

**STATYSTYKA W FINANSACH I UBEZPIECZENIACH**  
**LABORATORIUM**

**SPRAWOZDANIE**

OPRACOWAŁA:  
**ALEKSANDRA GRZESZCZUK**  
NUMER ALBUMU: 255707

## SPIS TREŚCI

<b>1</b>	<b>CEL</b>	<b>2</b>
<b>2</b>	<b>CZĘŚĆ I</b>	<b>2</b>
2.1	ZADANIE 1 . . . . .	2
2.2	ZADANIE 2 . . . . .	10
2.3	ZADANIE 3 . . . . .	14
<b>3</b>	<b>CZĘŚĆ II</b>	<b>17</b>
3.1	ZADANIE 1 . . . . .	17
3.2	ZADANIE 2 . . . . .	19
<b>4</b>	<b>PODSUMOWANIE</b>	<b>21</b>
<b>5</b>	<b>DODATEK</b>	<b>22</b>
<b>6</b>	<b>BIBLIOGRAFIA</b>	<b>26</b>

# 1 CEL

Celem pierwszej części naszego sprawozdania jest zdefiniowanie oraz omówienie procesu znakowanego na podstawie dwóch sposobów implementacji - ze stałą intensywnością oraz zmienną, zależną od  $t$ , przy jednorodnym oraz niejednorodnym procesie Poissona. Interesować nas także będzie jakość aproksymacji złożonego rozkładu Poissona rozkładem normalnym. W oparciu o powyższą aproksymację będziemy badać własności przedziałów predykcyjnych.

Druga część sprawozdania opierać się na będzie na zdefiniowaniu oraz zrozumieniu składki netto  $H(R)$  kontraktu *stop – loos* oraz oszacowaniu prawdopodobieństwa nieznaney wartości  $H(R)$  przez wybrany wcześniej przedział ufności.

## 2 CZĘŚĆ I

### 2.1 ZADANIE 1

Punktowy proces znakowany jest procesem stochastycznym<sup>1</sup> złożonym z procesu punktowego  $T_1, T_2, \dots$  i związanych z nim marek  $X_1, X_2, \dots$  zwanych inaczej znakami. Punktowy proces znakowany (markowany) jest procesem punktowym  $N$  na przestrzeni  $\mathcal{T} \times \mathcal{M}$ , gdzie  $\mathcal{T}$  jest dziedziną punktów, zaś  $\mathcal{M}$  dziedziną marek taką, że dla każdego ograniczonego zbioru  $A \subset \mathcal{T}$  mamy  $N(A \times \mathcal{M}) < \infty$ , gdzie  $N(A \times \mathcal{M})$  oznacza losową liczbę punktów w zbiorze  $A \times \mathcal{M}$ .

Procesem Poissona z parametrem (intensywnością)  $\lambda > 0$  ( $\mathbf{P}(\lambda)$ ) nazywamy proces stochastyczny  $N = (N_t)_{t \geq 0}$  który spełnia następujące warunki:

- $N_0 = 0$
- $N$  ma przyrosty niezależne
- Dla  $0 \leq s \leq t$  zmienna  $N_t - N_s$  ma rozkład  $\mathbf{P}(\lambda(t - s))$
- Trajektorie procesu  $N$  są prawostronnie ciągłe

Nadmienmy również, że jeżeli proces Poissona jest jednorodny z intensywnością  $\lambda$ , to odstęp między kolejnymi skokami procesu mają rozkład wykładniczy o wartości oczekiwanej  $1/\lambda$ . Ponadto, zakładając, że  $X_1, X_2, \dots$  są kolejnymi skokami niejednorodnego procesu Poissona o funkcji intensywności  $\lambda$  i skumulowanej funkcji intensywności  $\Lambda(t) = \int_0^t \lambda(u) du$ , to  $\Lambda(X_1), \Lambda(X_2), \dots$  są skokami jednorodnego procesu Poissona o intensywności 1.

Rozkład logarytmicznie normalny (zwany także log - normalnym  $\mathbf{LN}(\mu, \sigma^2)$ ) jest ciągłym rozkładem statystycznym wartości logarytmicznych z powiązanego rozkładu normalnego. Jeśli zmienna losowa  $X$  ma rozkład logarytmicznie - normalny, to zmienna losowa  $Y = \ln(X) \sim N(\mu, \sigma^2)$ . Zmienna losowa o rozkładzie log - normalnym przyjmuje tylko dodatnie wartości rzeczywiste. Warto podkreślić, że rozkłady te są dodatnio skośne z długimi prawymi ogonami, ze

---

<sup>1</sup>Procesem stochastycznym nazywamy rodzinę zmiennych losowych określonych na pewnej przestrzeni probabilistycznej o wartościach w pewnej przestrzeni mierzalnej. Procesem stochastycznym nazywa się funkcję zależną od czasu, której wartości w każdym momencie czasowym są zmiennymi losowymi.

względu na niskie wartości średnie i duże wariancje zmiennych losowych.

Zadanie pierwsze polega na napisaniu programu do generowania zdefiniowanego wyżej punktowego procesu znakowego, w którym punkty i znaki są niezależne oraz punkty na odcinku  $[0, t_0]$  generowane są zgodnie z rozkładem Poissona w dwóch wariantach

1. o stałej intensywności  $\lambda$
2. o intensywności  $\lambda(t) = \alpha\beta t^{\beta-1}$

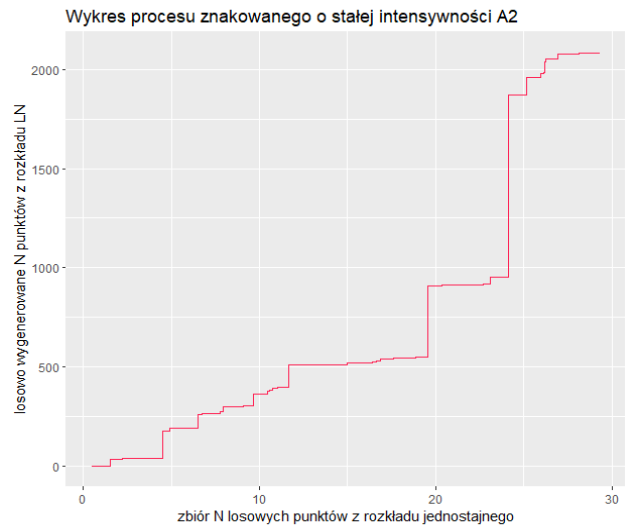
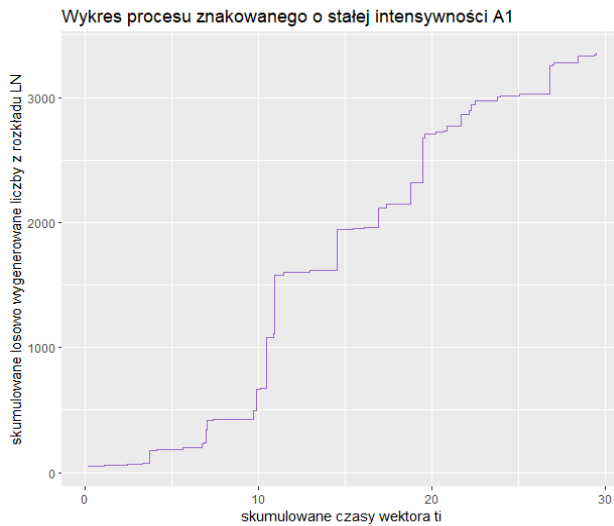
Znaki, w obydwu przypadkach generowane są z rozkładu logarytmicznie normalnego  $\mathbf{LN}(\mu, \sigma^2)$ .

Do generowania trajektorii jednorodnego procesu Poissona możemy wykorzystać dwa sposoby

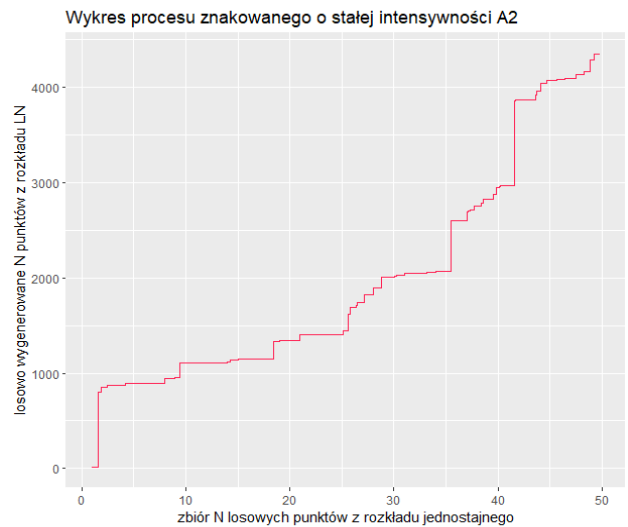
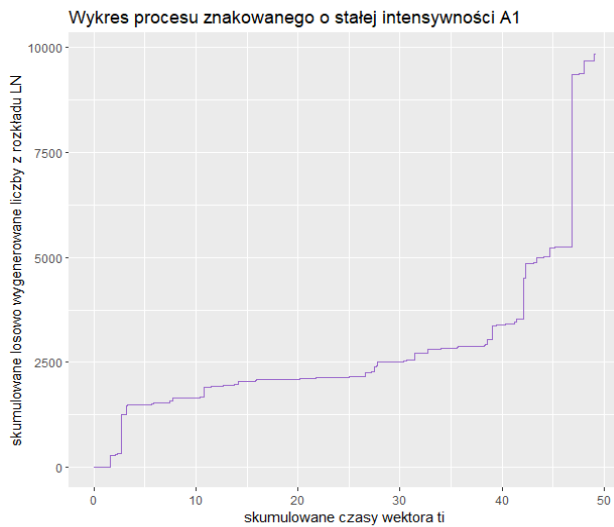
- Sposób 1 (Został zaimplementowany w funkcji *A1* (17))
  1. Losujemy wektor czasów oczekiwania na kolejne zdarzenie  $t_i$  z rozkładu wykładniczego z parametrem  $\lambda$  do momentu, gdy ich suma przekroczy ustalone  $t_0$
  2. Odrzucamy ostatni element wektora  $t_i$
  3. Wyznaczamy czas pojawia się kolejnych skoków procesu jako skumulowane sumy  $t_i$
  4. Wyznaczonym punktom przypisujemy znaki, czyli skumulowane, losowo wygenerowane liczby z rozkładu logarytmicznie normalnego  $\mathbf{LN}(\mu, \sigma^2)$
  5. Generujemy wykres otrzymanego procesu znakowego
- Sposób 2 (Został zaimplementowany w funkcji *A2* (18))
  1. Losujemy liczbę skoków procesu do czasu  $t$ . Liczba skoków oczywiście ma rozkład Poissona z parametrem  $\lambda t$ . Określamy ją literką  $N$
  2. W oparciu o zbiór  $N$  losowych punktów z rozkładu jednostajnego na odcinku  $[0, t_0]$  wybieramy, gdzie będą umieszczone skoki (punkty)  $N$
  3. Rosnąco sortujemy wybrane umiejscowienia
  4. Posortowanym punktom przypisujemy znaki, czyli losowo wygenerowane  $N$  punktów skumulowanych z rozkładu logarytmicznie normalnego  $\mathbf{LN}(\mu, \sigma^2)$
  5. Generujemy wykres otrzymanego procesu znakowanego

Można dostrzec, że drugi sposób jest potencjalnie szybszy, ponieważ już po pierwszym losowaniu wiemy, ile wartości z rozkładu jednostajnego musimy wylosować. Warto dodać również, że losowanie z rozkładu jednostajnego jest znacznie szybsze niż losowanie z rozkładu wykładniczego. W celu porównania opisanych sposobów, zaimplementowaliśmy obydwie funkcje. Zostały one zamieszczone w rozdziale (5).

Poniżej zaprezentowane zostało porównanie wykresów metody *A1* oraz *A2*. Każdy rząd ma zadane te same parametry.

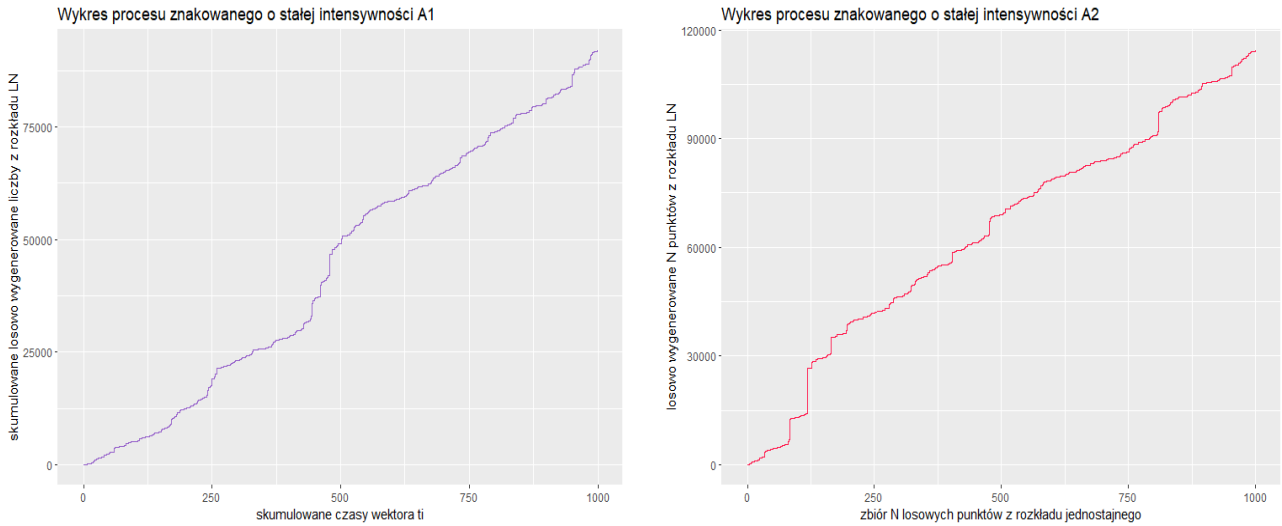


Rysunek 1: Wykres procesu znakowanego A1 i A2 z parametrami  $t_0 = 30, \lambda = 2, \mu = 2, \sigma = 2$   
- opracowanie własne



Rysunek 2: Wykres procesu znakowanego A1 i A2 z parametrami  $t_0 = 50, \lambda = 2, \mu = 2, \sigma = 2$   
- opracowanie własne

Już po pierwszych dwóch rysunkach (1) oraz (2) widzimy, że funkcje dla małego  $t_0$  nie pokrywają się. Nie wyglądają one jakby właściwie były nawet zbliżone do siebie. Może się tak dzieć tak z uwagi właśnie na zbyt małą liczbę obserwacji. Ustalamy zatem większe  $t_0 = 1000$  i spodziewamy się, że tym razem wykresy będą bardziej zbliżone do siebie.



Rysunek 3: Wykres procesu znakowanego A1 i A2 z parametrami  $t_0 = 1000$ ,  $\lambda = 2$ ,  $\mu = 2$ ,  $\sigma = 2$  - opracowanie własne

Zgodnie z naszymi przypuszczeniami, wykresy metod A1 i A2 dla wystarczająco dużego  $t_0$  znacznie się do siebie upodobniły.

Rzeczywistemu modelowaniu punktowego procesu znakowanego lepiej odpowiada metoda pierwsza (A1). Przykładem może być tutaj obserwowanie przystanku autobusowego, na którym autobusy przyjeżdżają ze stałą częstotliwością  $\lambda$ . Metoda druga (A2), mimo że potencjalnie jest szybsza, niekoniecznie pasuje nam do odzwierciedlania rzeczywistości.

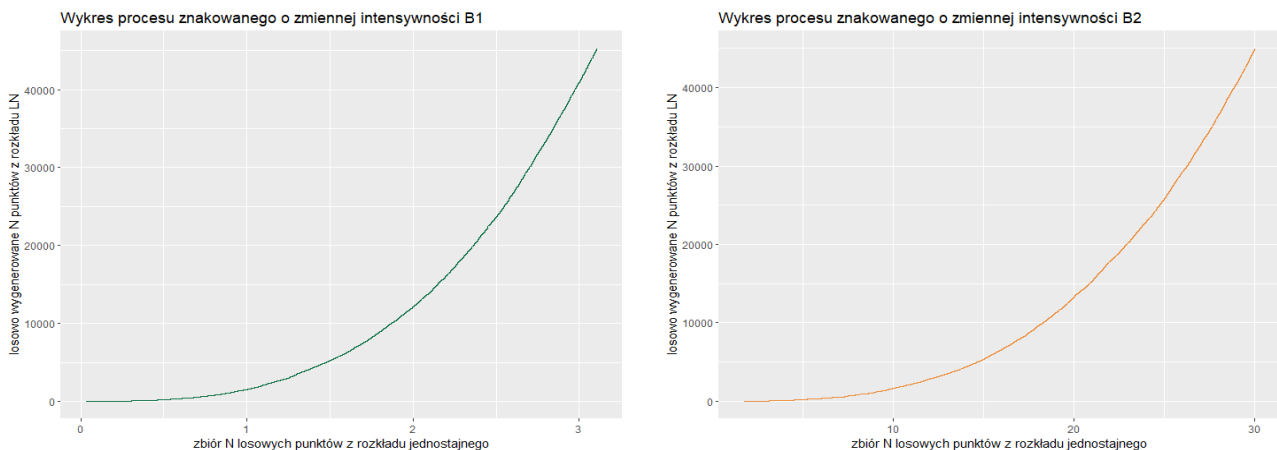
Proces Poissona, którego częstość skoków nie jest wartością stałą, nazywamy niejednorodnym procesem Poissona<sup>2</sup>. Proces ten bardzo często jest wykorzystywany w modelowaniu liczby szkód, gdy częstość ich występowania najczęściej zmienia się z czasem. W drugiej części zadania stała intensywność  $\lambda$  zmienia nam się na zmienną intensywność  $\lambda(t) = \alpha\beta t^{\beta-1}$  zależną od czasu. Zadanie ponownie można rozwiązać na dwa sposoby.

- Sposób 1 (Został zaimplementowany w funkcji B1 (19))
  1. Tworzymy wektor, nazwany jako *LAMBDA*, funkcji skumulowanych intensywności, dla zmiennych  $t$ ,  $\alpha$ ,  $\beta$
  2. Tworzymy funkcję odwrotną do funkcji *LAMBDA* i przechodzimy do generowania jednorodnego procesu Poissona
  3. Za  $\lambda_{MAX}$  przyjmujemy maksymalną wartość *LAMBDA* na przedziale  $[0, t_0]$
  4. Generujemy liczbę punktów  $N$  z rozkładu Poissona z parametrem  $\lambda_{MAX}$
  5. Generujemy punkty jako  $N$  liczb z rozkładu jednostajnego na przedziale  $[0, t_0]$

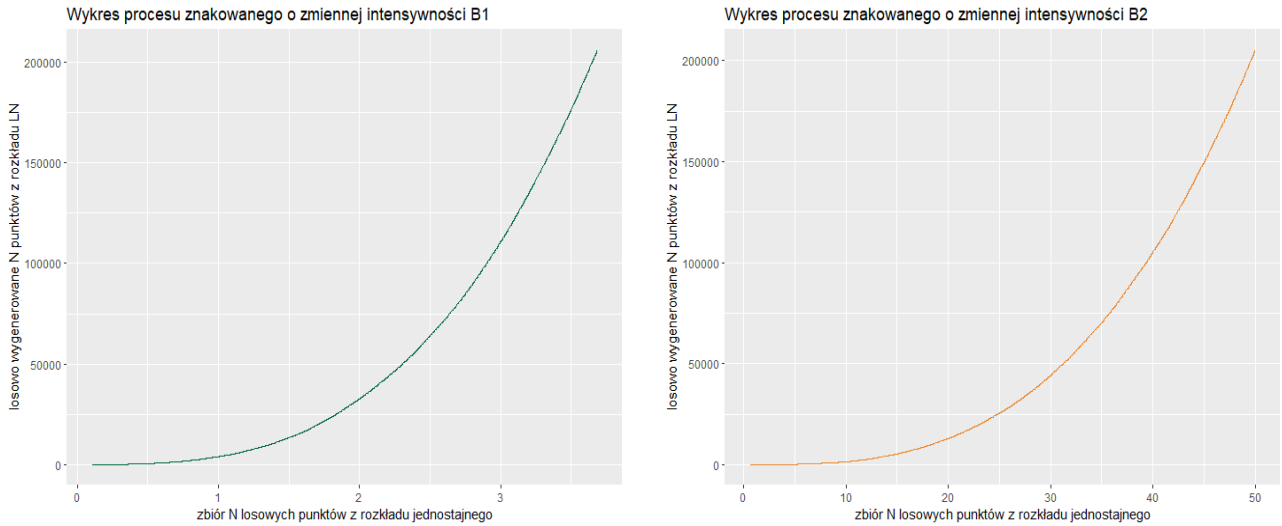
<sup>2</sup>Niejednorodny (niestacjonarny) proces Poissona jest liczącym procesem stochastycznym spełniającym następujące warunki 1.  $N_0 = 0$ , 2.  $N_t$  ma niezależne przyrosty, 3.  $P(N_{t+h} - N_t = 1) = \lambda(t)h + o(h)$ , 4.  $P(N_{t+h} - N_t \leq 2) = o(h)$ , 5. Gdy  $\lambda(t) = \lambda$ , to otrzymujemy jednorodny proces Poissona.

6. Sortujemy otrzymane punkty i przechodzimy do generowania niejednorodnego procesu Poissona (przez zmianę czasu)
  7. Tworzymy nowe punkty nakładając na funkcję odwrotną do  $LAMBDA$  wcześniej wyznaczone punkty
  8. Generujemy znaki jako skumulowaną sumę liczb z rozkładu logarytmicznie normalnego  $LN(\mu, \sigma^2)$
  9. Generujemy wykres otrzymanego procesu znakowanego
- Sposób 2 (Został zaimplementowany w funkcji  $B2$  (20))
    1. Tworzymy funkcję  $lambda$  intensywności dla zmiennych  $t, \alpha, \beta$  i przechodzimy do generowania jednorodnego procesu Poissona
    2. Za  $\lambda_{MAX}$  przyjmujemy maksymalną wartość  $lambda$  na przedziale  $[0, t_0]$
    3. Generujemy liczbę punktów  $N$  z rozkładu Poissona z parametrem  $\lambda_{MAX} \cdot t_0$
    4. Generujemy punkty jako  $N$  liczb z rozkładu jednostajnego na przedziale  $[0, t_0]$
    5. Sortujemy otrzymane punkty i przechodzimy do generowania niejednorodnego procesu Poissona (przez przerzedzenie)
    6. Tworzymy wektor  $u$  składający się z  $N$  liczb z rozkładu jednostajnego na  $[0, 1]$
    7. Tworzymy pusty wektor, do którego będą trafiać odrzucane punkty
    8. Odrzucamy punkty z zadaniem prawdopodobieństwem w pętli
    9. Ustalmy, że  $k$  są długością nowego wektora punktów
    10. Generujemy znaki jako skumulowaną sumę  $k$  liczb z rozkładu  $LN(\mu, \sigma^2)$
    11. Generujemy wykres otrzymanego procesu znakowanego

Metoda pierwsza nazywana jest również metodą zmiany czasu, podczas gdy druga metoda określana jest niekiedy metodą odsiewania. Możemy spodziewać się, że lepszą (szybszą) metodą będzie metoda odsiewania.



Rysunek 4: Wykres procesu znakowanego B1 i B2 z parametrami  $t_0 = 30, \alpha = 1, \beta = 3, \mu = 0, \sigma = 1$  - opracowanie własne



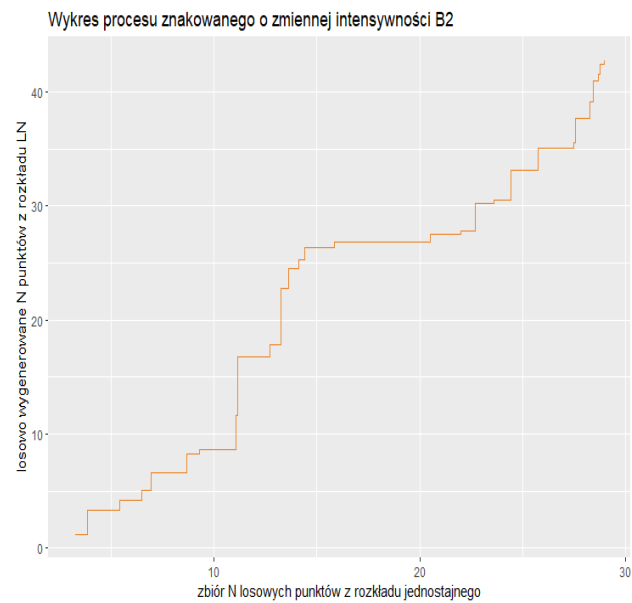
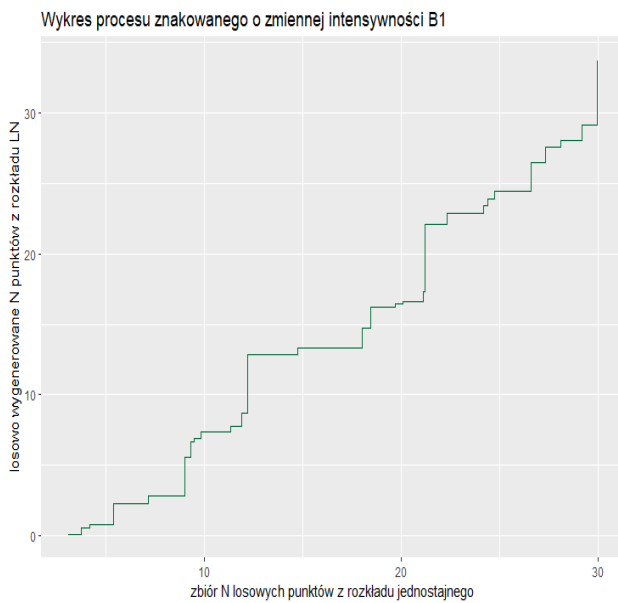
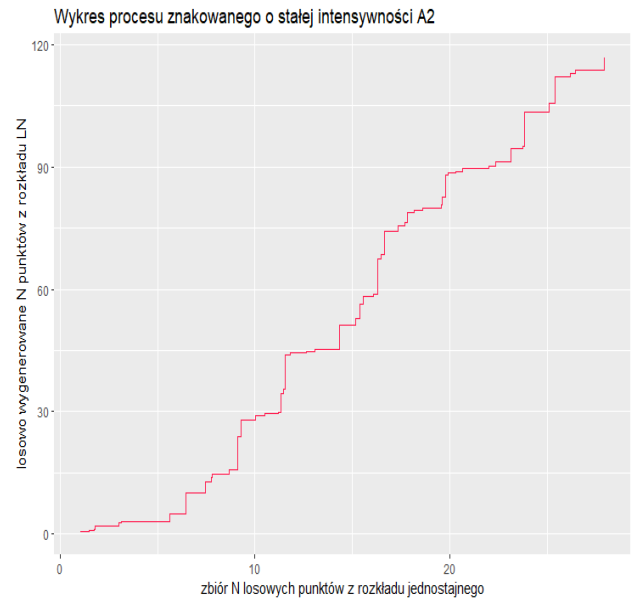
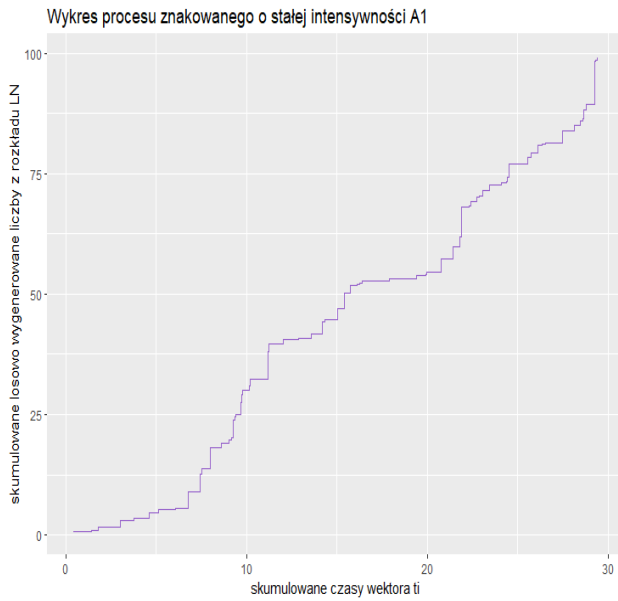
Rysunek 5: Wykres procesu znakowanego B1 i B2 z parametrami  $t_0 = 50, \alpha = 1, \beta = 3, \mu = 0, \sigma = 1$  - opracowanie własne

Powyższe wykresy metod  $B1$  i  $B2$  są funkcjami gładkimi, ciągłymi. Nie posiadają żadnych "schodków" ani skoków wartości, tak jak wcześniej widzieliśmy je metod  $A1$  oraz  $A2$ .

Rzeczywistemu modelowaniu punktowego procesu znakowanego lepiej odpowiada metoda nazywana metodą zmiany czasu  $B1$ , ponieważ właściwie nigdy czas nie jest wartością stałą, tylko zawsze ulega zmianie.

Warto zauważyć, że funkcja  $LAMBDA$  (początek wzoru (19)) oraz  $lambda$  określona na początku wzoru (20) sprowadza się do funkcji stałej dla parametrów  $\alpha = 1, \beta = 1$ . To znaczy, że przypadek procesu Poissona o zmiennej intensywności jesteśmy w stanie sprowadzić, do procesu Poissona o stałej intensywności. Porównamy teraz funkcje  $A1, A2, B1, B2$  dla parametrów  $t_0 = 30, \lambda = 2, \alpha = 1, \beta = 1, \mu = 0, \sigma = 1$ . Spodziewamy się, że wszystkie funkcje będą funkcjami schodkowymi zbliżonymi wglądowo do siebie.





Rysunek 6: Porównanie funkcji  $A1, A2, B1, B2$  dla parametrów  $t_0 = 30, \lambda = 2, \alpha = 1, \beta = 1, \mu = 0, \sigma = 1$  - opracowanie własne

Przyglądając się powyższemu (6) widzimy, że wykresy funkcji  $A1$  oraz  $A2$  są do siebie zbliżone, podczas gdy wykresy  $B1$  oraz  $B2$  znacząco od siebie odbiegają. Zbadamy teraz, które funkcje są lepsze od pozostałych dla różnych parametrów.

W celu sprawdzenia, które funkcje są lepsze, sprawdzimy szybkość ich działania, korzystając funkcji `system.time`<sup>3</sup>. Funkcja ta zwraca nam trzy liczby

- *elapsed* - ile czasu minęło
- *system* - ile czasu procesora zużyły w tym czasie procesy poza *R*
- *user* - ile czasu procesora zużył *R*

	<i>user</i>	<i>system</i>	<i>elapsed</i>
<i>A1</i> (30)	0.09	0.02	0.51
<i>A2</i> (30)	0.3	0.00	0.35
<i>A1</i> (50)	0.22	0.02	0.4
<i>A2</i> (50)	0.28	0.00	0.36
<i>A1</i> (1000)	1.20	0.39	2.29
<i>A2</i> (1000)	0.36	0.00	0.38
<i>A1</i> (10000)	64.33	17.26	99.51
<i>A2</i> (10000)	0.64	0.03	0.74

Tabela 1: Porównanie czasu działania funkcji *A1*, *A2*, dla  $t_0 = 30, 50, 1000, 10000$ ,  $\lambda = 2$ ,  $\sigma = 2$ ,  $\mu = 2$

Najbardziej interesuje nas wiersz *user* w powyższej tabeli (1). Widzimy na niej, że dla mniejszej wartości  $t_0$  funkcja pierwsza działa szybciej. Jednakże podczas zwiększania  $t_0$  widzimy, zacierającą się różnicę, aby w końcu dla dużego  $t_0 = 1000, 10000$  dostrzec, że funkcja druga jest zdecydowanie i bardzo widocznie szybsza od pierwszej. Stwierdzamy zatem, że zgodnie z wcześniejszymi przypuszczeniami, metoda *A2* jest metodą zdecydowanie lepszą od metody *A1*.

	<i>user</i>	<i>system</i>	<i>elapsed</i>
<i>B1</i> (30)	0.22	0.00	0.34
<i>B2</i> (30)	8.08	0.85	24.44
<i>B1</i> (50)	0.76	0.05	1.70
<i>B2</i> (50)	301.62	76.15	614.00
<i>B1</i> (100)	6.48	0.16	12.31
<i>B2</i> (100)	niewykonalne	niewykonalne	niewykonalne

Tabela 2: Porównanie czasu działania funkcji *B1*, *B2*, dla  $t_0 = 30, 50, 100$ ,  $\alpha = 1$ ,  $\beta = 3$ ,  $\mu = 0$ ,  $\sigma = 1$

Ponownie przyglądamy się wierszowi *user* w tabeli (2). Widzimy, że dla małej wielkości  $t_0 = 30$ , różnica w czasie między funkcją *B1* a *B2* jest prawie 40-krotnie większa. Lepszą funkcją jest funkcja *B1*. Jesteśmy dla niej w stanie wywołać zdecydowaną większość kompilacji. Funkcja *B2* (20) zawiera więcej pętli, danych do sprawdzenia, przez co jej czas działania mocno się

<sup>3</sup>Funkcja `system.time` zlicza, jak długo w sekundach wykonuje się określony zbiór instrukcji. Więcej informacji dostępnych pod adresem <https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/system.time>

wydłuża. W praktyce potrzebujemy sprawdzić wiele funkcji w dużej ilości iteracji. Zależy nam na optymalizowaniu swojego czasu pracy, dlatego z funkcji  $B1$  zwanej metodą zmiany czasu rzeczywiście powinniśmy korzystać częściej.

	<i>user</i>	<i>system</i>	<i>elapsed</i>
$A1(30)$	0.37	0.01	0.39
$A2(30)$	0.28	0.02	0.36
$B1(30)$	0.33	0.00	0.38
$B2(30)$	0.31	0.00	0.37

Tabela 3: Porównanie czasu działania funkcji  $A1, A2, B1, B2$ , dla  $t_0 = 30, \lambda = 2, \alpha = 1, \beta = 1, \mu = 0, \sigma = 1$

W powyższym przypadku (3) nie widać znaczących różnic w czasach działania funkcji. Możemy określić zatem, że dla małego  $t_0 = 30$  jest to właściwie obojętne, którą metodę wybierzemy, ponieważ wszystkie działają (szybkościowo) bardzo podobnie.

## 2.2 ZADANIE 2

Jeżeli  $N$  ma rozkład Poissona, gdzie  $\lambda$  jest znane, oraz zmienne losowe  $X_i, i = 1, \dots, N$  są niezależne o jednakowym rozkładzie o dystrybuancie  $F_X$ , to zmienna losowa  $S_N = X_1 + \dots + X_N$  ma złożony rozkład Poissona, co zapisujemy jako  $S_N \sim \mathbf{CP}(\lambda, \mathbf{F}_X)$ .

Warto również dodać, że jeżeli  $S_N$  ma złożony rozkład Poissona, to

$$\frac{S_N - \lambda E(X_1)}{\sqrt{\lambda E(X_1^2)}}$$

dąży według rozkładu do standardowego rozkładu normalnego, dla  $\lambda \rightarrow \infty$ . Ten wzór określamy jako aproksymację złożonego rozkładu Poissona rozkładem normalnym. Zadanie drugie polega właśnie na zbadaniu jakości wskazanej aproksymacji.

Ustalmy, że  $X_i, i = 1, 2, \dots$  są niezależnymi zmiennymi losowymi z rozkładu logarytmicznie normalnego  $\mathbf{LN}(\mu, \sigma^2)$  oraz  $N$  ma rozkład Poissona  $\mathbf{P}(\lambda)$ . Napišemy program, który dla zadanych wartości parametrów  $\mu, \sigma, \lambda$  oraz zadanego poziomu  $\alpha$  będzie wyznaczał przedział ufności  $(a, b)$  taki, że

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

Wiemy, że nasze  $a$  i  $b$  są następującej postaci:

$$a = \lambda E(X_1) + z(\alpha_1) \sqrt{\lambda E(X_1)^2}$$

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

Gdzie jako  $z(q)$  określony jest kwantyl rzędu  $q$  rozkładu  $\mathbf{N}(0, 1)$ <sup>4</sup>

---

<sup>4</sup>Napisaną funkcję wyznaczającą przedziały ufności można znaleźć w rozdziale DODATEK pod numerem (21)

Wyznamy teraz drugi moment naszej zmiennej  $X$ . Wiemy, że  $X_1 \sim \mathbf{LN}(\mu, \sigma^2)$ , więc  $\ln X_1 \sim \mathbf{N}(\mu, \sigma^2)$ . Korzystając z tej zależności wyznaczamy wzór ogólny na  $k$ -ty moment rozkładu oraz na interesujący nas pierwszy i drugi moment zwykły.

$$E(X_1)^k = E\left(e^{\ln(X_1)^k}\right) = \int_{-\infty}^{\infty} e^{kx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$E(X_1) = e^{\mu+\sigma^2/2}$$

$$E(X_1)^2 = e^{2\mu+2\sigma^2}$$

Czyli ostatecznie, krańce naszych przedziałów ufności są postaci

$$a = \lambda e^{\mu+\sigma^2/2} + z(\alpha_1) \sqrt{\lambda e^{\mu+\sigma^2/2}}$$

$$b = \lambda e^{\mu+\sigma^2/2} + z(1 - \alpha_2) \sqrt{\lambda e^{2\mu+2\sigma^2}}$$

Przyglądając się powyższym wzorom możemy przypuszczać, że zwiększenie parametrów  $\lambda$  oraz  $\mu$  wpłynie na zwiększenie przedziałów ufności.

Następnie napiszemy program<sup>5</sup>, który dla wybranych wartości parametrów  $\mu, \sigma, \lambda$  oraz poziomu  $\alpha$ , oszacuje prawdopodobieństwo pokrycia wartości

$$S_N = \sum_{i=1}^N X_i$$

przez przedział  $(a, b)$ .

wartości parametrów funkcji	$a$	$b$	$b - a$
$\mu = 1, \sigma = 0.5, \lambda = 10, \alpha = 0.05$	9.169196	52.435141	43
$\mu = 2, \sigma = 0.5, \lambda = 10, \alpha = 0.05$	24.92446	142.53349	118
$\mu = 5, \sigma = 0.5, \lambda = 10, \alpha = 0.05$	500.6212	2862.8617	500
$\mu = 1, \sigma = 0.01, \lambda = 10, \alpha = 0.05$	10.33472	44.03364	34
$\mu = 1, \sigma = 0.3, \lambda = 10, \alpha = 0.05$	9.999582	46.868388	37
$\mu = 1, \sigma = 1, \lambda = 10, \alpha = 0.05$	2.883018	78.627517	76
$\mu = 1, \sigma = 0.5, \lambda = 100, \alpha = 0.05$	239.6122	376.4311	137
$\mu = 1, \sigma = 0.5, \lambda = 1000, \alpha = 0.05$	2863.887	3296.547	433
$\mu = 1, \sigma = 0.5, \lambda = 0.01, \alpha = 0.05$	0	0.7148968	0.7

Tabela 4: Tabela przedstawiająca porównanie przedziałów ufności dla różnych parametrów

Pierwsze trzy wiersze naszej tabeli (6) zawierają takie same parametry  $\sigma, \lambda, \alpha$  oraz zmienny parametr  $\mu$ . Wzory na krańce przedziałów  $a$  oraz  $b$  zależą wprost proporcjonalnie od parametru  $\mu$ , zatem zgodnie z naszymi przewidywaniami, zwiększenie tego parametru skutkuje zwiększeniem przedziału ufności. Następne trzy wiersze (ale także pierwszy wiersz tabeli) zawierają takie same parametry  $\mu, \lambda, \alpha$  oraz zmienny parametry  $\sigma$ . Wiemy, że  $0.01 < 0.3 < 0.5 < 1$ . Największe

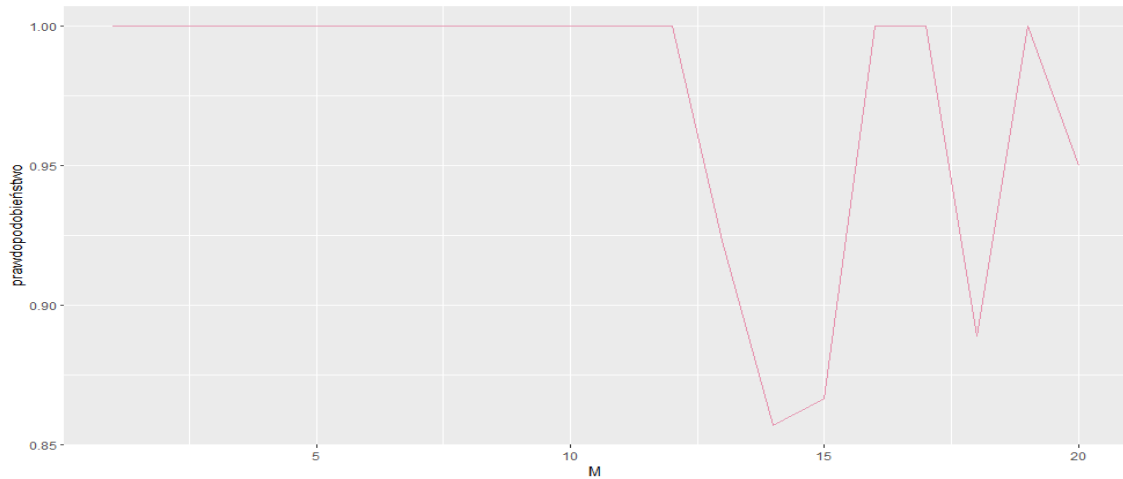
<sup>5</sup>Napisaną funkcję można znaleźć w rozdziale DODATEK pod numerem (22)

lewe przedziały ufności otrzymaliśmy dla najmniejszego parametru  $\sigma = 0.01$ , zaś najmniejsze dla największego parametru  $\sigma = 1$ . Równocześnie, największy prawy przedział ufności otrzymaliśmy dla największego parametru  $\sigma = 1$ , zaś najmniejszy - dla najmniejszego parametru  $\sigma = 0.01$ . Stwierdzamy zatem, że zwiększenie, bądź zmniejszenie parametru  $\sigma$  nie wpływa na szerokość przedziału ufności. Ostatnie trzy wiersze (oraz ponownie pierwszy wiersz tabeli) zawierają stałe parametry  $\mu, \sigma, \alpha$  oraz zmienny parametr  $\lambda$ . Stwierdzamy, że zgodnie z naszymi przypuszczeniami, zwiększenie parametru  $\lambda$  wpływa na zwiększenie przedziałów ufności.

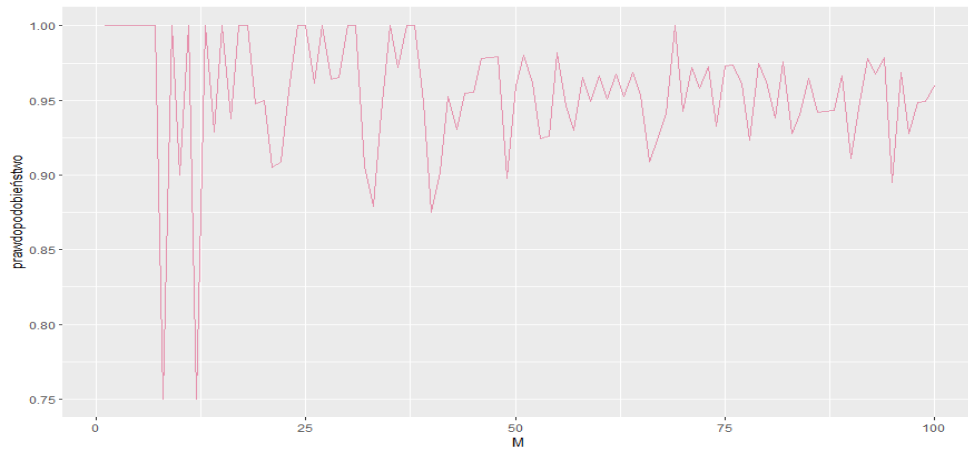
wartości parametrów funkcji	prawdopodobieństwo pokrycia
$\mu = 1, \sigma = 0.5, \lambda = 10, \alpha = 0.05$	0.94
$\mu = 2, \sigma = 0.5, \lambda = 10, \alpha = 0.05$	0.98
$\mu = 5, \sigma = 0.5, \lambda = 10, \alpha = 0.05$	0.92
$\mu = 1, \sigma = 0.01, \lambda = 10, \alpha = 0.05$	0.98
$\mu = 1, \sigma = 0.3, \lambda = 10, \alpha = 0.05$	1
$\mu = 1, \sigma = 1, \lambda = 10, \alpha = 0.05$	0.96
$\mu = 1, \sigma = 0.5, \lambda = 100, \alpha = 0.05$	0.96
$\mu = 1, \sigma = 0.5, \lambda = 1000, \alpha = 0.05$	0.9
$\mu = 1, \sigma = 0.5, \lambda = 0.01, \alpha = 0.05$	0

Tabela 5: Tabela przedstawiająca prawdopodobieństwo pokrycia dla różnych parametrów i stałego  $M = 50$

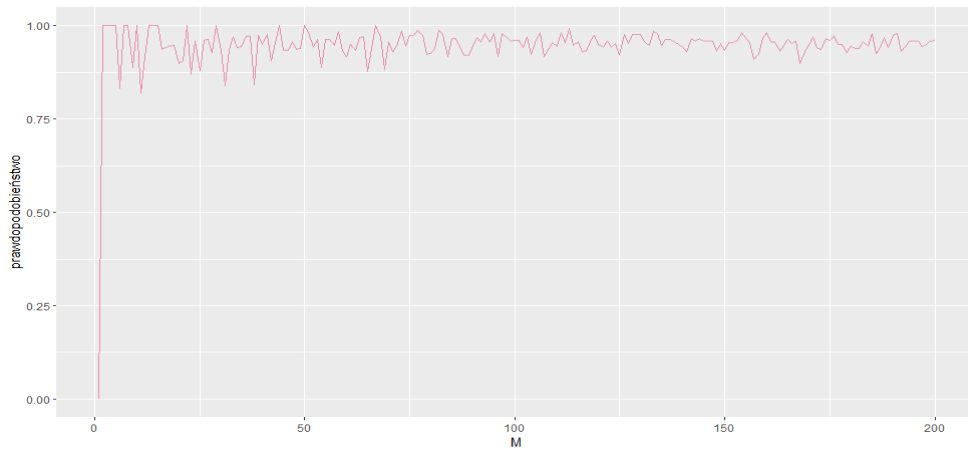
Prawdopodobieństwem pokrycia nazywamy ustalony poziom ufności. Powinien mieć on szerokość wynoszącą około  $1 - \alpha$ . Analizując tabelę (5) widzimy, że zmiana parametru  $\mu$  oraz  $\sigma$  losowo wpływa na zadane prawdopodobieństwo pokrycia. Istotną zmianą, jaką możemy zauważyć, to fakt, że podczas znacznego zmniejszenia parametru  $\lambda$  następuje spadek prawdopodobieństwa pokrycia do 0. Prezentujemy poniżej wykresy zależności prawdopodobieństwa pokrycia od liczby replikacji  $M$ .



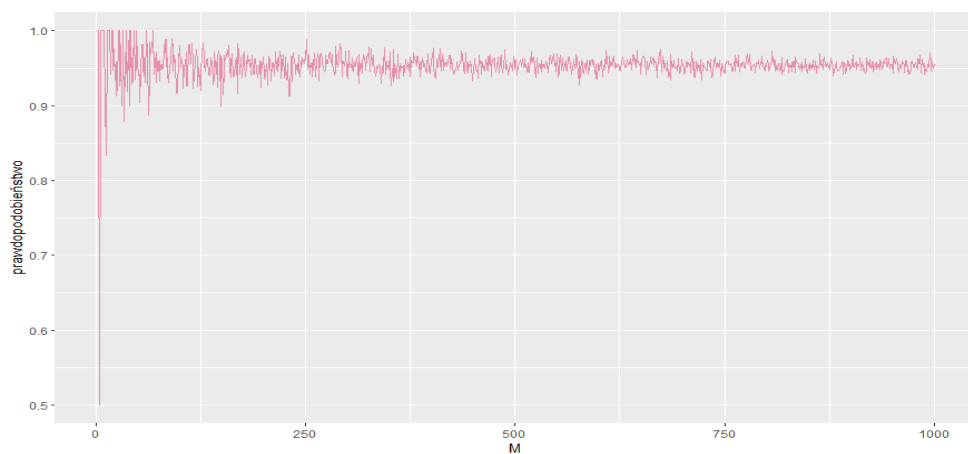
Rysunek 7: Wykres zależności prawdopodobieństwa pokrycia dla  $M = 20$  - opracowanie własne



Rysunek 8: Wykres zależności prawdopodobieństwa pokrycia dla  $M = 100$  - opracowanie własne



Rysunek 9: Wykres zależności prawdopodobieństwa pokrycia dla  $M = 200$  - opracowanie własne



Rysunek 10: Wykres zależności prawdopodobieństwa pokrycia dla  $M = 1000$  - opracowanie własne

Przyglądając się powyższemu wykresowi, widzimy, że minimalną wartość  $M$  jaką powinniśmy wziąć jest 300. Poniżej tej wartości dostrzegamy bardzo duże wahania naszej funkcji i brak jakiegokolwiek ustabilizowania.

## 2.3 ZADANIE 3

Ustalmy, że  $X_i$  są niezależnymi zmiennymi losowymi z rozkładu logarytmicznie normalnego  $\text{LN}(\mu, \sigma^2)$  oraz niech zmienna  $N$  na rozkład Poissona. Przy założeniu, że wartości parametrów nie są nam znane, będziemy chcieli napisać program, który na podstawie realizacji zmiennej losowej  $N$  oraz zmiennych  $X_1, \dots, X_N$  będzie wyznaczał przedział  $(a', b')$  taki, że

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

Celem wskazanego zadania jest zbadanie oraz określenie własności przedziałów predykcyjnych skonstruowanych na podstawie aproksymacji złożonego rozkładu Poissona rozkładem normalnym  $\text{N}(\mu, \sigma^2)$ .

Korzystając ze związku między rozkładem logarytmicznie normalnym a rozkładem normalnym, za estymatory  $\mu$  oraz  $\sigma$  przyjmujemy odpowiedni średnią z próby, oraz pierwiastek z wariancji próbkowej. Zadane są one wzorami:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Warto wspomnieć tutaj, że sposób wyznaczania przedziałów ufności oraz prawdopodobieństwa pokrycia jest analogiczny do wcześniejszego zadania, jednakże nie mamy tutaj podanych parametrów  $\mu$  oraz  $\sigma$ . Parametry te są nieznane i będziemy je estymować.

Ciekawym zagadnieniem jest przypadek, gdy mamy próbkę składającą się tylko z jednej obserwacji. Wtedy, przy estymatorze odchylenia standardowego, będącego (jak już zostało wspomniane) pierwiastkiem z wariancji próbkowej, mamy dzielenie przez 0. W takim przypadku błąd ignorujemy i losujemy próbkę tak długo, aż będziemy mieć co najmniej dwa elementy. Na ich podstawie dokonujemy dalszych analiz i oszacowań nieznanymi parametrów.

Poniżej przedstawiamy przykładowe estymowane przedziały ufności, wyznaczone na podstawie programu (23).

wartości parametrów funkcji	$a$	$b$	$b - a$
$X1 \sim \mathbf{LN}(0, 1), \alpha = 0.05$	3.119473	20.945111	17
$X1 \sim \mathbf{LN}(0, 0.5), \alpha = 0.05$	2.706484	17.956579	15
$X1 \sim \mathbf{LN}(0, 1/8), \alpha = 0.05$	3.75221	16.31035	13
$X1 \sim \mathbf{LN}(3, 1), \alpha = 0.05$	29.78883	459.02533	430
$X1 \sim \mathbf{LN}(1/8, 1), \alpha = 0.05$	0.3732287	20.1400207	19
$X1 \sim \mathbf{LN}(0.01, 1), \alpha = 0.05$	1.110278	13.764137	12

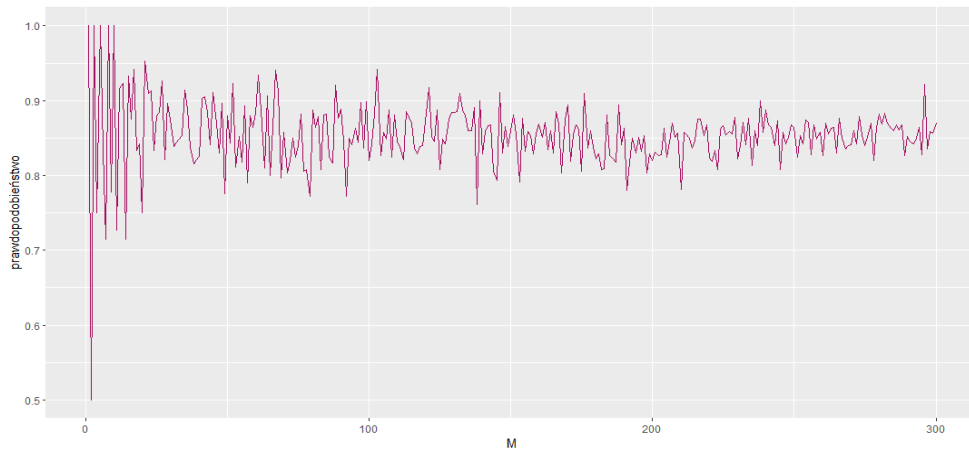
Tabela 6: Tabela przedstawiająca porównanie przedziałów ufności dla różnych parametrów

Przyglądając się trzem pierwszym wierszom, widzimy, że zmniejszenie parametru  $\sigma^2$  z 0.5 do 0.125 nieznacznie wpływa na zmniejszenie się naszego przedziału ufności. Analogicznie, zwiększenie parametru  $\sigma^2$  do 1 również minimalnie wpływa na zwiększenie się przedziałów ufności. Kolejne trzy wiersze przedstawiają stały parametr  $\sigma^2 = 1$  oraz zmienne  $\mu$ . Widzimy, że zwiększenie  $\mu$  do 3 bardzo wpływa na nasz prawy kraniec przedziału. W porównaniu z wartością  $\mu = 0.125$  zwiększa się on niemalże 23 - krotnie.

wartości parametrów funkcji	prawdopodobieństwo pokrycia
$\mu = 0, \sigma = 1, \lambda = 10, M = 10$	0.9
$\mu = 0, \sigma = 1, \lambda = 10, M = 100$	0.94
$\mu = 0, \sigma = 1, \lambda = 10, M = 10000$	0.909
$\mu = 10, \sigma = 1/8, \lambda = 100, M = 1000$	0.84
$\mu = 0.01, \sigma = 1/8, \lambda = 100, M = 1000$	0.833
$\mu = 30, \sigma = 0.0001, \lambda = 10, M = 100000$	0.82797

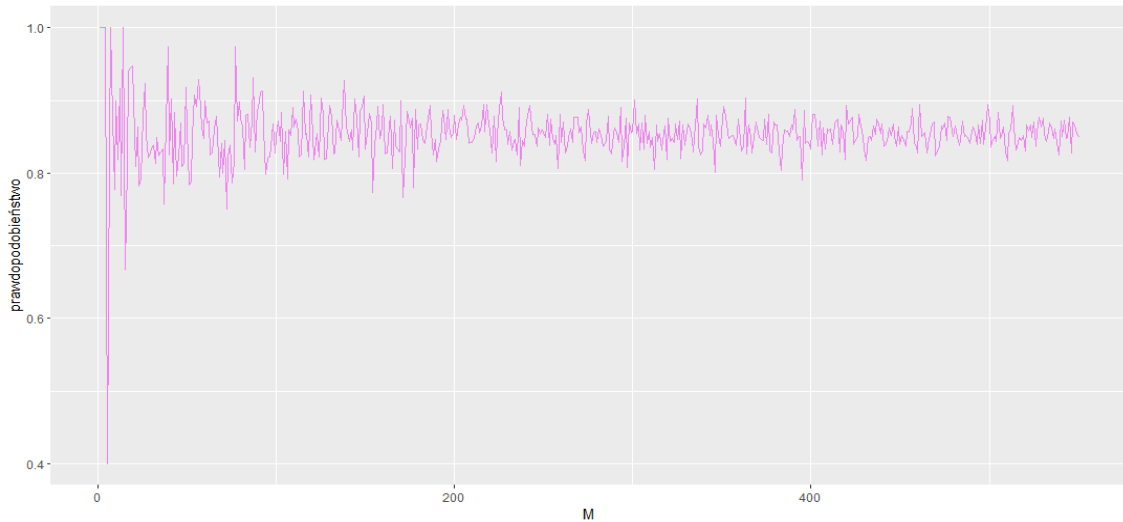
Tabela 7: Tabela przedstawiająca estymowane prawdopodobieństwo pokrycia, w każdym przypadku dla  $\alpha = 0.05$

Przyglądając się powyższej tabelce (7) widzimy, że zmiana parametrów nieznacznie wpływa na nasze prawdopodobieństwo pokrycia. Poniżej zaprezentujemy wykresy estymowanego prawdopodobieństwa pokrycia w zależności od wskazanego parametru  $M$ .

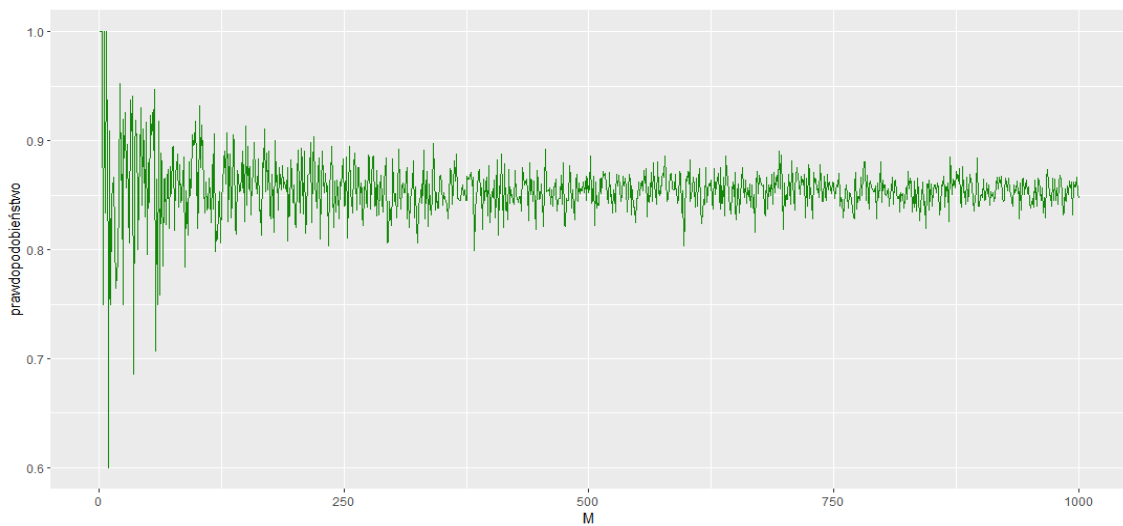


Rysunek 11: Wykres zależności estymowanego prawdopodobieństwa pokrycia dla  $M = 300$  - opracowanie własne





Rysunek 12: Wykres zależności estymowanego prawdopodobieństwa pokrycia dla  $M = 550$  - opracowanie własne



Rysunek 13: Wykres zależności estymowanego prawdopodobieństwa pokrycia dla  $M = 1000$  - opracowanie własne

Przyglądając się powyższym wykresom, widzimy, że minimalną rozsądną wartością  $M$  wydaje się 550. Poniżej tej wartości wykres jest większe wahania. Ponadto wykres dla największej wartości  $M = 1000$  stabilizuje się w okolicy 0.95. Zatem, dla zadanego poziomu istotności  $1 - \alpha = 1 - 0.05 = 0.95$  widzimy, że oscyluje wokół teoretycznej wartości prawdopodobieństwa.

## 3 CZĘŚĆ II

### 3.1 ZADANIE 1

Składka jest opłatą, którą podmiot narażony na ryzyko płaci ubezpieczycielowi za przejęcie części ryzyka związanego ze swoją działalnością. Składkę  $H = E(X)$  nazywamy składką netto. Reasekuracja jest przekazaniem przez ubezpieczyciela części ryzyka do ubezpieczenia w innym towarzystwie ubezpieczeniowym. Reasekurację dzielimy na dwie klasy:

- reasekuracje proporcjonalne
- reasekuracje nieproporcjonalne

Jedną z typowych reasekuracji nieproporcjonalnych jest reasekuracja nadwyżkowa, czyli kontrakty stop-loss, opisana jako  $R = (S_n - d)_+$ , gdzie  $d > 0$ . Widzimy, że kontrakty reasekuracyjne wymagają wyceny składki. Składki netto  $H(R)$  kontraktu stop-loss definiujemy jako

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

Gdzie  $F_{S_n}$  jest dystrybuantą rozkładu  $S_n$ . Dla ryzyka wykładniczego (czyli gdy  $S_n \sim \mathbf{E}(\lambda)$ ) składkę netto kontraktu stop-loss możemy zdefiniować jako

$$H(R) = \int_d^\infty \exp(-\lambda s) \, ds = \exp(-\lambda d) / \lambda$$

W zadaniu, przyjmując, że parametr  $\lambda$  nie jest znany, proponujemy oszacowanie punktowe i przedziałowe wartości  $H(R)$  na podstawie  $m$  realizacji  $s_n^{(1)}, \dots, s_n^{(m)}$  zmiennych losowych  $S_n^{(1)}, \dots, S_n^{(m)}$

W naszym przykładzie, w przypadku oszacowania punktowego, za estymator parametru  $\lambda$  przyjmujemy odwrotność wartości średniej z próby, czyli  $\frac{1}{\bar{S}_n}$ . Oczywiście, przez  $\bar{S}_n$  oznaczamy średnią z realizacji  $S_n$ . Trochę bardziej zaawansowane obliczeniowo jest oszacowanie przedziałowe. Szukamy takich wartości  $a$  oraz  $b$ , że  $P(a < H(R) < b) \approx 1 - \alpha$ . Ustalmy, że  $g(s) = se^{-d/s}$ , wtedy otrzymujemy, że  $H(R) = \bar{g}(\bar{S}_n)$ . Rozkładem średniej z realizacji  $S_n$  jest  $m\bar{S}_n \sim \text{Gamma}(m, \lambda)$ . Stąd, po pewnych obliczeniach, otrzymujemy wartość oczekiwania oraz wariancję naszego rozkładu.

$$E(m\bar{S}_n) = \frac{m}{\lambda} \Rightarrow E(\bar{S}_n) = \frac{1}{m} \cdot E(m\bar{S}_n) = \frac{1}{\lambda}$$

$$\text{Var}(m\bar{S}_n) = \frac{m}{\lambda^2} \Rightarrow \text{Var}(\bar{S}_n) = \frac{1}{m^2} \cdot \text{Var}(m\bar{S}_n) = \frac{1}{m} \cdot \frac{1}{\lambda^2}$$

Widzimy, że każde  $\bar{S}_n$  spełnia Centralne Twierdzenie Graniczne, czyli otrzymujemy

$$\frac{\bar{S}_n - E(\bar{S}_n)}{\sqrt{\text{Var}(\bar{S}_n)}} \rightarrow \mathbf{N}(0, 1)$$

Dalej do wyznaczania przedziałów ufności przyda nam się Metoda Delta <sup>6</sup>. Na jej podstawie otrzymujemy wzór na przedziały ufności

$$1 - \alpha \approx P(a < H(R) < b) \approx P(a < g(\bar{S}_n < b)) = \dots = F_Y(b_1) - F_Y(a_1) = \Phi(b_1) - \Phi(a_1)$$

Wystarczy wziąć zatem takie  $a_1, b_1$  aby  $\Phi(a_1) = \alpha/2$  oraz  $\Phi(b_1) = 1 - \alpha/2$ . Ostatecznie otrzymujemy przedziały ufności, postaci

$$a = g\left(\frac{1}{\lambda}\right) + z_{\alpha/2} \frac{1}{\sqrt{m}} \frac{1}{\lambda} g'\left(\frac{1}{\lambda}\right)$$

$$b = g\left(\frac{1}{\lambda}\right) + z_{1-\alpha/2} \frac{1}{\sqrt{m}} \frac{1}{\lambda} g'\left(\frac{1}{\lambda}\right)$$

Gdzie przez  $z_q$  oznaczamy kwantyl rzędu  $q$  standardowego rozkładu normalnego  $\mathbf{N}(0, 1)$

Sprawdźmy teraz, jak zmieniają się przedziały ufności w przypadku gdy mamy jeden zmienny parametr oraz trzy ustalone parametry. Naszymi parametrami są  $d, \alpha, \lambda, m$ .

Pierwszym sprawdzonym przypadkiem jest zmienny parametr  $d = 1, 2, 3$  oraz stałe parametry  $m = 30, \alpha = 0.05, \lambda = 1$ , który wprost występuje w definicji naszego kontraktu oraz we wzorze na składkę netto.

d	$\lambda$	m	punktowe	dolne	górne
1	1	30	0.2120	0.0380	0.3861
2	1	30	0.0582	0.0000	0.1328
3	1	30	0.0160	0.0000	0.0438

Tabela 8: Oszacowania w zależności od  $d = 1, 2, 3$  ( $\lambda = 1, m = 30, \alpha = 0.05$ )

Przypomnijmy, że  $H(R) = g\left(\frac{1}{\lambda}\right) = \frac{1}{\lambda} \exp(-d\lambda)$ , więc jest to funkcja malejąca, czyli nasze oszacowania punktowe powinny maleć. Zgodnie z założeniami, wraz ze wzrostem parametru  $d$  oszacowania punktowe faktycznie maleją. Ponadto, dla coraz większych wartości  $d$  lewe krańce przedziałów ufności są już zerem. Widzimy zatem, że środek naszego przedziału ufności zmierza w lewą stronę.

d	$\lambda$	m	punktowe	dolne	górne
1	1	30	0.3371	0.0905	0.5836
1	2	30	0.0296	0.0000	0.0674
1	3	30	0.0098	0.0000	0.0252

Tabela 9: Oszacowania w zależności od  $\lambda = 1, 2, 3$  ( $d = 1, m = 30, \alpha = 0.05$ )

W przypadku powyższej tabeli (9) otrzymaliśmy analogicznie wnioski jak przy tabeli (8). Ponownie, nasze oszacowania punktowe wraz ze wzrostem parametru  $\lambda$  maleją, zaś środek przedziału ufności dąży do lewej strony.

---

<sup>6</sup>Jeśli dla ciągu zmiennych losowych  $T_m$  mamy  $\sqrt{m}(T_m - \mu) \rightarrow \mathbf{N}(0, \sigma^2)$ , to zachodzi  $\sqrt{m}(h(T_m) - h(\mu)) \rightarrow \mathbf{N}(0, \sigma^2(h'(\mu))^2)$

d	$\lambda$	m	punktowe	dolne	górne
1	1	30	0.3234	0.0844	0.5625
1	1	50	0.2965	0.1231	0.4700
1	1	100	0.3144	0.1862	0.4426

Tabela 10: Oszacowania w zależności od  $m = 30, 50, 100$  ( $\lambda = 1, d = 1, \alpha = 0.05$ )

Wzrost parametru  $m$  nie wpływa na oszacowania punktowe. Widzimy, że dla  $m = 30$  oszacowanie wyniosło 0.32, zaś dla większego  $m = 100$  mamy 0.31. Nie widać tutaj żadnej zależności. W przypadku dolnych krańców przedziału ufności dostrzegamy, że wraz ze wzrostem parametru  $m$  one również rosną. Odwrotna sytuacja występuje dla górnych przedziałów ufności - dla coraz większego  $m$  prawe krańce maleją.

d	$\alpha$	m	punktowe	dolne	górne
1	0.01	30	0.2948	0.0017	0.5878
1	0.05	30	0.5837	0.2113	0.9562
1	0.1	30	0.3719	0.1492	0.5947

Tabela 11: Oszacowania w zależności od  $\alpha = 0.01, 0.05, 0.1$  ( $\lambda = 1, m = 30, d = 1$ )

Tabela (11) zawiera zmienny parametry  $\alpha$ . Można łatwo zorientować się, że w tym przypadku brakuje jakiejkolwiek zależności. Zarówno oszacowania punktowe jak i przedziałowe zmieniają się losowo, niezależnie od większego bądź mniejszego parametru  $\alpha$ .

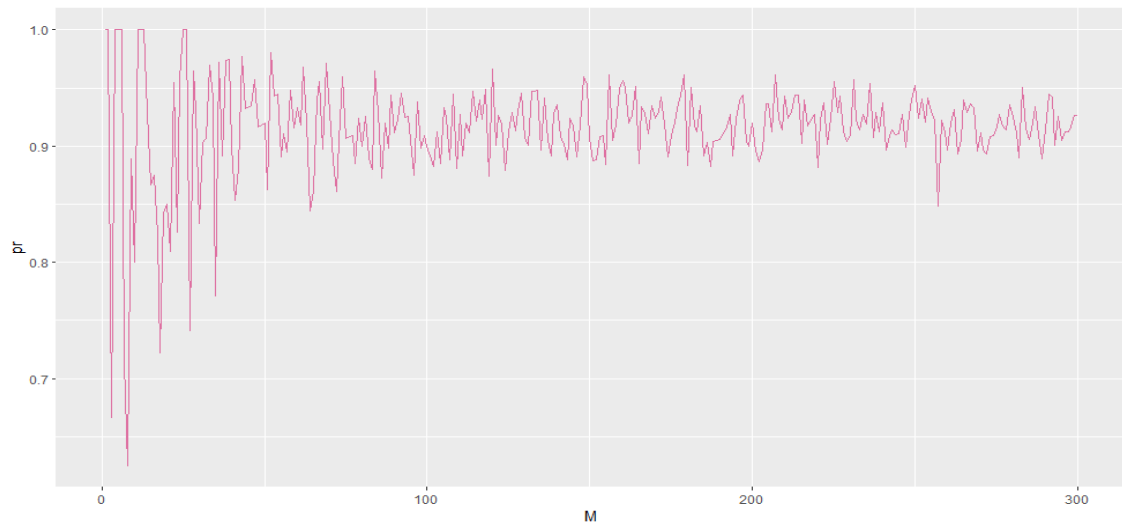
## 3.2 ZADANIE 2

Zadanie drugie jest kontynuacją zadania pierwszego. Naszym celem będzie przeprowadzenie symulacji, które oszacują prawdopodobieństwo pokrycia nieznannej wartości  $H(R)$  przez zaproponowany przedział ufności. Przyjmijmy, że  $m = 30, \lambda = d = 1$  oraz kilka wartości parametru  $\alpha = 0.01, 0.05, 0.1, 0.2, 0.5, 0.8, 0.99$ .

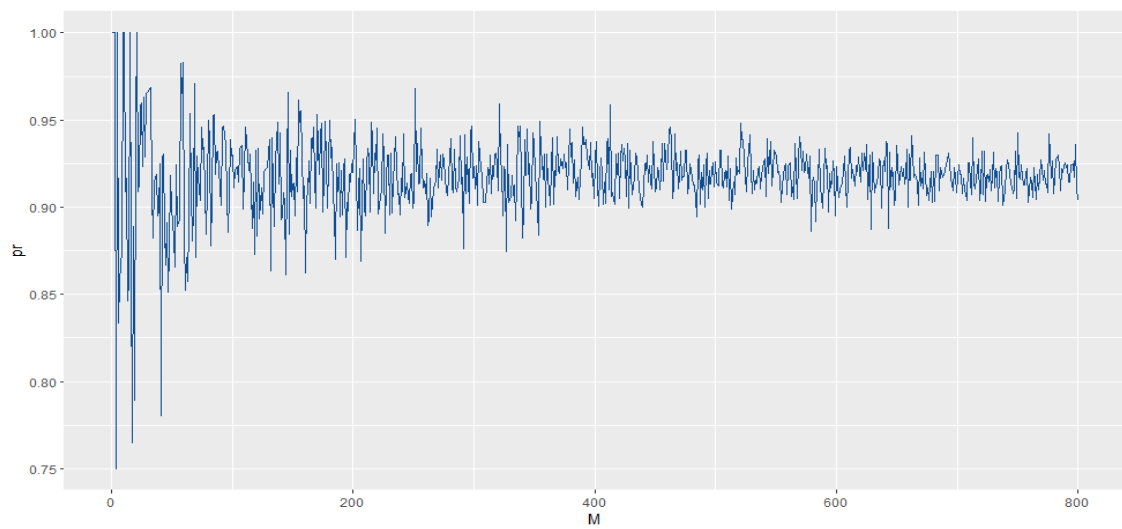
$\alpha$	$1 - \alpha$	prawdopodobieństwo	różnica bezwzględna
0.01	0.99	0.95	0.04
0.05	0.95	0.92	0.03
0.10	0.90	0.88	0.02
0.20	0.80	0.79	0.01
0.50	0.50	0.49	0.01
0.80	0.20	0.20	0.00
0.99	0.01	0.01	0.00

Tabela 12: Wyniki dla różnych poziomów ufności

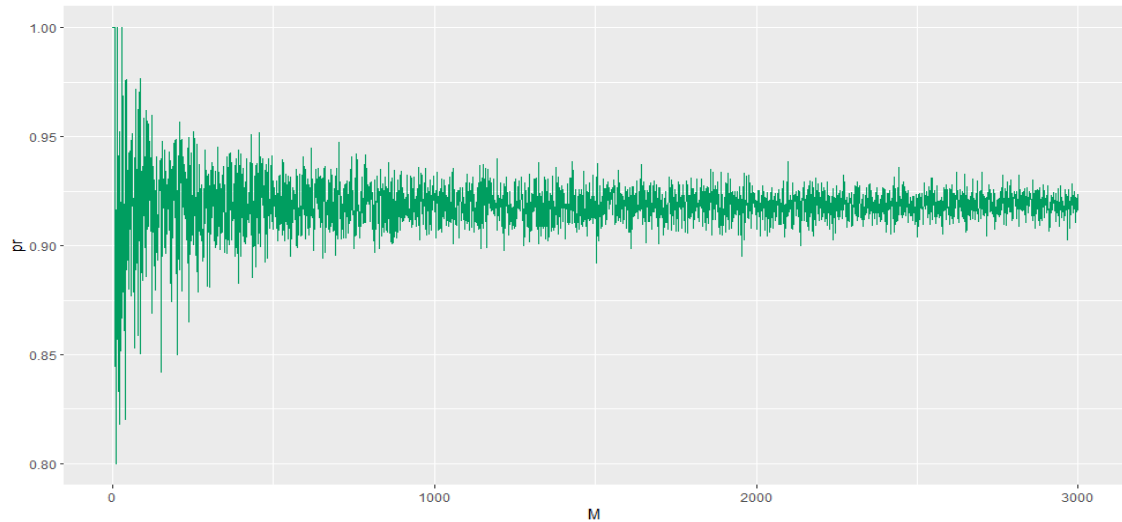
Pierwszą rzeczą rzucają się w oczy, jest fakt, że różnice bezwzględne nie są dużymi wartościami. Widzimy, że wraz ze wzrostem parametru  $\alpha$  dostajemy bardzo dokładne oszacowanie z różnicą bezwzględna równą 0. Dla  $\alpha \leq 0.5$  otrzymujemy oszacowania z niedomiarem, to znaczy, że przy wartości na przykład  $1 - \alpha = 0.5$  prawdopodobieństwo wynosi 0.49.



Rysunek 14: Wykres zależności prawdopodobieństwa pokrycia dla  $M = 300$  - opracowanie własne



Rysunek 15: Wykres zależności prawdopodobieństwa pokrycia dla  $M = 800$  - opracowanie własne



Rysunek 16: Wykres zależności prawdopodobieństwa pokrycia dla  $M = 3000$  - opracowanie własne

Przyglądając się powyższym trzem wykresom, widzimy, że minimalną rozsądną wartością  $M$  jest 800. Poniżej tej wartości wykres zależności prawdopodobieństwa pokrycia ma duże wahania. Ponadto, na wykresie (16) możemy dostrzec, że nasz wykres stabilizuje się wokół poziomu 0.92 – 0.93. Wracając do tabeli (12) widzimy, że dla  $\alpha = 0.05$  (dla takiej wartości  $\alpha$  generowany był wykres) właśnie wartość 0.92 jest naszym prawdopodobieństwem. Oznacza to, że nasze oscylacje wahają się wokół szacowanej, a nie teoretycznej wartości prawdopodobieństwa.

Dodatkowo przypomnijmy, że w części I w zadaniu 3 dokonywaliśmy podobnych analiz. Jednakże wyszło nam tam, że wykres stabilizuje się wokół teoretycznej wartości prawdopodobieństwa, czyli dokładnie odwrotnie, jak w powyższym zadaniu.

## 4 PODSUMOWANIE

W sprawozdaniu nauczyliśmy definiować, analizować oraz generować proces znakowany na podstawie różnych sposobów implementacji. Zauważyliśmy, że nie zawsze oczekiwanie lepsza funkcja jest funkcją faktycznie lepszą czy bardziej nadającą się do danych rzeczywistych. Nauczyliśmy się także badać jakość aproksymacji złożonego rozkładu Poissona przez rozkład normalny. Na podstawie zależności między rozkładem logarytmicznie normalnym a normalnym znajdowaliśmy estymatory parametrów  $\mu$  oraz  $\sigma$ . Warto dodać, że nauczyliśmy się oszacowywać punktowo i przedziałowo wartości składki netto kontraktu stop-loss, a także zauważyliśmy, że wykres zależności prawdopodobieństwa nie zawsze oscyluje wokół wartości teoretycznej - niekiedy jest to również wartość szacowana.

## 5 DODATEK

W tym podrozdziale zostały przedstawione niektóre funkcje, do których odwoływałam się w sprawozdaniu.

```
A1 <- function(t0, lambda, mu, sigma){
  ti <- c()
  while(sum(ti) <= t0){
    next_ti <- rexp(1, lambda)
    ti <- c(ti, next_ti)
  }
  ti <- head(ti, length(ti) - 1)
  punkty <- cumsum(ti)
  znaki <- cumsum(rlnorm(length(punkty), mu, sigma))
  dane_A1 <- cbind(punkty = c(0, punkty), znaki = c(0, znaki))
  plot_A1 <- ggplot(data = as.data.frame(dane_A1), aes(x = punkty, y = znaki)) +
    geom_step(data = as.data.frame(znaki), aes(y = znaki), col = "#9966CC") +
    xlab("skumulowane czasy wektora ti") +
    ylab("skumulowane losowo wygenerowane liczby z rozkładu LN") +
    ggtitle("Wykres procesu znakowanego o stałej intensywności A1")
  return(plot_A1)
}
```

Rysunek 17: Funkcja sposobu A1 z zadania pierwszego części I - opracowanie własne

```
A2 <- function(t0, lambda, mu, sigma){
  N <- rpois(1, lambda*t0)
  punkty <- runif(N, min = 0, max = t0)
  punkty <- sort(punkty)
  znaki <- cumsum(rlnorm(N, mu, sigma))
  dane_A2 <- cbind(punkty = c(0, punkty), znaki = c(0, znaki))
  plot_A2 <- ggplot(data = as.data.frame(dane_A2), aes(x = punkty, y = znaki)) +
    geom_step(data = as.data.frame(znaki), aes(y = znaki), col = "#FF2052") +
    xlab("zbiór N losowych punktów z rozkładu jednostajnego") +
    ylab("losowo wygenerowane N punktów z rozkładu LN") +
    ggtitle("Wykres procesu znakowanego o stałej intensywności A2")
  return(plot_A2)
}
```

Rysunek 18: Funkcja sposobu A2 z zadania pierwszego części I - opracowanie własne

```

LAMBDA <- function(t, alpha, beta){
  alpha * t^(beta)
}

LAMBDAinverse <- function(t, alpha, beta){
  (t/alpha)^(1/beta)
}

B1 <- function(t0, alpha1, beta1, mu, sigma){
  lambda_max_B1 <- LAMBDA(t0, alpha = alpha1, beta = beta1)
  N <- rpois(1, lambda_max_B1)
  punkty_j <- runif(N, min = 0, max = t0)
  punkty_j <- sort(punkty_j)
  punkty <- sapply(punkty_j, LAMBDAinverse, alpha = alpha1, beta = beta1)
  znaki <- cumsum(rlnorm(N, mu, sigma))
  dane_B1 <- cbind(punkty = c(0, punkty), znaki = c(0, znaki))
  plot_B1 <- ggplot(data = as.data.frame(dane_B1), aes(x = punkty, y = znaki)) +
    geom_step(data = as.data.frame(znaki), aes(y = znaki), col = "#006B3C") +
    xlab("Zbiór N losowych punktów z rozkładu jednostajnego") +
    ylab("losowo wygenerowane N punktów z rozkładu LN") +
    ggtitle("Wykres procesu znakowanego o zmiennej intensywności B1")
  return(plot_B1)
}

```

Rysunek 19: Funkcja sposobu B1 z zadania pierwszego części I - opracowanie własne

```

lambda <- function(t, alpha, beta){
  alpha * beta * (t^(beta - 1))
}

B2 <- function(t0, alpha, beta, mu, sigma){
  lambda_max_B2 <- lambda(t0, alpha, beta)
  N <- rpois(1, lambda_max_B2 * t0)
  punkty <- runif(N, min = 0, max = t0)
  punkty <- sort(punkty)
  u <- runif(N)
  t_delete <- c()
  for (i in 1:N){
    if(u[i] >= lambda(punkty[i], alpha, beta)/lambda_max_B2){
      t_delete <- c(t_delete, punkty[i])
    }
  }
  punkty <- punkty[! punkty %in% t_delete]
  k <- length(punkty)
  znaki <- cumsum(rlnorm(k, mu, sigma))
  dane_B2 <- cbind(punkty = c(0, punkty), znaki = c(0, znaki))
  plot_B2 <- ggplot(data = as.data.frame(dane_B2), aes(x = punkty, y = znaki)) +
    geom_step(data = as.data.frame(znaki), aes(y = znaki), col = "#006B3C") +
    xlab("Zbiór N losowych punktów z rozkładu jednostajnego") +
    ylab("losowo wygenerowane N punktów z rozkładu LN") +
    ggtitle("Wykres procesu znakowanego o zmiennej intensywności B2")
  return(plot_B2)
}

```

Rysunek 20: Funkcja sposobu B2 z zadania pierwszego części I - opracowanie własne



```

pu <- function(mu, sigma, lambda, alpha, a1 = NULL){
  if (is.null(a1)){a1 <- alpha/2}
  ex <- exp(mu + (sigma^2)/2)
  ex2 <- exp(2 * mu + 2*(sigma^2))
  a <- lambda * ex + qnorm(a1) * sqrt(lambda*ex2)
  b <- lambda * ex + qnorm(1 - (alpha - a1)) * sqrt(lambda * ex2)
  if(a < 0){a <- 0}
  return(c(a, b))
}

```

Rysunek 21: Funkcja wyznaczająca przedziały ufności z zadania drugiego części I - opracowanie własne

```

pr_p <- function(mu, sigma, lambda, alpha, M, a1 = NULL){
  if(is.null(a1)){a1 <- alpha/2}
  pu1 <- pu(mu, sigma, lambda, alpha, a1)
  w <- rep(0, M)
  for(i in 1:M){
    N <- rpois(1, lambda)
    S_N <- sum(rlnorm(N, mu, sigma))
    w[i] <- ifelse(pu1[1] < S_N & S_N < pu1[2], 1, 0)
  }
  prawd <- sum(w)/M
  return(prawd)
}

```

Rysunek 22: Funkcja wyznaczająca prawdopodobieństwo pokrycia z zadania drugiego części I - opracowanie własne

```

pu_est <- function(X1, alpha, a1 = NULL){
  if(is.null(a1)){a1 <- alpha/2}
  l <- length(X1)
  mu_e <- mean(log(X1))
  sigma_e <- sd(log(X1))
  ex_e <- exp(mu_e + (sigma_e^2)/2)
  ex2_e <- exp(2*mu_e + 2*(sigma_e^2))
  a_e <- l*ex_e + qnorm(a1) * sqrt(l*ex2_e)
  b_e <- l*ex_e + qnorm(1 - (alpha - a1))* sqrt(l*ex2_e)
  return(c(a_e, b_e))
}

```

Rysunek 23: Funkcja wyznaczająca estymowane przedziały ufności z zadania trzeciego części I - opracowanie własne

```

pr_p_est <- function(mu, sigma, lambda, alpha, M, a1 = NULL){
  if(is.null(a1)){a1 <- alpha/2}
  w <- rep(0, M)
  for (i in 1:M){
    while(TRUE){
      N1 <- rpois(1, lambda)
      if (N1 > 1){break}
    }
    X1 <- rlnorm(N1, mu, sigma)
    pu1_e <- pu_est(X1, alpha, a1)
    mu_e <- mean(log(X1))
    sigma_e <- sd(log(X1))
    N1 <- rpois(1, lambda)
    S_N1 <- sum(rlnorm(N1, mu_e, sigma_e))
    w[i] <- ifelse(pu1_e[1] < S_N1 & S_N1 < pu1_e[2], 1, 0)
  }
  prawd <- sum(w)/M
  return(prawd)
}

```

Rysunek 24: Funkcja wyznaczająca estymowane prawdopodobieństwo pokrycia z zadania trzeciego części I - opracowanie własne

## 6 BIBLIOGRAFIA

1. Alicja Jokiel - Rokita *Wykłady ze Statystyki w Finansach i Ubezpieczeniach*
2. Kamil Bogus *Materiały udostępnianie po laboratoriach*
3. Rafał Latała *Procesy Stochastyczne I* [https://www.mimuw.edu.pl/~rlatala/testy/proc/wykps1\\_1.pdf](https://www.mimuw.edu.pl/~rlatala/testy/proc/wykps1_1.pdf)
4. Przemysław Biecek *Przewodnik po pakiecie R* <http://pbiecek.github.io/Przewodnik/index.html>