

Sprawozdanie 1

Mateusz Cieślak

2025-11-10

Spis treści

1	Lista 1	1
1.1	Zadanie 1	1
1.2	Zadanie 2	2
1.3	Zadanie 3	3
1.4	Zadanie 4	3
1.5	Zadanie 5	4
2	Lista 2	5
2.1	Zadanie 1	5
2.2	Zadanie 2	6
2.3	Zadanie 3	6
3	Lista 3	7
3.1	Zadanie 1	7
3.2	Zadanie 2	8
3.3	Zadanie 3	9
4	Lista 4	10
4.1	Zadanie 1	10
4.2	Zadanie 2	10
4.3	Zadanie 3	12

Spis rysunków

1	Ilustracje hazardu dla różnych parametrów	3
2	Porównanie wygenerowanych histogramów do teoretycznych gęstości dla różnych parametrów	4

Spis tabel

1	Statystyki empiryczne prób z rozkładu Exponentiated-Weibull	4
2	Wartości teoretyczne rozkładu Exponentiated-Weibull	5
3	Statystyki opisowe – cenzurowanie typu I	6
4	Statystyki opisowe – cenzurowanie typu II	6
5	Statystyki opisowe – cenzurowanie losowe	6
6	Statystyki opisowe dla grup A i B	6
7	Oszacowania parametrów rozkładu wykładniczego	7
8	Przedziały ufności dla średniego czasu do remisji choroby dla pacjentów leczonych lekiem A i B dla różnych poziomów α	8
9	Oszacowania parametrów rozkładu wykładniczego	9

10	Przedziały ufności dla średniego czasu do remisji choroby dla pacjentów leczonych lekiem A i B dla różnych poziomów α	9
11	Porównanie dokładności estymatorów na podstawie obciążenia i błędu średniokwadratowego .	10
12	Moc testu LRT dla $n = 20$	11
13	Moc testu LRT dla $n = 50$	12
14	Weryfikacja hipotezy o średnim czasie do remisji w grupach A i B	12

Wykorzystane biblioteki:

```
library(knitr)
library(kableExtra)
library(binom)
```

1 Lista 1

1.1 Zadanie 1

W tym zadaniu piszemy deklaracje czterech funkcji.

Funkcja gęstości:

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

Dystrybuanta:

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

Funkcja kwantylowa:

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

Funkcja hazardu:

$$h(t; \alpha, \beta, \gamma) = \frac{f(t; \alpha, \beta, \gamma)}{1 - F(t; \alpha, \beta, \gamma)}$$

```
# gęstość
gestosc <- function(t, a, b, g) {
  if (any(c(a, b, g) <= 0)) stop("Parametry a, b, g muszą być dodatnie (> 0).")
  if (any(t < 0)) stop("Czas t musi być nieujemny (t >= 0).")

  ft <- (a * g / b) * ((t / b)^(a - 1)) *
    exp(-(t / b)^a) * ((1 - exp(-(t / b)^a))^(g - 1))

  return(ft)
}

# dystrybuanta
dystrybuanta <- function(t, a, b, g) {
  if (any(c(a, b, g) <= 0)) stop("Parametry a, b, g muszą być dodatnie (> 0).")
```

```

    if (any(t < 0)) stop("Czas t musi być nieujemny (t >= 0).")

    Ft <- (1 - exp(-(t / b)^a))^g
    return(Ft)
}

# kwantyl
kwantyl <- function(p, a, b, g) {
  if (any(c(a, b, g) <= 0)) stop("Parametry a, b, g muszą być dodatnie (> 0).")
  if (any(p < 0 | p > 1)) stop("Prawdopodobieństwa p muszą być w zakresie [0, 1].")

  Qt <- b * (-log(1 - p^(1 / g)))^(1 / a)
  return(Qt)
}

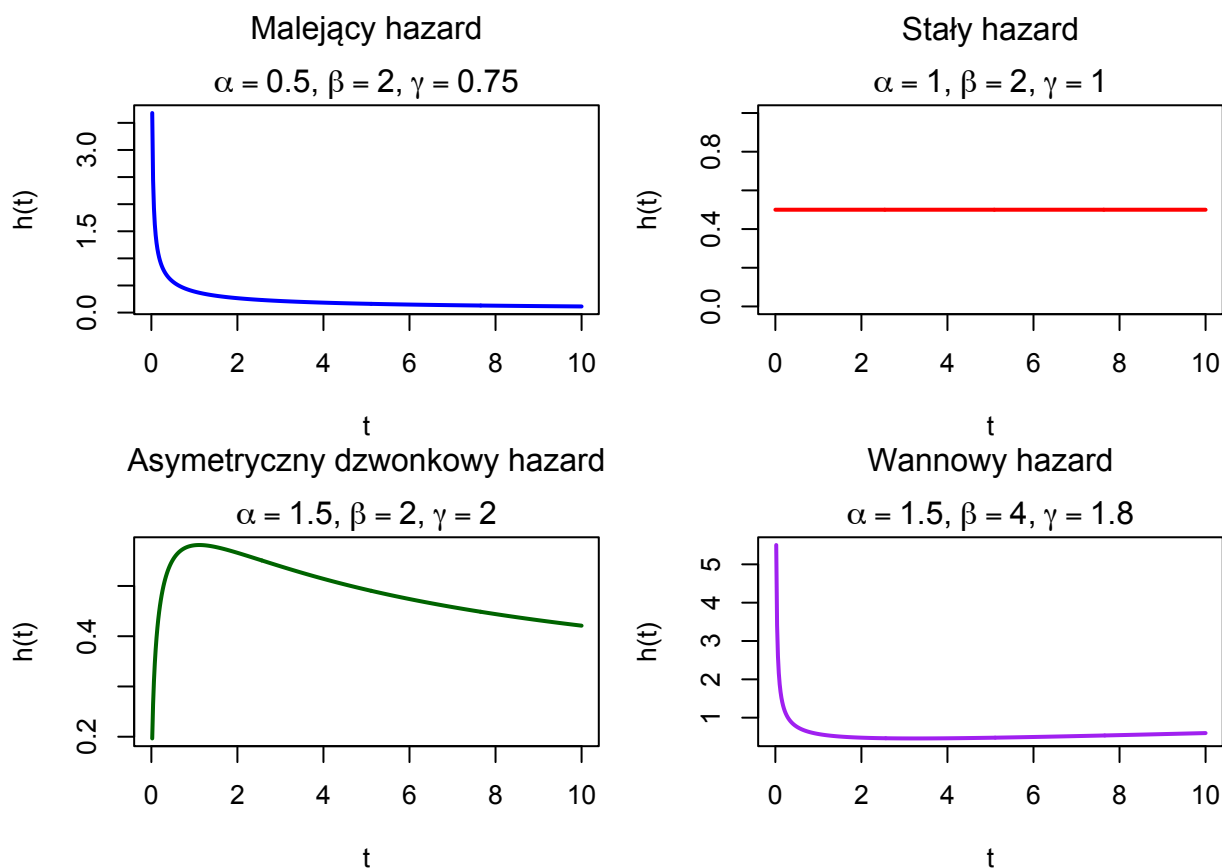
# hazard (intensywność ryzyka)
hazard <- function(t, a, b, g) {
  if (any(c(a, b, g) <= 0)) stop("Parametry a, b, g muszą być dodatnie (> 0).")
  if (any(t < 0)) stop("Czas t musi być nieujemny (t >= 0).")

  ft <- gestosc(t, a, b, g)
  Ft <- dystrybuanta(t, a, b, g)
  ht <- ft / (1 - Ft)
  return(ht)
}

```

1.2 Zadanie 2

W niniejszym zadaniu testujemy różne parametry, aby zilustrować różne kształty wykresów funkcji hazardu.



Rysunek 1: Ilustracje hazardu dla różnych parametrów

1.3 Zadanie 3

Deklarujemy funkcję do generowania zmiennych z rozkładu $\mathcal{EW}(\alpha, \beta, \gamma)$.

```
generuj_EW <- function(n, a, b, g) {
  u <- runif(n)
  t <- kwantyl(u, a, b, g)
  return(t)
}
```

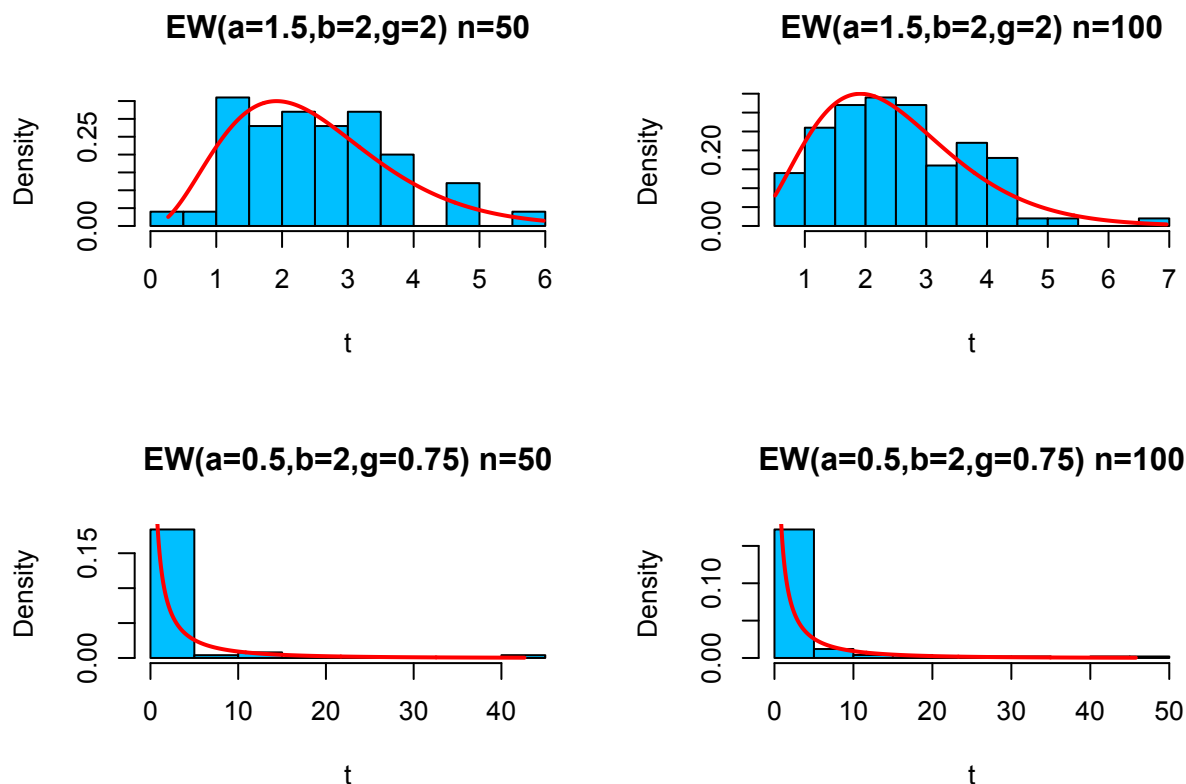
1.4 Zadanie 4

Będziemy analizować dwie trójki parametrów:

- $p1 = (\alpha = 1.5, \beta = 2, \gamma = 2)$
- $p2 = (\alpha = 0.5, \beta = 2, \gamma = 0.75)$

dla $n = 50$ i $n = 100$ obserwacji.

Zobaczymy również jak ich gęstości wyglądają w porównaniu do teoretycznych gęstości.



Rysunek 2: Porównanie wygenerowanych histogramów do teoretycznych gęstości dla różnych parametrów

Możemy zauważyć, że kształty histogramów są podobne do teoretycznych gęstości, szczególnie dla większego $n = 100$.

1.5 Zadanie 5

Ponownie analizujemy próbki wygenerowane w poprzednim zadaniu dla różnych trójek parametrów

- $p1 = (\alpha = 1.5, \beta = 2, \gamma = 2)$
- $p2 = (\alpha = 0.5, \beta = 2, \gamma = 0.75)$

dla $n = 50$ i $n = 100$ obserwacji.

Tym razem zobaczymy jak wyglądają statystyki opisowe wyznaczone na podstawie wygenerowanych danych w porównaniu z teoretycznymi wartościami.

Tabela 1: Statystyki empiryczne prób z rozkładu Exponentiated-Weibull

	n	α	β	γ	średnia	mediana	sd	q1	q3	rozstęp	min	max
p1	50	1.5	2	2.00	2.498	2.334	1.184	1.573	3.111	5.707	0.27	5.977
p1'	100	1.5	2	2.00	2.558	2.396	1.159	1.706	3.381	6.445	0.51	6.956
p2	50	0.5	2	0.75	2.370	0.430	6.365	0.038	1.969	42.631	0.00	42.631
p2'	100	0.5	2	0.75	3.310	0.690	7.839	0.094	2.272	45.764	0.00	45.764

Zazwyczaj dla $n = 100$ wyniki są bliższe teoretycznych wartości.

Tabela 2: Wartości teoretyczne rozkładu Exponentiated-Weibull

	α	β	γ	mediana teoretyczna	q1 teoretyczny	q3 teoretyczny
p1	1.5	2	2.00	2.293	1.566	3.185
p2	0.5	2	0.75	0.511	0.059	2.617

2 Lista 2

2.1 Zadanie 1

Poniżej znajdują się funkcje odpowiadające za generowanie zmiennych cenzurowanych z ogólnionego rozkładu wykładniczego w przypadku cenzurowania I typu, II typu i losowego.

```
#kwantyl
rGE <- function(n, l, a) {
  if(any(c(n,l,a) <= 0)) stop("n, lambda i alpha muszą być > 0")
  u <- runif(n)
  x <- - (1 / l) * log(1 - u^(1/a))
  return(x)
}

generuj_typI <- function(n, l, a, t0, return_full = FALSE) {
  if(t0 < 0) stop("t0 musi być >= 0")
  T <- rGE(n, l, a)
  C <- rep(t0, n)
  time <- pmin(T, C)
  delta <- as.integer(T <= C) # 1 = zdarzenie zaobserwowane, 0 = cenzurowane
  out <- data.frame(time = time, delta = delta)
  if(return_full) out <- cbind(out, T = T, C = C)
  return(out)
}

generuj_typII <- function(n, l, a, m, return_full = FALSE) {
  if(m <= 0 || m > n) stop("m musi spełniać 1 <= m <= n")
  T <- rGE(n, l, a)
  T_sorted <- sort(T)
  censor_time <- T_sorted[m] # czas zatrzymania
  C <- rep(censor_time, n)
  time <- pmin(T, C)
  delta <- as.integer(T <= C)
  out <- data.frame(time = time, delta = delta)
  if(return_full) out <- cbind(out, T = T, C = C)
  return(out)
}

generuj_losowo <- function(n, l, a, eta, return_full = FALSE) {
  if(eta <= 0) stop("eta musi być > 0")
  T <- rGE(n, l, a)
  rate <- 1 / eta
  C <- rexp(n, rate = rate)
  time <- pmin(T, C)
}
```

```

delta <- as.integer(T <= C)
out <- data.frame(time = time, delta = delta)
if(return_full) out <- cbind(out, T = T, C = C)
return(out)
}

```

2.2 Zadanie 2

Korzystając z napisanych funkcji wygenerujemy po jednym zbiorze danych cenzurowanych z każdego typu z rozkładu $\mathcal{GE}(\lambda, \alpha)$ i sprawdzimy sensowne dla nich statystyki opisowe.

Tabela 3: Statystyki opisowe – cenzurowanie typu I

liczba_observacji	kompletne	cenzurowane	min	Q1	mediana	Q3
100	61	39	0.0653	1.117	2.1458	3

Tabela 4: Statystyki opisowe – cenzurowanie typu II

liczba_observacji	kompletne	cenzurowane	min	Q1	mediana	Q3	max
100	60	40	0.0463	0.6701	1.6688	2.1681	2.1681

Tabela 5: Statystyki opisowe – cenzurowanie losowe

liczba_observacji	kompletne	cenzurowane	min	Q1	mediana	Q3	max
100	47	53	0.0172	0.5257	0.895	1.7544	4.2538

Dla danych cenzurowanych min to czas pierwszego zaobserwowanego zdarzenia. Natomiast max w typie drugim to czas ostatniego niecenzurowanego zdarzenia, a w typie trzecim max może oznaczać zarówno czas zdarzenia jak i cenzury.

2.3 Zadanie 3

W grupie otrzymującej lek A uzyskano następujące dane. U dziesięciu pacjentów remisja choroby nastąpiła w chwilach: 0.03345514, 0.08656403, 0.08799947, 0.24385821, 0.27755032, 0.40787247, 0.58825664, 0.64125620, 0.90679161, 0.94222208, natomiast u pozostałych dziesięciu pacjentów w ciągu roku nie zaobserwowano remisji. W grupie otrzymującej lek B uzyskano następujące dane. U dziesięciu pacjentów remisja choroby nastąpiła w chwilach: 0.03788958, 0.12207257, 0.20319983, 0.24474299, 0.30492413, 0.34224462, 0.42950144, 0.44484582, 0.63805066, 0.69119721, natomiast u pozostałych dziesięciu pacjentów w ciągu roku nie zaobserwowano remisji.

Poniższa tabela 6 prezentuje statystyki opisowe dla tych danych.

Tabela 6: Statystyki opisowe dla grup A i B

grupa	liczba_observacji	kompletne	cenzurowane	min	Q1	mediana	Q3
Lek A	20	10	10	0.0335	0.3753	0.9711	1

Lek B	20	10	10	0.0379	0.3329	0.8456	1
-------	----	----	----	--------	--------	--------	---

Na podstawie $Q1$ i *mediany* widzimy, że w grupie otrzymującej lek B remisja ogólnie występowała trochę szybciej, chociaż pierwsza remisja $\min = 0.0335$ wystąpiła szybciej w grupie przyjmującej lek A.

3 Lista 3

3.1 Zadanie 1

Poniższe funkcje odpowiadają za oszacowanie średniego czasu oczekiwania i wyznaczenie realizacji przedziału ufności dla danych cenzurowanych typu I.

```

censor_time <- 1.0

# Funkcja 1: estymacja theta i mu
oszacowanie_sredniego_czasu <- function(event_times, cens_count, censor_time) {
  R <- length(event_times)
  T_total <- sum(event_times) + cens_count * censor_time
  theta_hat <- R / T_total
  mu_hat <- 1 / theta_hat
  list(theta_hat = theta_hat, mu_hat = mu_hat)
}

# Funkcja 2: realizacja przedzialow ufności typu I
przedzial_ufnosci_cenzI <- function(r, n, t0, alpha = 0.05, metoda = "exact") {

  ci_p <- binom.confint(x = r, n = n, conf.level = 1 - alpha, methods = metoda)
  TL <- ci_p$lower
  TU <- ci_p$upper

  TL_teta <- -log(1 - TL) / t0
  TU_teta <- -log(1 - TU) / t0

  mu_lower <- 1 / TU_teta
  mu_upper <- 1 / TL_teta

  return(data.frame(alpha = alpha, lower = mu_lower, upper = mu_upper))
}

```

Tabela 9 przedstawia oszacowanie $\hat{\theta}$ jak i również oszacowanie średniego czasu do remisji $\hat{\mu}$ dla obu leków.

Tabela 7: Oszacowania parametrów rozkładu wykładniczego

Grupa	$\hat{\theta}$	$\hat{\mu}$
Lek A	0.70344	1.42158
Lek B	0.74302	1.34587

Tabela 10 przedstawia realizację przedziałów ufności dla średniego czasu do remisji choroby leku A i B dla $\alpha = 0.01$ i $\alpha = 0.05$.

Tabela 8: Przedziały ufności dla średniego czasu do remisji choroby dla pacjentów leczonych lekiem A i B dla różnych poziomów α

Grupa	α	dolna granica	górną granica
Lek A	0.05	0.768	3.1506
Lek A	0.01	0.656	4.0720
Lek B	0.05	0.768	3.1506
Lek B	0.01	0.656	4.0720

3.2 Zadanie 2

W tym zadaniu zmieniamy czas cenzurowania z jedynki na czas zaobserwowania dziesiątej remisji czyli zmieniamy dane cenzurowane typu I na typ II.

```
# Funkcja 1: estymacja theta i mu
oszacowanie_sredniego_czasu <- function(event_times, cens_count) {

  # Zmiana czasu cenzurowania
  censor_time<-max(event_times)

  R <- length(event_times)
  T_total <- sum(event_times) + cens_count * censor_time
  theta_hat <- R / T_total
  mu_hat <- 1 / theta_hat
  list(theta_hat = theta_hat, mu_hat = mu_hat)
}

# Funkcja 2: realizacja przedzialow ufnosci typu II.
oszacowanie_typeII <- function(event_times, n_total, alpha = 0.05) {
  m <- length(event_times)
  n <- n_total
  q_low <- qgamma(alpha/2, shape = m, rate = m)
  q_up <- qgamma(1 - alpha/2, shape = m, rate = m)

  suma <- sum(event_times) + event_times[m] * (n - m)

  TL <- m * q_low / suma
  TU <- m * q_up / suma

  MU <- 1/TL
  ML <- 1/TU

  data.frame(
    alpha = alpha,
    lower = ML,
    upper = MU
  )
}
```

Tabela 9: Oszacowania parametrów rozkładu wykładniczego

Grupa	$\hat{\theta}$	$\hat{\mu}$
Lek A	0.73324	1.36380
Lek B	0.96426	1.03706

Tabela 10: Przedziały ufności dla średniego czasu do remisji choroby dla pacjentów leczonych lekiem A i B dla różnych poziomów α

Grupa	α	dolna granica	górną granica
Lek A	0.05	0.7983	2.8440
Lek A	0.01	0.6820	3.6692
Lek B	0.05	0.6070	2.1626
Lek B	0.01	0.5186	2.7901

3.3 Zadanie 3

W tym zadaniu porównamy dokładność estymatorów $\hat{\vartheta} = \frac{R}{T_1}$ oraz $\tilde{\vartheta} = \frac{-\log(1 - \frac{R}{n})}{t_0}$ poprzez sprawdzenie obciążeń i błędów średniokwadratowych. Wykonamy 10000 symulacji.

```
# Funkcja do obliczenia estymatorów dla jednej próby
simulate_estimators <- function(n, t0, theta) {
  X <- rexp(n, rate = theta)

  # Dane cenzurowane typu I
  R <- sum(X <= t0)
  T1 <- sum(X[X <= t0]) + t0 * (n - R)

  theta_hat <- R / T1

  if (R == n) {
    theta_tilde <- 1 / mean(X)
    wszystkie_pelne <- TRUE
  } else {
    theta_tilde <- -log(1 - R/n) / t0
    wszystkie_pelne <- FALSE
  }

  return(c(theta_hat = theta_hat, theta_tilde = theta_tilde, wszystkie_pelne = wszystkie_pelne))
}
```

W przypadku $n = 10$ i $t_0 = 2.0$ estymator $\tilde{\vartheta}$ ma mniejsze obciążenie i błąd średniokwadratowy, ale we wszystkich pozostałych przypadkach to estymator $\hat{\vartheta}$ ma mniejsze obciążenie i błąd średniokwadratowy, więc można stwierdzić, że to właśnie $\hat{\vartheta}$ jest dokładniejszy.

Tabela 11: Porównanie dokładności estymatorów na podstawie obciążenia i błędu średniokwadratowego

n	t_0	\widehat{MSE}	\widehat{MSE}	\tilde{b}	\hat{b}	Kompletne przypadki
10	0.5	0.3383	0.3078	0.0789	0.0669	1
30	0.5	0.0912	0.0891	0.0229	0.0212	0
10	1.0	0.2454	0.2063	0.0963	0.0706	77
30	1.0	0.0668	0.0577	0.0315	0.0232	0
10	2.0	0.1608	0.1756	0.0661	0.0864	2296
30	2.0	0.0727	0.0444	0.0624	0.0311	120

4 Lista 4

4.1 Zadanie 1

```
test_lrt_cenzI <- function(r, s, n, t0, theta0, typ = "dwustronny", alpha = 0.05) {
  theta_hat <- r / (s + (n - r) * t0)
  W <- 2 * (r * log(theta_hat / theta0) + (theta0 - theta_hat) * (s + (n - r) * t0))

  # wartość p
  if (typ == "dwustronny") {
    pval <- 1 - pchisq(W, df = 1)
  } else if (typ == "prawostronny") {
    pval <- if (theta_hat > theta0) 1 - pchisq(W, df = 1) else 1
  } else if (typ == "lewostronny") {
    pval <- if (theta_hat < theta0) 1 - pchisq(W, df = 1) else 1
  } else stop("typ musi być 'dwustronny', 'prawostronny' lub 'lewostronny'")

  # poziom krytyczny
  c_alpha <- qchisq(1 - alpha, df = 1)

  list(theta_hat = theta_hat, statystyka = W, p_value = pval, krytyczny = c_alpha)
}
```

4.2 Zadanie 2

W tym zadaniu wykonamy oszacowanie mocy dla 10 różnych alternatyw i oszacowanie rozmiaru testu. Poziom istotności ustalamy $\alpha = 0.05$, Hipotezę zerową jako $\theta_0 = 0.5$, natomiast czas obserwacji wybieramy $t_0 = 3$.

```
symulacja_mocy <- function(n, t0, theta0, alternatywy, alpha = 0.05, B = 1000) {

  # Rozmiar testu (theta = theta0)
  odrzucenia_H0 <- 0
  for (b in 1:B) {
    dane <- generuj_typI(n, l = theta0, a = 1, t0 = t0)
    r <- sum(dane$delta == 1)
    s <- sum(dane$time[dane$delta == 1])
    test <- test_lrt_cenzI(r, s, n, t0, theta0, typ = "dwustronny", alpha = alpha)

    if (test$statystyka > test$krytyczny) odrzucenia_H0 <- odrzucenia_H0 + 1
  }
  rozmiar_testu <- odrzucenia_H0 / B
}
```

```

# Moc testu dla alternatyw
moc <- numeric(length(alternatywy))

for (i in seq_along(alternatywy)) {
  theta <- alternatywy[i]
  odrzucenia <- 0
  for (b in 1:B) {
    dane <- generuj_typI(n, l = theta, a = 1, t0 = t0)
    r <- sum(dane$delta == 1)
    s <- sum(dane$time[dane$delta == 1])
    test <- test_lrt_cenzI(r, s, n, t0, theta0, typ = "dwustronny", alpha = alpha)

    if (test$statystyka > test$krytyczny) odrzucenia <- odrzucenia + 1
  }
  moc[i] <- odrzucenia / B
}

return(list(
  rozmiar_testu = rozmiar_testu,
  moc = moc,
  alternatywy = alternatywy
))
}

theta0 <- 0.5
t0 <- 3
alternatywy <- seq(0.2, 0.8, length.out = 10) # 10 wybranych alternatyw
B <- 1000

set.seed(123)
wyniki20 <- symulacja_mocy(n = 20, t0 = t0, theta0 = theta0, alternatywy = alternatywy, B = B)

set.seed(123)
wyniki50 <- symulacja_mocy(n = 50, t0 = t0, theta0 = theta0, alternatywy = alternatywy, B = B)

```

Rozmiar testu dla $n = 20$ to 0.06

Tabela 12 przedstawia oszacowanie mocy testu dla różnych alternatyw.

Tabela 12: Moc testu LRT dla $n = 20$

theta	moc_testu
0.200	0.881
0.267	0.631
0.333	0.343
0.400	0.146
0.467	0.060
0.533	0.056
0.600	0.117
0.667	0.219
0.733	0.383
0.800	0.458

Rozmiar testu dla $n = 50$ to 0.046

Tabela 13 przedstawia oszacowanie mocy testu dla różnych alternatyw.

Tabela 13: Moc testu LRT dla $n = 50$

theta	moc_testu
0.200	1.000
0.267	0.956
0.333	0.677
0.400	0.291
0.467	0.068
0.533	0.073
0.600	0.199
0.667	0.469
0.733	0.708
0.800	0.883

W obu przypadkach rozmiar testu jest bliski założonemu $\alpha = 0.05$, ale dla większego n jest bliżej. Możemy również zauważyć, że testy mają większą moc dla θ , które są bardziej różne od wybranego przez nas $\theta_0 = 0.5$. Ponadto dla $n = 50$ testy mają znacznie większą moc niż dla $n = 20$.

4.3 Zadanie 3

Zweryfikujemy hipotezę, że czas do remisji wynosi $\theta_0 = 1$. Poziom istotności ustalamy $\alpha = 0.05$, a o czasie obserwacji wiemy, że wynosił $t_0 = 1$.

Tabela 14: Weryfikacja hipotezy o średnim czasie do remisji w grupach A i B

Grupa	$\hat{\theta}$	Statystyka	Wartość krytyczna	p -value	Odrzucamy H_0
A	0.70344	1.39624	3.84146	0.23735	FALSE
B	0.74302	0.97657	3.84146	0.32305	FALSE

W obu grupach p -wartość jest większa od ustalonego przez nas poziomu istotności $\alpha = 0.05$, zatem w żadnej z grup nie mamy podstaw do odrzucenia hipotezy zerowej.