

Raport 2

Bartłomiej Gintowt

4 kwietnia 2023

1 Zadanie 1

Celem w tym zadaniu jest dopasowanie modelu procesu ryzyka, do badanych przez nas danych liczbowych. Klasyczny model procesu ryzyka prezentuje następujący wzór:

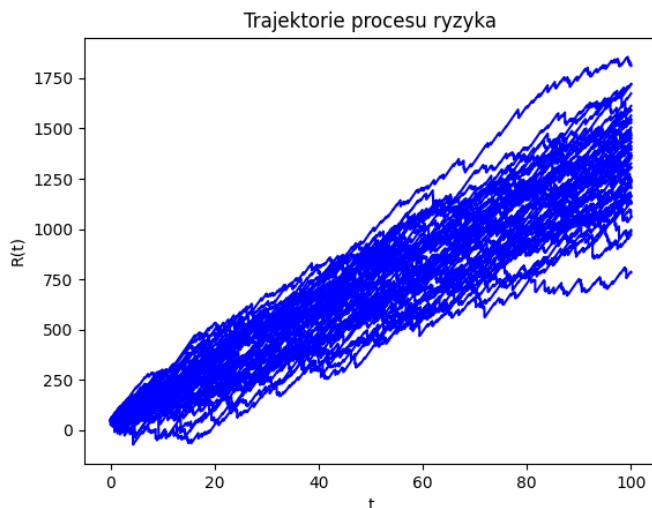
$$R(t) = u + ct - \sum_{i=1}^{N(t)} X_i.$$

Parametr $u > 0$ odpowiada za kapitał początkowy. $N(t)$ jest jednorodnym procesem Poissona z intensywnością $\lambda > 0$, a X_i jest ciągiem i.i.d. niezależnym od $N(t)$ o dodatnich wartościach i średniej $EX_i = \mu$. Odpowiada on za straty w budżecie. Natomiast zmienna $c > 0$ odpowiada za wysokość premii. Można ją również zapisać w nieco innej postaci:

$$c = (1 + \theta)\mu\lambda,$$

gdzie $0 < \theta < 1$. Przebieg procesu ryzyka będzie więc zależał od dwóch parametrów: μ i λ , a także od rozkładu prawdopodobieństwa ciągu X_i .

Rozważane przez nas dane liczbowe zawierają wartości jakie przyjmuje 50 trajektorii pewnego procesu ryzyka, o horyzoncie czasowym $T = 100$, oraz skoku czasowym $\Delta t = 0.01$. Wykresy tych trajektorii są zaprezentowane na poniższym rysunku:

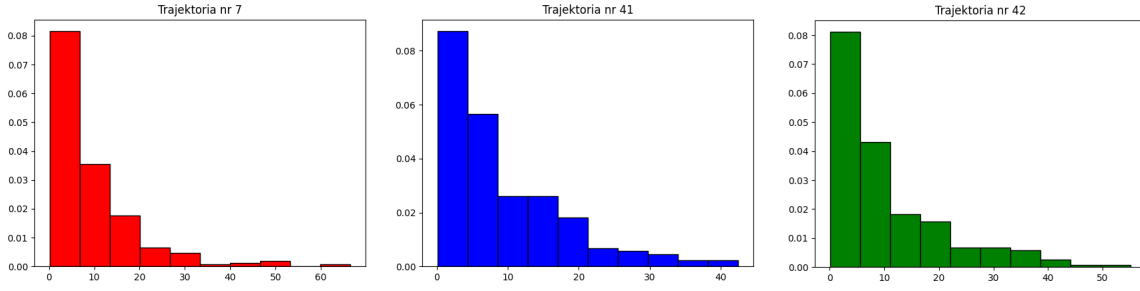


Rysunek 1: Wykres trajektorii wygenerowanych z danych.

Aby dopasować model ryzyka, do danych, należy odszukać przybliżone wartości newralgicznych parametrów i podstawienie ich do klasycznego wzoru na proces ryzyka.

1.1 Rozkład zmiennych X_i

Próba X_i odpowiada za straty w budżecie. Aby określić z jakiego rozkładu jest ta próba, musimy przeanalizować wszystkie wartości "spadków" dla każdej trajektorii. Rozważmy histogramy długości spadków dla kilku losowo wybranych trajektorii.



Rysunek 2: Histogramy wartości strat dla wybranych trajektorii.

Jak widać, histogramy te są do siebie bardzo podobne, a ich kształt przypomina krzywą gęstości dla rozkładu wykładniczego. Możemy zatem przyjąć, że rozkład wartości spadków jest rzeczywiście z rozkładu wykładniczego. Jaka jest zatem wartość parametru λ tego rozkładu? Wiemy, że wartość oczekiwana tego rozkładu ($EX = \frac{1}{\lambda}$), oraz wariancja ($VarX = \frac{1}{\lambda^2}$) zależą jedynie od parametru λ . Wykorzystajmy ten fakt i dla każdej trajektorii wyznaczmy wartość średnią i wariancję. Następnie z otrzymanych wyników wyliczymy wartości λ_i dla każdej trajektorii procesu ryzyka. Ostatecznie określimy średnią arytmetyczną z otrzymanych parametrów (osobną dla wartości średniej i wariancji). Rezultat jest następujący:

- λ wyznaczona z wartości średnich:

$$\bar{\lambda} = \frac{1}{50} \sum_{i=1}^{50} \lambda_i = \frac{1}{50} \sum_{i=1}^{50} \frac{1}{EX_i} = 0.0999184 \approx 0.1$$

- λ wyznaczona z wariancji:

$$\bar{\lambda} = \frac{1}{50} \sum_{i=1}^{50} \lambda_i = \frac{1}{50} \sum_{i=1}^{50} \sqrt{\frac{1}{VarX_i}} = 0.0994190 \approx 0.1$$

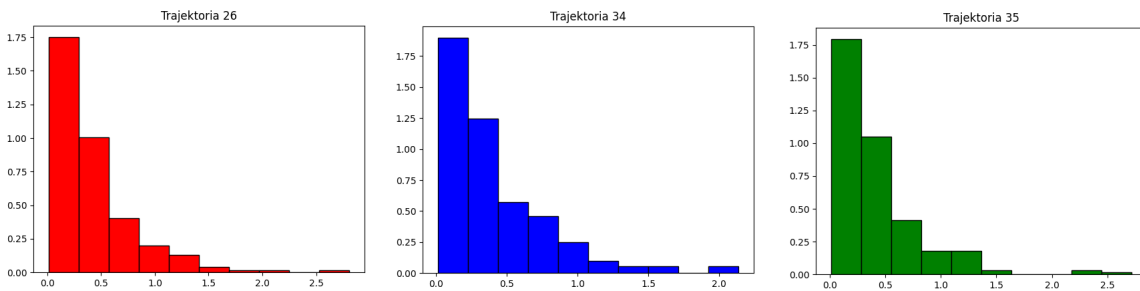
Jak możemy zauważyć, wartość parametru λ jest w przybliżeniu równa 0.1 dla obu metod. Stąd można stwierdzić, że wartości strat są z rozkładu wykładniczego $X_i \sim \mathcal{Exp}(0.1)$.

1.2 Parametr λ

Kolejnym parametrem procesu ryzyka, który chcemy wyznaczyć, jest λ - intensywność jednorodnego procesu Poissona, na którym opiera się cały proces ryzyka. Jednym ze sposobów wyznaczenia λ jest policzenie wszystkich odległości między sąsiednimi skokami, a następnie wyznaczenie rozkładu otrzymanych wartości. Wiemy bowiem, że:

$$T_i \sim \mathcal{Exp}(\lambda),$$

gdzie λ jest parametrem intensywności jednorodnego procesu Poissona. Zatem wystarczy odnaleźć wartość parametru rozkładu wykładniczego (analogicznie jak przy wyznaczaniu rozkładu X_i). Poniżej prezentujemy kilka przykładowych histogramów dla wybranych trajektorii.



Rysunek 3: Histogramy odległości między kolejnymi skokami dla wybranych trajektorii.

Parametr λ wyznaczamy analogicznie do wyznaczania parametru rozkładu próby X_i . Jego wartość jest równa $\lambda \approx 2.36$.

1.3 Parametr θ

Aby odnaleźć wartość parametru θ musimy najpierw obliczyć c . Wiemy, że c odpowiada za wzrost premii, a ten jest liniowy. Można powiedzieć, że parametr c odpowiada za tempo wzrostu wartości w procesie ryzyka. Można go wyznaczyć z poniższego wzoru:

$$c = \frac{R(t + \Delta t) - R(t)}{\Delta t},$$

gdzie Δt jest jednostkowym skokiem w czasie (w naszym przypadku $\Delta t = 0.01$). Należy jednak pamiętać, że ten wzór będzie prawdziwy jedynie wtedy, gdy $R(t + \Delta t) > R(t)$. Dla naszych danych $c \approx 37.5$. Jak znamy już wartość c wystarczy skorzystać z poniższego wzoru:

$$c = (1 + \theta)\mu\lambda,$$

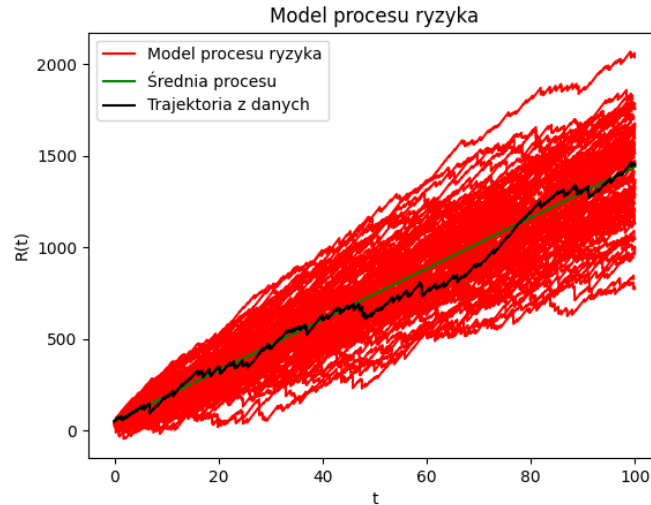
gdzie $\lambda = 2.36$ jest intensywnością jednorodnego procesu Poissona, a $\mu = EX_i = 10$ to wartość oczekiwana wysokości strat. Ostatecznie otrzymujemy, że $\theta \approx 0.589$.

1.4 Dopasowanie modelu do danych

Po wyestymowaniu wszystkich newralgicznych parametrów otrzymujemy model procesu ryzyka o poniższym wzorze:

$$R(t) = 50 + 37.5t - \sum_{i=1}^{N(t)} X_i,$$

gdzie $X_i \sim \text{Exp}(0.1)$, a $N(t)$ jest jednorodnym procesem Poissona, o intensywności $\lambda = 2.36$. Sprawdźmy teraz, czy zaproponowany przez nas model rzeczywiście dobrze przybliża proces ryzyka dla wyrysowanych trajektorii. W tym celu wygenerujemy 100 trajektorii z zaproponowanego procesu i porównamy z rozważanymi przez nas danymi. Wynik jest następujący:



Rysunek 4: Model procesu ryzyka.

Jak widać, nasz model dobrze przybliża proces ryzyka zadany za pomocą danych. Premia rośnie w takim samym tempie, a straty występują z podobną intensywnością. Końcowa wysokość budżetu mieści się między wartością 1000, a 2000 - podobnie jak dla badanych przez nas danych.

1.5 Prawdopodobieństwo ruiny

Moment ruiny (bankructwa) definiujemy jako czas spadku kapitału poniżej zera:

$$\tau(u) = \inf\{t \geq 0 : R(t) < 0\}.$$

Prawdopodobieństwo ruiny w skończonym czasie wyrażamy następująco:

$$\Psi(u, T) = P(\tau(u) < T),$$

gdzie u jest kapitałem początkowym, a T horyzontem czasowym. Aby numerycznie wyliczyć prawdopodobieństwo ruiny, należy się posłużyć metodą Monte-Carlo. Poniższy algorytm prezentuje krok po kroku, co należy wykonać:

1. Generuj N trajektorii procesu ryzyka $R^{(1)}(t), \dots, R^{(N)}(t)$ na przedziale $[0, T]$
2. Wyznacz $n = \#\{i \in \{1, \dots, N\} : \min_{t \in [0, T]} R^{(i)}(t) < 0\}$
3. Wstaw $\Psi(u, T) = \frac{n}{N}$

Wykorzystując ten algorytm, wyliczyliśmy prawdopodobieństwo ruiny dla dwóch skończonych horyzontów czasowych: $T_1 = 100$ i $T_2 = 200$. Otrzymaliśmy następujące wyniki: $\Psi(50, 100) = 0.1018$ i $\Psi(50, 200) = 0.0986$. Okazuje się, że wyniki są do siebie bardzo zbliżone i zwiększenie horyzontu czasowego dwukrotnie nie miało większego wpływu na zmianę.

1.6 Wzór Pollaczka-Chinczyna

Poprzednia metoda sprawdza się tylko wtedy gdy horyzont czasowy jest skończony. Jeżeli chcemy rozważyć przypadek nieskończonego czasu, należy skorzystać ze wzoru Pollaczka-Chinczyna:

$$\Psi(u) = \frac{\theta}{1 + \theta} \sum_{i=1}^{\infty} \left(\frac{1}{1 + \theta} \right)^i B_n(u),$$

gdzie $B_n(u) = P(Y_1 + \dots + Y_n > u)$, a Y_i to ciąg zmiennych losowych i.i.d. o gęstości:

$$f(x) = \frac{1 - F_{X_i}}{EX_i}.$$

Procedura symulacyjna wygląda następująco:

1. Wstaw $I = 0$
2. Dla każdego $i = 1, \dots, N$
 - (a) Generujemy $k \sim \text{Geom}(\theta/(1 + \theta))$
 - (b) Generujemy Y_1, \dots, Y_k
 - (c) Jeśli $Y_1 + \dots + Y_k > u$ to $I = I + 1$
3. Wstaw $\phi(u) = I/N$

Zmienne Y_i są zmiennymi i.i.d. o gęstości

$$f(x) = \frac{1 - P(X_i \leq x)}{\mu} = \frac{1 - F_{X_i}(x)}{\mu} = \frac{1 - (1 - e^{-\frac{1}{\mu}x})}{\mu} = \frac{1}{\mu} e^{-\frac{1}{\mu}x} \sim \text{Exp}\left(\frac{1}{\mu}\right)$$

Wykorzystując powyższy algorytm otrzymaliśmy $\Psi(u) = 0.098642$

Do wyliczenia wartości teoretycznej korzystamy z alternatywnej postaci wzoru:

$$\Psi(u) = \frac{1}{1 + \theta} \exp\left(-\frac{\theta}{EX_i(1 + \theta)}u\right)$$

z której otrzymaliśmy wartość $\Psi(u) = 0.09862092833531873$

Jak możemy zauważyć wartość teoretyczna i wyliczona symulacyjnie są bardzo podobne, tak więc prawdopodobieństwo ruiny w nieskończonym czasie wynosi ok. 0.1.

2 Zadanie 2

Celem zadania drugiego jest oszacowanie średniego czasu wyjścia z przedziału $[a, b]$ w zależności od $x \in (a, b)$ dla procesu $(B_t^x)_{t \geq 0}$, będącego ruchem Browna startującym z $x \in (a, b)$, gdzie $B_t^x = B_t^0 + x$ dla B_t^0 , który z kolei jest klasycznym ruchem Browna/procesem Wienera startującym z 0. Oszacujemy również prawdopodobieństwo, że wyjście z przedziału nastąpiło przez b , tzn. $P(B_{\tau^x}^x = b)$ dla $x \in (a, b)$ oraz dopasujemy odpowiednie funkcje dla uzyskanych symulacyjnie wyników.

2.1 Teoretyczny ruch Browna

Ruch Browna jest procesem stochastycznym $B(t)_{t \geq 0}$, który spełnia następujące warunki:

1. $B(0) = 0$
2. $B(t)$ ma niezależne przyrosty
3. $B(t)$ ma stacjonarne przyrosty
4. $B(t)$ ma ciągłe trajektorie
5. $B(t) \sim N(0, t)$

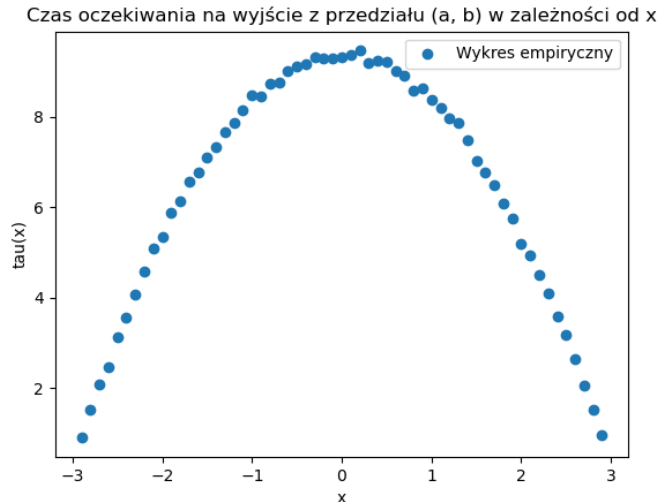
2.2 Algorytm generowania ruchu Browna

1. Ustalamy horyzont czasowy T , który podzielimy na n równych odcinków
2. $t_i = i \cdot h$, gdzie $i = 0, 1, 2, \dots, n$ oraz $h = \frac{T}{n}$
3. $B(t_0) = B(0) = 0$
4. $B(t_{i+1}) = B(t_i) + h^{\frac{1}{2}} \cdot \zeta_i$, gdzie $\zeta_i \sim N(0, h)$

2.3 Oszacowanie średniego czasu wyjścia z przedziału

Do oszacowania średniego czasu wyjścia z przedziału $[a, b]$ w zależności od $x \in (a, b)$ wykorzystamy symulację Monte Carlo dla każdego x . Pojedynczy krok Monte Carlo będzie polegał na wysymulowaniu trajektorii, z której wyznaczamy czas opuszczenia przedziału $[a, b]$. Całą procedurę powtarzamy N (liczba kroków Monte Carlo) razy i wyznaczamy średnią z otrzymanych wyników.

W celu przeprowadzenia symulacji za przedział $[a, b]$ wybierzemy przedział $[-3, 3]$. Zmienną x będziemy generować na przedziale $(-3, 3)$ dla wzrostu wartości co 0.1. Bazując na nieskończonym horyzoncie czasowym weźmiemy $h = 0.01$ oraz wykorzystamy $N = 10000$ kroków Monte Carlo. Po wykonaniu symulacji otrzymujemy wykres czasu średniego wyjścia z przedziału w zależności od x .



Rysunek 5: Wykres średniego czasu wyjścia z przedziału.

2.4 Dopasowanie funkcji

Kształt otrzymanego wykresu przypomina wykres funkcji kwadratowej. W celu wyznaczenia przybliżonego współczynnika c paraboli o równaniu $f(x) = ax^2 + bx + c$ bierzemy przedział danych dla wartości $x \in [-\frac{1}{2}, \frac{1}{2}]$ oraz obliczamy ich średnią arytmetyczną. Otrzymujemy:

$$c = 9.271224799999999.$$

Używamy takiej metodologii, ponieważ przy wierzchołku funkcji występują znaczne odchylenia mogące zaburzyć prawidłowość wyniku. W celu znalezienia współczynnika b wykorzystamy założenie, iż wartości funkcji w $x = 3$, $x = -3$ są pomijalnie małe, zatem uzyskujemy:

$$\lim_{x \rightarrow -3^+} f(x) = 0,$$

$$\lim_{x \rightarrow 3^-} f(x) = 0.$$

Możemy wówczas przyrównać wartości $f(-3)$ i $f(3)$:

$$f(-3) = a \cdot 9 - b \cdot 3 + c = a \cdot 9 + b \cdot 3 + c = f(3),$$

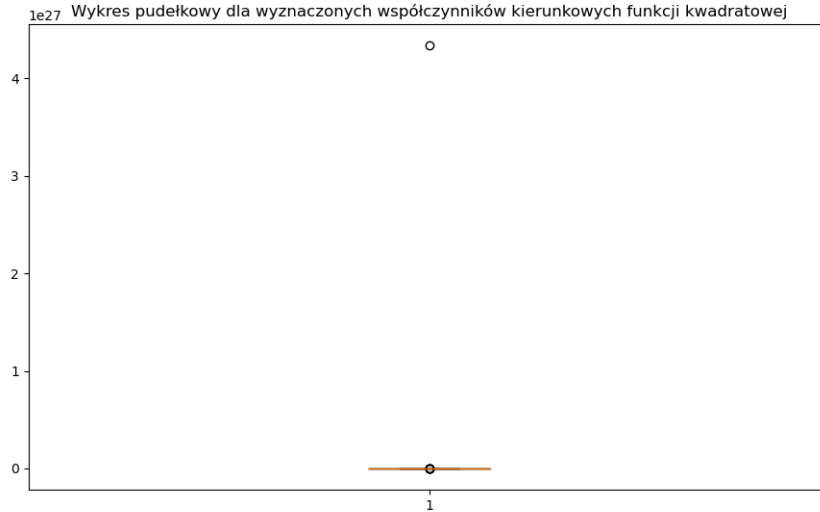
z czego otrzymujemy równość

$$-b = b.$$

Z tego równania ostatecznie wynika, iż $b = 0$. Wiedząc, że $b = 0$ współczynnik kierunkowy funkcji wyznaczamy ze wzoru

$$a_i = \frac{y_i - c}{x_i^2},$$

gdzie y_i to wartość czasu oczekiwania dla $x_i \in (-3, 3)$. Jako, iż otrzymane współczynniki a_i zawierają znaczące wartości odstające zaburzające przybliżenie wyniku, które możemy zaobserwować na wykresie pudełkowym:



Rysunek 6: Wykres pudełkowy wyznaczonych współczynników a .

W celu usunięcia zbyt mocno odstających wartości a_i wyznaczmy wpierw wąsy dla wykresu pudełkowego, które są maksymalnymi wartościami nieodstającymi dla danych. Korzystamy ze wzorów:

1. $C_1 = Q_{0.25} - \frac{3}{2} \cdot (Q_{0.75} - Q_{0.25})$, gdzie C_1 oznacza wąs dolny

2. $C_2 = Q_{0.75} + \frac{3}{2} \cdot (Q_{0.75} - Q_{0.25})$, gdzie C_2 oznacza wąs górny

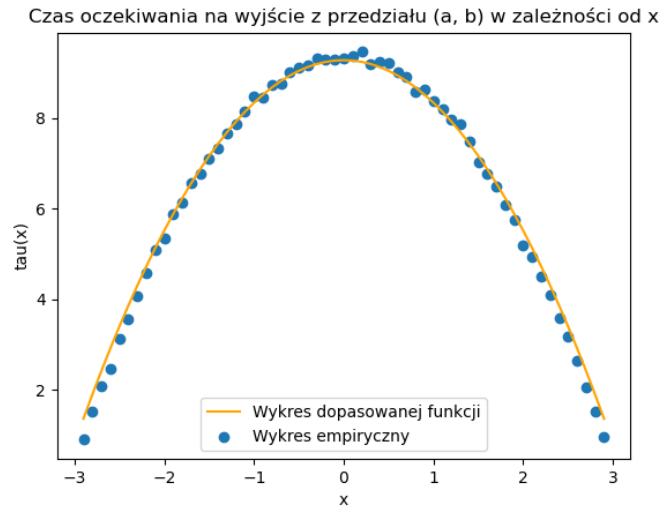
Dla naszych danych otrzymujemy wówczas wartości:

1. $C_1 = -1.2248702053355291$
2. $C_2 = -0.5873417032286518$

Następnie wszystkie współczynniki a_i , których wartości nie znalazły się w przedziale $[C_1, C_2]$ odrzucamy, gdyż traktujemy je jako odstające. Z pozostałych współczynników liczymy średnią arytmetyczną i otrzymujemy:

$$a = -0.9396145564949128.$$

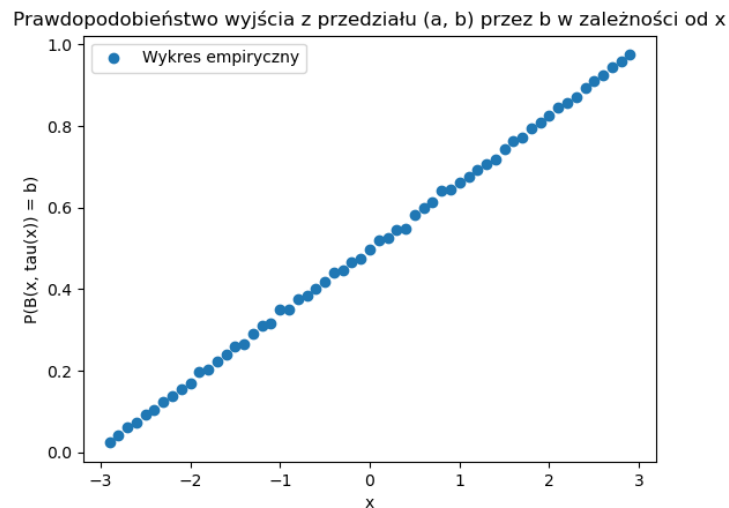
Wykres wyznaczonej dopasowanej funkcji $f(x) = -0.9396145564949128 \cdot x^2 + 9.271224799999999$:



Rysunek 7: Wykres średniego czasu wyjścia z przedziału wraz z dopasowaną funkcją.

2.5 Prawdopodobieństwo, że wyjście z przedziału nastąpiło przez b

Podczas zliczania średniego czasu wyjścia z przedziału (a, b) trajektorii zliczamy również ilość razy, kiedy to wyjście nastąpiło przez b . Następnie uzyskany wynik dzielimy przez łączną ilość trajektorii i otrzymujemy prawdopodobieństwo wyjścia z przedziału poprzez b dla pojedynczego x . Procedurę powtarzamy dla każdego $x \in (a, b)$ i uzyskujemy wykres prawdopodobieństwa wyjścia przez b w zależności od x .



Rysunek 8: Wykres prawdopodobieństwa wyjścia z przedziału przez b w zależności od x .

2.6 Dopasowanie funkcji

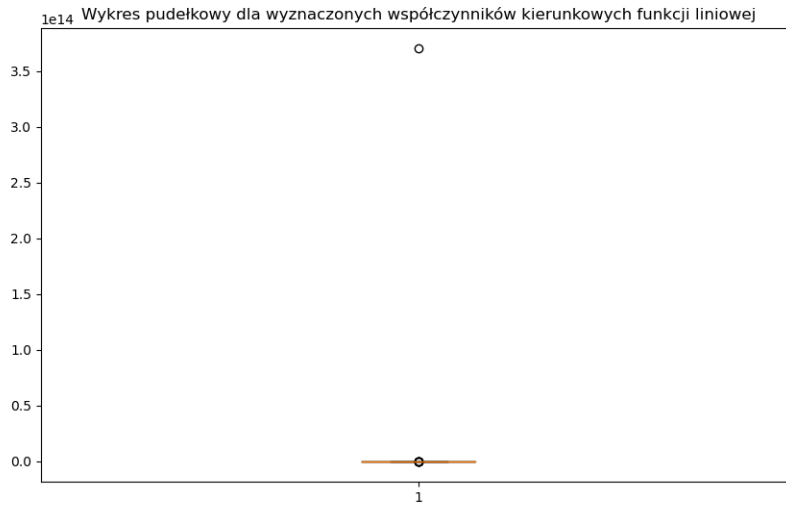
Kształt otrzymanego wykresu przypomina wykres funkcji liniowej. W celu wyznaczenia współczynnika b prostej o równaniu $f(x) = ax + b$ wyznaczmy średnią arytmetyczną wartości funkcji dla $x \in (-0.2, 0.2)$, gdyż jest on wyrazem wolny prostej, natomiast wartości w okolicach $x = 0$ mogą cechować się pewną niedokładnością. Otrzymujemy współczynnik:

$$b = 0.490375.$$

W celu wyznaczenia współczynnika kierunkowego prostej $f(x)$ skorzystamy ze wzoru

$$a_i = \frac{y_i - b}{x_i},$$

gdzie y_i to wartość prawdopodobieństwa wyjścia trajektorii przez b dla $x \in (-3, 3)$. Otrzymane współczynniki zawierają niepożądane wartości odstające, które można zaobserwować na wykresie pudełkowym:



Rysunek 9: Wykres pudełkowy współczynników a prostej $ax + b$.

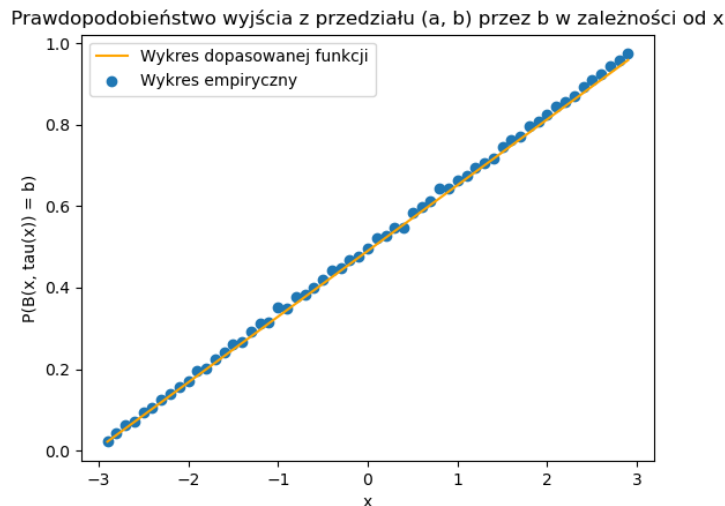
W celu pozbycia się zbyt mocno odstających współczynników a_i wyznaczmy wartości wąsów dla wykresu pudełkowego:

1. $C_1 = -2.46083017113095$, gdzie C_1 jest wąsem dolnym
2. $C_2 = 2.826093340773804$, gdzie C_2 jest wąsem górnym

Współczynniki, których wartości a_i znajdują się poza przedziałem $[C_1, C_2]$ traktujemy jako zbyt odstające i niefortunnie wpływające na wynik, zatem są one odrzucane. Z pozostałych współczynników liczymy średnią arytmetyczną i otrzymujemy:

$$a = 0.1616317391578326.$$

Wykres wyznaczonej dopasowanej funkcji $f(x) = 0.1616317391578326x + 0.490375$:



Rysunek 10: Wykres prawdopodobieństwa wyjścia z przedziału przez b wraz z dopasowaną funkcją.

3 Podsumowanie

Bazując na zadaniu pierwszym udowodniliśmy, że możliwa jest estymacja parametrów procesu ryzyka opierając się jedynie na wartościach z danych z pewnego okresu czasu. Za pomocą otrzymanych parametrów jesteśmy w stanie przewidzieć ruch trajektorii w przyszłości. Fakt ten umożliwia nam wyznaczenie przybliżonego prawdopodobieństwa ruiny w pewnym okresie czasu oraz w czasie nieskończonym.

Wykorzystując ruch Browna startujący z pewnej wartości możemy w przybliżeniu oszacować czas średniego wyjścia jego trajektorii z zadanego przedziału. Zliczanie ile razy wyjście każdej trajektorii nastąpiło przez górną wartość przedziału, umożliwia wyznaczenie prawdopodobieństwa tego zdarzenia. Do uzyskanych wartości jesteśmy również w stanie odpowiednio dopasować funkcję po wcześniejszym symulacyjnym wyznaczeniu jej współczynników.