasyczne (np. mediana, kwartyle) mogą być zniekształcone przez obecność danych cenzurowanych, dlatego ich interpretacja powinna być ostrożna. Mimo to funkcja stanowi użyteczne narzędzie do wstępnego opisu i porównania zbiorów danych zawierających obserwacje niepełne.

### 2.3 zadanie 3 - Zadanie praktyczne

Tabela 4: Statystyki opisowe

dane	min	dolny_kwartyl	gorny_kwartyl	mediana	maks	rozstep	liczba cenz	odsetek
grupa_A	0.03346	0.37529	1.00000	0.97111	1.00000	0.96654	10.00000	0.50000
grupa_B	0.03789	0.33291	1.00000	0.84560	1.00000	0.96211	10.00000	0.50000

W przypadku obu grup odsetek danych cenzurowanych jest taki sam - w zestawach A i B występuje po 10 obserwacji cenzurowanych oraz 10 niecenzurowanych. Patrząc na wartości minimalne, można zauważyc, że czas do osiągnięcia remisji w grupie A był nieznacznie krótszy niż w grupie B. Dane cenzurowane mają wartość  $t_0 = 1$ , ponieważ obserwacja trwała przez rok, dlatego w obu przypadkach maksimum jest jednakowe i równe 1. Mediana w grupie A wynosi 0,97, a w grupie B - 0,85. Dolne kwartyle dla grup A i B wynoszą odpowiednio 0,38 i 0,33, natomiast górny kwartyl w obu przypadkach wynosi 1. Wyniki te można interpretować w ten sposób, że choć minimalny czas wystąpienia remisji był krótszy w grupie A, to większa liczba pacjentów osiągnęła remisję szybciej w grupie B. Można więc wnioskować, że mimo szybszych początkowych efektów leku w grupie A, ostatecznie lek B okazał się skuteczniejszy.

## 3 LISTA 3

### 3.1 Zadanie 1a - Oszacowania największej wiarygodności

Dane bierzeny z zadania 3 z litsy 2. Wyznaczmy 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

### 3.2 Zadanie 1b - Realizacje przedziału ufności

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  $[T_L, T_U] = [0.25, 1.52]$

### 3.3 Zadanie 2 - Oszacowania największej wiarygodności i realizacja przedziału ufności dla danych z listy 2. zad.3

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=20
theta_a <- round(m/(sum(grupa_A[1:m])+(n-m)*grupa_A[m]),2)
theta_b <- round(m/(sum(grupa_B[1:m])+(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=20
Ta<-sum(grupa_A[1:m])+(n-m)*grupa_A[m]
Tb<-sum(grupa_B[1:m])+(n-m)*grupa_B[m]
## dla alfy = 0.05
alfa=0.05
qm_a1 = qgamma(alfa/2,m,rate=m)
TLa <- round((m*qm_a1)/Ta,2)
qm_a2 = qgamma((1-alfa/2),m,rate=m)
TUa <-round((m*qm_a2)/Ta,2)

qm_b1 = qgamma(alfa/2,m,rate=m)
TLb <- round((m*qm_b1)/Tb,2)
qm_b2 = qgamma((1-alfa/2),m,rate=m)
TUb <-round((m*qm_b2)/Tb,2)

```

Dla  $\alpha=0.05$  realizacjami przedziałów ufności dla grupy A jest przedział [0.35, 1.25] a dla grupy B: [0.46, 1.65].

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

```

m=10
n=20
## dla alfy = 0.01
alfa=0.01
qm_a1 = qgamma(alfa/2,m,m)
TLa <- round((m*qm_a1)/Ta,2)
qm_a2 = qgamma((1-alfa/2),m,m)
TUa <-round((m*qm_a2)/Ta,2)

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

```

Stąd otrzymujemy, że realizacjami przedziałów ufności dla grupy A jest przedział [0.27, 1.47] a dla grupy B: [0.36, 1.93].

### 3.4 Zadanie 3 - Porównanie dokładności estymatorów

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

$$\hat{\vartheta} = \frac{R}{T_1} \quad (5)$$

$$\tilde{\vartheta} = -\frac{\log(1 - R/n)}{t_0} \quad (6)$$

gdzie  $R = \sum_{i=1}^n 1(X_i \leq t_0)$  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.

Tabela 5: Wyniki symulacji dla różnych wartości n i t0

n	$t_0$	Bias $\hat{\vartheta}$	Bias $\tilde{\vartheta}$	MSE $\hat{\vartheta}$	MSE $\tilde{\vartheta}$
10	0.5	0.0669	0.0789	0.3035	0.3347
10	1.0	0.0615	0.0864	0.1900	0.2278
10	2.0	-0.0474	-0.0739	0.0791	0.0608
30	0.5	0.0166	0.0186	0.0875	0.0900
30	1.0	0.0236	0.0331	0.0580	0.0674
30	2.0	0.0248	0.0570	0.0409	0.0699

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  $\hat{\vartheta} = \frac{R}{T_1}$  charakteryzuje się mniejszym błędem średniokwadratowym, co wskazuje na jego większą efektywność w porównaniu z estymatorem  $\tilde{\vartheta} = \frac{-\log(1-R/n)}{t_0}$ .

## 4 LISTA 4

### 4.1 Zadanie 1 - Funkcja ilorazu wiarygodności

Zadanie zaczynamy od wygenerowania danych cenzurowanych I-go typu, które są realizacjami zmiennych losowych z rozkładu wykładniczego.

```
generowanie_1 <- function(n,lambda){
  x <- rexp(n,lambda)
  x <- sort(x)
  t0 <- mean(x)
  delta <- numeric(n)
  delta[x<=t0]=1
  x[x>t0]=t0
  dane <- data.frame(x,delta)
  names(dane)<-c("czas","dane")
  return(dane)
}

##generowanie_1(10,1/2)
```

Następnie piszemy funkcję wyliczającą funkcję wiarygodności:

```
L <- function(x, t0, r, theta) {
  n <- length(x$czas)
  return (factorial(n) / factorial(n - r) * theta^r *
```

```

        exp(-theta * (sum(x$czas[1:r]) + t0 * (n - r)))
}

```

Korzystając z wcześniejszych funkcji deklarujemy następną, w której wartością zwracaną będzie wartość poziomu krytycznego w teście ilorazu wiarygodności do testowania hipotez dwustronnych (type="two.sided"), jednostronnych prawostronnie (type="greater"), jednostronnych lewostronnie (type="less"):

```

IW <- function(x,t0,r,theta0,type){
  n <- length(x$czas)
  T1 <- sum(x$czas[1:r])+t0*(n-r)
  t <- r/T1
  mianownik <- L(x,t0, r, t)
  if(type=="two.sided"){
    licznik = L(x,t0, r, theta0)
  }
  if(type=="greater"){
    if(mianownik <= theta0){
      licznik=mianownik
    }else{licznik=L(x,t0, r, theta0)}
  }
  if(type=="less"){
    if(mianownik>theta0){
      licznik=mianownik
    }else{licznik=L(x,t0, r, theta0)}
  }
  lambda = licznik/mianownik
  G=-2*log(lambda)
  return(1-pchisq(G,1))
}

```

## 4.2 Zadanie 2 - Oszacowanie testu mocy za pomocą symulacji

W tym zadaniu przeprowadzono symulacje mające na celu oszacowanie mocy testu ilorazu wiarygodności oraz jego rozmiaru dla wybranych wartości parametrów.

Założono, że badane dane pochodzą z rozkładu wykładniczego, a cenzurowanie jest typu I przy ustalonym czasie  $t\_0 = 0.5$ . Hipoteza zerowa dotyczyła wartości parametru  $\theta_0 = 3$ , natomiast alternatywy — dziesięć różnych wartości parametru  $\theta$ : 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5. Dla każdej z tych wartości przeprowadzono  $M = 1000$  replikacji testu.

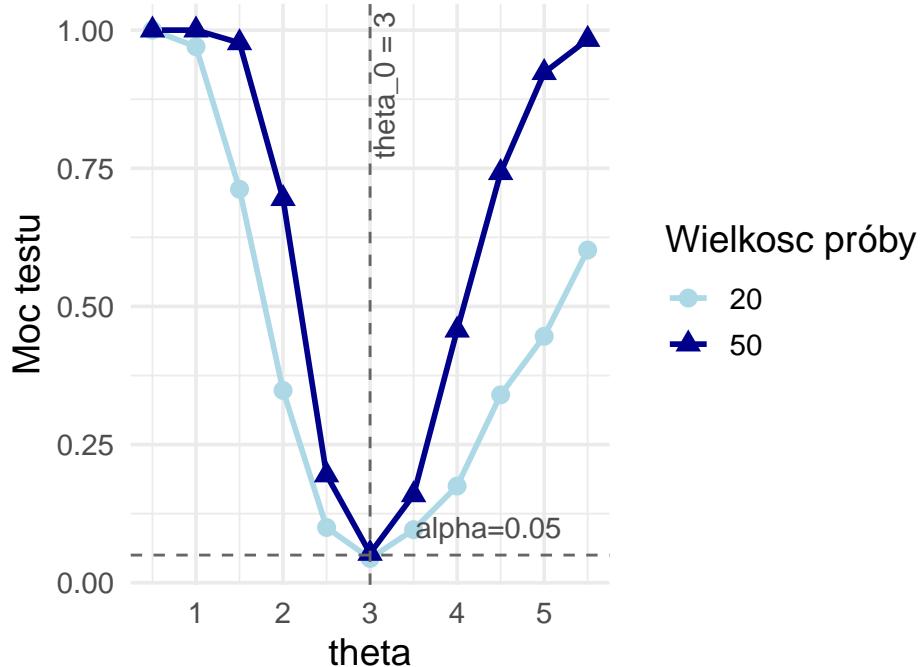
Symulacje wykonano dla dwóch liczebności prób:  $n = 20$  oraz  $n = 50$ .

Na podstawie otrzymanych wartości p-value wyznaczono estymator mocy testu oraz rozmiaru testu.

Na wykresie poniżej przedstawiono krzywe mocy testu dla obu wielkości prób.

Dodatkowo zaznaczono:  
- przerywaną linię poziomą odpowiadającą poziomowi istotności  $\alpha = 0.05$ ,  
- przerywaną linię pionową dla  $\theta_0 = 3$

## Krzywe mocy testu ilorazu wiarygodnosci



Można zaobserwować, że wraz z oddalaniem się od naszego  $\theta_0$  moc testu rośnie, a dla  $\theta = \theta_0$  empiryczny rozmiar testu jest zbliżony do nominalnego poziomu  $\alpha$  i wynosi dla  $n = 20$ : 0.039 oraz dla  $n = 50$ : 0.048. Dodatkowo możemy zauważać, że wraz ze wzrostem liczebności próby moc testu również rośnie.

### 4.3 Zadanie 3 - Zadanie praktyczne

Zakładamy, że  $\alpha=0.05$ . Korzystając z wcześniej napisanej funkcji IW, liczymy p-value dla danych pochodzących z listy 2. z zadania 3. Weryfikujemy hipotezę, że średni czas do remisji w grupie A i B wynosi 1.

```
p1 <- round(IW(dane_A, 1, 10, 1, "two.sided"), 2)
p2 <- round(IW(dane_B, 1, 10, 1, "two.sided"), 2)
```

Otrzymaliśmy wyniki  $p=0.24$  dla grupy A oraz  $p=0.32$  dla grupy B. Przy założeniu ustalonej  $\alpha > 0.05$  dla obu przypadków. Zatem nie ma podstaw do odrzucenia hipotezy ani w grupie A, ani w grupie B.