Kolokwium Nr 1 Forma Zdalna

Krystian Baran 145000

14 kwietnia 2021

Spis treści

1 Cześć I - 15 punktów

Modelem czasu zdatności T (w godz.) pewnych elementów jest nieujemna zmienna losowa o gęstości:

$$f_T(t) = At\exp(-0,0000005t^2)\mathbb{I}_{[0,\infty)}(t)$$

a) Rozpoznać rozkład oraz ustalić wartość stałej A oraz parametry tego rozkładu.

b) Wyznaczyć wartość oczekiwaną i drugi moment zwykły.

c) Wyznaczyć odchylenie standardowe i współczynnik zmienności.

d) Wyznaczyć dominantę czasu zdatności.

e) Wyznaczyć współczynnik skośności czasu zdatności.

f) Wyznaczyć dystrybuantę czasu zdatności T.

g) Obliczyć prawdopodobieństwa zdarzeń: $T > 500$, $|T - \mathbb{E}T| < \mathbb{D}T$.

h) Obliczyć prawdopodobieństwa zdarzeń: $T > 1500$, $(T > 1500|T > 1000)$.

i) Wyznaczyć funkcję kwantylową czasu zdatności T.

j) Wyznaczyć kwartyle oraz kwantyle rzędu 0,1 i 0,9.

k) Sporządzić krzywą gęstości i zaznaczyć na wykresie kwartyle, dominantę i wartość oczekiwaną.

l) Obliczyć dla jakiej wartości stałej a zachodzi równość $P(a < T < t_{0,95}) = 0,90$.

m) Przyjmując, że elementy są wycofywane z eksploatacji po uszkodzeniu lub przepracowaniu 2500 godzin obliczyć prawdopodobieństwo najbardziej prawdopodobnej liczby elementów sprawnych wśród 50 wycofanych z eksploatacji.

n) Przyjmując, że elementy po przepracowaniu 500 godzin poddawane są kontroli sprawności, obliczyć prawdopodobieństwo, że trzeci niesprawny element nie znajdzie się wśród pierwszych 100 sprawdzanych.

o) Ustalić najbardziej prawdopodobną liczbę sprawdzanych elementów do natrafienia na trzeci uszkodzony. Ile to prawdopodobieństwo wynosi?

1.1 a)

Podana funkcja gęstości przypomina mocno funkcje gęstości rozkładu Rayleigha która wygląda następująco:

$$f(x|\sigma) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, x \in [0, \infty)$$

Zatem stosując przekształcenia można wyznaczyć stałą A.

$$\begin{aligned} f_T(t) &= At \cdot e^{-0.5 \cdot 10^{-6} x^2} \\ &= At \cdot e^{-x^2/(2 \cdot 10^6)} = \frac{t}{\sigma^2} e^{-t^2/(2\sigma^2)} \end{aligned}$$

$$\sigma^2 = 10^6; A = 10^{-6}$$

Zatem podany rozkład jest rozkładem $Rayleigh(10^3)$, i stała A wynosi 0.0000001.

1.2 b)

Skorzystamy z gotowego wzoru na wartość oczekiwaną, natomiast wyprowadzimy wzór na drugi moment zwykły.

$$\mathbb{E}X = \sigma\sqrt{\frac{\pi}{2}}$$

$$\begin{aligned} \mathbb{E}(X^2) &= \int_{\mathbb{R}} \frac{x^3}{\sigma^2} e^{-x^2/(2\sigma^2)} \mathbb{I}_{[0,\infty]}(x) dx \\ &= \int_0^\infty \frac{x^3}{\sigma^2} e^{-x^2/(2\sigma^2)} dx \end{aligned}$$

	D	I
$+$	x^2	$\frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}$
$-$	$2x$	$-e^{-x^2/(2\sigma^2)}$

$$\begin{aligned} &= -x^2 e^{-x^2/(2\sigma^2)} + \int_0^\infty 2x e^{-x^2/(2\sigma^2)} dx \\ &= \lim_{a\to\infty} -x^2 e^{-x^2/(2\sigma^2)} - 2\sigma^2 e^{-x^2/(2\sigma^2)} \Big|_0^a \\ &= -0 + 0 - 0 + 2\sigma^2 = 2\sigma^2 \end{aligned}$$

Zatem podstawiając znany parametr $\sigma = 10^3$ otrzymujemy szukane wartości.

$$\begin{aligned} \mathbb{E}T &= 10^3\sqrt{\frac{\pi}{2}} \approx 1253.3141 \\ \mathbb{E}(T^2) &= 2 \cdot 10^6 = 2000000 \end{aligned}$$

1.3 c)

Korzystając z gotowego wzoru na wartość wariancji można obliczyć wartość odchylenia standardowego jako jej pierwiastek.

$$\mathbb{D}^2(T) = \frac{4-\pi}{2}\sigma^2 = \frac{4-\pi}{2}10^6 \approx 429203.6732$$
$$\mathbb{D}T = \sqrt{429203.6732} \approx 655.1364$$

Współczynnik zmienności (V) określa się jako stosunek odchylenia standardowego (σ_t) do średniej arytmetycznej próby ($\overline{x}$). Jego estymator (v) określa się jako stosunek odchylenia standardowego do wartości oczekiwanej.

$$V = \frac{\sigma_t}{\overline{x}} \sim v = \frac{\sigma_t}{\mathbb{E}T} = \frac{655.1364}{1253.3141} \approx 0.5227$$

1.4 d)

Korzystając z gotowego wzoru na Dominantę (Wartość modalną) rozkładu Rayleigha można łatwo tę wartość wyznaczyć:

$$Dominanta = \sigma = 10^3$$

Zatem wartość dominująca wynosi 1000.

1.5 e)

Współczynnik skośności określa się jako stosunek wartość oczekiwanej obniżonej o wartości modalnej (D) do odchylenia standardowego czyli:

$$A_d = \frac{\mathbb{E}T - D}{\sigma_t} = \frac{1253.3141 - 1000}{655.1364} = \frac{253.3141}{655.1364} \approx 0.3867$$

Zatem rozkład jest prawostronnie asymetryczny.

1.6 f)

Obliczymy dystrybuantę całkując funkcję gęstości:

$$\begin{aligned} F_T(x) &= \int_{\mathbb{R}} \frac{t}{\sigma^2} e^{-t^2/(2\sigma^2)} \mathbb{I}_{[0,\infty]}(t) dt \\ &= \int_0^x \frac{t}{\sigma^2} e^{-t^2/(2\sigma^2)} dt \\ &= -e^{-t^2/(2\sigma^2)} \Big|_0^x \\ &= -e^{-x^2/(2\sigma^2)} \end{aligned}$$

Gdzie $x \in [0, \infty)$.

1.7 g)

Mając dystrybuantę można łatwo obliczyć szukane prawdopodobieństwa, lub za pomocą oprogramowania R gdzie jest rozkład Rayleigha dostępny w pakiecie *extraDistr*.

$$P(T > 500) = 1 - P(T < 500) \stackrel{R}{=} 1 - prayleigh(500, 1000) \approx 0.8824969$$

$$\begin{aligned} P(|T - \mathbb{E}T| < \mathbb{D}T) &= P(-\mathbb{D}T < T - \mathbb{E}T < \mathbb{D}T) \\ &= P(\mathbb{E}T - \mathbb{D}T < T < \mathbb{E}T + \mathbb{D}T) \\ &= P(1253.3141 - 655.1364 < T < 1253.3141 + 655.1364) \\ &= P(598.1777 < T < 1908.4505) \stackrel{R}{=} prayleigh(1908.4505, 1000) - prayleigh(598.1777, 1000) \\ &\approx 0.6743336 \end{aligned}$$

Zatem jest bardzo prawdopodobne że element będzie sprawny przez więcej niż 500 godzin.

1.8 h)

Podobnie jak wcześniej łatwo obliczyć prawdopodobieństwo korzystając z dystrybuanty:

$$P(T > 1500) = 1 - P(T < 1500) \stackrel{R}{=} 1 - prayleigh(1500, 1000) \approx 0.3246525$$

Korzystając z definicji prawdopodobieństwa uwarunkowanego można wyliczyć szukane prawdopodobieństwo:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\begin{aligned} P(T > 1500|T > 1000) &= \frac{P(T > 1500 \wedge T > 1000)}{P(T > 1000)} \\ &= \frac{P(T > 1500)}{P(T > 1000)} \stackrel{R}{=} 0.3246525/(1 - prayleigh(1000, 1000)) \\ &\approx 0.5352615 \end{aligned}$$

Zatem prawdopodobieństwo że element będzie sprawny przez 1500 godzin pod warunkiem że był sprawny już przez 1000 godzin wynosi 0.54.

1.9 i)

Aby wyznaczyć funkcję kwantylową wystarczy obrócić już wyliczoną dystrybuantę, czyli:

$$\begin{aligned} y &= 1 - e^{-x^2/(2\sigma^2)} \\ 1 - y &= e^{-x^2/(2\sigma^2)} \\ \ln(1-y) &= -\frac{x^2}{2\sigma^2} \\ -2\sigma^2 \cdot \ln(1-y) &= x^2 \\ x &= \sqrt{-2\sigma^2 \cdot \ln(1-y)} = F_T^{-1}(y) \end{aligned}$$

Gdzie $y \in [0, 1]$.
Zatem dla naszego rozkładu funkcja kwantylowa jest następująca:

$$F_T^{-1}(y) = \sqrt{-2 \cdot 10^6 \cdot \ln(1-y)}$$

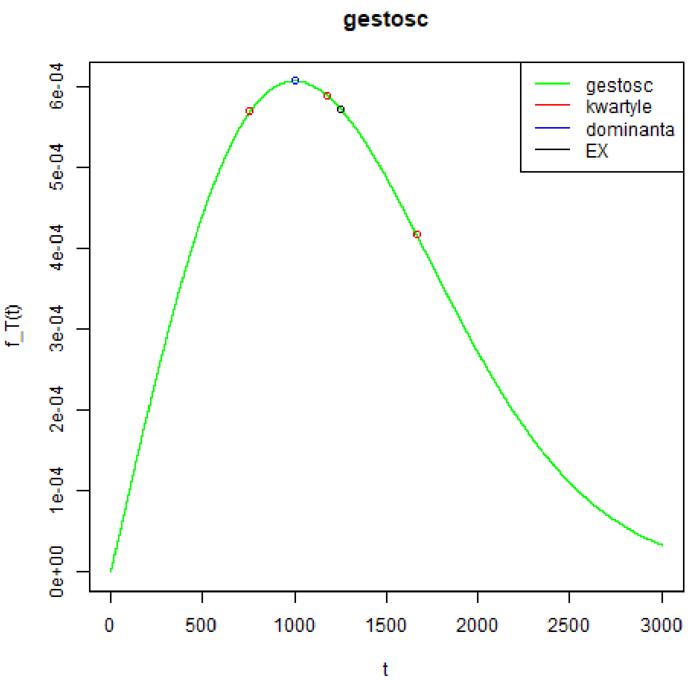
1.10 j)

Mając już obliczoną funkcję kwantylowa można wyznaczyć kwartyle i szukane kwantyle rzędu 0.1 i 0.9.
Natomiast skorzystamy z gotowej funkcji w R *qrayleigh(p, σ)*.

$$\begin{aligned} t_{0.25} &\stackrel{R}{=} qrayleigh(0.25, 1000) \approx 758.5276 \\ t_{0.5} &\stackrel{R}{=} qrayleigh(0.5, 1000) \approx 1177.41 \\ t_{0.75} &\stackrel{R}{=} qrayleigh(0.75, 1000) \approx 1665.109 \\ t_{0.1} &\stackrel{R}{=} qrayleigh(0.1, 1000) \approx 459.0436 \\ t_{0.9} &\stackrel{R}{=} qrayleigh(0.9, 1000) \approx 2145.966 \end{aligned}$$

1.11 k)

Poniżej przedstawiony został wykres gęstości w którym zaznaczono kwartyle kolorem czerwonym, dominantę kolorem niebieskim, i wartość oczekiwaną kolorem czarnym.



1.12 l)

Aby znaleźć wartość a, obliczymy kwantyl 0.95.

$$t_{0.95} = F_T^{-1}(0.95) \stackrel{R}{=} qrayleigh(0.95, 1000) \approx 2447.747$$

Następnie przekształcimy równanie tak aby wyznaczyć stałą a

$$\begin{aligned}
P(a < T < T_{0.95}) &= 0.9 \\
F_T(t_{0.95}) - F_T(a) &= 0.9 \\
F_T(a) &= F_T(t_{0.95}) - 0.9 \\
F_T(a) &= F_T(F_T^{-1}(0.95)) - 0.9 = 0.95 - 0.9 \\
F_T(a) &= 0.05 \\
a &= F_T^{-1}(0.05) \stackrel{R}{=} qrayleigh(0.05, 1000) \approx 320.2914
\end{aligned}$$

Zatem dla wartości $a = 320.2914$ spełniona jest podana równość.

1.13 m)

Załóżmy że każdy element pracuje niezależnie od każdego innego elementu. Obliczmy prawdopodobieństwo że element straci sprawność przed upływem 2500 godzin pracy, czyli:

$$P(T_i < 2500) \stackrel{R}{=} praylegh(2500, 1000) \approx 0.9560631 = p_i$$

Każdy element z pośród $n = 50$ odrzuconych ma prawdopodobieństwo p_i bycia uszkodzonym, a możliwe są tylko dwa wyniki, sukces lub porażka, zatem zmienną Y, opisującą liczbę uszkodzonych elementów z pośród 50 odrzuconych ma w przybliżeniu rozkład dwumianowy.

$$Y = \sum Y_i \sim b(50, 0.96)$$

Aby obliczyć najbardziej prawdopodobną liczbę, czyli wartość modalna sprawdzimy czy $(n + 1)p$ jest liczbą całkowitą.

$$(50 + 1) \cdot 0.96 = 48.96$$

Nie jest to liczba całkowita, zatem wartość modalna wynosi 48.
Wtedy można obliczyć prawdopodobieństwo najbardziej prawdopodobnej liczby jako:

$$P(Y = 2) = \binom{50}{48} 0.96^{48} \cdot 0.04^2 \stackrel{R}{=} choose(50, 48) * 0.96^4 8 * 0.04^2 \approx 0.2762328$$

Zatem najbardziej prawdopodobna liczba sprawnych elementów z pośród 50 odrzuconych wynosi $50 - 48 = 2$, a jej prawdopodobieństwo 0.28.

1.14 n)

Czas oczekiwania na $k - ty$ sukces opisuję się rozkładem Pascala. W naszym przypadku szukany jest trzeci sukces, gdzie sukces oznacz że element będzie nie sprawny. Aby obliczyć prawdopodobieństwo szukane potrzebujemy prawdopodobieństwo sukcesu, to znaczy prawdopodobieństwo że element straci zdatność w ciągu 500 godzin, czyli:

$$P(T < 500) = \stackrel{R}{=} prayleigh(500, 1000) \approx 0.1175031 \approx 0.12$$

Wtedy możemy wyznaczyć zmienną losową Z czasu oczekiwania za trzeci sukces jako, $Z \sim nbiom(x|3, 0.12), x \in \{0, 1, 2, \dots\}$.
Można teraz obliczyć prawdopodobieństwo że trzeci niesprawny element nie znajdzie się w pierwszych 100 jako:

$$P(Z > 100) = 1 - P(Z \leqslant 100) \stackrel{R}{=} 1 - nbinom(100, 3, 0.12) \approx 0.0002156435$$

Prawdopodobieństwo to jest bardzo małe, zatem można powiedzieć że prawie na pewno znajdziemy trzeci uszkodzony element w pierwszych 100 badanych.

1.15 o)

Korzystając z rozkładu przedstawionego w podpunkcie **n** i korzystając z gotowego wzoru na wartość modalną rozkładu Pascala można tę wartość łatwo obliczyć.

$$mo(Z) = \left\lfloor \frac{(k-1)\cdot(1-p)}{p} \right\rfloor = \left\lfloor \frac{2\cdot 0.88}{0.12} \right\rfloor \approx \lfloor 14.66666667 \rfloor = 14$$

Zatem trzeci niesprawny element znajdzie się, najbardziej prawdopodobnie w 14 próbie. Prawdopodobieństwo to wynosi:

$$P(Z=14) \stackrel{R}{=} dnbinom(14, 3, 0.12) \approx 0.03463238$$

2 Cześć II - 5 punktów

p) Modelem czasu zdatności X (w godz.) pewnych elementów jest dwuparametrowa rodzina rozkładów określona przez dystrybuantę

$$F(X; k, \lambda) = 1 - e^{-(x/\lambda)^k} \mathbb{I}_{[0,\infty)}(x)$$

Dla $x = x_1$ dystrybuanta przyjmuje wartość p_1 a dla $x = x_2 > x_1$ wartość $p_2 > p_1$, czyli $F(x_1; k, \lambda) = p_1$, $F(x_2; k, \lambda) = p_2$ Wyznaczyć parametry k, λ rozważanej rodziny rozkładów czasu zdatności elementów jako funkcji zmiennych x_1, x_2, p_1, p_2. Przyjąć pewne wartości tych zmiennych, obliczyć parametry i sporządzić krzywą gęstości.

q) Czas zdatności X elementów ma rozkład z punktu p), a wartość oczekiwana spełnia warunek $0 < a \leqslant \mathbb{E}X \leqslant b < \infty$. Oszacować parametr λ dla $k = 1, 2, 3, 4, 5$ oraz sporządzić krzywe gęstości i wykresy dystrybuant dla otrzymanych oszacowań.

r) Tym razem oprócz warunku $0 < a \leqslant \mathbb{E}X \leqslant b < \infty$ dodatkowo narzucony jest warunek na wariancję $\mathbb{D}^2T \leqslant c < \infty$. Czy przy tych warunkach można oszacować obydwa parametry czasu zdatności? Rozważyć szczególny przypadek.

2.1 p)

Podana dystrybuanta jest dystrybuantą rozkładu Weibulla. Aby wyznaczyć parametry λ i k z podanych warunków obrócimy najpierw dystrybuantę.

$$\begin{aligned}
y &= 1 - e^{-(x/\lambda)^k} \\
1 - y &= e^{-(x/\lambda)^k} \\
\ln(1-y) &= -\frac{x^k}{\lambda^k} \\
x^k &= -\lambda^k \ln(1-y) \\
x &= \lambda(-\ln(1-y))^{1/k} = F^{-1}(y)
\end{aligned}$$

Wtedy można wygodniej wyznaczyć szukane parametry:

$$\begin{cases} x_1 = \lambda(-\ln(1-p_1))^{1/k} \\ x_2 = \lambda(-\ln(1-p_2))^{1/k} \end{cases}$$
$$\begin{cases} \lambda = \frac{x_1}{(-\ln(1-p_1))^{1/k}} \\ \lambda = \frac{x_2}{(-\ln(1-p_2))^{1/k}} \end{cases}$$
$$\begin{cases} \lambda = \frac{x_1}{(-\ln(1-p_1))^{1/k}} \\ \frac{x_1}{(-\ln(1-p_1))^{1/k}} = \frac{x_2}{(-\ln(1-p_2))^{1/k}} \end{cases}$$
$$\begin{cases} \lambda = \frac{x_1}{(-\ln(1-p_1))^{1/k}} \\ \frac{x_1}{x_2} = \left(\frac{\ln(1-p_1)}{\ln(1-p_2)}\right)^{1/k} \end{cases}$$
$$\begin{cases} \lambda = \frac{x_1}{(-\ln(1-p_1))^{1/k}} \\ \ln(\frac{x_1}{x_2}) = \frac{1}{k}\ln(\frac{\ln(1-p_1)}{\ln(1-p_2)}) \end{cases}$$
$$\begin{cases} \lambda = \frac{x_1}{(-\ln(1-p_1))^{1/k}} \\ k = \ln(\frac{\ln(1-p_1)}{\ln(1-p_2)})/\big(\ln(x_1) - \ln(x_2)\big) \end{cases}$$
$$\begin{cases} \lambda = \frac{x_1}{(-\ln(1-p_1))^{(\ln(x_1)-\ln(x_2))/\ln(\frac{\ln(1-p_1)}{\ln(1-p_2)})}} \\ k = \ln(\frac{\ln(1-p_1)}{\ln(1-p_2)})/\big(\ln(x_1) - \ln(x_2)\big) \end{cases}$$

Zatem parametry są następujące:

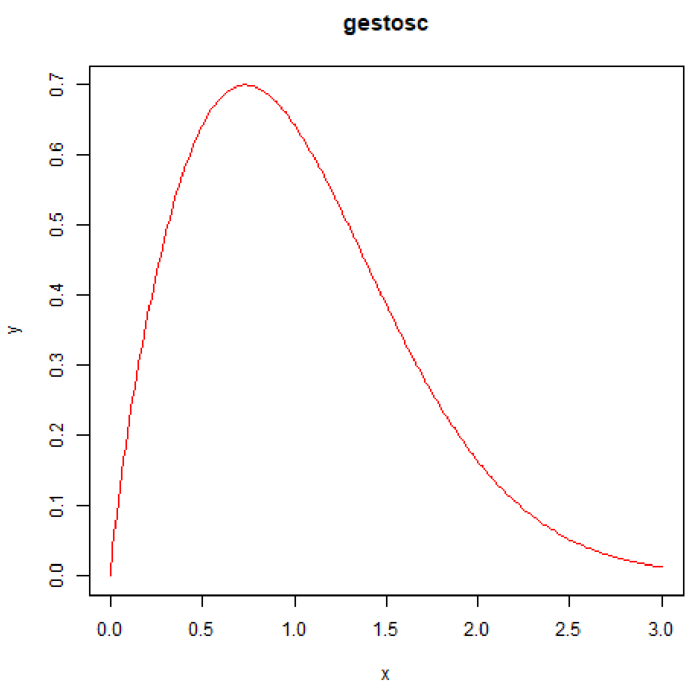
$$\lambda = \frac{x_1}{(-\ln(1-p_1))^{(\ln(x_1)-\ln(x_2))/\ln(\frac{\ln(1-p_1)}{\ln(1-p_2)})}}$$

$$k = \ln\left(\frac{\ln(1-p_1)}{\ln(1-p_2)}\right)/\big(\ln(x_1) - \ln(x_2)\big)$$

Załóżmy ze $x_1 = 0.5$, $x_2 = 1.5$, $p_1 = 0.2$, $p_2 = 0.8$. Wtedy krzywa gęstości wygląda następująco, a parametry obliczone w R są następujące:

$$\lambda \overset{R}{=} x1/(-log(1-p1, exp(1)))\hat{}(1/k) \approx 1.151264$$

$k \stackrel{R}{=} log(log(1-p1, exp(1))/log(1-p2, exp(1)))/(log(x1, exp(1)) - log(x2, exp(1)))$
≈ 1.798473



2.2 q)

Skorzystamy z funkcji wiarygodności rozkładu Weibulla.

$$\begin{aligned} L(x_1, x_2, \ldots, x_n|\lambda, k) &= \prod_{i=1}^{n} \frac{k}{\lambda}\left(\frac{x_i}{\lambda}\right)^{k-1} e^{-(x_i/\lambda)^k} \\ &= \frac{k^n}{\lambda^{kn}} e^{-\sum_{i=1}^{n}(x_i/\lambda)^k} \prod_{i=1}^{n} x_i^{k-1} \\ \ln(L(x_1, x_2, \ldots, x_n|\lambda, k)) &= n\ln(k) - kn\ln(\lambda) - \sum_{i=1}^{n}(x_i/\lambda)^k + (k-1)\sum_{i=1}^{n} x_i \end{aligned}$$

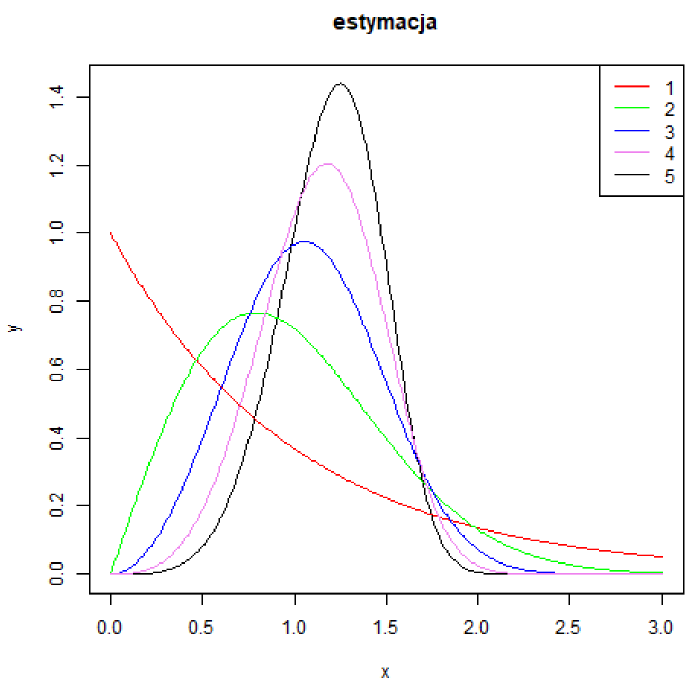
Pochodna tej funkcji po parametrze λ powinna się równać zero, czyli szukamy wartość maksymalną dla parametru λ.

$$\begin{aligned} \frac{d}{d\lambda}\ln(L) = 0 - \frac{kn}{\lambda} + \sum_{i=1}^{n} x_i^k \frac{k}{\lambda^{k+1}} + 0 = 0 \\ \frac{k}{\lambda^{k+1}} \sum_{i=1}^{n} x_i^k = \frac{kn}{\lambda} \\ \frac{\sum_{i=1}^{n} x_i^k}{n} = \lambda^k \\ \lambda = \left(\frac{\sum_{i=1}^{n} x_i^k}{n}\right)^{1/k} \end{aligned}$$

Skorzystamy z przykładowych czterech parametrów z poprzedniego podpunktu i dla każdego k dokonamy estymacje λ z wyprowadzonego wzoru. Uzyskane wartości są następujące:

k	λ
1	1
2	1.118034
3	1.205071
4	1.26522
5	1.306899

Wykresy estymowanych parametrów są następujące:



2.3 r)

Rozkład Weibulla jest rozkładem dwuparametrowym, zatem żeby oszacować parametry potrzebujemy co najmniej dwa momenty zwykłe, to znaczy wartość oczekiwana i drugi moment zwykły, ponieważ, dana nam jest wartość oczekiwana i Wariancja można obliczyć drugi moment zwykły z definicji wariancji.

Niech wartość oczekiwana i wariancja będą wynosić, odpowiednio $\mathbb{E}(X) = 7$, $\mathbb{D}^2(X) = 0.2$. Wzory na wariancje i wartość oczekiwaną są następujące:

$$\mathbb{E}(X) = \lambda\Gamma(1 + \frac{1}{k})$$

$$\mathbb{D}^2(X) = \lambda^2(\Gamma(1 + \frac{2}{k}) - \Gamma^2(1 + \frac{1}{k}))$$

Stosując przekształcenia można wyznaczyć szukane parametry.

$$\begin{cases} \lambda = \frac{7}{\Gamma(1+\frac{1}{k})} \\ \lambda = \sqrt{\frac{0.2}{\Gamma(1+\frac{2}{k})-\Gamma^2(1+\frac{1}{k})}} \end{cases}$$

$$\begin{cases} \lambda = \frac{7}{\Gamma(1+\frac{1}{k})} \\ \frac{7}{\Gamma(1+\frac{1}{k})} = \sqrt{\frac{0.2}{\Gamma(1+\frac{2}{k})-\Gamma^2(1+\frac{1}{k})}} \end{cases}$$

$$\begin{cases} \lambda = \frac{7}{\Gamma(1+\frac{1}{k})} \\ \frac{7}{\sqrt{0.2}} = \frac{\Gamma(1+\frac{1}{k})}{\sqrt{\Gamma(1+\frac{2}{k})-\Gamma^2(1+\frac{1}{k})}} \end{cases}$$

$$\begin{cases} \lambda = \frac{7}{\Gamma(1+\frac{1}{k})} \\ 0.06389 = \frac{\sqrt{\Gamma(1+\frac{2}{k})-\Gamma^2(1+\frac{1}{k})}}{\Gamma(1+\frac{1}{k})} \end{cases}$$

$$\begin{cases} \lambda = \frac{7}{\Gamma(1+\frac{1}{k})} \\ 0.06389 = \sqrt{\frac{\Gamma(1+\frac{2}{k})-\Gamma^2(1+\frac{1}{k})}{\Gamma^2(1+\frac{1}{k})}} \end{cases}$$

Dalsze rozważania nie wykonane na czas.

3 Bibliografia

- https://en.wikipedia.org/wiki/Rayleigh_distribution
- https://pl.wikipedia.org/wiki/Wsp%C3%B3%C5%82czynnik_zmienno%C5%9Bci
- https://pl.wikipedia.org/wiki/Wsp%C3%B3%C5%82czynnik_sko%C5%9Bno%C5%9Bci
- https://en.wikipedia.org/wiki/Negative_binomial_distribution
- https://en.wikipedia.org/wiki/Binomial_distribution