

Sprawozdanie 1

Analiza przeżycia

Marta Stankiewicz (282244)

Spis treści

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

Spis rysunków

Spis tabel

1	Statystyki dla rozkładu Weibulla	5
2	Statystyki opisowe dla wygenerowanych danych cenzurowanych	7
3	Statystyki opisowe dla danych dotyczących leków	7
4	Realizacja Przedziału ufności dla średniego czasu do remisji choroby	8
5	Szacunki parametrów $\hat{\theta}$ i $\hat{\mu}$ oraz przedziały ufności $[TL, TU]$ dla leków A i B	8
7	Porównanie mocy testu dla różnych wartości theta i liczebności prób (n = 20, 50)	9

1 Lista 1

1.1 Zadanie 1

Gęstość funkcji

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

```
density_weib <- function(x, alpha, beta, gamma) {  
  dens <- ifelse(  
    x > 0,  
    (alpha * gamma / beta) *  
    (x / beta)^(alpha - 1) *  
    (1 - exp(-(x / beta)^alpha))^(gamma - 1) *  
    exp(-(x / beta)^alpha),  
    0  
  )  
  return(dens)  
}
```

Dystrybuanta funkcji

$$F(t, \alpha, \beta, \gamma) = \begin{cases} \left[1 - \exp\left(-\left(\frac{t}{\beta}\right)^{\alpha}\right)\right]^{\gamma}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

```
distrib_weib <- function(x, alpha, beta, gamma){
  dist <- (1-exp(-(x/beta)^alpha))^gamma
  return(dist)
}
```

Funkcja hazardu

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

```
haz_weib <- function(x, alpha, beta, gamma){
  haz <- density_weib(x, alpha, beta, gamma) /
    (1 - distrib_weib(x, alpha, beta, gamma))
  return(haz)
}
```

Funkcja kwantylowa (odwrotnej dystrybucyjnej)

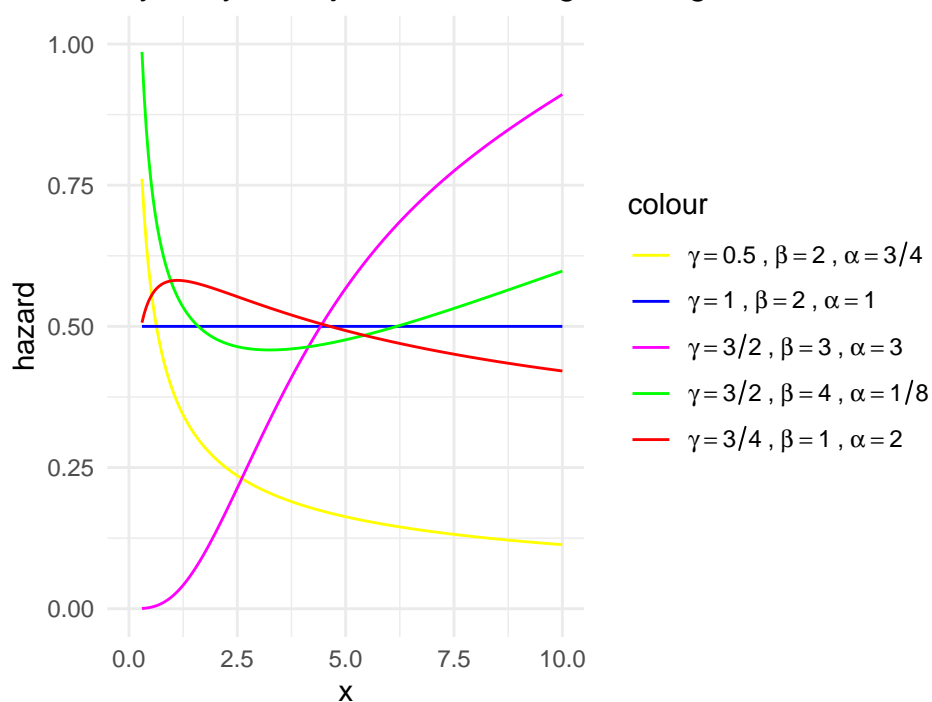
$$Q(p; \alpha, \beta, \gamma) = \beta \left[-\ln(1 - p^{1/\gamma}) \right]^{1/\alpha}, \quad 0 < p < 1,$$

```
quant_weib <- function(p, alpha, beta, gamma){
  quant <- beta * (-log(1 - p^(1 / gamma)))^(1 / alpha)
  return(quant)
}
```

Funkcje zostały napisane zgodnie z definicjami

1.2 Zadanie 2

Wykresy funkcji hazardu z uogólnionego rozkładu Weib



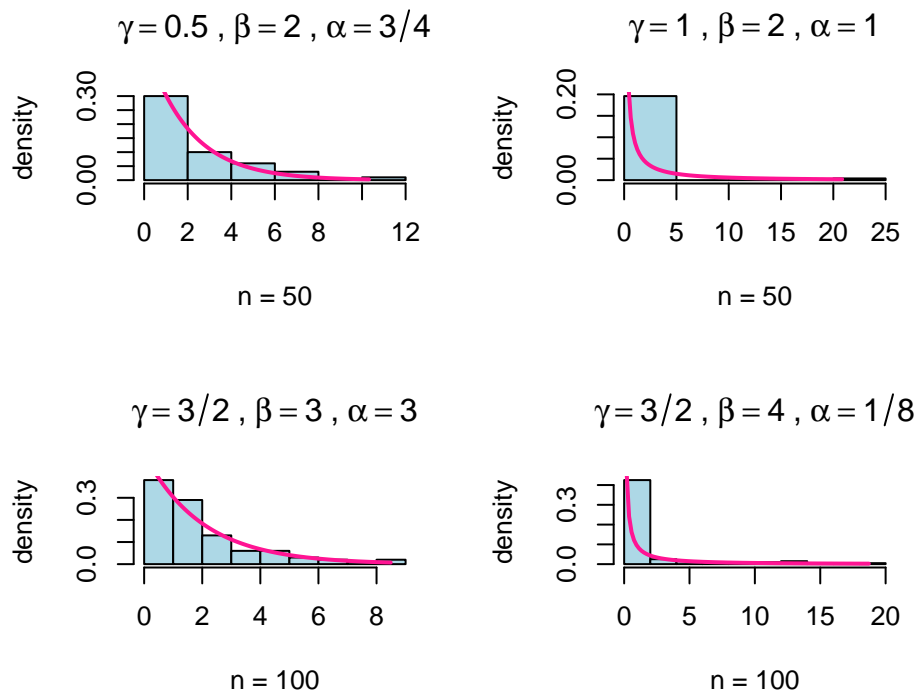
Wnioski : Wykresy dla odpowiednich parametrów są zgodne z oczekiwaniami

1.3 Zadanie 3

Dane są generowane przy użyciu metody odwrotnej dystrybucyjności (u nas funkcji kwatylowej)

```
number_generator <- function(n, alpha, beta, gamma){  
  u <- runif(n) # losujemy z U(0,1)  
  x <- quant_weib(u, alpha, beta, gamma)  
  return(x)  
}
```

1.4 Zadanie 4



Wnioski : Dla $n = 100$ histogramy są lepiej dopasowane do wykresów gęstości rozkładów. Co jest zgodne z intuicją, biorąc pod uwagę fakt, że dane są generowane z tych rozkładów.

1.5 Zadanie 5

Tabela 1: Statystyki dla rozkładu Weibulla

Args	Mean	Med_emp	Med	SD	Q1_em	Q1_tr	Q3_em	Q3_tr	Min	Max	rozstęp
Zestaw 1	2.17	1.40	1.39	2.15	0.63	0.58	3.04	2.77	0.05	10.32	10.27
Zestaw 2	0.90	0.03	0.06	3.02	0.00	0.00	0.74	0.78	0.00	20.85	20.85
Zestaw 3	1.99	1.39	1.39	1.85	0.73	0.58	2.49	2.77	0.02	8.49	8.47
Zestaw 4	1.37	0.04	0.06	3.35	0.00	0.00	0.64	0.78	0.00	18.72	18.72

Zestaw 1 : $\alpha = 1, \beta = 2, \gamma = 1, n = 50$ | Zestaw 3 : $\alpha = 1, \beta = 2, \gamma = 1, n = 100$

Zestaw 2 : $\alpha = \frac{1}{2}, \beta = 2, \gamma = \frac{3}{8}, n = 50$ | Zestaw 4 : $\alpha = \frac{1}{2}, \beta = 2, \gamma = \frac{3}{8}, n = 100$

Wnioski : Dla większej próby widać dokładniejsze kwantyle bardziej zbliżone do ich teoretycznej wartości, niż dla próby mniejszej. Jedynie w zestawie pierwszym, Q3 jest bliższe teoretycznej wartości niż w zestawie 3, jest to zapewne spowodowane losowością danych oraz dość małą próbą (50 i 100).

2 Lista 2

2.1 Zadanie 1

Funckja została napisana zgodnie z wytycznymi zadania

```
#-----Generatory-----
gen1 <- function(n, alpha, lambda, t0){
  y <- runif(n)
  x <- - (1/lambda)*log(1- y^(1/alpha))

  delta <- ifelse(x <= t0, 1, 0)
  x_obs <- pmin(x, t0)

  return(data.frame(x = x_obs, delta = delta))
}

gen2 <- function(n, alpha, lambda, m) {
  y <- runif(n)
  x <- - (1/lambda)*log(1- y^(1/alpha))

  x <- sort(x)
  x_cenz <- x[m]

  x <- c(x[1:m], rep(x_cenz, n-m))
  delta <- c(rep(1,m), rep(0, n-m))

  return(data.frame(x, delta))
}

gen3 <- function(n, alpha, lambda, eta){
  y <- runif(n)
  x <- - (1/lambda)*log(1- y^(1/alpha))
  c <- rexp(n, rate=1/eta)
  x <- pmin(x, c)
  delta <- ifelse(x==c, 0, 1)

  return(data.frame(x, delta))
}
#-----
```

2.2 Zadanie 2

Tabela 2: Statystyki opisowe dla wygenerowanych danych cenzurowanych

	Q1	Q2	Q3	RANGE	IQR	CENSORED	NCENSORED
Cenzura 1 typu	0.4996	0.7673	1.5163	2.0954	1.0167	0.0000	20.0000
Cenzura 2 typu	0.4463	1.1395	1.1807	1.1195	0.7344	8.0000	12.0000
Cenzura 3 typu	0.3205	0.4799	0.8105	4.3109	0.4899	7.0000	13.0000

Wnioski : Dane cenzurowane typu 1 i 2 mają podobne Q1 i Q2, różnice dopiero widać w Q3, dla typu drugiego jest ono niższe, co jest zgodne z intuicją. W pierwszym typie wszystkie dane są pełne za to w drugim na końcu pojawia się jakaś ilość danych cenzurowanych. W cenzurze 3 typu, czas do którego danego są cenzurowane generowany jest z rozkładu wykładniczego, czyli dane n pierwszych obserwacji są cenzurowane. Powoduje to zmniejszenie IQR, ze względu na zastąpienie brakujących danych czasem od którego obserwacje zaczęły być całkowite.

2.3 Zadanie 3

Tabela 3: Statystyki opisowe dla danych dotyczących leków

	Q1	Q2	Q3	RANGE	IQR
lek A	0.3753	0.9711	1.0000	0.9665	0.6247
lek B	0.3329	0.8456	1.0000	0.9621	0.6671

Wnioski: Q2 i zakres (range) są niższe w grupie B, co może wskazywać na szybsze działanie leku oraz mniejsze zróżnicowanie wyników związane z indywidualnymi predyspozycjami biologicznymi badanych osób. W obu grupach wartość Q3 wynosi 1, co jest zrozumiałe, ponieważ połowa obserwacji została ocenzurowana.

3 Lista 3

3.1 Zadanie 1a

Na podstawie danych opisanych w zadaniu 3 z listy 2, przyjmując, że są one realizacjami zmiennych z rozkładu wykładniczego można wyznaczyć oszacowania największej wiarygodności średniego czasu do remisji choroby dla pacjentów leczonych lekiem A oraz pacjentów leczonych lekiem B. W tym celu skorzystano ze wzorów podanych na Wykładzie 3:

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

gdzie

$$T_1 = \sum_{i=1}^R X_{(i)} + t_0(n - R)$$

gdzie zmienna losowa R jest losową liczbą obserwacji niecenzurowanych (kompletnych) w przedziale $(0, t_0]$. Oszacowanie NW średniego czasu do remisji choroby pacjentów leczonych lekiem A wynosi 0.7034 natomiast oszacowanie to w przypadku leku B wynosi 0.743

3.2 Zadanie 1b

Tabela 4: Realizacja Przedziału ufności dla średniego czasu do remisji choroby

Lek	Alpha	Realizacja.przedziału.ufności.
A	0.05	[0.7680, 3.1506]
A	0.01	[0.6560, 4.0720]
B	0.05	[0.7680, 3.1506]
B	0.01	[0.6560, 4.0720]

3.3 Zadanie 2

Tabela 5: Szacunki parametrów $\hat{\theta}$ i $\hat{\mu}$ oraz przedziały ufności $[TL, TU]$ dla leków A i B

Lek	α	$\hat{\theta}$	$\hat{\mu}$	T_L	T_U
A	0.05	0.7332	1.3638	0.3516	1.2527
A	0.01	0.7332	1.3638	0.2725	1.4664
B	0.05	0.9643	1.0371	0.4624	1.6474
B	0.01	0.9643	1.0371	0.3584	1.9284

3.4 Zadanie 3

n	t0	Bias_hat	MSE_hat	Bias_wave	MSE_wave
10	0.5	0.0694319	0.3162310	0.0809274	0.3430579
10	1.0	0.0602460	0.1863513	0.0849757	0.2252779
10	2.0	-0.0535022	0.0806411	-0.0816322	0.0622232
30	0.5	0.0224046	0.0886669	0.0246007	0.0917522
30	1.0	0.0251901	0.0580875	0.0330171	0.0671793
30	2.0	0.0228089	0.0411371	0.0522263	0.0677856

4 Lista 4

4.1 Zadanie 1

4.2 Zadanie 2

Tabela 7: Porównanie mocy testu dla różnych wartości theta i liczebności prób ($n = 20, 50$)

Theta	Moc ($n = 20$)	Moc ($n = 50$)
0.2	0.8074	0.9943
0.3	0.3982	0.7584
0.4	0.1228	0.2280
0.5	0.0552	0.0523
0.6	0.1050	0.1824
0.7	0.2414	0.4948
0.8	0.4173	0.8036
0.9	0.6120	0.9473
1.0	0.7688	0.9912
1.5	0.9961	1.0000
2.0	1.0000	1.0000

4.3 Zadanie 3

[1] 0.2373546

[1] 0.3230468