

Statystyka w finansach i ubezpieczeniach

Sprawozdanie 3

Justyna Niedźwiedzka 229877

10 lutego 2021

Część 1

W tej części sprawozdania wygenerujemy punktowy proces znakowany, w którym punkty i znaki są niezależne. Punkty na odcinku $[0, t_0]$ generowane są zgodnie z procesem Poissona, natomiast znaki generowane są z rozkładu wykładniczego $\mathcal{E}(\theta)$. Rozpatrzmy dwa sposoby generowania punktów:

- (a) proces Poissona o stałej intensywności λ ,
- (b) proces Poissona o intensywności $\lambda(t) = \alpha\beta t^{\beta-1}$.

Punktowy proces znakowany jest procesem stochastycznym złożonym z procesu punktowego T_1, T_2, \dots i związanych z nim marek (znaków) X_1, X_2, \dots , tzn. jest procesem postaci $(T_1, X_1), (T_2, X_2), \dots$

- (a) Punktowy proces znakowany o stałej intensywności λ .

W przypadku stałej intensywności korzystamy z następującego algorytmu:

1. Generujemy pojedynczą obserwację N z rozkładu Poissona $\mathcal{P}(t_0\lambda)$. Odpowiada ona za liczbę skoków.
2. Losujemy N obserwacji z rozkładu jednostajnego na odcinku $[0, t_0]$.
3. Sortujemy obserwacje.
4. Generujemy N obserwacji z rozkładu wykładniczego $\mathcal{E}(\theta)$.
5. Wyznaczamy skumulowaną sumę dla każdej obserwacji. Odpowiada ona za znaki.

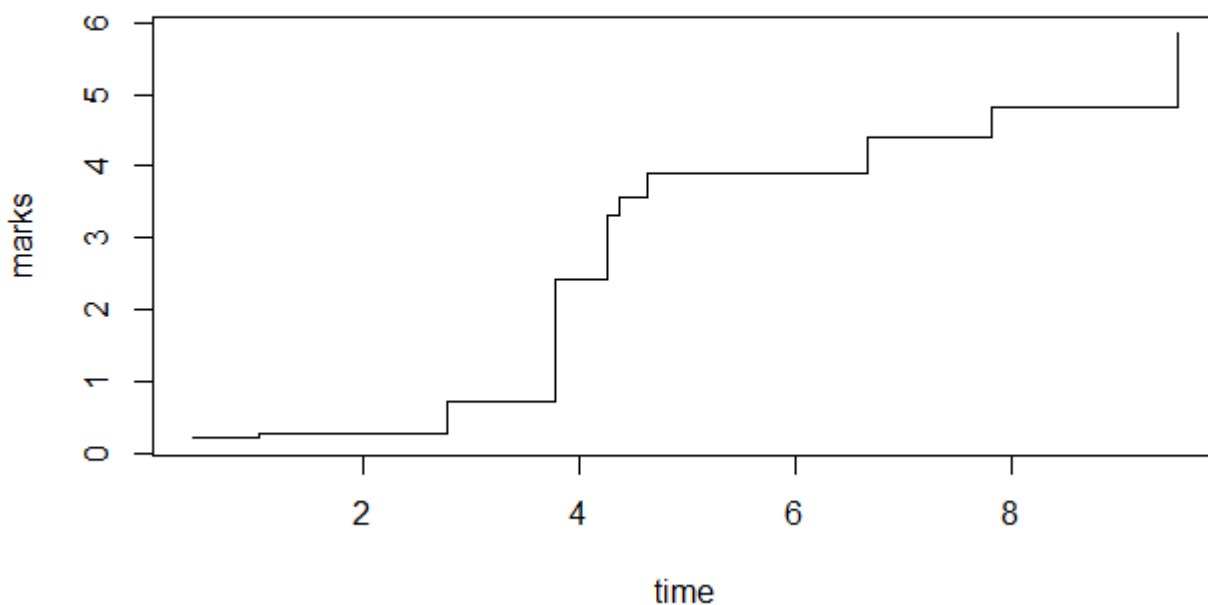
Poniżej przedstawiono funkcję generującą punktowy proces znakowany o stałej intensywności λ .

```
marked_process_a <- function(t0, lambda, theta){  
  N <- rpois(1, t0*lambda)  
  process <- sort(runif(N, 0, t0))  
  signs <- cumsum(rexp(n = length(process), rate = theta))  
  return(list(process, signs))  
}
```

Rozpatrzmy przypadek, gdy $t_0 = 10$, $\lambda = 1$ oraz $\theta = 2$. Zilustrujemy wygenerowaną realizację procesu na wykresie.

```
results1 <- marked_process_a(t0 = 10, lambda = 1, theta = 2)  
plot(x = results1[[1]], y = results1[[2]], xlab = "time", ylab = "marks",  
     main = "Punktowy proces znakowany", type = 's')
```

Punktowy proces znakowany



Rysunek 1: Punktowy proces znakowany o stałej intensywności λ dla $t_0 = 10$, $\lambda = 1$, $\theta = 2$

(b) Punktowy proces znakowany o intensywności $\lambda(t) = \alpha\beta t^{\beta-1}$.

Teraz rozpatrzmy przypadek, gdy punktowy proces znakowany ma intensywność daną przez $\lambda(t) = \alpha\beta t^{\beta-1}$.

Niech $\Lambda(t) = \int_0^t \lambda(u) du$. Po obliczeniu całki otrzymujemy $\Lambda(t) = \alpha t^\beta$. Wyznaczamy też funkcję odwrotną i otrzymujemy $\Lambda^{-1}(t) = \left(\frac{t}{\alpha}\right)^{\frac{1}{\beta}}$.

Poniżej przedstawiono zaimplementowane funkcje $\Lambda(t)$ oraz $\Lambda^{-1}(t)$. Będą one potrzebne do wygenerowania procesu o zmiennej intensywności.

```
lambda <- function(t0, alpha, beta){
  return(alpha*t0^(beta))
}

lambda_inverse <- function(alpha, beta){
  funkcja <- function(t){
    return((t/alpha)^(1/beta))
  }
}
```

W celu wygenerowania punktowego procesu znakowanego o intensywności $\lambda(t) = \alpha\beta t^{\beta-1}$ korzystamy z następującego algorytmu:

1. Generujemy pojedynczą obserwację N z rozkładu Poissona $\mathcal{P}(\Lambda(t_0))$. Odpowiada ona za liczbę skoków.
2. Losujemy N obserwacji z rozkładu jednostajnego na odcinku $[0, \Lambda(t_0)]$.
3. Sortujemy obserwacje.
4. Generujemy N obserwacji z rozkładu wykładniczego $\mathcal{E}(\theta)$.
5. Wyznaczamy skumulowaną sumę dla każdej obserwacji. Odpowiada ona za znaki.

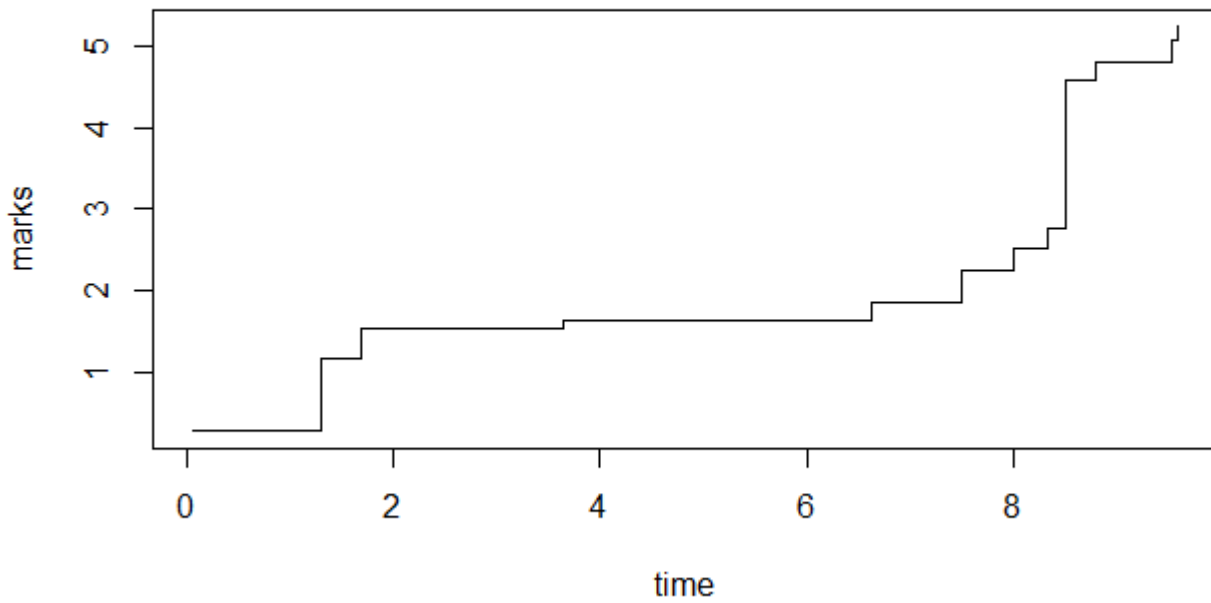
Poniżej przedstawiono funkcję generującą punktowy proces znakowany o intensywności $\lambda(t) = \alpha\beta t^{\beta-1}$.

```
marked_process_b <- function(t0, alpha, beta, theta){  
  N <- rpois(1, lambda(t0, alpha, beta))  
  sorted <- sort(runif(N, 0, lambda(t0, alpha, beta)))  
  process <- lapply(sorted, lambda_inverse(alpha, beta))  
  signs <- cumsum(rexp(n = length(process), rate = theta))  
  return(list(process, signs))  
}
```

W tym przypadku także rozpatrujemy parametry $t_0 = 10$, $\lambda = 1$ oraz $\theta = 2$. Wygenerowaną realizację procesu przedstawiono na poniższym wykresie.

```
results2 <- marked_process_b(t0 = 10, theta = 2, alpha = 1, beta = 1)  
plot(x = results2[[1]], y = results2[[2]], xlab = "time", ylab = "marks",  
     main = "Punktowy proces znakowany", type = 's')
```

Punktowy proces znakowany



Rysunek 2: Punktowy proces znakowany o intensywności $\lambda(t) = \alpha\beta t^{\beta-1}$ dla $t_0 = 10$, $\lambda = 1$, $\theta = 2$

W obu sposobach generowania punktów procesy są rosnące i zawarte w przedziale $[0, 10]$.

Część 2

W tej części zbadamy jakość aproksymacji złożonego rozkładu Poissona rozkładem normalnym oraz własności przedziałów predykcyjnych skonstruowanych w oparciu o aproksymację złożonego rozkładu Poissona rozkładem normalnym i estymatory nieznanymi parametrów.

Zadanie 1

Niech X_i , $i = 1, 2, \dots$ będą niezależnymi zmiennymi losowymi z rozkładu logarytmiczno-normalnego $\mathcal{LN}(\mu, \sigma^2)$ a N ma rozkład Poissona $\mathcal{P}(\lambda)$.

Poniżej przedstawiono zaimplementowaną funkcję, która dla zadanych wartości parametrów μ , σ , λ oraz zadanego poziomu istotności α wyznacza przedział (a, b) taki, że

$$P\left(\sum_{i=1}^N X_i \in (a, b)\right) \approx 1 - \alpha.$$

Końce przedziału ufności (a, b) wyznaczamy ze wzorów:

$$a = \lambda \mathbb{E}(X_1) + z(\alpha_1) \sqrt{\lambda \mathbb{E}(X_1^2)},$$
$$b = \lambda \mathbb{E}(X_1) + z(1 - \alpha_2) \sqrt{\lambda \mathbb{E}(X_1^2)},$$

gdzie $z(q)$ oznacza kwantyl rzędu q standardowego rozkładu normalnego $\mathcal{N}(0, 1)$.

We wszystkich poniższych analizach będziemy rozpatrywać przypadek, gdy $\alpha_1 = \alpha_2$. Oczywiście $\alpha = \alpha_1 + \alpha_2$.

```
confidence_interval <- function(mi, sigma, lambda, alpha){
  repeat{
    N <- rpois(1, lambda)
    if (N != 0)
      break
  }
  X <- rlnorm(N, meanlog = mi, sdlog = sigma)
  Sn <- sum(X)
  EX <- exp(mi + 1/2*sigma^2)
  VarX <- (exp(sigma^2) - 1)*exp(2*mi + sigma^2)
  EX2 <- VarX + EX^2
  a <- lambda*EX + qnorm(alpha/2)*sqrt(lambda*EX2)
  b <- lambda*EX + qnorm(1 - alpha/2)*sqrt(lambda*EX2)
  return(c(Sn, a, b))
}
```

Funkcja ta zwraca wartość $S_N = \sum_{i=1}^N X_i$ oraz końce przedziału ufności (a, b) .

Poniżej przedstawiono funkcję, która dla wybranych wartości parametrów μ , σ , λ , α oraz liczby powtórzeń M szacuje prawdopodobieństwo pokrycia wartości $S_N = \sum_{i=1}^N X_i$ przez przedział (a, b) . Funkcja zwraca średnią wartość $\bar{S}_N = \frac{1}{N} \sum_{i=1}^N X_i$, końce przedziału ufności (a, b) oraz prawdopodobieństwo pokrycia.

```
coverage_probability <- function(mi, sigma, lambda, alpha, M){
  Sn <- c()
  a <- c()
  b <- c()
  for (i in 1:M){
    conf_int <- confidence_interval(mi, sigma, lambda, alpha)
    Sn[i] <- conf_int[1]
    a <- conf_int[2]
    b <- conf_int[3]
  }
  logical <- between(Sn, left = a, right = b)
  prob <- sum(logical)/M
  return(c(mean(Sn), a, b, prob))
}
```

Najpierw ustalamy parametry $\mu = 1$, $\sigma = 0.5$, $\lambda = 10$, $\alpha = 0.05$ i rozpatrujemy przypadki, gdy $M = 100$, $M = 1000$ oraz $M = 10000$.

```
cov_prob1 <- coverage_probability(mi = 1, sigma = 0.5, lambda = 10,
                                   alpha = 0.05, M = 100)
cov_prob2 <- coverage_probability(mi = 1, sigma = 0.5, lambda = 10,
                                   alpha = 0.05, M = 1000)
cov_prob3 <- coverage_probability(mi = 1, sigma = 0.5, lambda = 10,
                                   alpha = 0.05, M = 10000)
```

Wyniki w zależności od parametru M przedstawiono w tabeli 1.

Tabela 1: Średnia S_N , końce przedziału ufności (a, b) oraz prawdopodobieństwo pokrycia w zależności od ilości powtórzeń M

	$\bar{S}_N = \frac{1}{N} \sum_{i=1}^N X_i$	a	b	prawdopodobieństwo pokrycia
M = 100	32.285442	9.169196	52.435141	0.960000
M = 1000	30.764061	9.169196	52.435141	0.957000
M = 10000	30.817272	9.169196	52.435141	0.953500

Możemy zauważyć, że im większe M , tym prawdopodobieństwo pokrycia jest bliższe $1 - \alpha = 0.95$.

Teraz przeprowadzimy analizę w przypadku, gdy ustalone są parametry $\mu = 1$, $\sigma = 0.5$, $\alpha = 0.05$, $M = 1000$, a parametr λ jest zmienny. Rozpatrujemy $\lambda = 20$, $\lambda = 30$, $\lambda = 40$ oraz $\lambda = 100$.

```
cov_prob_lambda1 <- coverage_probability(mi = 1, sigma = 0.5, lambda = 20,
                                           alpha = 0.05, M = 1000)
cov_prob_lambda2 <- coverage_probability(mi = 1, sigma = 0.5, lambda = 40,
                                           alpha = 0.05, M = 1000)
cov_prob_lambda3 <- coverage_probability(mi = 1, sigma = 0.5, lambda = 100,
                                           alpha = 0.05, M = 1000)
```

Wyniki w zależności od parametru λ przedstawiono w tabeli 2.

Tabela 2: Średnia S_N , końce przedziału ufności (a, b) oraz prawdopodobieństwo pokrycia w zależności od parametru λ

	$\bar{S}_N = \frac{1}{N} \sum_{i=1}^N X_i$	a	b	prawdopodobieństwo pokrycia
$\lambda = 20$	61.91463	31.01069	92.19798	0.96000
$\lambda = 40$	122.21623	79.94273	166.47462	0.96300
$\lambda = 100$	307.1704	239.6122	376.4311	0.95200

Rozważmy jeszcze przypadek, gdy parametr α jest zmienny, a pozostałe parametry są ustalone i wynoszą $\mu = 1$, $\sigma = 0.5$, $\lambda = 10$, $M = 1000$. Rozpatrujemy $\alpha = 0.005$, $\alpha = 0.01$ oraz $\alpha = 0.1$.

```
cov_prob_alpha1 <- coverage_probability(mi = 1, sigma = 0.5, lambda = 10,
                                          alpha = 0.005, M = 1000)
cov_prob_alpha2 <- coverage_probability(mi = 1, sigma = 0.5, lambda = 10,
                                          alpha = 0.01, M = 1000)
cov_prob_alpha3 <- coverage_probability(mi = 1, sigma = 0.5, lambda = 10,
                                          alpha = 0.1, M = 1000)
```

Wyniki w zależności od parametru α przedstawiono w tabeli 3.

Tabela 3: Średnia S_N , końce przedziału ufności (a, b) oraz prawdopodobieństwo pokrycia w zależności od parametru α

	$\bar{S}_N = \frac{1}{N} \sum_{i=1}^N X_i$	a	b	prawdopodobieństwo pokrycia
$\alpha = 0.005$	30.3259016	-0.1802802	61.7846172	0.99500
$\alpha = 0.01$	30.582389	2.371624	59.232713	0.99100
$\alpha = 0.1$	30.61544	12.64721	48.95713	0.90000

Zarówno w przypadku zmiennego parametru λ jak i α otrzymane prawdopodobieństwa pokrycia są bliskie wartości $1 - \alpha$. Możemy zauważyć, że wraz ze wzrostem λ szerokość przedziałów ufności rośnie, natomiast im większy parametr α , tym mniejsza szerokość przedziałów ufności.

Zadanie 2

Niech X_i , $i = 1, 2, \dots$ będą niezależnymi zmiennymi losowymi z rozkładu logarytmiczno-normalnego $\mathcal{LN}(\mu, \sigma^2)$ oraz N ma rozkład Poissona $\mathcal{P}(\lambda)$. Przyjmując, że wartości parametrów nie są znane, napiszemy program, który na podstawie realizacji zmiennej losowej N oraz realizacji zmiennych X_1, \dots, X_N będzie wyznaczał przedział (a', b') taki, że

$$P\left(\sum_{i=1}^{N'} X'_i \in (a', b')\right) \approx 1 - \alpha,$$

gdzie N' , X'_i są zmiennymi losowymi o rozkładach odpowiednio jak N i X_i .

```
confidence_interval2 <- function(mi, sigma, lambda, alpha){
  repeat{
    N <- rpois(1, lambda)
    if (N != 0)
      break
  }
  X <- rlnorm(N, meanlog = mi, sdlog = sigma)
  mi_est <- mean(log(X))
  if (length(X) == 1){
    sigma_est <- 0
  } else {
    sigma_est <- sd(log(X))
  }
  lambda_est <- mean(N)
  EX <- exp(mi_est + 1/2*sigma_est^2)
  VarX <- (exp(sigma_est^2) - 1)*exp(2*mi_est + sigma_est^2)
  EX2 <- VarX + EX^2
  repeat{
    N_prim <- rpois(1, lambda_est)
    if (N_prim != 0)
      break
  }
  X_prim <- rlnorm(N_prim, meanlog = mi_est, sdlog = sigma_est)
  Sn_prim <- sum(X_prim)
  a_prim <- lambda_est*EX + qnorm(alpha/2)*sqrt(lambda_est*EX2)
  b_prim <- lambda_est*EX + qnorm(1 - alpha/2)*sqrt(lambda_est*EX2)
  return(c(Sn_prim, a_prim, b_prim, mi_est, sigma_est, lambda_est))
}
```

Powyższa funkcja zwraca wartość $S'_{N'} = \sum_{i=1}^{N'} X'_i$, końce przedziału ufności (a', b') oraz estymowane parametry μ , σ i λ . Estymując nieznanne parametry korzystamy ze związku rozkładu logarytmiczno-normalnego z rozkładem normalnym, tzn. jeśli zmienna losowa X ma rozkład $\mathcal{LN}(\mu, \sigma^2)$, to zmienna losowa $Y = \ln X$ ma rozkład $\mathcal{N}(\mu, \sigma^2)$.

Zaimplementowano też funkcję, która dla wybranych wartości parametrów μ , σ , λ , α oraz liczby powtórzeń M szacuje prawdopodobieństwo pokrycia wartości $S'_{N'} = \sum_{i=1}^{N'} X'_i$ przez przedział (a', b') . Funkcja zwraca średnią wartość $S'_{N'} = \frac{1}{N'} \sum_{i=1}^{N'} X'_i$, końce przedziału ufności (a', b') oraz prawdopodobieństwo pokrycia.

```
coverage_probability2 <- function(mi, sigma, lambda, alpha, M){
  Sn_prim <- c()
  logical <- c()
  a_prim <- c()
  b_prim <- c()
  for(i in (1:M)){
    conf_int <- confidence_interval2(mi, sigma, lambda, alpha)
    a_prim <- conf_int[2]
    b_prim <- conf_int[3]
    mi_est <- conf_int[4]
    sigma_est <- conf_int[5]
    lambda_est <- conf_int[6]
    repeat{
      N_prim <- rpois(1, lambda_est)
      if (N_prim != 0)
        break
    }
    X_prim <- rlnorm(N_prim, meanlog = mi_est, sdlog = sigma_est)
    Sn_prim[i] <- sum(X_prim)
    logical[i] <- between(Sn_prim[i], left = a_prim, right = b_prim)
  }
  prob <- sum(logical)/M
  return(c(mean(Sn_prim), a_prim, b_prim, prob))
}
```

Najpierw przeprowadzimy analizę w zależności od parametru M i przy ustalonych $\mu = 1$, $\sigma = 0.5$, $\lambda = 10$ oraz $\alpha = 0.05$. Będziemy rozpatrywać $M = 100$, $M = 1000$ i $M = 10000$.

```
cov_prob4 <- coverage_probability2(mi = 1, sigma = 0.5, lambda = 10,
                                   alpha = 0.05, M = 100)
cov_prob5 <- coverage_probability2(mi = 1, sigma = 0.5, lambda = 10,
                                   alpha = 0.05, M = 1000)
cov_prob6 <- coverage_probability2(mi = 1, sigma = 0.5, lambda = 10,
                                   alpha = 0.05, M = 10000)
```

Wyniki w zależności od parametru M przedstawiono w tabeli 4.

Tabela 4: Średnia $S'_{N'}$, końce przedziału ufności (a', b') oraz prawdopodobieństwo pokrycia w zależności od ilości powtórzeń M

	$S'_{N'} = \frac{1}{N'} \sum_{i=1}^{N'} X'_i$	a'	b'	prawdopodobieństwo pokrycia
$M = 100$	29.771690	8.703794	58.161674	0.97000
$M = 1000$	31.390354	6.057107	46.519413	0.95200
$M = 10000$	31.14648	13.24067	62.42405	0.95290

Ogólnie im większy parametr M , tym prawdopodobieństwo pokrycia powinno być bliższe wartości $1 - \alpha = 0.95$. Tutaj w przypadku $M = 10000$ otrzymaliśmy wynik gorszy niż dla $M = 1000$, ale różnica jest niewielka. Jest to spowodowane tym, jakie próby X'_i , $i = 1, \dots, N'$ zostały wygenerowane.

Rozpatrzmy jeszcze prawdopodobieństwo pokrycia w zależności od parametru λ przy ustalonych $\mu = 1$, $\sigma = 0.5$, $\alpha = 0.05$, $M = 1000$. Rozważamy przypadki $\lambda = 20$, $\lambda = 40$ oraz $\lambda = 100$.

```
cov_prob_lambda4 <- coverage_probability2(mi = 1, sigma = 0.5, lambda = 20,
                                           alpha = 0.05, M = 1000)
cov_prob_lambda5 <- coverage_probability2(mi = 1, sigma = 0.5, lambda = 40,
                                           alpha = 0.05, M = 1000)
cov_prob_lambda6 <- coverage_probability2(mi = 1, sigma = 0.5, lambda = 100,
                                           alpha = 0.05, M = 1000)
```

Wyniki w zależności od parametru λ przedstawiono w tabeli 5.

Tabela 5: Średnia $S'_{N'}$, końce przedziału ufności (a', b') oraz prawdopodobieństwo pokrycia w zależności od parametru λ

	$S'_{N'} = \sum_{i=1}^{N'} X'_i$	a'	b'	prawdopodobieństwo pokrycia
$\lambda = 20$	62.57872	32.23622	101.77715	0.94500
$\lambda = 40$	123.55027	93.06567	189.85150	0.95600
$\lambda = 100$	310.9025	238.4033	372.7215	0.94700

Tutaj także widzimy, że w przypadku zmiennego parametru λ prawdopodobieństwo pokrycia jest bliskie wartości $1 - \alpha = 0.95$. Różnica jest tylko w końcach przedziałów ufności oraz ich szerokościach. Możemy zauważyć, że szerokość przedziału ufności rośnie wraz ze wzrostem parametru λ .

Część 3

Zadanie 1

Wiemy, że składka netto kontraktu *stop-loss* jest równa

$$H(R) = E(R) = \int_d^\infty (1 - F_{S_n}(s)) ds,$$

gdzie F_{S_n} jest dystrybucją rozkładu S_n . Dla ryzyka wykładniczego, czyli gdy $S_n \sim \mathcal{E}(\lambda)$ mamy

$$H(R) = \int_d^\infty e^{-\lambda s} ds = \frac{1}{\lambda} e^{-\lambda d}.$$

Przyjmując, że λ jest nieznane, zaproponujemy oszacowanie punktowe i przedziałowe (na poziomie ufności w przybliżeniu $1 - \alpha$) wartości $H(R)$ na podstawie m realizacji $s_n^{(1)}, \dots, s_n^{(m)}$ zmiennych losowych $S_n^{(1)}, \dots, S_n^{(m)}$.

Oszacowanie punktowe wartości $H(R)$ wyznaczymy korzystając z faktu, że estymatorem parametru $\frac{1}{\lambda}$ jest średnia z próby m realizacji $s_n^{(1)}, \dots, s_n^{(m)}$ zmiennych losowych $S_n^{(1)}, \dots, S_n^{(m)}$, ponieważ $S_n \sim \mathcal{E}(\lambda)$. W oszacowaniu przedziałowym skorzystamy z metody delta.

Twierdzenie 1 (Metoda delta) Niech (X_n) będzie ciągiem zmiennych losowych i niech (c_n) będzie ciągiem liczbowym rozbieżnym do ∞ . Przypuśćmy, że istnieją zmienna losowa V i liczba rzeczywista a takie, że

$$c_n(X_n - a) \xrightarrow{D} V.$$

Jeśli funkcja $g : \mathbb{R} \rightarrow \mathbb{R}$ jest różniczkowalna w punkcie a i $g'(a) \neq 0$, to

$$c_n[g(X_n) - g(a)] \xrightarrow{D} g'(a) \cdot V.$$

W naszym przypadku

$$\begin{aligned} \sqrt{n}[X_n - \theta] &\xrightarrow{D} \mathcal{N}(0, \sigma^2), \\ \sqrt{n}[g(X_n) - g(\theta)] &\xrightarrow{D} \mathcal{N}(0, \sigma^2 \cdot [g'(\theta)]^2). \end{aligned}$$

Średnia w rozkładzie wykładniczym wynosi $\frac{1}{\lambda}$. Oznaczmy $\theta = \frac{1}{\lambda}$.

$$\begin{aligned} g(\lambda) &= \frac{1}{\lambda} e^{-\lambda d}, \\ g(\theta) &= \theta e^{-\frac{d}{\theta}}, \\ g'(\theta) &= e^{-\frac{d}{\theta}} \left(1 + \frac{d}{\theta}\right) \neq 0, \end{aligned}$$

zatem spełnione są założenia Twierdzenia (1).

Estymatorem parametru θ jest średnia, więc

$$\hat{\theta} = \bar{X}_n \sim \mathcal{AN}(\theta, \frac{1}{n}\theta^2).$$

Korzystając z Twierdzenia (1)

$$g(\bar{X}_n) \sim \mathcal{AN}(g(\theta), \frac{1}{n}\theta^2 \cdot [g'(\theta)]^2).$$

Końce przedziału ufności wyznaczmy przekształcając poniższe wyrażenie

$$P\left(z_{\frac{\alpha}{2}} \leq \frac{g(\bar{X}_n) - g(\theta)}{\sqrt{\frac{1}{n}\theta^2 \cdot [g'(\theta)]^2}} \leq z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha.$$

Zastępując θ przez \bar{X}_n otrzymujemy ostatecznie

$$P\left(g(\bar{X}_n) - \frac{z_{1-\frac{\alpha}{2}} \cdot \bar{X}_n \cdot |g'(\bar{X}_n)|}{\sqrt{n}} \leq g(\theta) \leq g(\bar{X}_n) - \frac{z_{\frac{\alpha}{2}} \cdot \bar{X}_n \cdot |g'(\bar{X}_n)|}{\sqrt{n}}\right) \approx 1 - \alpha.$$

Zatem końce przedziału ufności (a, b) są postaci

$$a = g(\bar{X}_n) - \frac{z_{1-\frac{\alpha}{2}} \cdot \bar{X}_n \cdot |g'(\bar{X}_n)|}{\sqrt{n}},$$

$$b = g(\bar{X}_n) - \frac{z_{\frac{\alpha}{2}} \cdot \bar{X}_n \cdot |g'(\bar{X}_n)|}{\sqrt{n}}.$$

Poniżej przedstawiono funkcję, która szacuje punktowo i przedziałowo wartości $H(R)$.

```
confidence_interval3 <- function(alpha, lambda, d, m){
  Sm <- rexp(m, rate = lambda)
  mean <- mean(Sm)
  g <- mean*exp(-d/mean)
  g_prim <- exp(-d/mean)*(1+d/mean)
  a <- max(0, g - (qnorm(1-alpha/2)*mean*abs(g_prim))/sqrt(m))
  b <- g - (qnorm(alpha/2)*mean*abs(g_prim))/sqrt(m)
  return(c(g, a, b))
}
```

Zadanie 2

Oszacujemy teraz prawdopodobieństwa pokrycia nieznanej wartości $H(R)$ przez zaproponowany przedział ufności. Przeprowadzimy symulacje z parametrami $\lambda = d = 1$, $m = 30$. W tym celu skorzystamy z zaimplementowanej funkcji. Zwraca ona rzeczywistą wartość, średnią estymowaną punktowo wartość, końce przedziału ufności (a, b) oraz prawdopodobieństwo pokrycia.

```
coverage_probability3 <- function(alpha, lambda, d, m, M){
  real_value <- exp(-d/lambda)*lambda
  g <- c()
  logical <- c()
  for(i in (1:M)){
    conf_int <- confidence_interval3(alpha, lambda, d, m)
    a <- conf_int[2]
    b <- conf_int[3]
    g[i] <- conf_int[1]
    logical[i] <- between(real_value, left = a, right = b)
  }
  prob <- sum(logical)/M
  return(c(real_value, mean(g), a, b, prob))
}
```

Najpierw przeprowadzimy analizę w zależności od parametru M i ustalonego α , a później ustalimy M i będziemy zmieniać poziom istotności.

Wyniki w zależności od parametru M przedstawiono w tabeli 6. Przyjmujemy $\alpha = 0.05$.

```
cov_prob3_M1 <- coverage_probability3(alpha = 0.05, lambda = 1, d = 1,
                                     m = 30, M = 100)
cov_prob3_M2 <- coverage_probability3(alpha = 0.05, lambda = 1, d = 1,
                                     m = 30, M = 1000)
cov_prob3_M3 <- coverage_probability3(alpha = 0.05, lambda = 1, d = 1,
                                     m = 30, M = 10000)
```

Tabela 6: Wartość rzeczywista, estymator punktowy, końce przedziału ufności (a, b) oraz prawdopodobieństwo pokrycia w zależności od ilości powtórzeń M

	wart. rzecz.	est. punktowy	a	b	prawd. pokrycia
M = 100	0.36787944	0.36811900	0.04359571	0.40937512	0.9100
M = 1000	0.36787944	0.37125520	0.12778770	0.70630590	0.9140
M = 10000	0.36787944	0.37424011	0.04094325	0.39840125	0.9162

Wyniki w zależności od poziomu istotności α przedstawiono w tabeli 7. Przyjmujemy $M = 1000$.

```
cov_prob3_alpha1 <- coverage_probability3(alpha = 0.005, lambda = 1, d = 1,
                                           m = 30, M = 1000)
cov_prob3_alpha2 <- coverage_probability3(alpha = 0.01, lambda = 1, d = 1,
                                           m = 30, M = 1000)
cov_prob3_alpha3 <- coverage_probability3(alpha = 0.1, lambda = 1, d = 1,
                                           m = 30, M = 1000)
```

Tabela 7: Wartość rzeczywista, estymator punktowy, końce przedziału ufności (a, b) oraz prawdopodobieństwo pokrycia w zależności od parametru α

	wart. rzecz.	est. punktowy	a	b	prawd. pokrycia
$\alpha = 0.005$	0.36787944	0.37834773	0.00024183	0.81379771	0.9630
$\alpha = 0.01$	0.36787944	0.36780123	0.00241309	0.59235984	0.9490
$\alpha = 0.1$	0.36787944	0.37804614	0.07636011	0.39028418	0.8690

W przypadku zmiennego parametru M , im większa jest jego wartość, tym estymator punktowy jest bliższy wartości rzeczywistej. Prawdopodobieństwo pokrycia rośnie wraz ze wzrostem M , jednak odbiega minimalnie od wartości $1 - \alpha = 0.05$. Analizując tabelę 7 możemy zauważyć, że oszacowanie estymatora punktowego nie zależy proporcjonalnie od parametru α , bowiem dla $\alpha = 0.01$ otrzymujemy wynik najbliższy wartości rzeczywistej. Widzimy także, że szerokość przedziału ufności maleje wraz ze wzrostem parametru α . Prawdopodobieństwo pokrycia w każdym przypadku wynosi nieco mniej niż $1 - \alpha$.