

Raport 1

Analiza Przeżycia

Dominik Kowalczyk 282229 i Matylda Mordal 282240

2025-11-12

Spis treści

1	Lista 1	2
1.1	Zadanie 1	2
1.2	Zadanie 2	3
1.3	Zadanie 3	4
1.4	Zadanie 4	5
1.5	Zadanie 5	8
2	Lista 2	11
2.1	Zadanie 1	11
2.2	Zadanie 2	13
2.3	Zadanie 3	15
3	Lista 3	17
3.1	Zadanie 1	17
3.2	Zadanie 2	18
3.3	Zadanie 3	20
4	Lista 4	23
4.1	Zadanie 1	23
4.2	Zadanie 2	24
4.3	Zadanie 3	26
5	Zadania dodatkowe	28
5.1	Zadanie 1 z Listy 1	28
5.2	Zadanie 1 z Listy 3	28

1 Lista 1

1.1 Zadanie 1

Celem zadania jest implementacja (deklaracja) czterech oddzielnych funkcji w pakiecie R definiujących funkcję gęstości, dystrybuantę, funkcję kwantylową (dystrybuantę odwrotną) oraz funkcję hazardu rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$.

Funkcja gęstości

Wzór na funkcja gęstości rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$ wyraża się wzorem:

$$f(t; \alpha, \beta, \gamma) = \frac{\alpha\gamma}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} \left\{1 - \exp\left[-\left(\frac{t}{\beta}\right)^\alpha\right]\right\}^{\gamma-1} \exp\left[-\left(\frac{t}{\beta}\right)^\alpha\right] \mathbf{1}_{(0,\infty)}(t),$$

gdzie $(\alpha, \beta, \gamma > 0)$ są parametrami.

```
#Funkcja gęstości dla EW
funkcja_gestosci <- function(t, alfa, beta, gamma){
  if (any(t <= 0)) stop("t musi być dodatnie") #indykator
  wyrazenie1 <- ((alfa*gamma)/beta)*((t/beta)^(alfa-1))
  wyrazenie2 <- ((1-exp(-(t/beta)^alfa)))^(gamma-1)*exp(-(t/beta)^alfa)
  wzor_gestosc <- wyrazenie1*wyrazenie2
  return (wzor_gestosc)}
```

Dystrybuanta

Dystrybuanta rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$ wyraża się wzorem:

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

gdzie $(\alpha, \beta, \gamma > 0)$ są parametrami.

```
#Dystrybuanta dla rozkładu EW
dystrybuanta <- function(t, alfa, beta, gamma){
  wzor <- (1-exp(-(t/beta)^alfa))^gamma
  return(wzor)}
```

Funkcja kwantylowa

Funkcja kwantylowa (dystrybuanta odwrotna) rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$ wyraża się wzorem:

Aby wyznaczyć funkcję kwantylową $Q(p)$ rozwiązujemy równanie:

$$p = \left\{1 - \exp\left[-\left(\frac{Q(p)}{\beta}\right)^\alpha\right]\right\}^\gamma.$$

Po przekształceniu powyższego wzoru względem $Q(p)$ otrzymujemy:

$$Q(p) = \beta \left[-\ln(1 - p^{1/\gamma}) \right]^{1/\alpha},$$

gdzie $(\alpha, \beta, \gamma > 0)$ są parametrami, a p - kwantylem.

```
#Funkcja kwantylowa dla rozkładu EW
funkcja_kwantylowa <- function(p, alfa, beta, gamma){
  if (any(p<=0) | (p>=1)) stop("p - kwantyl ma wartości z przedziału [0,1]")
  Q <- beta*(-log(1-p^(1/gamma)))^(1/alfa)
  return(Q)}
```

Funkcja hazardu

Mając wcześniej zdefiniowane oraz zaimplementowane wzory na gęstość oraz dystrybuantę otrzymujemy, że funkcja hazardu rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$ wyraża się wzorem:

$$h(t) = \frac{f(t)}{(1 - F(t))}.$$

```
#Funkcja hazardu dla rozkładu EW
funkcja_hazardu <- function(t, alfa, beta, gamma){
  f_t <- funkcja_gestosci(t, alfa, beta, gamma)
  S_t <- 1 - dystrybuanta(t, alfa, beta, gamma)
  h_t <- f_t/S_t
  return(h_t)}
```

1.2 Zadanie 2

Celem zadania jest narysowanie wykresów funkcji hazardu dla wybranych parametrów rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$ i zilustrowanie i opisanie różnic pomiędzy każdym z nich.

Rozkłady i parametry

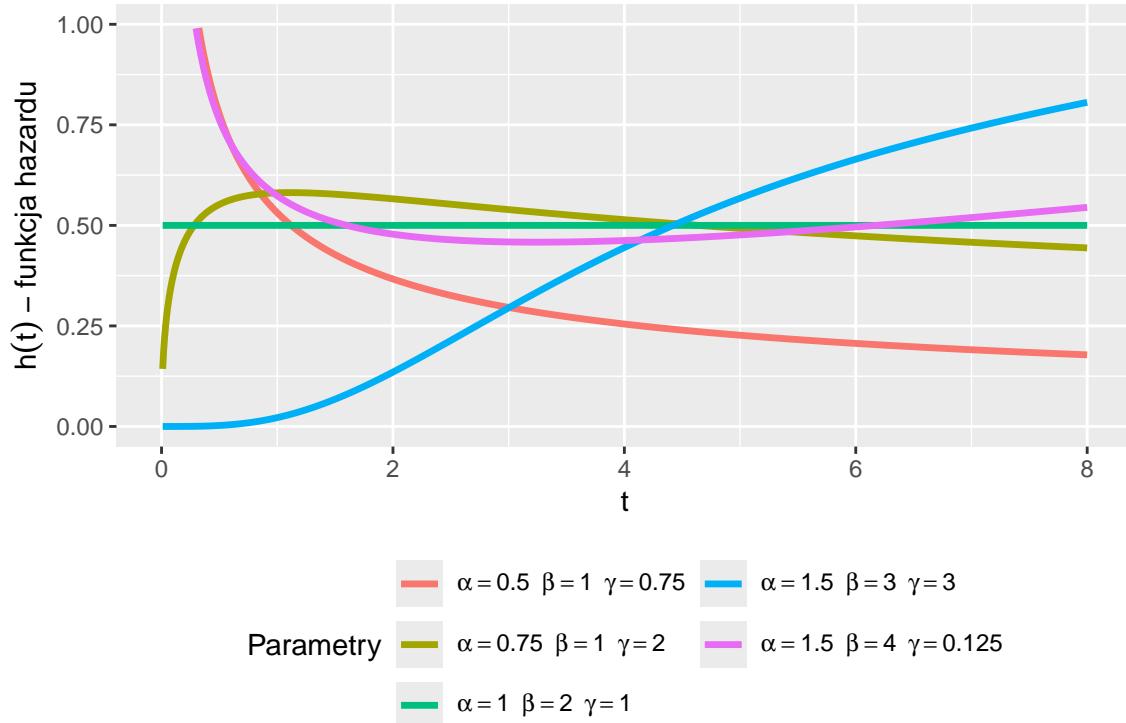
1. $\mathcal{EW}(0.5, 1, 0.75)$
2. $\mathcal{EW}(1, 2, 1)$
3. $\mathcal{EW}(1.5, 3, 3)$
4. $\mathcal{EW}(1.5, 4, 0.125)$
5. $\mathcal{EW}(0.75, 1, 2)$

Kod

Kod do tego zadania jest stosunkowo prosty. Generujemy 1000 równomiernie rozmieszczonych wartości z pewnego przedziału (w naszym przypadku $[0.01, 8]$) dla których obliczamy

funkcję hazardu rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$ przy różnych zestawach parametrów. Następnie wyniki wizualizujemy za pomocą biblioteki ggplot2.

Wizualizacja



Rysunek 1: Wykres funkcji hazardu dla różnych parametrów rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$

Interpretacja

Na rysunku 1 obserwujemy zróżnicowane przebiegi funkcji hazardu w zależności od wartości parametrów. Dla niskich wartości α i $\gamma > 1$ (funkcja czerwona) funkcja hazardu maleje (DFR) na początku i wzrasta dalej. Zdecydowanie rosnącą funkcję hazardu (IFR) jest krzywa przedstawiona dla parametrów o dużych wartościach $\alpha, \beta, \gamma > 1$ (funkcja niebieska). Dla $\alpha = \gamma = 1$ (funkcja zielona) wykres jest funkcją liniową, prostą na stałej wartości funkcji hazardu. W przypadku bardzo małego $\gamma > 0$ (funkcja żółta) hazard przyjmuje kształt wannowy – wysoki na początku, następnie malejący, a w końcowej fazie wzrasta. Natomiast funkcja przyjmująca $\alpha < 1$ oraz $\gamma > 1$ przyjmuje kształt jednomodalny – z początku osiąga swoje maksimum, a następnie maleje wraz z wzrastającym czasem.

1.3 Zadanie 3

Celem zadania jest napisanie programu (w pakiecie R) do generowania zmiennych losowych z rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$.

Kod

Pamiętając o tym, że parametry rozkładu muszą być dodatnie, oraz n (liczba generowanych próbek) musi być dodatnia i całkowita, generujemy n niezależnych zmiennych losowych z rozkładu jednostajnego i poprzez zastosowanie funkcji kwantylowej otrzymujemy zmienne losowe pożądanego przez nas rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$.

#Zadanie 3

```
funkcja_generacji <- function(n, alfa, beta, gamma){  
  if (n <= 0 || n != round(n)) stop("n - liczba dodatnia i całkowita")  
  p <- runif(n)  
  Q <- beta*(-log(1-p^(1/gamma)))^(1/alfa)  
  return(Q)  
}
```

1.4 Zadanie 4

Celem zadania jest wygenerowanie dla wybranych dwóch trójków parametrów wykresów gęstości rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$ i histogramy uzyskane w oparciu o wygenerowane dane. Punkty generujemy dla wielkości $n = 50$ oraz $n = 100$ za pomocą funkcji napisanej w zadaniu 1.3.

Rozkłady i parametry

Weźmy dwa wybrane (dowolne) rozkłady zaprezentowane w zadaniu 1.2:

1. $\mathcal{EW}(0.75, 1, 2)$
2. $\mathcal{EW}(1.5, 1, 3)$

Kod

#Zadanie 4

```
#parametry  
pA <- c(0.75, 1, 2)  
pB <- c(1.5, 1, 3)  
rozmiary <- c(50, 100)  
wszystkie_parametry <- list(pA, pB)  
  
par(mfrow = c(2, 2))  
dane_zadanie4 <- list()  
  
for (i in 1:length(wszystkie_parametry)){  
  parametr <- wszystkie_parametry[[i]]
```

```

#przypisanie standardowe zmiennych
alfa <- parametr[1]
beta <- parametr[2]
gamma <- parametr[3]

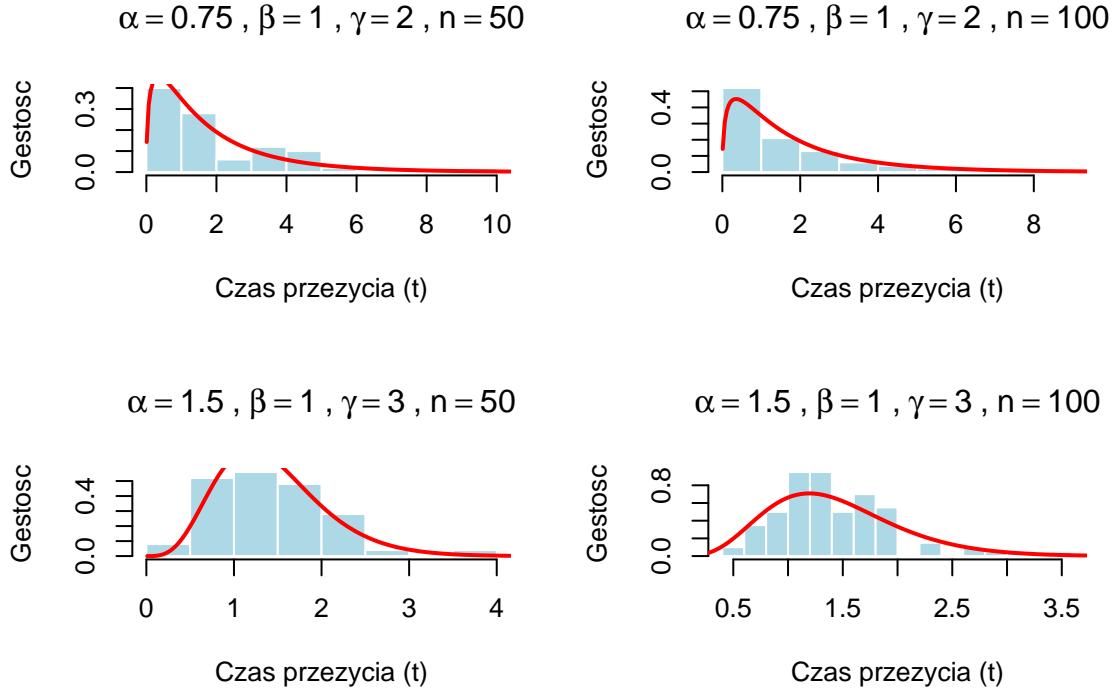
#generowanie probek
for (n in rozmiary){
  probki <- funkcja_generacji(n, alfa, beta, gamma)
  nazwa_grupy <- ifelse(i == 1, "pA", "pB")
  dane_zadanie4[[paste0(nazwa_grupy, n)]] <- list(
    n = n,
    alfa = alfa,
    beta = beta,
    gamma = gamma,
    dane = probki
  )
}

#wizualizacja histogramem
print(hist(probki,
            breaks = 12,
            freq = FALSE, #normalizacja
            main = bquote(alpha == .(alfa) ~ "," ~ beta == .(beta) ~
                          ",," ~ gamma == .(gamma) ~ "," ~ n == .(n))
            ,
            xlab = "Czas przeżycia (t)",
            ylab = "Gęstość",
            col = "lightblue",
            border = "white"))

#teoretyczna gestosc (na stworzonym histogramie)
t_przedzialy <- seq(0.01, max(probki) * 1.2, length.out = 200)
lines(t_przedzialy, funkcja_gestosci(t_przedzialy, alfa, beta, gamma),
      col = "red",
      lwd = 2)
}

par(mfrow = c(1,1))

```



Rysunek 2: Histogramy oraz funkcje gęstości dla różnych parametrów rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$

Interpretacja

Na rysunku 2 przedstawiono histogramy oraz gęstości rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$ dla dwóch wybranych trójkąt parametrów oraz dwóch liczebności $n = 50$ i $n = 100$.

1. $\mathcal{EW}(0.75, 1, 2)$

- Dla małej wartości parametru $\alpha < 1$ rozkład przyjmuje mocno skośny kształt z dużą koncentracją obserwacji przy małych wartościach czasu przeżycia oraz długim "ogonem" po prawej stronie.
- Widoczne jest osiągnięcie maksimum gęstości dla małego t , a następnie stopniowy spadek funkcji
- Zwiększenie liczebności prób z $n = 50$ do $n = 100$ powoduje, że histogram lepiej odwzorowuje kształt gęstości rozkładu.

2. $\mathcal{EW}(1.5, 1, 3)$

- Dla wartości parametru $\alpha > 1$ rozkład jest mniej skośny, a jego maksimum jest bardziej przesunięte w prawo niż dla $\mathcal{EW}(0.75, 1, 2)$.
- Zwiększenie liczebności prób z $n = 50$ do $n = 100$ powoduje, że histogram lepiej odwzorowuje kształt.

Dodatkowo, z analizy rysunku 2 oraz rysunku 1 możemy wyciągnąć pewne wnioski co do wpływu α , β oraz γ na kształt funkcji. Parametr α odpowiada za występowanie trendu,

stopniowi skośności wykresu, parametr γ wpływa na rozciąganie lub ścislenie kształtu funkcji, a β można określić jako parametr skali.

1.5 Zadanie 5

Celem zadania jest wyznaczenie podstawowych statystyk opisowych dla danych wygenerowanych w zadaniu 1.4. Dla tych danych należy również wyznaczyć teoretyczne wartości mediany oraz kwartyli. Wszystkie wyniki zamieścić w odpowiednich tabelach.

Rozkłady i parametry

1. pA - próbka A reprezentująca rozkład $\mathcal{EW}(0.75, 1, 2)$ - rosnący hazard
2. pB - próbka B reprezentująca rozkład $\mathcal{EW}(1.5, 1, 3)$ - jednomodalny hazard

Wszystkie dane i rozkłady są takie same jak w zadaniu 1.4

1.5.1 Kod

```
#Zadanie 5
Wyniki_Statystyczne <- data.frame()

for (nazwa in names(dane_zadanie4)) {
  przypadek <- dane_zadanie4[[nazwa]]

  n <- przypadek$n
  probka <- przypadek$dane
  alfa <- przypadek$alfa
  beta <- przypadek$beta
  gamma <- przypadek$gamma
  nazwa_grupy_czysta <- gsub("[0-9]+", "", nazwa)

  #statystyki empiryczne
  kwantyle_emp <- quantile(probka, probs = c(0, 0.25, 0.50, 0.75, 1))

  #statystyki teoretyczne
  mediana_teor <- funkcja_kwantylowa(0.50, alfa, beta, gamma)
  q1_teor <- funkcja_kwantylowa(0.25, alfa, beta, gamma)
  q3_teor <- funkcja_kwantylowa(0.75, alfa, beta, gamma)

  #tworzymy nowy wiersz do tabeli
  nowy_wiersz <- data.frame(
    "Nazwa Próbki" = nazwa_grupy_czysta,
    N = n,
    Min = kwantyle_emp[1],
    "Q1 (Emp.)" = kwantyle_emp[2],
```

```

    "Q2 (Emp.)" = kwantyle_emp[3],
    "Q3 (Emp.)" = kwantyle_emp[4],
    Max = kwantyle_emp[5],
    "Rozstęp" = diff(range(probka)),
    "Średnia" = mean(probka),
    "Odch. Std." = sd(probka),
    "Q1 (Teor.)" = q1_teor,
    "Q2 (Teor.)" = mediana_teor,
    "Q3 (Teor.)" = q3_teor,
    check.names = FALSE
)
}

Wyniki_Statystyczne <- rbind(Wyniki_Statystyczne, nowy_wiersz)
}

Wyniki_Statystyczne[, 3:ncol(Wyniki_Statystyczne)] <-
  round(Wyniki_Statystyczne[, 3:ncol(Wyniki_Statystyczne)], 3)

```

Tabela 1: Wybrane statystyki opisowe (z danych) rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$ dla próbek A i B

Statystyki Empiryczne (z próbek)						
Nazwa Próbki	N	Min	Max	Rozstęp	Średnia	Odch. Std.
pA	50	0.032	9.490	9.458	1.859	1.823
pA	100	0.011	8.431	8.420	1.510	1.608
pB	50	0.311	3.702	3.391	1.410	0.680
pB	100	0.449	3.523	3.075	1.417	0.549

Tabela 2: Porównanie kwantyli teoretycznych (z wzoru) i empirycznych (z danych) rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$ dla próbek A i B

Statystyki Porównawcze (Kwantyle)							
Nazwa Próbki	N	Medianą		Kwantyl Q1		Kwantyl Q3	
		Q2 (Emp.)	Q2 (Teor.)	Q1 (Emp.)	Q1 (Teor.)	Q3 (Emp.)	Q3 (Teor.)
pA	50	1.425	1.315	0.462	0.613	2.861	2.537
pA	100	0.929	1.315	0.490	0.613	2.144	2.537
pB	50	1.278	1.356	0.855	0.996	1.859	1.789
pB	100	1.351	1.356	1.026	0.996	1.659	1.789

1.5.2 Interpretacja

W tabeli 1 i tabeli 2 zestawiono wybrane statystyki opisowe (średnia, odchylenie standardowe, rozstęp, wartości skrajne) dla dwóch zestawów parametrów rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$ oznaczonych jako pA oraz pB wygenerowych dla $n = 50$ i $n = 100$.

Zestawione statystyki wskazują, że próbki pA i pB dobrze odwzorowują własności rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$. Dla próbki pA obserwuje się większą zmienność i szerszy rozstęp wartości, co odpowiada przypadkowi rosnącego hazardu, natomiast próbka pB (jednomodalny hazard) cechuje się mniejszym odchyleniem standardowym i bardziej skupionym rozkładem.

Porównanie kwantyli empirycznych i teoretycznych pokazuje ich zgodność, co potwierdza poprawność generacji danych i stabilność rozkładu.

2 Lista 2

2.1 Zadanie 1

Celem zadania było napisanie funkcji w języku R do generowania danych z uogólnionego rozkładu wykładniczego $\mathcal{GE}(\lambda, \alpha)$ z trzema typami cenzurowania:

- **I-go typu** – obserwacja do ustalonego czasu t_0 ,
- **II-go typu** – obserwacja do momentu uzyskania m niecenzurowanych danych,
- **Losowe (niezależne)** – czas cenzurowania pochodzi z rozkładu wykładniczego $\mathcal{E}(\eta)$.

Wzory użyte w zadaniu

1. Dystrybuanta rozkładu:

$$F(x) = 1 - (1 - e^{-\lambda x})^\alpha, \quad x > 0$$

2. Funkcja kwantylowa (metoda odwrotnej dystrybuanty):

$$Q(u) = -\frac{1}{\lambda} \ln(1 - u^{1/\alpha}), \quad u \sim U(0, 1)$$

3. Wskaźnik cenzurowania:

$$\delta_i = \begin{cases} 1, & \text{gdy dane niecenzurowane,} \\ 0, & \text{gdy dane cenzurowane.} \end{cases}$$

Na podstawie przedstawionych wzorów funkcja `funkcja_kwantylowa` generuje wartości z rozkładu $\mathcal{GE}(\lambda, \alpha)$, wykorzystując metodę dystrybuanty odwrotnej. Kod implementacji zaprezentowano poniżej.

```
# Funkcja kwantylowa - metoda dystrybuanty odwrotnej
funkcja_kwantylowa <- function(n, lambda, alpha){
  u <- runif(n)
  x <- -(1/lambda) * log(1-u^(1/ alpha))
  return(x)
}
```

Cenzurowanie I-go typu

Funkcja `Ityp` realizuje generowanie danych z rozkładu $\mathcal{GE}(\lambda, \alpha)$ z cenzurowaniem I-go typu, czyli obserwacja trwa do ustalonego czasu t_0 . Zgodnie z teorią, dla każdej obserwacji czas T_i jest określony jako:

$$T_i = \begin{cases} X_i, & \text{gdy } X_i \leq t_0, \\ t_0, & \text{gdy } X_i > t_0, \end{cases}$$

gdzie X_i to rzeczywisty czas życia, a t_0 to czas obserwacji (cenzurowania).

Wskaźnik cenzurowania δ_i przyjmuje wartość:

$$\delta_i = \begin{cases} 1, & \text{gdy dane niecenzurowane } (X_i \leq t_0), \\ 0, & \text{gdy dane cenzurowane } (X_i > t_0). \end{cases}$$

Na podstawie powyższych zależności poniższa funkcja generuje obserwowane czasy oraz wskaźniki cenzurowania dla próby n elementów:

```
# Generowanie danych z cenzurowaniem I-go typu
Ityp <- function (n, alfa, lambda, t0){
  x <- funkcja_kwantylowa(n, lambda, alpha)
  t <- pmin(x, t0)
  delta <- as.integer(x <= t0)
  data.frame(Czas = t, Cenzura = delta)
}
```

Cenzurowanie II-go typu

Funkcja $IItyp$ realizuje generowanie danych z rozkładu $\mathcal{GE}(\lambda, \alpha)$ z cenzurowaniem II-go typu. W tym przypadku eksperyment kończy się w momencie zaobserwowania m zdarzeń, gdzie m jest ustalone z góry. Dane cenzurowane II-go typu składają się z m najmniejszych statystyk pozycyjnych $X_{(1)}, X_{(2)}, \dots, X_{(m)}$, natomiast pozostałe $n - m$ obserwacji przyjmują wartość $X_{(m)}$, co odpowiada czasowi zakończenia eksperymentu.

Obserwowany czas T_i dla próby n elementów jest zatem określony jako:

$$T_i = \begin{cases} X_{(i)}, & \text{dla } i \leq m, \\ X_{(m)}, & \text{dla } i > m, \end{cases}$$

Wskaźnik cenzurowania δ_i przyjmuje wartość:

$$\delta_i = \begin{cases} 1, & \text{gdy dane niecenzurowane } (i \leq m), \\ 0, & \text{gdy dane cenzurowane } (i > m). \end{cases}$$

Poniższa funkcja, bazując na przedstawionych zasadach, konstruuje obserwowane czasy i przypisuje im wskaźniki cenzurowania.

```
# Generowanie danych z cenzurowaniem II-go typu
IItyp <- function (m, alpha, lambda, n){
  if (m > n) stop ("m nie moze byc wieksze od n")
  x <- funkcja_kwantylowa(n, lambda, alpha)
  x_sorted <- sort(x)
  niecenzurowane <- x_sorted[1:m]
  cenzurowane <- rep(x_sorted[m], n - m)
```

```

t <- c(nieczurowane, czurowane)
delta <- c(rep(1, m), rep(0, n - m))
data.frame(Czas = t, Cenzura = delta)
}

```

Cenzurowanie losowe

Funkcja *losowe* generuje dane z rozkładu $\mathcal{GE}(\lambda, \alpha)$ z losowym cenzurowaniem. W tym przypadku każdej zmiennej X_i odpowiada zmienna cenzurująca C_i , która jest generowana z rozkładu wykładniczego $\mathcal{E}(\eta)$. Obserwowany czas T_i jest zdefiniowany jako:

$$T_i = \min(X_i, C_i),$$

natomiast wskaźnik cenzurowania δ_i przyjmuje wartość:

$$\delta_i = \begin{cases} 1, & \text{gdy } X_i \leq C_i, \\ 0, & \text{gdy } X_i > C_i. \end{cases}$$

Poniższa implementacja generuje obserwowane czasy oraz wskaźniki cenzurowania dla próby n elementów:

```

rExp <- function(n, eta) {
  rexp(n, rate = 1 / eta)
}

# Generowanie danych z losowym cenzurowaniem
losowe <- function(n, lambda, alpha, eta) {
  x <- funkcja_kwantylowa(n, lambda, alpha)
  c <- rExp(n, eta)
  t <- pmin(x, c)
  delta <- as.integer(x <= c)
  data.frame(Czas = t, Cenzura = delta)
}

```

2.2 Zadanie 2

Korzystając z zaimplementowanych funkcji: *Ityp*, *Ityp* oraz *losowe*, wygenerowano dane z rozkładu $\mathcal{GE}(\lambda, \alpha)$ dla trzech typów cenzurowania. Parametry symulacji przedstawiono w tabeli 3.

Tabela 3: Parametry symulacji danych cenzurowanych

n	λ	α	m	t_0	η
200	0.5	1.5	140	2	4

Na podstawie wygenerowanych danych obliczono podstawowe statystyki opisowe za pomocą funkcji *opisowe_statystyki*.

```
# Funkcja do statystyk opisowych
opisowe_statystyki <- function(data, type_label) {
  kompletne <- data %>% filter(Cenzura == 1)
  q <- quantile(data$Czas, probs = c(0.25, 0.75))
  wartosci_iqr <- IQR(data$Czas)
  data.frame(
    Typ = type_label,
    "Liczba obserwacji" = nrow(data),
    "Liczba kompletnych" = nrow(kompletne),
    Min = round(min(data$Czas), 3),
    Max = round(max(data$Czas), 3),
    Mediana = round(median(data$Czas), 3),
    "Rozstęp" = round(max(data$Czas), 3) - round(min(data$Czas), 3),
    "1d Qu." = round(q[1], 3),
    "3d Qu." = round(q[2], 3),
    IQR = round(wartosci_iqr, 3),
    check.names = FALSE
  )
}
```

Tabela 4: Statystyki opisowe dla danych cenzurowanych wygenerowanych z rozkładu $\mathcal{GE}(\lambda, \alpha)$ dla trzech typów cenzurowania

Typ	Liczba obserwacji	Liczba kompletnych	Min	Max	Mediana	Rozstęp	1d Qu.	3d Qu.	IQR
I-go typu	200	186	0.005	2.000	0.636	1.995	0.363	1.119	0.756
II-go typu	200	140	0.069	3.094	1.851	3.025	0.961	3.094	2.133
Losowe	200	119	0.003	9.841	1.283	9.838	0.576	2.367	1.791

Na podstawie danych przedstawionych w tabeli 4 można zauważyc istotne różnice pomiędzy trzema typami cenzurowania.

Dla danych I-go typu liczba obserwacji kompletnych jest największa (186), co oznacza, że wpływ cenzurowania jest stosunkowo niewielki. Dane te charakteryzują się najmniejszym rozrzutem (rozstęp 1.955, $IQR = 0.756$) oraz niskimi wartościami miar położenia: mediana wynosi 0.636, a kwartyle 0.363 oraz 1.119, co świadczy o przewadze krótkich czasów obserwacji.

W przypadku cenzurowania II-go typu liczba danych kompletnych spada do 140, co wskazuje na silniejsze oddziaływanie mechanizmu cenzurowania. Zwiększa się zarówno wartości maksymalne (3.094), jak i miary rozproszenia (rozstęp 3.025, $IQR = 2.133$), a mediana

(1.851) jest wyraźnie wyższa, co potwierdza większą zmienność i przesunięcie rozkładu w kierunku większych wartości.

Najśilniejsze cenzurowanie występuje w przypadku losowym, gdzie liczba obserwacji kompletnych wynosi jedynie 119. Wartości maksymalne osiągają najwyższy poziom (9.841), co przekłada się na bardzo duży rozrzut danych (rozstęp 9.838, $IQR = 1.791$). Mediana (1.283) znajduje się pomiędzy wartościami uzyskanymi dla cenzurowania I-go i II-go typu, co odzwierciedla większe rozproszenie danych wynikające z losowości procesu cenzurowania.

Podsumowując, wraz ze zmianą typu cenzurowania rośnie zróżnicowanie danych, przy czym największe jest w przypadku losowego cenzurowania. Cenzurowanie I-go typu daje najbardziej skompresowane dane z najmniejszym rozstępem i największą liczbą kompletnych obserwacji, natomiast II-go typu charakteryzuje się wyraźnym wzrostem mediany i miar rozproszenia, co świadczy o większej zmienności. Losowe cenzurowanie jest najbardziej nieprzewidywalne, co potwierdza duży rozstęp oraz najmniejsza liczba danych kompletnych.

2.3 Zadanie 3

Poniższy kod przygotowuje dane, zawierającą czasy do remisji, wskaźniki cenzurowania oraz informację o grupie (lek A lub lek B), co stanowi podstawę do dalszej analizy.

```
# Dane
czas_A <- c(0.03345514, 0.08656403, 0.08799947, 0.24385821, 0.27755032,
            0.40787247, 0.58825664, 0.64125620, 0.90679161, 0.94222208,
            rep(1, 10))
cenzura_A <- c(rep(1, 10), rep(0, 10))

czas_B <- c(0.03788958, 0.12207257, 0.20319983, 0.24474299, 0.30492413,
            0.34224462, 0.42950144, 0.44484582, 0.63805066, 0.69119721,
            rep(1, 10))
cenzura_B <- c(rep(1, 10), rep(0, 10))

Czas <- c(czas_A, czas_B)
Cenzura <- c(cenzura_A, cenzura_B)
Grupa <- factor(c(rep("Lek A", 20), rep("Lek B", 20)))

Dane_chorobowe <- data.frame(Czas, Cenzura, Grupa)
Dane_A <- subset(Dane_chorobowe, Grupa == "Lek A")
Dane_B <- subset(Dane_chorobowe, Grupa == "Lek B")
```

Tabela 5: Statystyki opisowe dla całej grupy pacjentów

Liczba obserwacji	Liczba kompletnych	Min	Max	Median	Rozstęp	1d Qu.	3d Qu.	IQR
40	20	0.033	1	0.971	0.967	0.333	1	0.667

Tabela 6: Statystyki opisowe dla grupy pacjentów leczonych lekiem A

Min	Max	Median	Rozstęp	1d Qu.	3d Qu.	IQR
0.033	1	0.971	0.967	0.375	1	0.625

Tabela 7: Statystyki opisowe dla grupy pacjentów leczonych lekiem B

Min	Max	Median	Rozstęp	1d Qu.	3d Qu.	IQR
0.038	1	0.846	0.962	0.333	1	0.667

Na podstawie statystyk przedstawionych w tabelach 6 i 7 można zauważać różnice pomiędzy grupami pacjentów leczonych lekami A i B.

W przypadku pacjentów leczonych lekiem A (tabela 6) przedstawiono rezultaty dla 20 pacjentów. W tej grupie remisję zaobserwowano u 10 pacjentów. Czasy do remisji u pacjentów mieszczą się w zakresie od 0.033 do 1, rozstęp wynosi 0.967, a $IQR = 0.625$. Mediana 0.971 sugeruje, że u połowy pacjentów poprawa zdrowia występowała późno, co wskazuje na umiarkowaną skuteczność leku w ciągu roku obserwacji.

Wyniki dla pacjentów leczonych lekiem B zostały przedstawione w tabeli 7. Podobnie jak w przypadku leku A, remisję zaobserwowano u połowy pacjentów. Zakres czasów jest lekko przesunięty, wartości mieszczą się od 0.038 do 1. Mediana wynosi 0.846, czyli mniej niż dla leku A. Wskazuje to, że remisja choroby następowała przeciętnie szybciej u pacjentów przyjmujących lek B. Wartości zmienności (rozstęp 0.962, $IQR = 0.667$).

Podsumowując w każdej grupie połowa pacjentów nie doświadczyła remisji w ciągu roku, co wskazuje na silne cenzurowanie danych. Mediana czasu do remisji dla grupy przyjmującej lek B jest niższa niż dla pacjentów przyjmujących lek A, co może sugerować większą skuteczność leku B pod względem szybkości działania. Zmienność czasów do remisji pozostaje na podobnym poziomie w obu grupach, dlatego trudno wskazać jednoznaczne różnice w stabilności efektów leczenia.

3 Lista 3

3.1 Zadanie 1

Oszacowanie estymatorów

Celem zadania jest oszacowanie średniego czasu do remisji choroby ($\mu = 1/\vartheta$) dla pacjentów leczonych dwoma różnymi lekami (A i B) przy założeniu, że czasy do remisji mają rozkład wykładniczy. Dane pochodzą z zadania 2.3 i zawierają obserwacje prawostronnie cenzurowane I-go typu.

Zgodnie z teorią przedstawioną na wykładzie, dla danych cenzurowanych I-go liczymy estymator największej wiarogodności i alternatywny estymator, oparty na liczbie obserwacji nieocenzurowanych.

Poniższy kod oblicza estymatory $\hat{\mu}$ i $\tilde{\mu}$ dla obu grup pacjentów.

```
# Estymator mu_hat
T1_A <- sum(czas_A[cenzura_A == 1]) + t0 * (n - R)
T1_B <- sum(czas_B[cenzura_B == 1]) + t0 * (n - R)
mu_1_A <- T1_A / R
mu_1_B <- T1_B / R

# Estymator mu_tilde
mu_2_A <- t0 / -log(1 - R / n)
mu_2_B <- t0 / -log(1 - R / n)
```

Tabela 8: Oszacowania średniego czasu do remisji

Lek	$\hat{\mu} = \frac{T_1}{R}$	$\tilde{\mu} = \frac{t_0}{-\log(1-R/n)}$
A	1.4216	1.4427
B	1.3459	1.4427

Z wyników z tabeli 8 możemy wywnioskować, że oszacowanie $\tilde{\mu}$ jest identyczne dla obu leków. Wynika to z faktu, że estymator zależy wyłącznie od R , n i t_0 , ignorując obserwoweble czasy zdarzeń, przez co jest on nieużyteczny. Natomiast dla leku A oszacowanie $\hat{\mu}$ wynosi 1.4216, a dla leku B jest niższa (1.3459), co wskazuje na trochę krótszy średni czas do remisji.

Wyznaczenie przedziałów ufności

W tej części zadania wyznaczamy przedziały ufności dla średniego czasu do remisji choroby dla pacjentów leczonych lekiem A oraz lekiem B, przy poziomach ufności $1 - \alpha = 0,95$ oraz $1 - \alpha = 0,99$.

Wykorzystując funkcję `binom.confint()`, można uzyskać dokładne przedziały, które następnie przekształcamy do postaci przedziałów według wzoru:

$$\tilde{T}_L = -\frac{t_0}{\log(1 - T_L)}, \quad \tilde{T}_U = -\frac{t_0}{\log(1 - T_U)}.$$

Finalny przedział dla μ wygląda następująco $[\tilde{T}_U, \tilde{T}_L]$.

Kod poniżej przedstawia implementację funkcji obliczającej przedziały ufności dla parametru μ :

```
# Funkcja do przedziału ufności dla średniego czasu
przedzialy_ufnosci <- function(R, n, alpha, t0) {
  TL <- binom.confint(R, n, methods = "exact", conf.level = 1 - alpha)$lower
  TU <- binom.confint(R, n, methods = "exact", conf.level = 1 - alpha)$upper
  TL_tylدا <- -t0 / log(1 - TL)
  TU_tylда <- -t0 / log(1 - TU)
  data.frame(
    alpha = alpha,
    "TL" = round(TU_tylда, 3),
    "TU" = round(TL_tylда, 3),
    "Poziom ufności" = paste0(100 * (1 - alpha), "%")
  )
}
```

Tabela 9: Przedziały ufności dla średniego czasu do remisji

α	\tilde{T}_L	\tilde{T}_U	Poziom ufności
0.05	0.768	3.151	95%
0.01	0.656	4.072	99%

Z tabeli 9 wynika, że dla $\alpha = 0,05$ średni czas do remisji choroby mieści się w przedziale od 0.768 do 3.151, natomiast dla $\alpha = 0,01$ w przedziale od 0.656 do 4.072. Zwiększenie poziomu ufności powoduje rozszerzenie przedziału, co jest zgodne z oczekiwaniami, ponieważ wyższy poziom ufności wiąże się z większą niepewnością co do dokładnej wartości parametru.

Otrzymane przedziały obejmują oszacowanie średniego czasu do remisji choroby z tabeli 8, co oznacza, że uzyskane estymatory są zgodne z przyjętym poziomem ufności. Szerokość przedziałów wskazuje na przeciętne oszacowanie, co może wynikać z liczby obserwacji i obecności danych cenzurowanych.

3.2 Zadanie 2

Oszacowanie estymatorów

W tym zadaniu przyjmujemy, że obserwacje czasu do remisji choroby były prowadzone do momentu, w którym u $m = 10$ pacjentów zostanie ona zaobserwowana, przy liczbie wszystkich pacjentów $n = 20$. Jest to przypadek cenzurowania II-go typu. Wykorzystując wzory podane na wykładzie wyznaczyliśmy estymatory średniego czasu do remisji.

Poniżej zamieszczony jest kod obliczający estymatory $\hat{\mu}$ dla obu grup:

```
# Estymatory
T2_A <- sum(czas_A_obs) + (n - m) * czas_A_obs[m]
T2_B <- sum(czas_B_obs) + (n - m) * czas_B_obs[m]

mu_A <- T2_A / m
mu_B <- T2_B / m
```

Tabela 10: Oszacowania estymatorów $\hat{\mu}$ (cenzurowanie II-go typu)

Lek	$\hat{\mu}$
A	1.3638
B	1.0371

Obliczone estymatory wynoszą odpowiednio $\hat{\mu}_A = 1.3638$ i $\hat{\mu}_B = 1.0371$, co możemy zobaczyć w tabeli 10. Lek B ma niższą wartość $\hat{\mu}$, co sugeruje krótszy średni czas do remisji w porównaniu z lekiem A. W odniesieniu do wyników uzyskanych w Zadaniu 1 (tabela 8), estymator dla Leku A pozostał zbliżony, natomiast estymator dla leku B jest nieco mniejszy.

Wyznaczenie przedziałów ufności

Teraz zajmiemy się wyznaczeniem przedziałów ufności. Skorzystamy w tym celu z własności rozkładu gamma $\mathcal{G}(m, \frac{1}{m})$ i wzorów:

$$\tilde{T}_L = \frac{T_2}{mq_m\left(\frac{\alpha}{2}\right)}, \quad \tilde{T}_U = \frac{T_2}{mq_m\left(1 - \frac{\alpha}{2}\right)},$$

gdzie $mq_m\left(\frac{\alpha}{2}\right)$ oraz $mq_m\left(1 - \frac{\alpha}{2}\right)$ są odpowiednimi kwantylami rozkładu gamma $\mathcal{G}(m, \frac{1}{m})$. Wówczas przedział ufności ma postać $[\tilde{T}_U, \tilde{T}_L]$.

W praktyce kwantyle te wyznaczono za pomocą funkcji *qgamma()*, a przedziały obliczono zgodnie ze wzorem:

```
# Funkcja do przedziału ufności
przedzialy_ufnosci2 <- function(T2, m, alpha) {
  q_1 <- qgamma(alpha/2, shape = m, scale = 1/m)
  q_2 <- qgamma(1 - alpha/2, shape = m, scale = 1/m)
  TL <- T2 / (m * q_1)
  TU <- T2 / (m * q_2)
  c(round(TU, 4), round(TL, 4))
}
```

Tabela 11: Przedziały ufności (cenzurowanie II-go typu)

Poziom ufności	Lek A $[\tilde{T}_L, \tilde{T}_U]$	Lek B $[\tilde{T}_L, \tilde{T}_U]$
95% ($\alpha = 0.05$)	[0.7983 ; 2.844]	[0.607 ; 2.1626]
99% ($\alpha = 0.01$)	[0.682 ; 3.6692]	[0.5186 ; 2.7901]

Wyniki przedstawione w tabeli 11 wynika, że dla poziomu ufności 95%, średni czas do remisji choroby dla pacjentów leczonych lekiem A mieści się w przedziale od 0.7983 do 2.844, natomiast dla leku B od 0.607 do 2.1626. Dla poziomu ufności 99%, przedziały są szersze, od 0.682 do 3.6692 dla leku A oraz od 0.5186 do 2.7901 dla leku B, co jest skutkiem niższego α . Widać, że zarówno dolne, jak i górne granice przedziałów dla leku B są niższe niż dla leku A, co potwierdza tendencję do krótszego czasu remisji w grupie B. Również w tym przypadku otrzymane przedziały obejmują oszacowanie średniego czasu do remisji choroby z tabeli 10.

3.3 Zadanie 3

Symulacja obciążen i błędów średniokwadratowych estymatorów

W tym zadaniu porównano dokładność dwóch estymatorów parametru ϑ rozkładu wykładniczego dla danych prawostronnie cenzurowanych I-go typu:

$$\hat{\vartheta} = \frac{R}{T_1},$$

gdzie

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

$$T_1 = \sum_{i=1}^R X_{(i)} + t_{(0)}(n - R)$$

oraz

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

Zgodnie z treścią zadania, należy przeprowadzić symulacje w celu porównania obciążenia (Bias) i błędu średniokwadratowego (MSE) obu estymatorów dla różnych wartości parametrów:

$$\vartheta = 1, \quad n \in \{10, 30\}, \quad t_0 \in \{0.5, 1, 2\}.$$

Poniżej przedstawiono kod wykorzystany do obliczeń.

```
# Funkcja do obliczenia estymatorów
estymatory <- function(dane, t0) {
  R <- sum(dane$Cenzura)
  T1 <- sum(dane$Czas[dane$Cenzura == 1]) + t0 * (nrow(dane) - R)
  theta_hat <- R / T1
```

```

theta_tilde <- -log(1 - R / nrow(dane)) / t0
c(theta_hat, theta_tilde)
}

# Funkcja do symulacji
symulacja_bias_mse <- function(n, t0, M = 10000, theta = 1) {
  theta_hat_vec <- numeric(M)
  theta_tilde_vec <- numeric(M)

  for (i in 1:M) {
    repeat {
      dane <- Ityp(n, theta, t0)
      R <- sum(dane$Cenzura)
      if (R != n) break
    }
    est <- estymatory(dane, t0)
    theta_hat_vec[i] <- est[1]
    theta_tilde_vec[i] <- est[2]
  }

  bias_hat <- mean(theta_hat_vec - theta)
  bias_tilde <- mean(theta_tilde_vec - theta)
  mse_hat <- mean((theta_hat_vec - theta)^2)
  mse_tilde <- mean((theta_tilde_vec - theta)^2)

  data.frame(
    n = n, t0 = t0,
    Bias_hat = round(bias_hat, 4),
    Bias_tilde = round(bias_tilde, 4),
    MSE_hat = round(mse_hat, 4),
    MSE_tilde = round(mse_tilde, 4)
  )
}
}

```

Tabela 12: Porównanie obciążień (Bias) i błędów średniokwadratowych (MSE) estymatorów $\hat{\vartheta}$ i $\tilde{\vartheta}$ dla parametru ϑ

n	t_0	$\text{Bias}(\vartheta, \hat{\vartheta})$	$\text{Bias}(\vartheta, \tilde{\vartheta})$	$\text{MSE}(\vartheta, \hat{\vartheta})$	$\text{MSE}(\vartheta, \tilde{\vartheta})$
10	0.5	0.0666	0.0786	0.3071	0.3376
10	1.0	0.0617	0.0862	0.1887	0.2271
10	2.0	-0.0498	-0.0767	0.0810	0.0614
30	0.5	0.0168	0.0191	0.0885	0.0917
30	1.0	0.0234	0.0330	0.0578	0.0676
30	2.0	0.0259	0.0582	0.0415	0.0707

Interpretacja wyników symulacji

Tabela 12 przedstawia wartości obciążenia ($Bias$) i błędu średniokwadratowego (MSE) dla dwóch estymatorów parametru ϑ , w zależności od liczby obserwacji n i czasu cenzurowania t_0 .

Zwiększenie liczby obserwacji z 10 do 30 poprawia dokładność obu estymatorów. Dla $t_0 = 1.0$ MSE dla $\hat{\vartheta}$ spada z 0.1887 do 0.0578, a dla $\tilde{\vartheta}$ z 0.2271 do 0.0676. Większa próba zmniejsza zarówno obciążenie, jak i wariancję. Przy $t_0 = 2.0$ dla $n = 10$ $Bias$ jest ujemny (-0.0498 dla $\hat{\vartheta}$ i -0.0767 dla $\tilde{\vartheta}$), co oznacza tendencję do niedoszacowania parametru ϑ .

Estymator $\hat{\vartheta}$ jest bardziej precyzyjny niż $\tilde{\vartheta}$. Zarówno $Bias$, jak i MSE są dla niego mniejsze w prawie każdej konfiguracji. Przykładowo, dla $n = 10$ i $t_0 = 1.0$ $Bias$ wynosi 0.0617 dla $\hat{\vartheta}$ i 0.0862 dla $\tilde{\vartheta}$, a MSE odpowiednio 0.1887 i 0.2271. Estymator $\hat{\vartheta}$ można uznać za bardziej stabilny i dokładny, natomiast estymator $\tilde{\vartheta}$ lepiej sprawdza się dopiero przy większej liczbie obserwacji.

4 Lista 4

4.1 Zadanie 1

Celem zadania jest napisanie deklaracji funkcji, implementacji testu ilorazu wiarygodności (IW) dla danych cenzurowanych I-go typu z rozkładu wykładniczego do testowania następujących hipotez:

1. $H_0^{(1)} : \vartheta = \vartheta_0$, przy hipotezie alternatywnej $H_1^{(1)} : \vartheta \neq \vartheta_0$,
2. $H_0^{(2)} : \vartheta \leq \vartheta_0$, przy hipotezie alternatywnej $H_1^{(2)} : \vartheta > \vartheta_0$,
3. $H_0^{(3)} : \vartheta \geq \vartheta_0$, przy hipotezie alternatywnej $H_1^{(3)} : \vartheta < \vartheta_0$.

Kod

```
#Zadanie 1
IW <- function(x, r, t0, n, theta0, type = "dwustronna") {
  L <- function(theta) {
    theta^r * exp(-theta*(sum(x)+t0*(n-r)))
  }
  s <- sum(x)+t0*(n-r)
  theta_szac <- r/s
  L1 <- L(theta_szac)

  #hipotezy
  if (type == "dwustronna") {
    licznik <- L(theta0)
  } else if (type == "prawostronna") {
    if (theta_szac <= theta0) licznik <- L1 else licznik <- L(theta0)
  } else if (type == "lewostronna") {
    if (theta_szac >= theta0) licznik <- L1 else licznik <- L(theta0)
  }
  lambda <- licznik/L1

  statystyka <- -2*log(lambda)
  p_wartosc <- 1-pchisq(statystyka, df = 1)

  list(theta_szac = theta_szac,
        statystyka = statystyka,
        p_wartosc = p_wartosc,
        type = type)
}
```

4.2 Zadanie 2

Celem zadania jest przeprowadzenie symulacji na podstawie której mamy oszacować moc dla 10 wybranych alternatyw oraz rozmiaru testu z punktu (a) zadania 1 dla wybranej wartości ϑ_0 i t_0 oraz n przyjmuje wartości $\{20, 50\}$.

Kod

Do napisania kodu użyjemy funkcji generowania danych cenzurowanych I typu, która została już przez nas napisana i zaprezentowana w zadaniu 2.1.

```
#Zadanie 2
symuluj_moc_IW <- function(M = 2000,
                                n_wartosci = c(20, 50),
                                theta0 = 2,
                                alternatywy = c(0.5, 0.8, 1.2, 1.5, 1.8, 2.0,
                                                2.2, 2.5, 3.0, 4.0, 5.0),
                                t0 = 3,
                                alfa = 0.05,
                                typ = "dwustronna") {

  wyniki <- expand.grid(n = n_wartosci, theta = alternatywy)
  wyniki$moc <- NA
  wyniki$rodzaj <- ifelse(wyniki$theta == theta0, "rozmiar testu", "moc")

  for (i in seq_len(nrow(wyniki))) {
    n <- wyniki$n[i]
    theta <- wyniki$theta[i]
    liczba_odrzucen <- numeric(M)

    for (m in 1:M) {
      dane <- Ityp(n = n, theta = theta, t0 = t0)
      x <- dane$Czas[dane$Cenzura == 1]
      r <- sum(dane$Cenzura)

      if (r == 0) {
        p_wartosc <- 1 #brak zdarzeń -> nie odrzucamy
      } else {
        wynik_testu <- IW(x = x, r = r, t0 = t0, n = n,
                            theta0 = theta0, typ = typ)
        p_wartosc <- wynik_testu$p_wartosc
      }
      liczba_odrzucen[m] <- as.integer(p_wartosc < alfa)
    }

    #Obliczenie mocy
  }
}
```

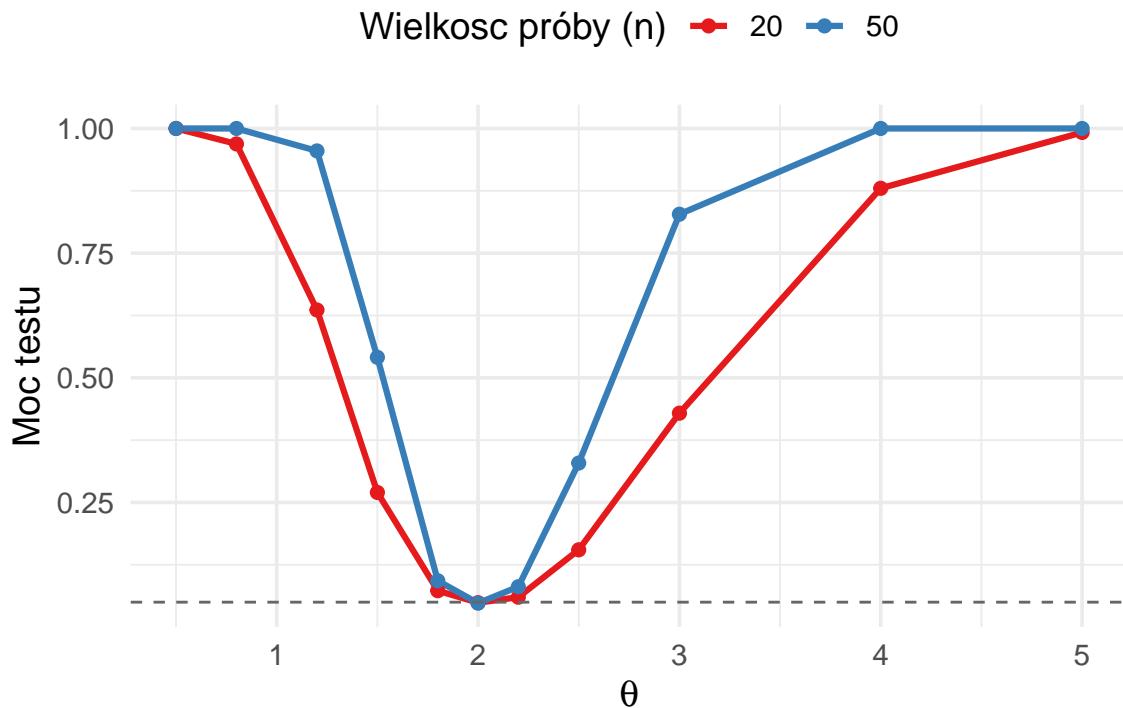
```

    wyniki$moc[i] <- mean(liczba_odrzucen)
}
wyniki
}

```

Tabela 13: Symulowana moc i rozmiar testu IW

n	theta	moc	rodzaj
20	0.5	1.000	moc
50	0.5	1.000	moc
20	0.8	0.969	moc
50	0.8	1.000	moc
20	1.2	0.636	moc
50	1.2	0.955	moc
20	1.5	0.270	moc
50	1.5	0.541	moc
20	1.8	0.073	moc
50	1.8	0.093	moc
20	2.0	0.049	rozmiar testu
50	2.0	0.048	rozmiar testu
20	2.2	0.060	moc
50	2.2	0.081	moc
20	2.5	0.155	moc
50	2.5	0.329	moc
20	3.0	0.429	moc
50	3.0	0.828	moc
20	4.0	0.880	moc
50	4.0	1.000	moc
20	5.0	0.992	moc
50	5.0	1.000	moc



Rysunek 3: Wykres mocy testu IW w zależności od parametru θ

Interpretacja

Jak widać na rysunku 3 oraz w tabeli 13 na podstawie wyników przeprowadzonej przez nas symulacji dla wybranych alternatyw ($\{0.5, 0.8, 1.2, 1.5, 1.8, 2.0, 2.2, 2.5, 3.0, 4.0, 5.0\}$) możemy wyciągnąć następujące informacje:

1. Moc testu jest najmniejsza dla rozmiaru testu i zwiększa się wraz ze wzrostem różnicy pomiędzy przyjętą ϑ a $\vartheta_0 = 2$.
2. Rozmiar testu (dla $\vartheta = \vartheta_0 = 2$) wahaje się pomiędzy 0.04 a 0.05).
3. Test ma większą moc przy większej liczbie obserwacji ($n = 50$).

Zatem, można stwierdzić, że przeprowadzona symulacja odzwierciedla teoretyczne własności badanego testu IW .

4.3 Zadanie 3

Celem zadania jest weryfikacja hipotezy mówiącej, że średni czas do remisji choroby w grupie A oraz średni czas do remisji choroby w grupie B wynosi 1. Dane A oraz B bierzemy z treści zadania 2.3.

Kod

```
#Dane (zadanie 3 lista 2)
A <- c(0.03345514, 0.08656403, 0.08799947, 0.24385821, 0.27755032,
```

```

    0.40787247, 0.58825664, 0.64125620, 0.90679161, 0.94222208)
B <- c(0.03788958, 0.12207257, 0.20319983, 0.24474299, 0.30492413,
      0.34224462, 0.42950144, 0.44484582, 0.63805066, 0.69119721)

n = 20      #length(A) + length(B) = 20
t0 = 1       #dowolne
theta0 = 1   #dowolne

#wyniki
wynik_A <- IW(x = A, r = length(A), t0 = t0, n = n,
                 theta0 = theta0, type = "dwustronna")
wynik_B <- IW(x = B, r = length(B), t0 = t0, n = n,
                 theta0 = theta0, type = "dwustronna")

```

Tabela 14: Wyniki testu ilorazu wiarygodności dla grup A i B

Grupa	$\hat{\theta}$	P-wartość	Typ testu	Wniosek
A	0.70344	0.23735	dwustronna	Brak podstaw do odrzucenia H0
B	0.74302	0.32305	dwustronna	Brak podstaw do odrzucenia H0

Interpretacja

Na podstawie testu ilorazu wiarygodności (wyniki w tabeli 14) zarówno dla grupy A (p-wartość ≈ 0.237) jak i dla grupy B (p-wartość ≈ 0.323) nie stwierdzono statystycznie istotnych dowodów przeciwko hipotezie mówiącej, że średni czas do remisji choroby wynosi 1 (nie ma powodu do odrzucenia hipotezy, gdyż p-wartość $> \alpha$).

5 Zadania dodatkowe

5.1 Zadanie 1 z Listy 1

```
## Warning: There was 1 warning in `mutate()`.  
## i In argument: `across(where(is.numeric), round, 4)`.  
## Caused by warning:  
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.  
## Supply arguments directly to `.fns` through an anonymous function instead.  
##  
## # Previously  
## across(a:b, mean, na.rm = TRUE)  
##  
## # Now  
## across(a:b, \((x) mean(x, na.rm = TRUE))
```

Tabela 15: Porównanie metod obliczania momentów rozkładu

	Zestaw	E.X..num	SD.X..num	E.X..MC	SD.X..MC
E_X	Zestaw A	1.9088	1.9363	1.9046	1.9319
E_X1	Zestaw B	1.4361	0.6059	1.4359	0.6061

5.2 Zadanie 1 z Listy 3

Tabela 16: Porównanie dwóch postaci statystyki T_2 dla różnych wartości parametru λ (cenzurowanie II-go typu)

λ	T_2 z definicji	T_2 z sumy D_i	Różnica
0.5	4.317884	4.317884	0
1.0	4.760065	4.760065	0
2.0	5.173695	5.173695	0
3.0	2.807090	2.807090	0

Dla różnych wartości parametru λ obliczono statystykę T_2 na dwa sposoby: według definicji oraz jako sumę składników D_i . Jak pokazują wyniki w tabeli 16, różnice między tymi wartościami wynoszą zero, co oznacza, że oba sposoby obliczania T_2 są równoważne i dają identyczne wyniki.

Rozwiążanie

Zauważmy, że

$$X_{(i)} = \sum_{j=1}^i (X_{(j)} - X_{(j-1)}) ,$$

więc

$$\sum_{i=1}^m X_{(i)} = \sum_{i=1}^m \sum_{j=1}^i (X_{(j)} - X_{(j-1)}) = \sum_{j=1}^m (m-j+1) (X_{(j)} - X_{(j-1)}).$$

Podstawiamy to do definicji statystyki T_2 :

$$T_2 = \sum_{i=1}^m X_{(i)} + (n-m) X_{(m)} = \sum_{j=1}^m (m-j+1) (X_{(j)} - X_{(j-1)}) + (n-m) \sum_{j=1}^m (X_{(j)} - X_{(j-1)}).$$

Porządkując wyrazy przy $(X_{(j)} - X_{(j-1)})$, otrzymujemy:

$$T_2 = \sum_{j=1}^m [(m-j+1) + (n-m)] (X_{(j)} - X_{(j-1)}) = \sum_{j=1}^m (n-j+1) (X_{(j)} - X_{(j-1)}).$$

Zatem

$$T_2 = \sum_{i=1}^m D_i, \quad \text{gdzie } D_i = (n-i+1) (X_{(i)} - X_{(i-1)}),$$

co należało dowieść.