

# Proces Ryzyka

Piotr Mikler

236895

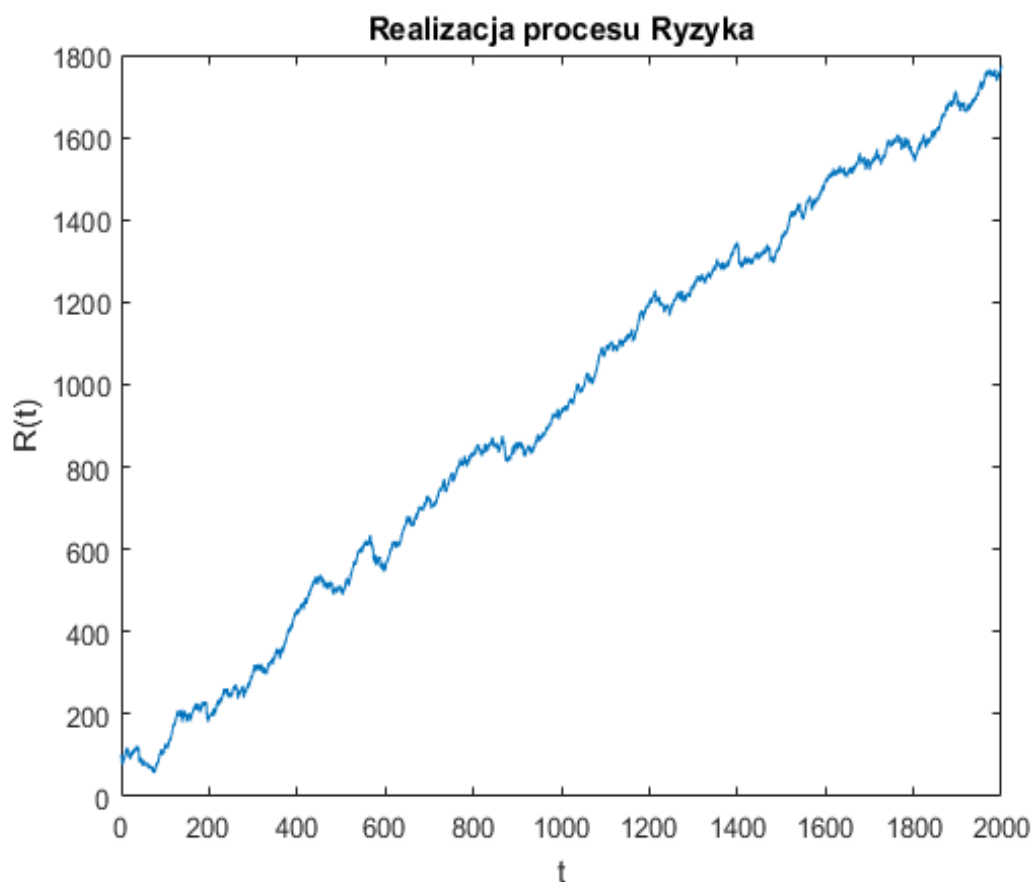
27 czerwca 2018

## Spis treści

<b>1</b>	<b>Wstęp</b>	<b>2</b>
<b>2</b>	<b>O Procesie Ryzyka</b>	<b>3</b>
<b>3</b>	<b>Estymacja parametrów</b>	<b>6</b>
3.1	Początkowy kapitał . . . . .	6
3.2	Funkcja przychodów . . . . .	6
3.3	Rozkład kwot wypłat . . . . .	8
3.4	Rozkład czasów oczekiwania na wypłaty . . . . .	10
<b>4</b>	<b>Symulacja</b>	<b>14</b>
<b>5</b>	<b>Prawdopodobieństwa Ruiny</b>	<b>16</b>
5.1	W skończonym czasie . . . . .	16
5.2	W nieskończonym czasie . . . . .	19
<b>6</b>	<b>Inne pomysły i niespełnione marzenia</b>	<b>21</b>
<b>7</b>	<b>Podsumowanie i wnioski</b>	<b>22</b>

# 1 Wstęp

Praca dotyczy zagadnienia klasycznego procesu ryzyka. Pokażę w niej w jaki sposób próbowałem estymować parametry tego procesu, bazując na otrzymanych danych, oraz w jaki sposób konstruowałem jego model. Dane które zostały mi udostępnione opisują wartości pewnej realizacji klasycznego procesu ryzyka, w spróbkowanych momentach czasu. Jest to próbka o liczności 40001 obserwacji, zawierająca kolejne wartości procesu od czasu  $t_0 = 0$  do  $T = 2000$  w odstępach 0.05 która po narysowaniu wygląda następująco (Rysunek 1):



Rysunek 1: Otrzymane dane

Przy procesie ryzyka, wartości  $R(t)$  na osi pionowej interpretujemy jako stan finansowy firmy w momencie  $t$ . Zadanie przed którym stajemy, przypomina zatem pracę analityka finansowego. Musimy skorzystać z dostępnych nam informacji, oraz nabytej wiedzy, aby zamodelować mechanizm stojący za tą konkretną realizacją procesu.

## 2 O Procesie Ryzyka

Zanim zaczniemy mówić o procesie Ryzyka, przypomnijmy sobie kilka definicji, które są z nim nierozłącznie związane. Punktem wyjściowym będzie dla nas następująca definicja:

**Definicja 1. Procesem liczącym** nazywamy proces stochastyczny, którego wartość  $N(t)$  oznacza liczbę losowych zdarzeń, które zaszły do momentu  $t$ .

Zgodnie z tą definicją, procesem liczącym jest na przykład liczba klientów sklepu, lub liczba smsów przychodzących na telefon. W naszym przypadku będziemy natomiast rozważać liczbę szkód zgłaszanych do towarzystwa ubezpieczeniowego. Jak wiemy z rzeczywistości czasy oczekiwania na takie kolejne zgłoszenia powinny być od siebie niezależne, oraz mieć własność braku pamięci- to znaczy, że prawdopodobieństwo nastąpienia zgłoszenia przed upływem czasu  $\tau$  jest takie samo liczone od dziś, od jutra, czy dowolnego innego momentu. Taką własność mają zmienne losowe z rozkładu wykładniczego, co prowadzi nas do definicji Jednorodnego Procesu Poissona.

**Definicja 2. Jednorodnym Procesem Poissona** z intensywnością  $\lambda$  nazywamy taki proces liczący, który spełnia warunki:

1.  $N(0) = 0$
2.  $N(t)$  ma niezależne przyrosty.
3.  $N(t)$  ma stacjonarne przyrosty.
4.  $N(t) \sim \mathcal{P}(\lambda t)$
5.  $N(t)$  ma trajektorie CADLAG

Tak zdefiniowany proces liczący, ma czasy oczekiwania na kolejne skoki wzięte z rozkładu wykładniczego z parametrem  $\lambda$ . Ważną dla nas własnością tego procesu którą wykorzystamy za chwilę jest fakt, że  $E[N(t)] = \lambda t$ .

**Definicja 3. Procesem Ryzyka** nazywamy proces stochastyczny, dany wzorem

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

gdzie:

1.  $u$  - początkowy kapitał
2.  $c(t)$ - funkcja premii (przychody)
3.  $X_i$  - wielkość  $i$ -tej wypłaty (zmienne losowe iid), niezależne od  $N(t)$
4.  $N(t)$  - Proces Odnowy, czyli proces liczący o czasach oczekiwania iid

Jak widzimy proces ryzyka składa się z 3 części. Pierwszą jest początkowy kapitał  $u$ . W modelach procesów ryzyka pełni on czysto rolę parametru. Często ważnym zadaniem dla osoby analizującej tę rodzinę procesów jest ustalenie jak duża musi być początkowa wartość kapitału, gdy chce się otrzymać prawdopodobieństwo bankructwa w pewnym czasie na jakimś zadanym, dostatecznie niskim poziomie.

Za wzrosty procesu odpowiada druga składowa, tak zwana funkcja premii, czyli składnik  $c(t)$ . Interpretacją tego składnika są wpływy związane z kupnem przez klientów przeróżnych polis, akcji spółki, czy innych usług. Z punktu widzenia spółki, chcielibyśmy aby  $c(t)$  radziła sobie łatwo z nieuniknionymi wypłatami środków, więc najlepiej, gdyby była funkcją szybko rosnącą. Kończąc opis procesu ryzyka należy powiedzieć o ostatnim składniku- sumie  $\sum_{i=1}^{N(t)} X_i$ , której efektem są losowe spadki wartości procesu. Zwróćmy uwagę na to, że wielkości spadków  $X_i$  są zmiennymi losowymi iid, zadanymi jakimś rozkładem, oraz następują w losowych, niezależnych od siebie momentach czasu, za co odpowiada pewien proces odnowy  $N(t)$ . Ten składnik będziemy interpretowali jako wypłaty odszkodowań o losowej wartości, następujące w losowych momentach.

Otrzymujemy w ten sposób całkiem logiczny i zrozumiały model działania spółki ubezpieczeniowej, która otrzymuje zyski w przewidywalny sposób, a traci w sposób na pierwszy rzut oka nieprzewidywalny.

Użyłem sformułowania 'na pierwszy rzut oka', ponieważ nie jest to do końca prawdą. Z jednej strony nie możemy przewidzieć jaki będzie wynik jednego rzutu kością, ale z drugiej możemy powiedzieć, że jeśli weźmiemy takich rzutów wystarczająco dużo, to ich średnia będzie zbliżać się do liczby 3.5. Tak samo o jednej, konkretnej trajektorii procesu ryzyka nie można nic powiedzieć z pełnym przekonaniem, jednak korzystając z odpowiednich praw i twierdzeń teorii prawdopodobieństwa, można bardzo dobrze opisać trendy i oczekiwane zachowania dużej grupy trajektorii.

Spróbujmy użyć tej teorii, aby przewidzieć jak będzie zachowywać się rozchód, tzn policzmy wartość oczekiwaną sumy wypłat. Będziemy do tego potrzebowali kilku założeń. Przede wszyst-

kim założymy, że kwoty wypłat pochodzą z pewnego rozkładu, o wartości oczekiwanej  $EX = \mu$ . Założymy też, że proces odnowy w definicji procesu ryzyka jest jednorodnym Procesem Poissona z intensywnością  $\lambda$ .

Wtedy otrzymamy:

$$E\left[\sum_{i=1}^{N(t)} X_i\right] \stackrel{N(t) \perp X_i}{=} E[N(t)] \cdot EX_1 = \lambda\mu t \quad (1)$$

Jeśli to jest nasz rozchód i zależy nam na utrzymaniu firmy w dobrej kondycji, to jasne jest, że przychód musi być co najmniej taki sam, a najlepiej większy, bo inaczej z prawa wielkich liczb, bankructwo będzie następować prawie na pewno. Można temu zapobiec, wstawiając jako funkcję premii coś delikatnie większego. Taka konstrukcja procesu ryzyka nazywana jest **Klasycznym Procesem Ryzyka**.

**Definicja 4. Klasycznym Procesem Ryzyka** nazywamy proces stochastyczny, dany wzorem

$$R(t) = u + (1 + \theta)\mu\lambda t - \sum_{i=1}^{N(t)} X_i,$$

gdzie:

1.  $u$  - początkowy kapitał
2.  $X_i$  - wielkość  $i$ -tej wypłaty (zmienne losowe iid)
3.  $\mu > 0$  - wartość oczekiwana wypłat
4.  $N(t)$  - jednorodny Proces Poissona z intensywnością  $\lambda$
5.  $\lambda > 0$  - intensywność  $N(t)$
6.  $\theta > 0$  - parametr ekspercki

Powyższa definicja jest kluczowa w naszych rozważaniach, bo otrzymane dane będziemy próbowali modelować właśnie przy jej pomocy. Zakłada ona liniowy przychód środków, przy czym za rzeczywisty dochód "na czysto" odpowiada parametr ekspercki  $\theta$ . Ma on niejako dwie funkcje- musi być na tyle duży, aby ubezpieczyć firmę na wypadek pechowych realizacji wypłat, ale nie może być za duży, aby nie odstraszyć klientów zbyt wysoką ceną usług.

Dobrze widać jego rolę wyliczając wartość oczekiwaną klasycznego procesu Ryzyka. Korzystając z równości (1) mamy:

$$E[R(t)] = E\left[u + (1 + \theta)\mu\lambda t - \sum_{i=1}^{N(t)} X_i\right] \stackrel{(1)}{=} u + \theta\lambda\mu \cdot t \quad (2)$$

Można również pokazać, że wariancja procesu ryzyka dana jest wzorem:

$$Var[R(t)] = Var\left[u + (1 + \theta)\mu\lambda t - \sum_{i=1}^{N(t)} X_i\right] = Var\left[\sum_{i=1}^{N(t)} X_i\right] = \lambda E[X^2]t \quad (3)$$

Te dwie własności Klasycznego Procesu Ryzyka przydadzą nam się później.

W tej pracy zajmę się najpierw próbą wyestymowania wszystkich parametrów niezbędnych do opisu klasycznego procesu ryzyka, następnie na ich podstawie zbuduję model, aby na końcu, na jego podstawie obliczyć prawdopodobieństwo ruiny (bankructwa) w skończonym i nieskończonym czasie.

## 3 Estymacja parametrów

### 3.1 Początkowy kapitał

Zdecydowanie najprostszym do podania parametrem jest początkowy kapitał. Aby go znaleźć wystarczy sprawdzić jaka jest wartość procesu w momencie  $t = 0$ . Szybkie sprawdzenie danych zwraca nam wartość  $u = 100$ . Jest to parametr który można wyznaczyć najdokładniej ze wszystkich, bez żadnych wątpliwości co do jego realnej wartości.

### 3.2 Funkcja przychodów

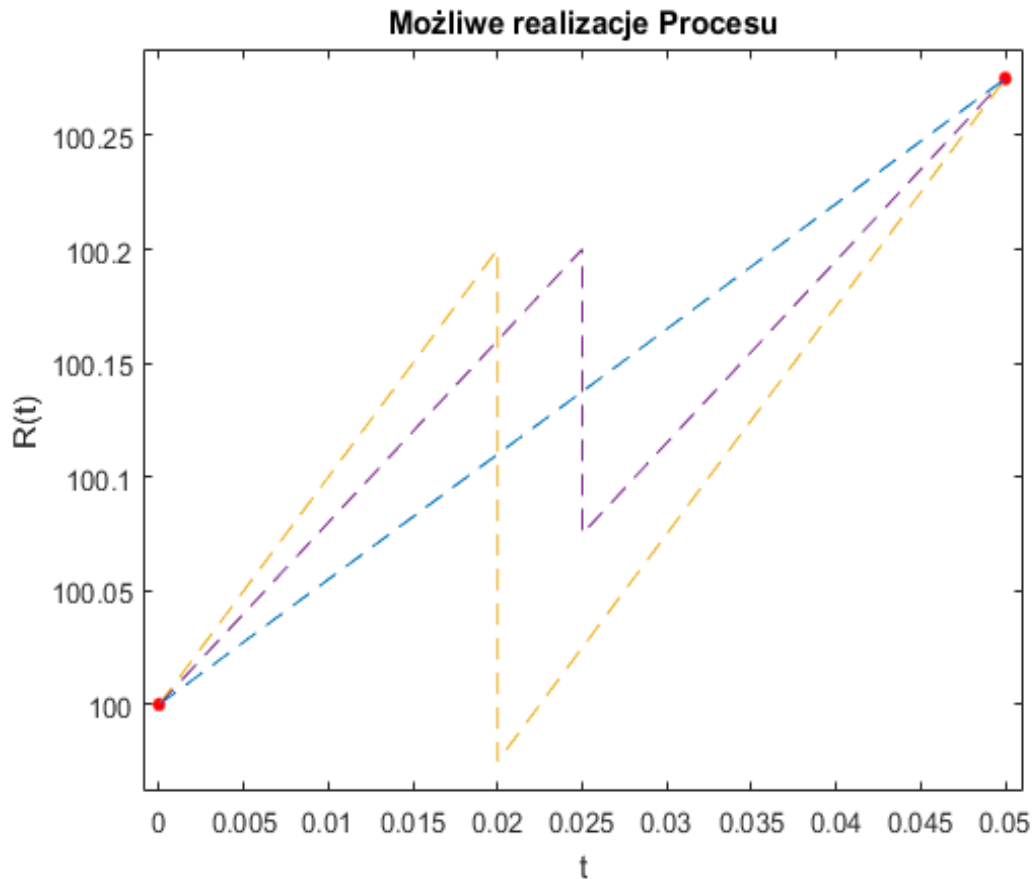
Przypomnijmy postać funkcji przychodów:

$$c(t) = (1 + \theta)\mu\lambda t = C \cdot t$$

Aby wyznaczyć dokładną postać tej funkcji musielibyśmy znać parametry  $\mu, \theta$ , oraz  $\lambda$ , a te będziemy znać dopiero gdy znajdziemy rozkład czasów oczekiwania na wypłaty, oraz rozkład wielkości tych wypłat. Tych jednak nie będziemy znać jeszcze przez najbliższą chwilę. Zajmiemy się więc najpierw estymacją całej wartości  $(1 + \theta)\mu\lambda$ , którą oznaczmy sobie jako  $C$ .

Naturalnym podejściem do estymowania parametru  $C$  byłoby sprawdzenie tangensa kąta nachylenia prostych, liniowo rosnących odcinków procesu Ryzyka. Przy takiej próbie estymacji tego parametru zaczynają się pojawiać jednak problemy, ponieważ nasze dane wychwytyują wartości procesu w dyskretnych momentach czasu. Oznacza to, że nie posiadamy informacji o tym, co dzieje się z procesem pomiędzy tymi momentami. Przypomnijmy, że proces Ryzyka

ma skoki występujące w losowych momentach, oraz o losowych wartościach. Metoda tangensa byłaby dobra, gdybyśmy mieli pewność, że na danym odcinku nie było żadnego skoku, ale tego pewni być nie możemy. Spójrzmy przykładowo na dwie pierwsze wartości procesu (Rysunek 2). Pokazują one trzy różne trajektorie, każda z nich wygląda inaczej, każda ma inny parametr  $C$ , ale przy naszym sposobie próbkowania, każda zwróciłaby te same wartości.



Rysunek 2: Różne możliwe trajektorie procesu

Musimy zatem estymować ten parametr w inny sposób. Zauważmy, że jeżeli pomiędzy dwoma najbliższymi wartościami procesu byłyby skoki, to przyrost wartości procesu automatycznie będzie mniejszy. Moją propozycją estymacji parametru  $C$  jest w związku z tym zebranie listy wszystkich przyrostów procesu pomiędzy najbliższymi punktami, i jako estymator parametru  $C$  przyjąć

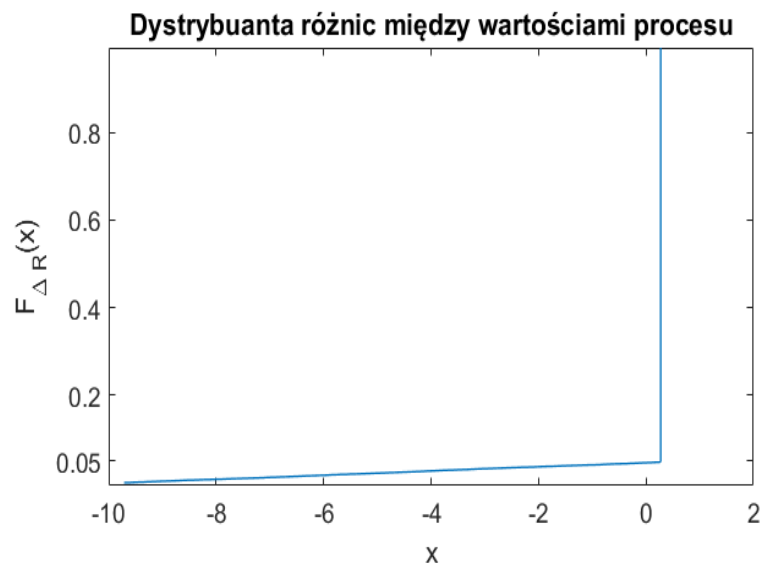
$$\hat{C} = \frac{\max\{\text{Przyrosty}\}}{\Delta t}$$

Otrzymana w ten sposób wartość wynosi  $\hat{C} \approx \frac{0.275}{0.05} = 5.5$ .

Dzięki temu wynikowi mamy już postać funkcji przychodów. Naszym celem jednak jest jej inna postać, taka, w której widać parametry  $\mu$ ,  $\lambda$  i  $\theta$ . Aby ją wyznaczyć, będziemy jednak potrzebowali najpierw znać rozkład wielkości kwot wypłat, oraz rozkład czasów oczekiwania na kolejne wypłaty- te rozkłady dadzą nam odpowiednio  $\mu$  i  $\lambda$ , a  $\theta$  obliczymy korzystając z nich, oraz  $\hat{C}$ .

### 3.3 Rozkład kwot wypłat

Zdecydowanie największą trudnością postawionego problemu estymacji parametrów jest estymacja rozkładu kwot wypłat, oraz rozkładu czasów oczekiwania na wypłaty. Wyznaczając te rozkłady po raz kolejny rozbijamy się o ścianę jaką jest problem spróbkowanych wartości. Gdybyśmy mieli dokładne momenty tych wypłat, pozostałoby nam dobre dopasowanie odpowiednich funkcji gęstości, lub dystrybuant. Wartości dokładnych jednak nie mamy, co więcej nie jesteśmy nawet w stanie powiedzieć ile spadków było pomiędzy kolejnymi wartościami procesu. Jednak obserwując z bliska wygląd danej nam trajektorii, możemy zauważyć, że gdyby podzielić ją na przedziały odpowiadające kolejnym pomiarom wartości procesu, to na wielu z nich nie widać śladu spadków. To znaczy wizualnie, trajektoria na wielu przedziałach rośnie liniowo tak, jakbyśmy tego od niej oczekiwali gdyby spadków nie było wogóle. To skłoniło mnie do sprawdzenia rozkładu różnic pomiędzy kolejnymi wartościami procesu. Dystrybuanta tych różnic prezentuje się w następujący sposób (Rysunek 3). Widzimy tutaj, że praktycznie 95% tych



Rysunek 3: Dystrybuanta różnic pomiędzy wartościami procesu



wartości jest bardzo blisko siebie. Po sprawdzeniu kwantyla rzędu 0.05 otrzymałem ciekawy wynik. Okazuje się, że  $q_{0.05} \approx 0.275$ . Dokładna wartość tego kwantyla, różni się od maksymalnego przyrostu między wartościami o liczbę rzędu  $10^{-12}$ . Co więcej, gdybyśmy dla każdego przedziału należącego do tych 95% policzyli różnicę pomiędzy wartością przyrostu na nim, a maksymalnym występującym przyrostem, to ich suma zamknęłaby się w wielkości rzędu  $10^{-8}$ . Jest to nieporównywalnie mało do wartości jakie przyjmuje proces, a nawet do przyrostów między przedziałami. W tej sytuacji zdecydowałem się na uznanie, że na przedziałach, na których różnica wartości procesu była większa od rzeczzonego kwantyla żadne spadki nie miały miejsca. Innymi słowy rozważając czy na przedziale był jakiś spadek, odrzucać będę jednostki, na których odchylenie od spodziewanej wartości jest prawie niezauważalne. Metody których używam w tej pracy są tak czy inaczej obarczone niepewnościami, bo poruszamy się w dziedzinie losowych procesów- więc nie powinno to być wielkim grzechem.

Ustaliłem w ten sposób przedziały, na których według mnie zaistniał spadek procesu. Niestety, nadal nie jestem w stanie stwierdzić, czy na danym przedziale nastąpił jeden duży skok, czy może było to kilka mniejszych. Postanowiłem na ten moment założyć, że był to tylko jeden skok i zobaczyć czy doprowadzi mnie to do jakiegoś sensownego modelu.

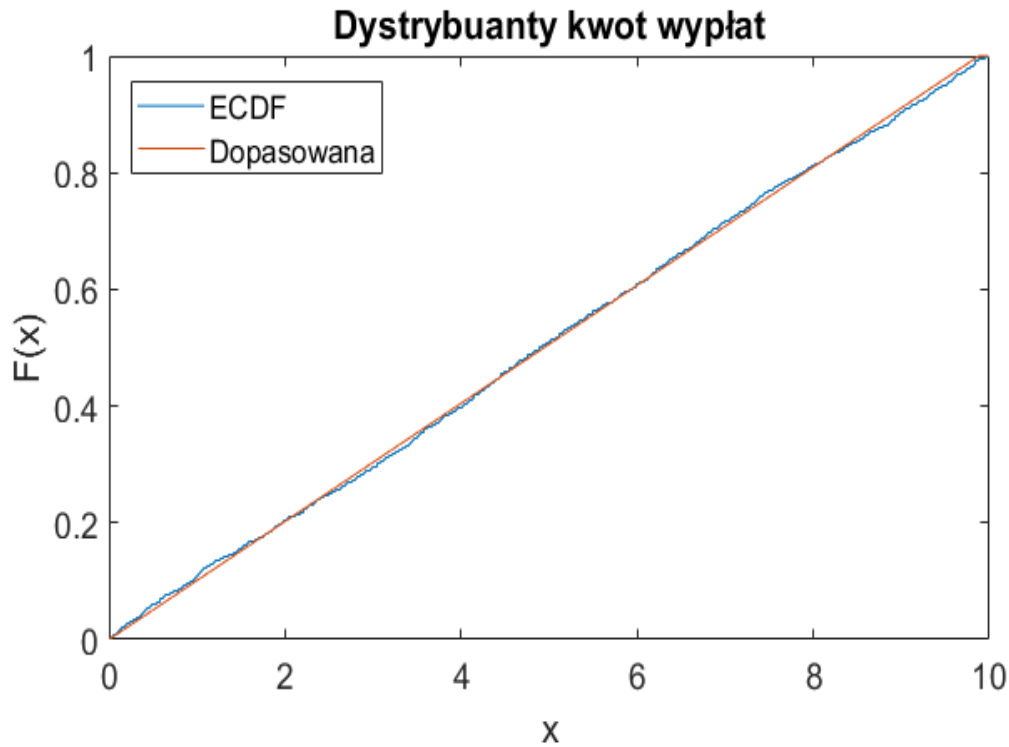
Stawiając sobie takie założenie, wartościami wypłat będą po prostu różnice pomiędzy wartością spodziewaną (bez spadków), a rzeczywistą procesu. Jeszcze inaczej: są to nieodrzucone różnice wartości procesu, lecz ze zmienionym znakiem i przesunięte o wartość spodziewaną, żeby kwota wypłat przyjmowała wartości dodatnie. Takich różnic mamy 1888, a ich dystrybuanta wygląda w następujący sposób (Rysunek 4). Dystrybuanta empiryczna ECDF wygląda na dystrybuantę rozkładu jednostajnego. Kierując się tym, wyliczyłem wartość średnią strat, oraz ich wariancję w próbie:

1.  $\bar{X} \approx 4.94$

2.  $\sigma_X^2 \approx 8.26$

Korzystając z tych wartości, dopasowałem model rozkładu jednostajnego z którego mogły pochodzić dane i naniosłem jego dystrybuantę na wykres. Parametry rozkładu  $U(a, b)$  otrzymałem wstawiając  $a = 0$ , ponieważ kwoty wypłat muszą być dodatnie, oraz rozwiązując równanie na wartość oczekiwaną.

$$E[X] = \frac{a + b}{2} \approx \bar{X}$$



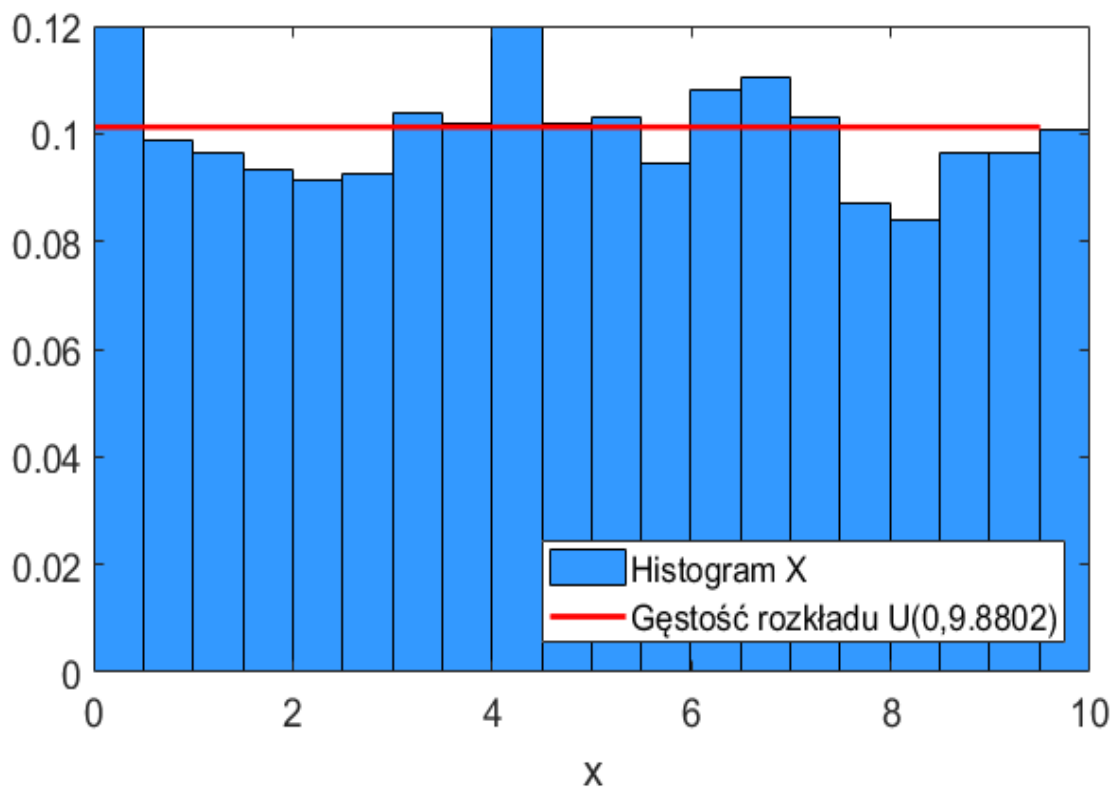
Rysunek 4: Dystrybuanta kwot wypłat

Otrzymałem w ten sposób  $b \approx 9.882$ .

Jak można zobaczyć, dystrybuanta o takich parametrach dobrze pasuje do dystrybuanty empirycznej. Także wbudowany w Matlaba test KS porównujący dystrybuanty nie zgłosił zastrzeżeń co do możliwości pochodzenia próbki z takiego rozkładu. P-wartość testu wyniosła aż 0.86, co oznacza bardzo dobre dopasowanie. Wariancja zmiennej o rozważanym rozkładzie wynosi  $\text{Var}[X] \approx 8.13$ , więc jest blisko jej empirycznego odpowiednika  $\sigma_X^2$ . Nie będę porównywał gęstości tej zmiennej z empirycznym odpowiednikiem, ponieważ dla rozkładów jednostajnych nie jest ona dobrym odnośnikiem. Zamiast tego, narysowałem histogram kwot wypłat, i naniosłem na niego gęstość rozważanego rozkładu jednostajnego (Rysunek 5). Na rysunku widzimy delikatne wahania, ale biorąc pod uwagę rząd wielkości na osi OY to dopasowanie jest. Potwierdzając już jedynie dla formalności wykresem kwantylowym (Rysunek 6), mogę powiedzieć, że zmienna losowa z rozkładu jednostajnego  $\mathcal{U}(0, 9.882)$  jest dobrym modelem kwoty wypłaty.

### 3.4 Rozkład czasów oczekiwania na wypłaty

Aby otrzymać rozkład czasów oczekiwania na wypłaty, trzymałem się założeń postawionych w poprzednim podrozdziale, a co za tym idzie, rozważałem istnienie skoków na takich samych



Rysunek 5: Histogram wartości wypłat oraz gęstość rozkładu jednostajnego

przedziałach.

Obrabiając dane w programie Matlab, wychwyciłem początki przedziałów, na których nachylenie odcinka jest odbiegające od normy. Nie są to jednak momenty spadków procesu, ponieważ sam spadek ma miejsce gdzieś wewnątrz przedziału. Niestety nie wiadomo ile czasu upływa między początkiem takiego przedziału, a faktycznym spadkiem. Wiemy jedynie że jest to mniej niż 0.05, ponieważ jest to długość przedziału, oraz, utrzymując założenia które sobie postawiliśmy, wiemy, że nastąpił tam tylko jeden spadek. Co więcej, wiemy też, że momenty spadków dane są jednorodnym procesem Poissona. Skorzystam więc z jednego z twierdzeń, którego używa się do konstrukcji procesu Poissona.

**Twierdzenie 1.** Niech  $N(t)$  będzie jednorodnym procesem Poissona z intensywnością  $\lambda > 0$ . Wtedy warunkowy jego wektora momentów skoków  $[S_1, S_2, \dots, S_n]$  na odcinku  $[0, T]$  pod warunkiem  $N(T) = n$ , jest równy rozkładowi wektora statystyk pozycyjnych  $[U_{1:n}, U_{2:n}, \dots, U_{n:n}]$ , gdzie  $U_1, \dots, U_n$  są iid,  $U \sim \mathcal{U}(0, T)$ , gdzie  $U_{k:n}$  to  $k$ -ta najmniejsza wartość ze zbioru  $U_1, \dots, U_n$ .

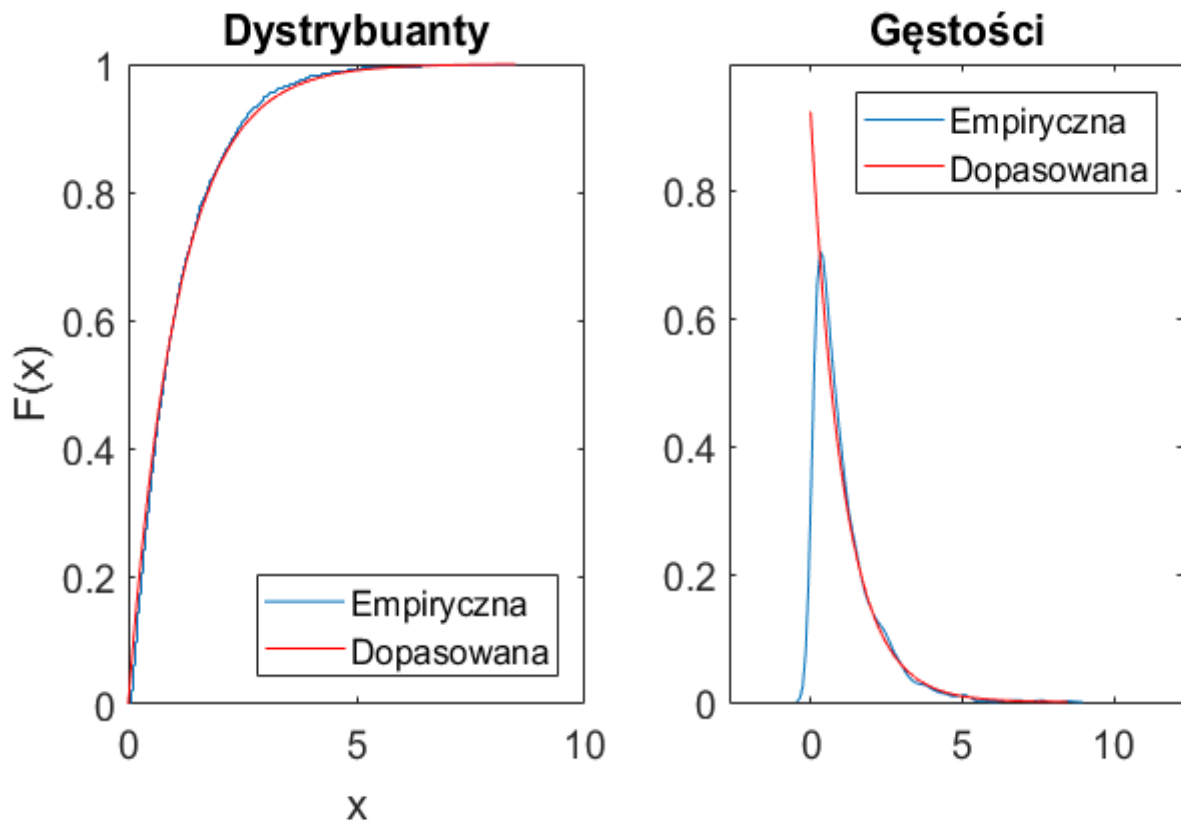
Mówiąc prostym językiem, w naszym przypadku, jeżeli początek przedziału ze spadkiem



Rysunek 6: Wykres kwantylowy kwot wypłat

niał miejsce w momencie  $t_p$ , oraz zakładamy, że spadek był tylko jeden, to moment tego spadku ma rozkład jednostajny na odcinku  $[t_p; t_p + 0.05]$ . Pozwala nam to otrzymać z wektora 'początków przedziałów na których jest spadek', wektor 'momentów spadków' poprzez dodanie do każdej wartości realizację zmiennej losowej z rozkładu  $\mathcal{U}(0; 0.05)$ .

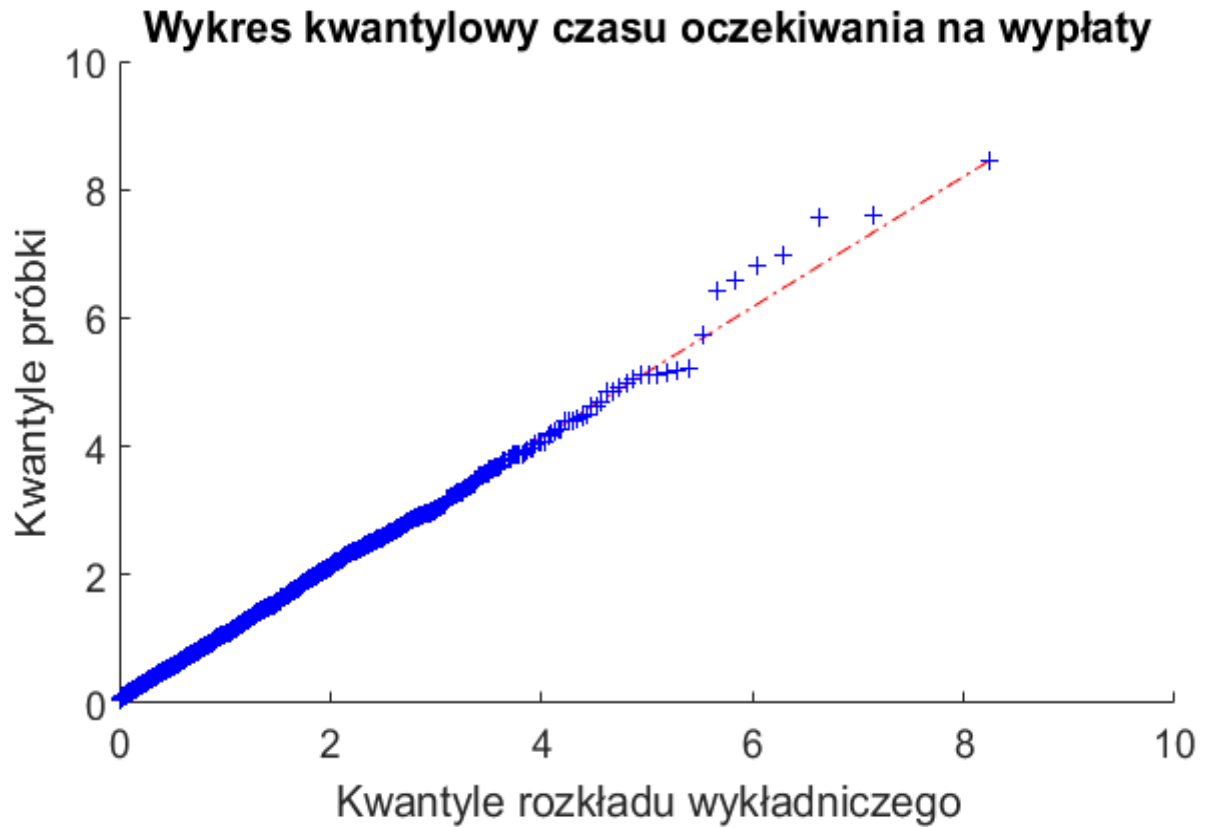
Czasy oczekiwania na kolejne skoki są dane jednorodnym Procesem Poissona, więc powinny mieć rozkład wykładniczy. Po obejrzeniu empirycznych: dystrybuanty oraz gęstości utworzonych czasów oczekiwania na skoki (Rysunek 7) rzeczywiście wszystko na to wskazuje. Aby dopasować odpowiedni parametr rozkładu wyliczyłem wartość średnią czasów oczekiwania na wypłaty. Wyniosła ona  $\bar{T} \approx 1.08$ , a ponieważ dla  $T \sim \text{Exp}(\lambda)$  mamy  $E[T] = 1/\lambda$ , to do danych dopasowywałem rozkład  $\text{Exp}(\frac{1}{1.08})$ . Dystrybuanty nie dopasowały się tak dobrze jak przy kwotach wypłat, ale biorąc pod uwagę długość próbki i inny charakter rozkładu, jest to możliwe. Wariancja w próbie czasów oczekiwania na kolejne wypłaty wyniosła  $\sigma_T^2 \approx 1.04$ , podczas gdy wariancja zmiennej losowej o rozkładzie wykładniczym do którego dopasowujemy wynosi  $\text{Var}[T] \approx 1.18$ . Wykres kwantylowy prezentuje się w następujący sposób (Rysunek 8). Widzimy na nim, że na pierwszej połowie odcinka, gdzie mamy sporo obserwacji kwantyle dobrze się pokrywają. Na drugiej połowie natomiast, obserwacje są coraz rzadsze, co skutkuje rzadziej



Rysunek 7: Porównanie dystrybuant i gęstości czasów oczekiwania na wypłaty

położonymi punktami. Odstępstwa od prostej w drugiej części odcinka są nie do uniknięcia przy rozkładzie takim jak wykładniczy. Przy dopasowywaniu tego rozkładu martwi jedynie test KS, który nie chce zgodzić się z hipotezą, że próbka pochodzi z zadanego rozkładu. Wynika to z braku obserwacji bliskich zeru, co powoduje zbyt dużą różnicę między dystrybuantami, aby KS test mógł się zgodzić z hipotezą zerową. Wykonałem z ciekawości KS test jeszcze raz, tym razem testując dane do przesuniętego o wartość  $\min\{T_i\}$  rozkładu wykładniczego. Test zgodził się, choć p-wartość była bardzo bliska poziomowi istotności testu.

Biorąc pod uwagę lekkie różnice dystrybuant i gęstości, oraz wyżej wymieniony fakt o Procesie Poissona, przyjmę ten rozkład i będę kontynuował korzystając z rozkładu  $\text{Exp}(0.9225)$  jako modelu z którego pochodzą czasy oczekiwania na kolejne spadki procesu.



Rysunek 8: Wykres kwantylowy czasów oczekiwania na wypłaty

## 4 Symulacja

Podsumujmy to, co otrzymaliśmy do tej pory. Nasz proces Ryzyka jest postaci:

$$R(t) = u + (1 + \theta)\mu\lambda \cdot t - \sum_{i=1}^{N(t)} X_i,$$

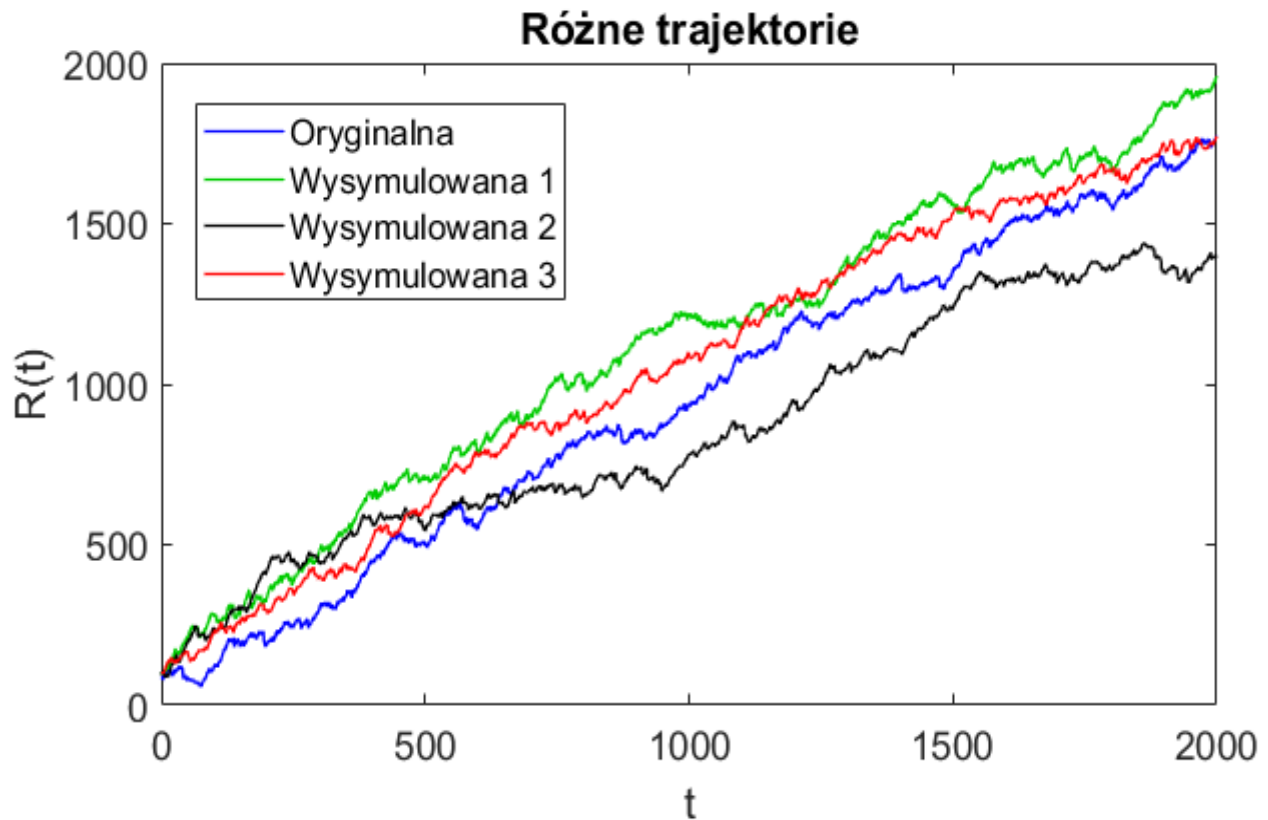
Gdzie

1.  $u = 100$
2.  $(1 + \theta)\mu\lambda = C = 5.5$
3.  $X_i \sim \mathcal{U}(0, 9.882)$ , *iid*
4.  $N(t)$ — Jednorodny Proces Poissona z intensywnością 0.9225,  $\perp X_i$

Mamy już zatem wszystkie parametry potrzebne do symulacji trajektorii według naszego modelu. Symulować proces będziemy w następujący sposób. Skorzystamy z przytoczonego wcześniej

Twierdzenia 1, aby wysymulować momenty spadków z jednorodnego procesu Poissona z odpowiednią intensywnością. Wysymulujemy następnie niezależnie wektor kwot wypłat, o tej samej długości. Następnie od prostej  $y(t) = u + C \cdot t$  będziemy odejmować kolejne wartości wektora wypłat, w momentach danych wektorem momentów spadków procesu. Ostatni szlif to próbkowanie danych, czyli otrzymanie wartości procesu tylko na zadanej siatce punktów (tę przyjąłem taką jak dla oryginalnej trajektorii).

Wysymulowałem w ten sposób kilka trajektorii, które można obejrzeć na Rysunku 9. Widzimy,



Rysunek 9: Przykładowe wysymulowane trajektorie, oraz trajektoria oryginalna

że trajektorie osiągają porównywalne wartości na horyzoncie czasowym. Zauważyłem jednak, że więcej trajektorii osiąga wyższą wartość od trajektorii oryginalnej. Metodą Monte Carlo obliczyłem, że jest to około 80% wszystkich obserwacji.

Postanowiłem sprawdzić jak rozkładają się wartości trajektorii na horyzoncie czasowym. Wysymulowałem okrągły milion realizacji procesu i spisałem ich wartości w  $t = 2000$ . Tak utworzona próba ma rozkład podobny do normalnego, średnią  $\overline{R(2000)} \approx 1980$ , oraz wariancję w próbie  $\sigma_{R(2000)}^2 \approx 60002$ , co dobrze odpowiada wartości oczekiwanej i wariancji procesu ryzyka przy

używanych przy mnie parametrach:

$$E[R(2000)] = u + \theta\mu\lambda \cdot 2000 \approx 1985$$

$$Var[R(2000)] = \lambda E[X^2] \cdot 2000 \approx 60038$$

Wartość naszej zadanej, pierwotnej trajektorii jest kwantylem rzędu 0.199 tego rozkładu, co potwierdza wcześniejsze estymacje- około 20% trajektorii generowanych przez ten model łąduje po czasie 2000 pod nią, a 80% nad. Zadana trajektoria osiąga więc tam dość typową jak na zaproponowany model wartość.

## 5 Prawdopodobieństwa Ruiny

Istotą analiz procesów Ryzyka często jest obliczenie tak zwanego prawdopodobieństwa ruiny. Jest to prawdopodobieństwo, że proces Ryzyka spadnie poniżej 0, co interpretujemy jako bankructwo firmy. Naturalnie rozważamy dwa rodzaje takiego prawdopodobieństwa. Może nas interesować prawdopodobieństwo bankructwa w pewnym skończonym czasie, lub w czasie nieskończonym. W kolejnych podrozdziałach rozważę te prawdopodobieństwa dla zaproponowanego wcześniej modelu.

### 5.1 W skończonym czasie

**Definicja 5.** Prawdopodobieństwo Ruiny w skończonym czasie  $T$  definiujemy jako

$$\Psi(u, T) \stackrel{df}{=} P(\tau(u) \leq T),$$

gdzie  $\tau = \inf\{t > 0 : R(t) < 0\}$  jest momentem ruiny, czyli pierwszym momentem w którym proces spada poniżej zera.

Takie prawdopodobieństwo można aproksymować w bardzo prosty sposób metodą Monte Carlo. Należy po prostu wysymulować dużo realizacji procesu na przedziale  $[0; T]$  i sprawdzić jaki ułamek zbankrutował, czyli jaki procent z nich spełnia nierówność:  $\min_{0 \leq t \leq T} \{R(t)\} < 0$ . Przeprowadziłem taką symulację dla 10000 trajektorii, oraz różnych czasów  $T$ . Wyniki można zobaczyć w poniższej tabeli.

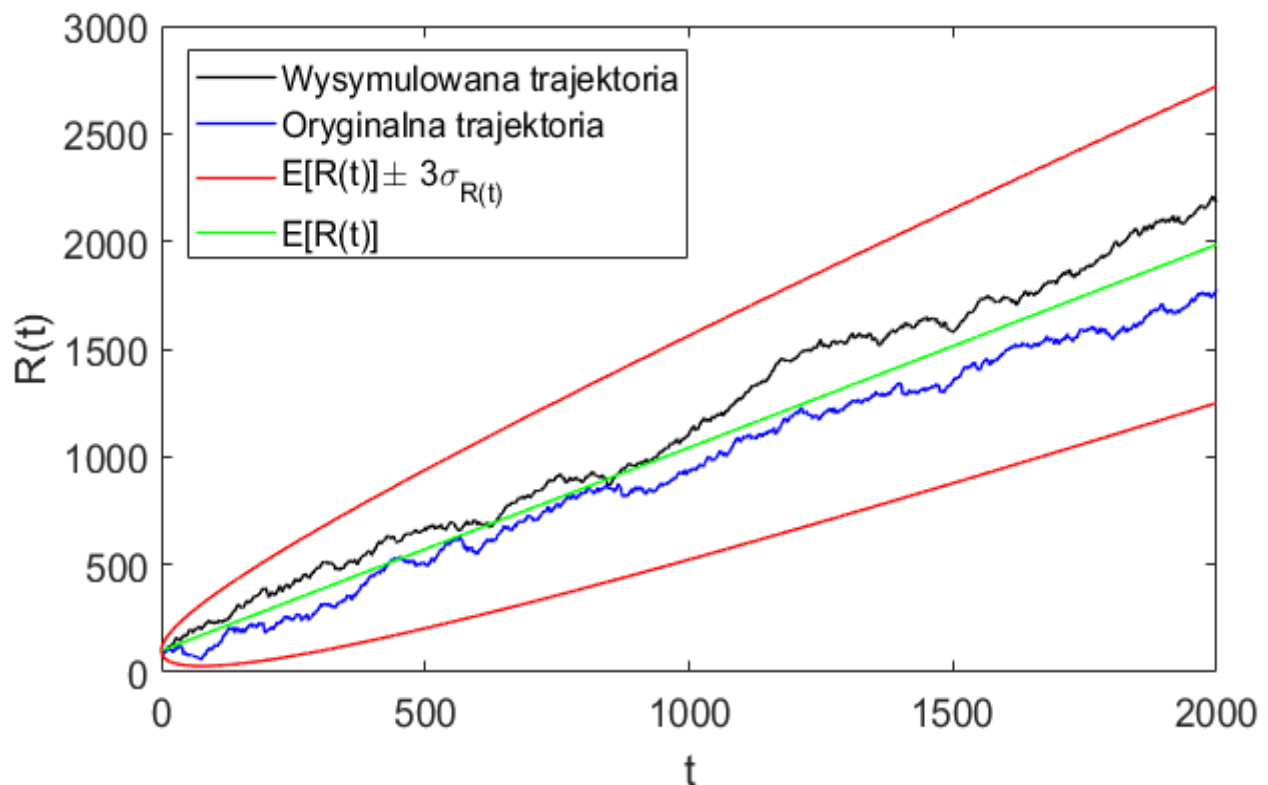


T	10	20	30	40	50	60	70	80	90	100
$\Psi(u, T)$	0	0	0.0002	0.0002	0.0009	0.0016	0.0024	0.0026	0.0023	0.0022
T	110	120	130	140	150	160	170	180	190	200
$\Psi(u, T)$	0.0028	0.0027	0.0029	0.0047	0.0027	0.0037	0.0031	0.0033	0.0042	0.0038
T	210	220	230	240	250	260	270	280	290	300
$\Psi(u, T)$	0.0027	0.0043	0.0029	0.0043	0.0042	0.0048	0.0047	0.0025	0.0033	0.0032
T	310	320	330	340	350	360	370	380	390	400
$\Psi(u, T)$	0.0047	0.0040	0.0031	0.0046	0.0041	0.0036	0.0036	0.0036	0.0035	0.0041
T	420	450	500	600	700	800	900	1000	1500	2000
$\Psi(u, T)$	0.0035	0.0035	0.0040	0.0035	0.0033	0.0048	0.0032	0.0046	0.0032	0.0037

Tabela przedstawia prawdopodobieństwa ruiny w różnych czasach  $T$ , przy ustalonym kapitale początkowym  $u = 100$ . Staralem się przekrojowo ująć horyzonty czasowe, aby pokazać pewną ciekawą obserwację. Można zauważyć, że prawdopodobieństwo rośnie w miarę zwiększania horyzontu czasowego, lecz tylko do pewnego momentu. Od około  $T = 160$ , gdy  $\Psi(u, 160) = 0.0037$ , wartość prawdopodobieństwa nie zwiększa się znacząco. Pojawiają się oczywiście wartości większe od 0.004, ale nie można oprzeć się wrażeniu, że i tak te obliczone prawdopodobieństwa oscylują wokół jakiejś rzeczywistej, nie zmieniającej się mimo zwiększania horyzontu czasowego wartości. Innymi słowy, wykonując te symulacje, odniosłem wrażenie, że jeśli miniemy już jakąś "magiczną" wartość w czasie i nie zbankrutujemy do tej pory, to prawdopodobieństwo bankructwa później jest już znikome. Pomyślałem, że być może można to powiązać z wartością oczekiwaną i wariancją procesu. W poprzednim rozdziale sprawdzałem rozkład wartości procesu w momencie  $t = 2000$ . Rozkład ten bardzo przypominał rozkład normalny o średniej  $\mu_R = E[R(t)]$  i wariancji  $\sigma_R^2 = Var[R(t)]$ . Wartości tego procesu powinny przestrzegać nierówności Czebyszewa, co oznacza, że

$$P\left(|R(t) - E[R(t)]| \geq 3\sigma_{R(t)}\right) \leq \frac{1}{9}$$

Około 88% wartości powinno więc trafić w przedział  $[\mu_R - 3\sigma_R; \mu_R + 3\sigma_R]$ . W procesie Ryzyka, zarówno wartość oczekiwana jak i wariancja są jednak funkcjami czasu. Dlatego postanowiłem iść o krok dalej i narysować ewolucję tego przedziału ufności dla procesu w czasie (Rysunek 10)



Rysunek 10: Nierówność Czebyszewa w czasie dla zaproponowanego modelu procesu Ryzyka

Obszar pomiędzy czerwonymi liniami, jest obszarem w którym z bardzo dużym prawdopodobieństwem może znaleźć się realizacja procesu Ryzyka pochodząca z mojego modelu. Widzimy tu przede wszystkim, że zadana trajektoria znajduje się pomiędzy nimi, co jest bardzo dobrym sygnałem. Być może model byłby lepszy, gdyby wartość oczekiwana procesu pokrywała się z wartością zadanej trajektorii w  $t = 2000$ , lecz z drugiej strony widzimy że zaproponowany model jest też poprawny- zadana trajektoria mogła powstać z niego. Widzimy też potwierdzenie tego, co było widać w prawdopodobieństwach ruiny. Rzeczywiście, jeśli proces ma spaść poniżej zera, to tylko na początku ewolucji. Im dalej się odsuwamy od początku osi, tym mniej prawdopodobne są małe wartości procesu. Ta obserwacja daje podstawę do myślenia, że prawdopodobieństwo ruiny w czasie nieskończonym, które będziemy liczyć za chwilę nie będzie się wiele różniło od prawdopodobieństwa ruiny w skończonym czasie  $\Psi(u, 2000)$ , ponieważ bankructwo dalej niż  $t = 2000$  przy takich parametrach modelu jest już praktycznie niespotykane.

## 5.2 W nieskończonym czasie

Prawdopodobieństwo ruiny w nieskończonym czasie jest intuicyjnie jasne, po omówieniu prawdopodobieństwa ruiny w czasie skończonym. Musimy jedynie przejść z horyzontem czasowym do nieskończoności.

**Definicja 6.** Prawdopodobieństwem ruiny w nieskończonym czasie nazywamy

$$\Psi(u) \stackrel{df}{=} P(\tau(u) < \infty),$$

gdzie  $\tau = \inf\{t > 0 : R(t) < 0\}$  jest momentem ruiny, czyli pierwszym momentem w którym proces spada poniżej zera.

Zupełnie inny natomiast musi być sposób wyliczania tego prawdopodobieństwa. Na skończonym horyzoncie czasowym mogliśmy symulować trajektorie. Potrzeba mniejszej, lub większej mocy obliczeniowej, ale w końcu dostaniemy żadaną ilość trajektorii, na dowolnie wybranym horyzoncie czasowym. Niestety w tym przypadku taka sztuka się już nie uda, ponieważ trajektorie musielibyśmy symulować w nieskończoność.

Pomysłem na pokonanie tej trudności jest skorzystanie ze wzoru Pollacza-Chińczyna:

**Twierdzenie 2.**

$$\Psi(u) = \frac{\theta}{1+\theta} \sum_{n=0}^{\infty} \left( \frac{1}{1+\theta} \right)^n B_n(u),$$

gdzie  $B_n(u) = P(Y_1 + \dots + Y_n > u)$ , a  $Y_i$  to zmienne losowe iid o gęstości  $f_Y(x) = \frac{P(X_i > x)}{E[X_i]}$

Co ciekawe, można pokazać, że zachodzi równość:

$$\Psi(u) = P(Y_1 + \dots + Y_K > u), \text{ gdzie } K \sim \text{Geo}\left(\frac{\theta}{1+\theta}\right) \quad (4)$$

Powyższa równość pozwala już skorzystać z metod Monte Carlo bez potrzeby symulowania trajektorii w nieskończoność. Jedyny problem jaki może nas tu spotkać to symulacja zmiennych losowych  $Y_i$  z żadanego rozkładu. Zobaczmy jak będzie to wyglądać w naszym przypadku.

Zmienne losowe  $X_i$  mają rozkład jednostajny na przedziale  $[0; 2\mu]$ , zatem  $E[X] = \mu$ , dystrybucja  $F_X(x)$  dana jest wzorem

$$P(X \leq x) = F_X(x) = \begin{cases} 1, & x > 2\mu, \\ \frac{x}{2\mu}, & x \in (0; 2\mu], \\ 0 & x \leq 0. \end{cases}$$

Do wyznaczenia gęstości  $f_Y$  użyjemy jednak tylko przedziału z którego  $X$  przyjmuje wartości, czyli  $(0; 2\mu]$ . Odejmując wartość dystrybuanty od jedynki i dzieląc przez  $E[X]$  otrzymujemy gęstość postaci:

$$f_Y(x) = \frac{2\mu - x}{2\mu^2} \cdot \mathbb{1}_{x \in (0; 2\mu]}$$

Do generowania realizacji zmiennej można użyć metody odwrotnej dystrybuanty, czyli najpierw z gęstości policzyć dystrybuantę:

$$F_Y(x) = \int_{-\infty}^x f_X(t)dt = \begin{cases} 1, & x > 2\mu, \\ \frac{x}{\mu} - \frac{x^2}{4\mu^2}, & x \in (0; 2\mu], \\ 0, & x \leq 0. \end{cases}$$

Stąd już możemy generować zmienne losowe o zadanej gęstości podstawiając  $Y = F^{-1}(U)$ , gdzie  $U \sim \mathcal{U}(0, 1)$ .

Korzystając zatem z równości (4) prawdopodobieństwo ruiny w nieskończonym czasie będziemy aproksymować losując  $K \sim \mathcal{P}(\frac{\theta}{1+\theta})$  zmiennych losowych  $Y$  z rozkładu o gęstości  $f_Y(x)$  i sprawdzając czy ich suma jest większa od początkowego kapitału  $u$ . Zrobimy tak wiele razy, za każdym razem losując nową  $K$  i sprawdzimy jaki ułamek prób zakończył się sukcesem (suma zmiennych była większa od  $u$ ).

Przeprowadziłem taką symulację dla liczby powtórzeń Monte Carlo równej  $10^6$  i początkowego kapitału  $u = 100$ . Wynik potwierdził moje przypuszczenia z poprzedniego podrozdziału, ponieważ okazało się, że  $\Psi(100) = 0.0037$ .

Z ciekawości sprawdziłem jakby takie prawdopodobieństwo wyglądało dla innych kapitałów początkowych. W poniższej tabeli zamieszczam wyniki symulacji:

u	10	20	50	200	300
$\Psi(u)$	0.5051	0.2926	0.0574	0.000019	0

Widzimy, że zgodnie z intuicyjnym oczekiwaniem, im mniejszy kapitał początkowy, tym większa szansa na bankructwo. Muszę jednak szczerze powiedzieć, że nie spodziewałem się aż takiego jego wpływu na wartość tego prawdopodobieństwa. Wystarczy podwojenie kapitału z  $u = 100$  do  $u_2 = 200$  żeby zmniejszyć prawdopodobieństwo bankructwa o kilka rzędów, a jego potrojenie powoduje jego praktyczne zniwelowanie! Jest to dla mnie zaskakujący i pouczający wynik.

Pokusiłem się również o dodatkowe sparametryzowanie tego prawdopodobieństwa i zabawę parametrem eksperckim  $\theta$ . Okazało się, że gdy dotychczasowe prawdopodobieństwo  $\Psi(100, \theta) =$

0.0037, to zmniejszenie dwukrotne parametru eksperckiego powoduje zwiększenie prawdopodobieństwa bankructwa do  $\Psi(100; \theta/2) = 0.0502$ , co według niektórych mogłoby być już sporym ryzykiem. Natomiast symulacja  $\Psi(100, \theta/4)$  zwraca pokąźną wartość 0.21, co byłoby absolutnie niedopuszczalne dla chyba każdego zdrowo działającego inwestora.

## 6 Inne pomysły i niespełnione marzenia

Zanim podsumuję pracę, chciałbym przypomnieć, że opierała się na założeniu które twierdziło, że pomiędzy kolejnymi wartościami procesu był maksymalnie jeden spadek. Jak mieliśmy szansę się przekonać, taki model zwrócił całkiem dobre wyniki, choć założenie wydaje się być dość mocne. Chciałbym więc przed zakończeniem opisać krótko inny pomysł estymacji parametrów procesu, którego nie wykonałem z powodu braku czasu.

Motywacją do podejścia które zaprezentuję niżej jest problem niewiadomej ilości skoków na przedziale. Uważam, że metodą, która mogłaby dać lepsze wyniki może być następujące podejście.

Mając oryginalną trajektorię można zgromadzić informację o okresach na których nie było skoków. Następnie, możnaby policzyć prawdopodobieństwo warunkowe, że ilość skoków na pojedynczym przedziale na którym coś się działo równa jest 1, 2, 3, ... pod warunkiem tego, co wiemy o okresach bez żadnego skoku. Takie podejście byłoby dokonaniem partycji, ponieważ jeśli na przedziale coś się działo, to był to albo 1 skok, albo 2, albo 3 itd. Dlatego prawdopodobieństwa te powinny sumować się do jedynki. Moją propozycją wobec tego faktu jest rozpatrzenie liczby skoków na każdym przedziale na którym coś się działo jako zmiennej losowej o rozkładzie zadany wektorem prawdopodobieństw takim jak ten wyliczony wyżej. Bardzo chciałbym napisać w tym miejscu do czego by takie podejście prowadziło, ale niestety na obecny moment tego nie wiem. Mogę tylko podejrzewać, że rolę jednorodnego procesu Poissona przejąłby mieszany proces Poissona, ze zmienną mieszającą z wyliczonego rozkładu, ale nie piszę tego z pełnym przekonaniem i bardzo żałuję że nie mam okazji przedstawić faktycznych wyników.

Przepraszam za ten dość nieprofesjonalny rozdział, ale różne, alternatywne podejścia i miliony pomysłów które przewinęły się podczas pisania tej pracy zainteresowały mnie na tyle, że po prostu musiałem zostawić choć cień tego co działo się w mojej głowie.

## 7 Podsumowanie i wnioski

W tej pracy, na podstawie zadanej trajektorii zaproponowałem model klasycznego procesu ryzyka danego wzorem:

$$R(t) = u + (1 + \theta)\mu\lambda t - \sum_{i=1}^{N(t)} X_i,$$

w którym:

1.  $u = 100$
2.  $\theta = 0.2068$
3.  $N(t)$  - jednorodny Proces Poissona z intensywnością  $\lambda = 0.9225$
4.  $X_i$  - zmienne losowe *iid* z rozkładu  $\mathcal{U}(0, 9.8802)$
5.  $\mu = 4.9401$

Tak zaproponowany model, sporządzony w oparciu o założenie, iż pomiędzy spisywanymi wartościami procesu był maksymalnie jeden spadek, zwraca trajektorie podobne do oryginalnej. Po czasie  $t = 2000$  otrzymujemy około 20% symulowanych trajektorii znajdujących się pod, a 80% nad zadaną trajektorią.

Prawdopodobieństwo Ruiny w czasie  $t = 2000$  wynosi  $\Psi(100, 2000) = 0.0037$ , tyle samo, co prawdopodobieństwo Ruiny w czasie nieskończonym ( $\Psi(100) = 0.0037$ ).

Ponadto, pokazaliśmy jak duży wpływ na prawdopodobieństwo ruiny ma odpowiedni dobór kapitału początkowego, oraz parametru eksperckiego. Ilustruje to zasadniczą rolę odpowiedzialnej kontroli ryzykiem w zarządzaniu firmą.

Myślę jednak, że najcenniejszą lekcją jaką można było wyciągnąć z tej pracy jest to, że nawet najprostsze modele potrafią zwracać zaskakująco dobre wyniki, więc czasem warto stawiać sobie proste założenia, bo może powstać z nich coś bardzo ciekawego.