

Adam Łojowski
Marzena Rybak

Raport nr 2
Symulacje komputerowe

Czerwiec 2022

Wstęp

Zadanie 1

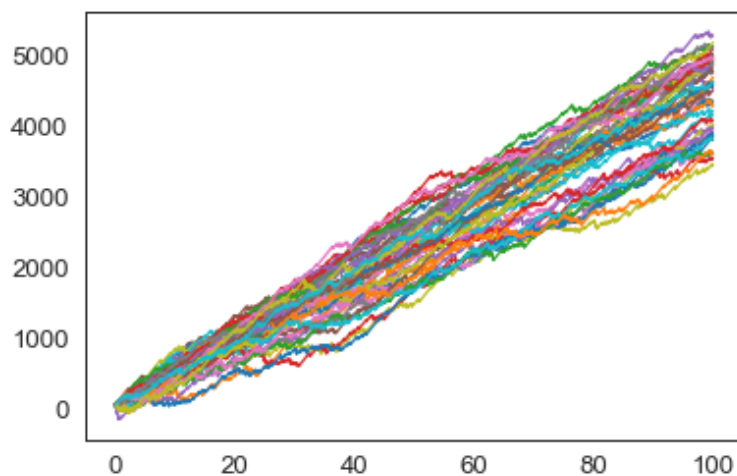
Proces ryzyka to proces opisany przez wzór

$$R_n = R_0 + ct - S_n, n = 0, 1, \dots,$$

gdzie R_0 jest kapitałem początkowym, c przyrostem otrzymywanym w jednym kroku czasowym, a $S_n = X_1 + \dots + X_n$ jest sumą wypłaconą do chwili n . Niezależne zmienne losowe X_i to kolejne wypłaty.

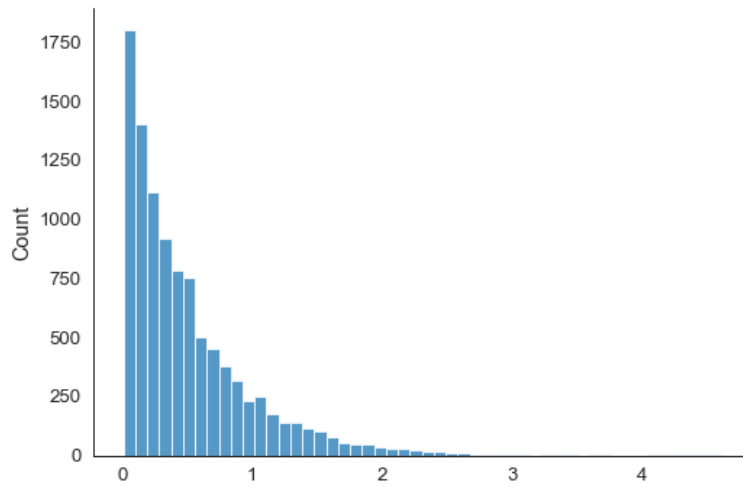
Dobór parametrów do otrzymanych trajektorii

Do zadania 1 otrzymaliśmy dane dotyczące 50 trajektorii pewnego klasycznego procesu Ryzyka. Trajektorie te zostały narysowane na rysunku 1.



Rysunek 1. Wykres 50 trajektorii z otrzymanych danych

Aby wyestymować parametry badanego procesu ryzyka zaczniemy od zbadania rozkładu częstości spadków. W tym celu na podstawie momentów spadków wszystkich 50 trajektorii stworzyliśmy histogram (rysunek 2).

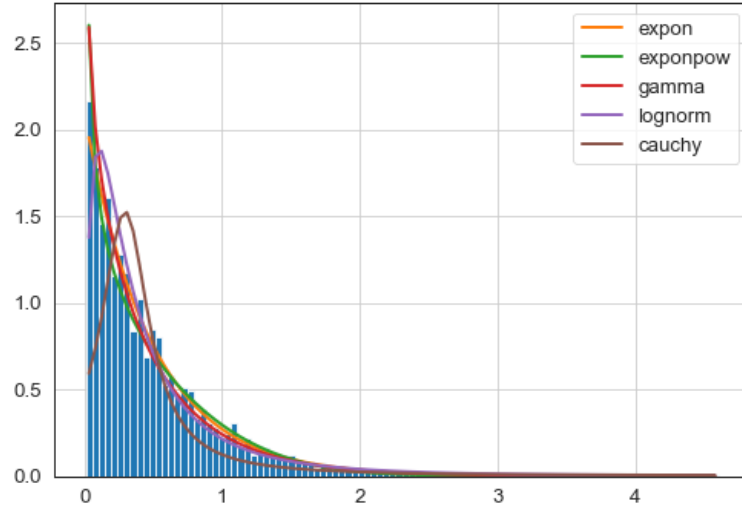


Rysunek 2. Histogram momentów spadków trajektorii

Do sprawdzenia z jakiego rozkładu są momenty i -tej szkody użyliśmy funkcji *Fitter* z biblioteki *fitter()*. Jako parametr dotyczący badanych rozkładów użyliśmy funkcji *get_common_distributions()* zawierającej większość popularnych rozkładów. Jako rezultat otrzymaliśmy tabelę 1 zawierającą rozkłady o najmniejszym błędzie kwadratowym oraz wykres porównujący histogram danych z gęstościami najlepiej dopasowanych rozkładów (rysunek 3).

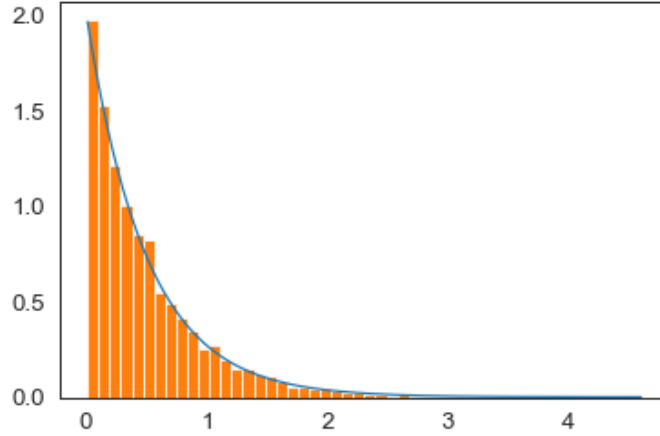
	sumsquare_error	aic	bic	kl_div
expon	0.251969	800.952414	-105197.522201	inf
exponpow	0.494767	1318.551584	-98479.632422	inf
gamma	0.518867	798.582852	-98006.780099	inf
lognorm	1.212969	667.171117	-89564.246721	inf
cauchy	5.188853	713.581192	-75123.334681	inf

Tabela 1. Dopasowania rozkładów za pomocą funkcji *fitter()*



Rysunek 3. Porównanie histogramu spadków trajektorii z dopasowaniami gęstości rozkładów

Z danych zawartych w tabeli 1 możemy wyciągnąć wniosek, że spadki w trajektoriach są z rozkładu wykładniczego, co zgadza się z przypuszczeniami. Aby uzyskać parametr λ tego rozkładu skorzystaliśmy z funkcji `stats.expon()` z biblioteki `scipy`. Aby uzyskać dopasowanie znanego nam rozkładu wykładniczego (z parametrem λ) jako parametr przesunięcia ustaliliśmy 0. Otrzymaliśmy w ten sposób intensywność jednorodnego procesu Poissona $\lambda \approx 2.00768$. Na rysunku 4 zostało przedstawione porównanie histogramu spadków z gęstością ostatecznego dopasowania rozkładu wykładniczego.

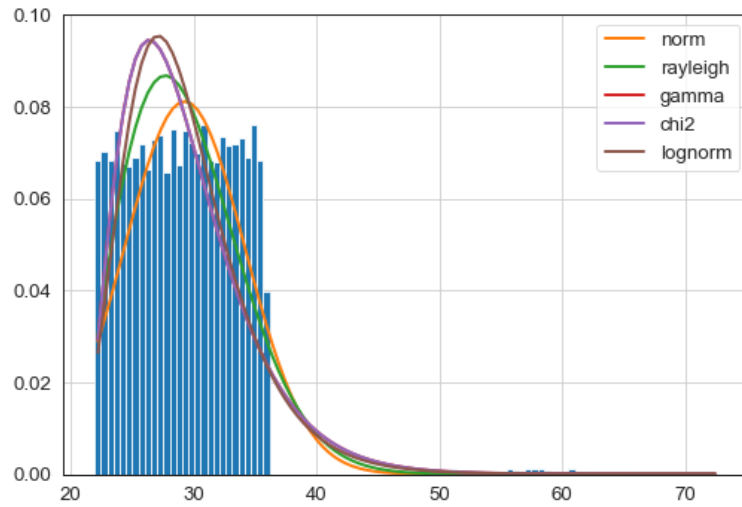


Rysunek 4. Porównanie histogramu momentów spadków trajektorii z gęstością dopasowanego rozkładu wykładniczego

Na podstawie danych możemy sprawdzić empiryczną wartość oczekiwaną $\mathbb{E}X_i = 29.26991$. Następnym krokiem jest sprawdzenie z jakiego rozkładu są wielkości spadków. W tym celu również skorzystamy z funkcji *fitter()*. Jako parametr dotyczący badanych rozkładów także użyliśmy funkcji *get_common_distributions()*. Jako rezultat otrzymaliśmy tabelę 2 zawierającą rozkłady o najmniejszym błędzie kwadratowym oraz wykres porównujący histogram danych z gęstościami najlepiej dopasowanych rozkładów (rysunek 5).

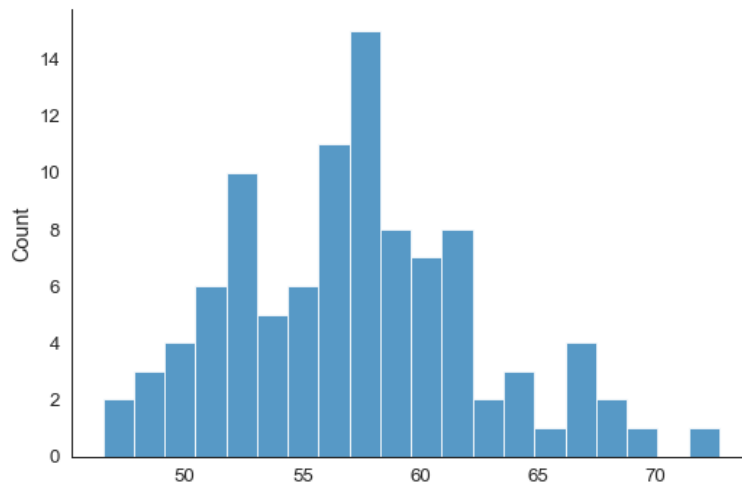
	sumsquare_error	aic	bic	kl_div
norm	0.012337	2742.876249	-135919.646527	inf
rayleigh	0.012866	2064.203459	-135500.328616	inf
gamma	0.017960	1414.953136	-132158.104936	inf
chi2	0.017960	1414.959899	-132158.037633	inf
lognorm	0.018201	1418.390039	-132024.800417	inf

Tabela 2. Tabela dopasowania rozkładów za pomocą funkcji *fitter()*

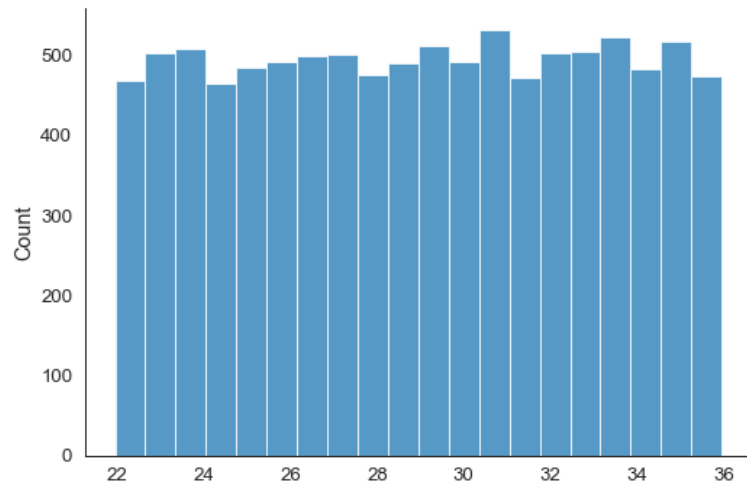


Rysunek 5. Porównanie histogramu momentów spadków trajektorii z dopasowaniami gęstości rozkładów

Wykres z rysunku 5 wskazuje na złe dopasowanie wszystkich ze sprawdzanych rozkładów. Sytuacja ta jest sugestią, że dane nie są z żadnego ze sprawdzanych rozkładów. Przyjrzyjmy się osobno histogramowi odstających wartości – nie mniejszych niż 40 (rysunek 6) oraz tych mniejszych od 40 (rysunek 7).



Rysunek 6. Histogram spadków trajektorii dla wartości nie mniejszych niż 40

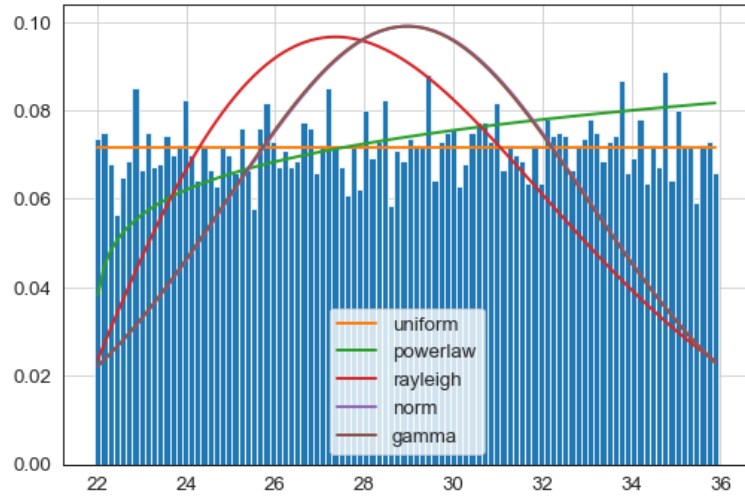


Rysunek 7. Histogram spadków trajektorii dla wartości mniejszych niż 40

Histogramy z rysunków 6 i 7 są bardziej podobne do gęstości znanych nam rozkładów, niż gdy wielkości spadków były rozpatrywane razem. Zastosujmy więc dopasowanie rozkładów za pomocą funkcji *fitter()*. Dla wartości mniejszych niż 40 otrzymane wyniki opisuje tabela 3 oraz rysunek 8, natomiast dla wartości nie mniejszych niż 40 wyniki są opisane przez tabelę 4 oraz rysunek 9.

	sumsquare_error	aic	bic	kl_div
uniform	0.004154	531.752088	-145242.280336	0.004043
powerlaw	0.010992	535.544926	-135607.048967	0.011265
rayleigh	0.063271	562.182020	-118300.985640	0.073525
norm	0.068990	567.050377	-117444.889800	0.081171
gamma	0.068993	569.050469	-117435.218081	0.081170

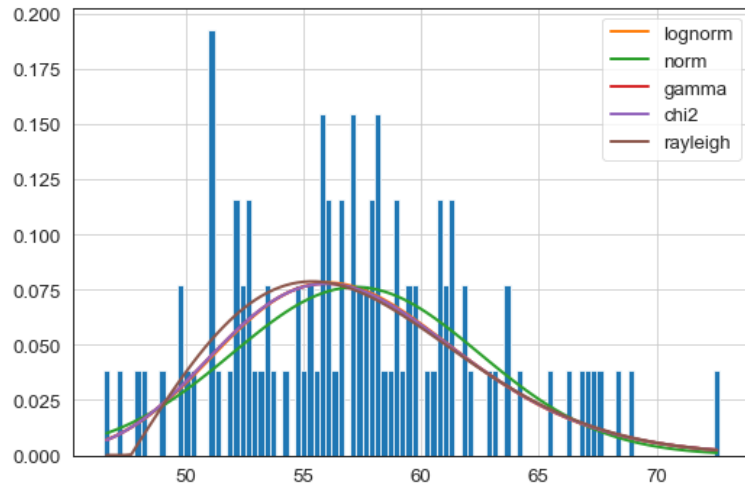
Tabela 3. Tabela dopasowania rozkładów za pomocą funkcji *fitter* dla wartości mniejszych niż 40



Rysunek 8. Porównanie histogramu wielkości spadków trajektorii dla wartości mniejszych niż 40 z dopasowanymi gęstościami rozkładów

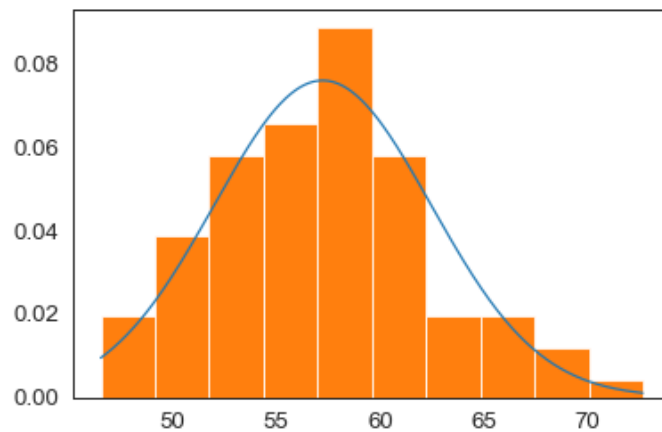
	sumsquare_error	aic	bic	kl_div
lognorm	0.130665	737.878538	-642.607886	inf
norm	0.130726	747.517737	-647.157219	inf
gamma	0.130784	736.792286	-642.517926	inf
chi2	0.130784	736.792298	-642.517925	inf
rayleigh	0.133657	inf	-644.962222	inf

Tabela 4. Tabela dopasowania rozkładów za pomocą funkcji *fitter* dla wartości nie mniejszych niż 40



Rysunek 9. Porównanie histogramu wielkości spadków trajektorii dla wartości nie mniejszych niż 40 z dopasowanymi gęstościami rozkładów

Wykres przedstawiony na rysunku 9 nie odzwierciedla dobrze dopasowania rozkładu z powodu niewystarczająco liczebnego zbioru danych. Aby lepiej odzwierciedlić dane na rysunku 10 przedstawiony został histogram z mniejszą ilością słupków oraz z ostatecznym dopasowaniem.



Rysunek 10. Porównanie histogramu wielkości spadków trajektorii dla wartości nie mniejszych niż 40 z dopasowanym rozkładem normalnym

Podsumowując dobieranie rozkładów dla spadków trajektorii procesu ryzyka mamy dla wartości mniejszych niż 40 rozkład jednostajny na odcinku [22; 36] oraz rozkład normalny z parametrami $\mu = 57.23180$ oraz $\sigma = 5.23996$ dla wartości

nie mniejszych niż 40. Ostateczną gęstość spadków trajektorii możemy zapisać za pomocą wzoru

$$f(x) = f_U(x) \cdot \frac{n_U}{n} + f_N(x) \cdot \frac{n_N}{n},$$

gdzie f_U jest gęstością rozkładu jednostajnego na odcinku [22; 36], f_N jest gęstością rozkładu normalnego z parametrami $\mu = 57.23180$ oraz $\sigma = 5.23996$, n_U jest liczebnością spadków trajektorii procesu ryzyka mniejszych niż 40, n_N jest liczebnością spadków trajektorii procesu ryzyka dla wartości nie mniejszych niż 40 oraz $n = n_U + n_N$. Wzór ten można równoważnie interpretować jako rozkład, którego zmienna losowa z prawdopodobieństwem 99.00921% jest z rozkładu jednostajnego na odcinku [22; 36] oraz z prawdopodobieństwem 0.99079% jest z rozkładu normalnego z parametrami $\mu = 57.23180$ oraz $\sigma = 5.23996$.

Kolejną częścią jest estymacja parametru odpowiedzialnego za wysokość premii θ . W tym celu sprawdziliśmy w kilku momentach trajektorii, gdzie nie występował spadek wzrost wartości otrzymując współczynnik kierunkowy wzrostów $c = 105$ oraz obliczając $\theta = \frac{c}{\mathbb{E}X_i \cdot \lambda} - 1 = 0.78679$.

Symulacja trajektorii procesu ryzyka

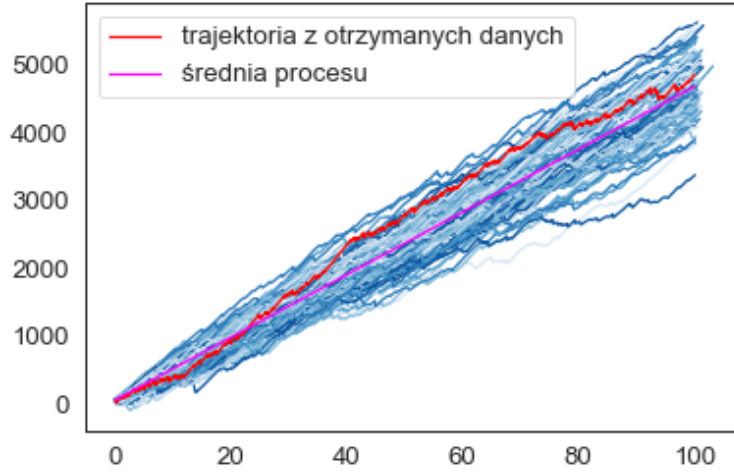
Do symulacji momentów spadków trajektorii procesu ryzyka skorzystamy z poniższego algorytmu:

1. wstawiamy $I = 0$, $t = 0$,
2. generujemy U z rozkładu jednostajnego na odcinku $[0, 1]$,
3. wstawiamy $t = t - \frac{1}{\lambda} \log(U)$, jeśli $t > T$ kończymy, w przeciwnym przypadku wstawiamy $I = I + 1$, $S_I = t$ (S_I jest momentem skoku),
4. wracamy do punktu 2 algorytmu.

Do generowania samego procesu ryzyka posłuży nam algorytm:

1. generujemy $N(t)$ (proces liczący stały) na przedziale $[0, T]$,
2. generujemy $X_1, \dots, X_{N(t)}$ (zmienne losowe reprezentujące wysokość strat),
3. wstawiamy $R(t) = u + c(t) - \sum_{i=1}^{N(t)} X_i$.

Na rysunku 11 przedstawione zostały 100 uzyskanych symulacyjnie trajektorii badanego procesu ryzyka wraz z średnią procesą (kolor różnowy) oraz jedną trajektorią z otrzymanych danych (kolor czerwony).



Rysunek 11. Sto wysymulowanych trajektorii wraz średnią procesu oraz trajektorią z otrzymanych danych

Prawdopodobieństwo ruiny

Zacznijmy od zdefiniowania paru pojęć.

Moment ruiny to moment czasowy, gdy wartość procesu ryzyka spadnie poniżej zera.

Prawdopodobieństwo ruiny jest równe prawdopodobieństwu, że wartość procesu ryzyka w pewnym ustalonym czasie (skończonym lub nieskończonym) spadnie poniżej zera.

Wzór Pollaczka-Chinczyna to zależność zadana przez równanie

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

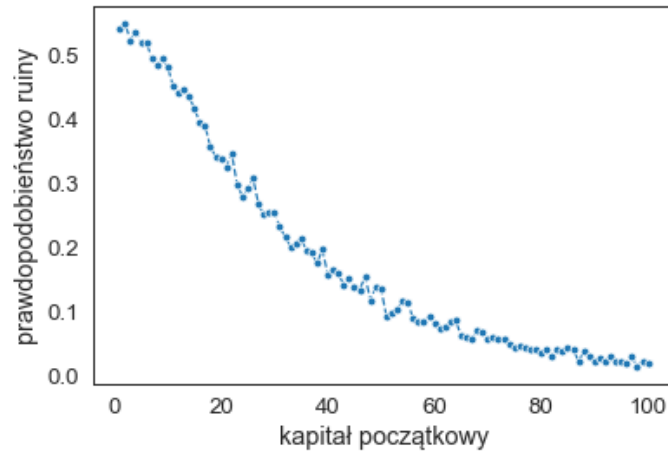
gdzie $B_n(u) = P(Y_1 + \dots + Y_n > u)$, Y_i to niezależne zmienne losowe z tego samego rozkładu o gęstości $f(x) = \frac{1 - F_{X_i}(x)}{\mu}$ oraz $\mathbb{E}X_i = \mu$.

Algorytm na podstawie wzoru Pollaczka-Chinczyna prezentuje się następująco:

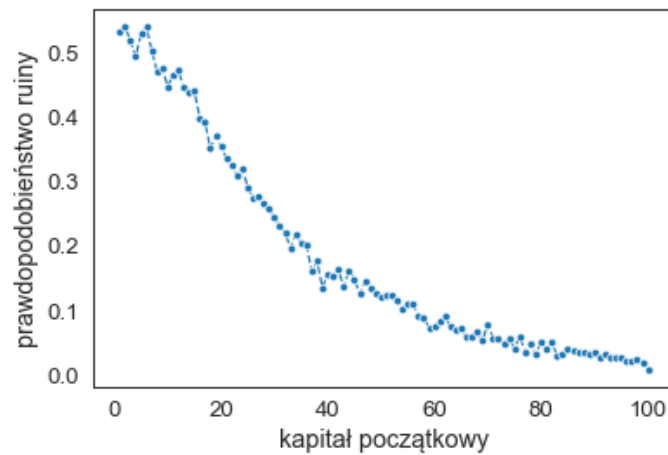
1. generujemy $K \sim \text{Geom}\left(\frac{\theta}{1+\theta}\right)$,
2. generujemy Y_1, Y_K – niezależne zmienne losowe o tym samym rozkładzie zadanym przez funkcję gęstości $f(x) = \frac{1 - F_{X_i}(x)}{\mu}$,
3. jeśli $Y_1 + \dots + Y_K > u$ wstawiamy $Z(i) = 1$, w przeciwnym razie $Z(i) = 0$,
4. zwiększamy i o 1 i powtarzamy kroki 1-3 do momentu gdy $i > N$,

5. wstawiamy $\psi(u) = \frac{Z(1) + \dots + Z(N)}{N}$.

Na rysunkach 12-13 przedstawione zostały prawdopodobieństwa ruiny w zależności od kapitału początkowego dla różnych czasów trwania symulacji.



Rysunek 12. Wyniki symulacji prawdopodobieństwa ruiny w zależności od kapitału początkowego dla $T = 100$



Rysunek 13. Wyniki symulacji prawdopodobieństwa ruiny w zależności od kapitału początkowego dla $T = 200$

Korzystając ze wzoru Pollaczka-Chinczyna mamy, że dla $u = 0$ (brak kapitału początkowego) $\psi(u) = \frac{1}{1+\theta}$. Wiemy również że wraz ze wzrostem kapitału początko-

wego prawdopodobieństwo ruiny spada, więc prawdopodobieństwo ruiny badanego procesu w nieskończoności jest mniejsze niż 55.96624%.

Podsumowanie i wnioski z zadania 1

Otrzymane dane opisują proces ryzyka o częstotliwości występowania wypłat $\lambda \approx 2.00768$. Zmienne losowe X_i (wielkości strat) są z rozkładu będącego mieszaniną rozkładu normalnego oraz jednostajnego. Parametr wysokości premii θ jest równy 0.78679.

Prawdopodobieństwo ruiny badane w zadaniu zmniejsza się tworząc wklęsły wykres prawdopodobieństwa od kapitału początkowego zbliżając się do 0. Otrzymane wyniki są zgodne z intuicją, która sugeruje zmniejszanie się prawdopodobieństwa ruiny wraz ze wzrostem kapitału początkowego.

Prawdopodobieństwo ruiny badanego procesu w nieskończoności jest mniejsze niż 55.96624%. Szacowanie zgadza się z wynikami otrzymanymi symulacyjne. Wyniki symulacji sugerują prawdopodobieństwo równe w przybliżeniu 14% dla kapitału początkowego 50.

Zadanie 2

Proces Wienera

Proces Wienera to proces stochastyczny $(B_t)_{t \geq 0}$ na przestrzeni probabilistycznej (Ω, F, P) , który spełnia następujące warunki:

1. $B_0 = 0$ prawie na pewno,
2. Proces ten ma przyrosty niezależne oraz stacjonarne,
3. Trajektorie procesu są ciągłe,
4. $B_t - B_s \sim N(0, t - s)$.

Warto zaznaczyć, że Proces Wienera nie jest różniczkowalny nigdzie.

Sposób generowania Procesu Wienera

Do generowania tego procesu stochastycznego wykorzystaliśmy metodę zaprezentowaną w trakcie zajęć laboratoryjnych. Zastosowaliśmy następujący wzór w języku *Python*

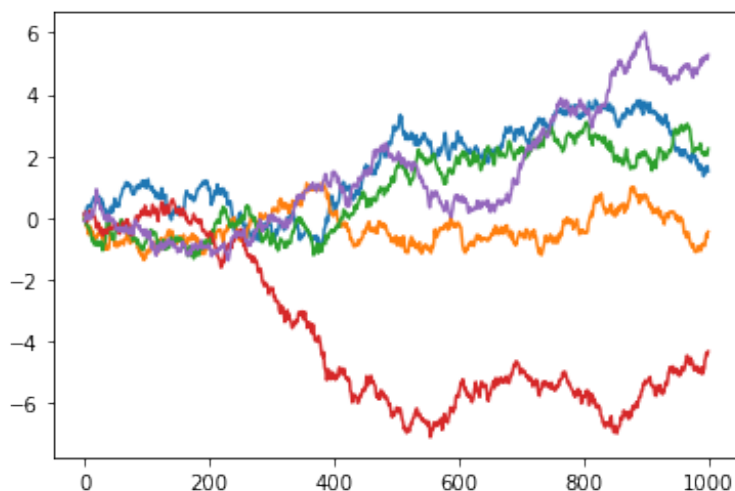
$$B_{t_k} = \sum_{j=1}^k Z_j \sqrt{\Delta}, Z_j \sim N(0, 1).$$

Wartość Δ jest zadana wzorem $\Delta = \frac{T}{n}$ gdzie n oznacza długość symulacji, a T oznacza wartość takiego kroku czasowego, że:

$$0 = t_0 < t_1 < \dots < t_n = T$$

Trajektorie Procesu Wienera

Dla powyżej zadanego wzoru wykonaliśmy 5 trajektorii procesu Wienera, które zostały zaprezentowane na rysunku 14.



Rysunek 14. Wykres 5 trajektorii Procesu Wienera dla $\Delta = 0.01$

Średni czas wyjścia z przedziału $[a, b]$

Weźmy proces Wienera $(B_t^x)_{t \geq 0}$ startujący z $x \in \mathbb{R}$. Zakładając, że B_t^0 jest procesem Wienera startującym z 0, jest on zdefiniowany następującym wzorem

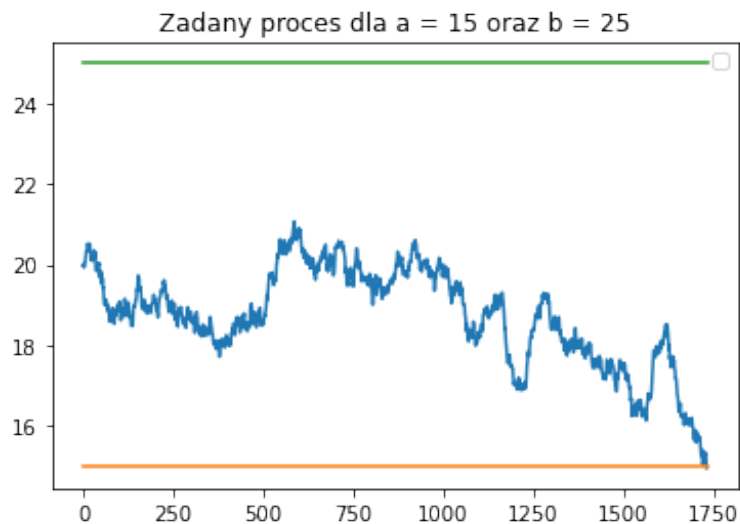
$$B_t^x = B_t^0 + x$$

Szacujemy średni czas wyjścia z przedziału $[a, b]$ w zależności od $x \in (a, b)$. Czas wyjścia oznaczamy symbolem τ^x . Obliczenia wykonujemy dla $n = 1000$ kroków.

Korzystając z Centralnego Twierdzenia Granicznego stwierdzamy, że średnia w próbie jest dobrym estymatorem wartości oczekiwanej $\mathbb{E}\tau^x$, więc wystarczy jedynie zasymulować wyżej zdefiniowany proces wielokrotnie i sprawdzić, dla których kroków wartości B_t^x wyjdą z przedziału $[a, b]$, a następnie wyznaczyć z nich średnią.

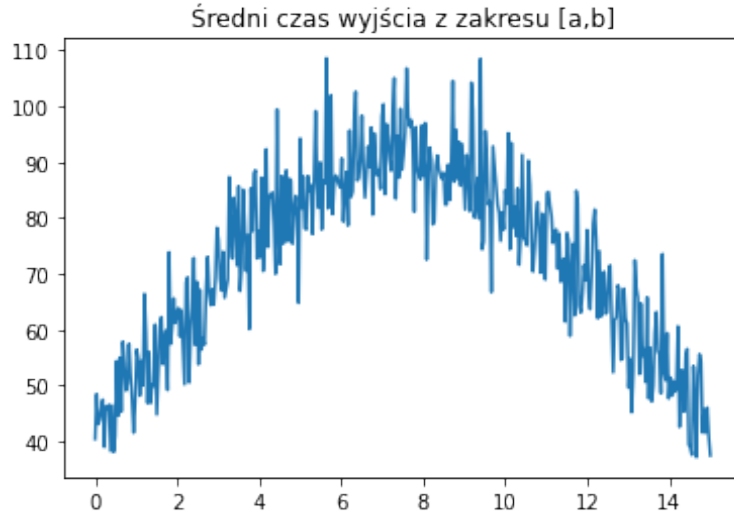
Wyniki $\mathbb{E}\tau^x$ dla różnych parametrów

Wykres przykładowej symulacji wskazanego procesu (B_t^x) został przedstawiony na rysunku 15. Dla uproszczenia zakładamy, że w poniższym przypadku x znajduje się w środku przedziału $[a, b]$, czyli w punkcie $\frac{a+b}{2}$.



Rysunek 15. $(B_t^x)_{t \geq 0}$ startujący z x pośrodku przedziału $[a, b]$

W kolejnej symulacji sprawdzamy ile średnio kroków będziemy potrzebowali, żeby wyjść z zakresu $[a, b]$ dla każdego x zawierającego się w zadanym przedziale.



Rysunek 16. Wykres przedstawiający średni czas wyjścia z zakresu w zależności wyboru od miejsca startowego

Paraboliczny kształt powyższego wykresu nie powinien nikogo dziwić. Wybierając miejsce bliższe którejś z granic będziemy się spodziewać średnio mniejszej ilości kroków potrzebnych do wyjścia z zadanego przedziału, co jest intuicyjne. Dlatego też chcąc otrzymać najdłuższą możliwą symulację powinniśmy wybrać właśnie środek przedziału $[a, b]$.

Zmieniając parametry takie jak długość symulacji (wpływając tym samym na wartość Δ) czy wielkość przedziału wpływamy na średnią wartość τ^x , czyli potrzebnych kroków potrzebnych do opuszczenia przedziału. Tabela 5 przedstawia $\mathbb{E}\tau^x$ dla zadanych różnych parametrów Δ, a oraz b , że proces startujemy ze środka przedziału. Gwarantujemy tym samym,

Δ	$[a, b]$	$\mathbb{E}\tau^x$
0.01	[18,20]	110.079
0.01	[17,23]	912.901
0.05	[18,20]	27.467
0.05	[17,23]	198.441
0.1	[17,23]	99.519
0.1	[15,25]	268.504

Tabela 5. Tabela wartości $\mathbb{E}\tau^x$ dla różnych parametrów

Dyskretyzacja

W trakcie procesu Wienera generujemy punkty co krok $Z_j\sqrt{\Delta}$, gdzie $Z_j \sim N(0, 1)$. Dlatego też praktycznie oraz zanedbywalnie nigdy nie zajdzie równość taka, że $P(B_{\tau^x}^x = b)$. Dlatego też możemy zmodyfikować równanie na

$$P(B_{\tau^x}^x \geq b).$$

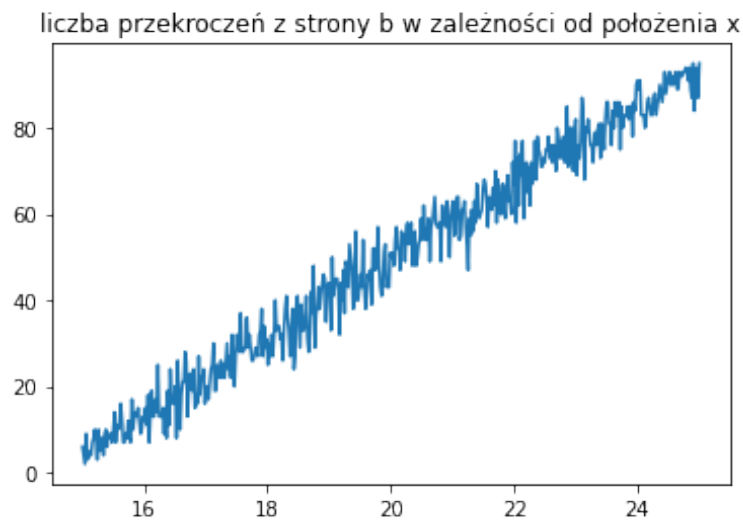
Wystarczy więc, że policzymy, ile razy dla danej symulacji wyszliśmy przez b oraz podzielimy tę liczbę przez ilość wszystkich prób, jaką generowaliśmy.

Na tabeli 6 przedstawiliśmy wyniki badań wartości $P(B_{\tau^x}^x \geq b)$ w zależności od wybranego punktu startowego x . Do symulacji wybraliśmy przedział $[15, 25]$.

x	$P(B_{\tau^x}^x \geq b)$
15.5	0.05
17.5	0.266
20	0.503
22.5	0.741
24.5	0.940

Tabela 6. Tabela wartości $P(B_{\tau^x}^x \geq b)$ w zależności od wybranego punktu startu x

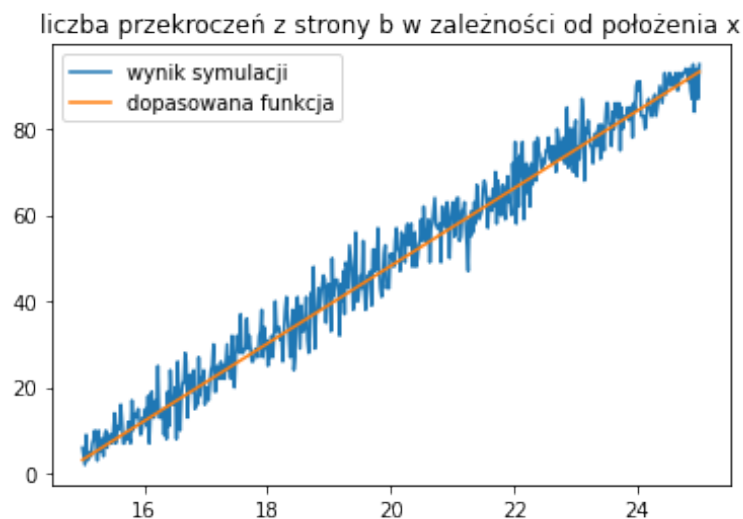
W drugiej części zadania mieliśmy zasymulować proces Wienera startującego z różnych punktów $x \in [a, b]$ oraz sprawdzić ile potrzebujemy średnio kroków, aby wyjść z tego przedziału ze strony b .



Rysunek 17. Wykres przedstawiający liczbę przekroczeń ze strony b przedziału $[15, 25]$ w zależności od wyboru punktu startowego x

Powyższy wykres przypomina wykres funkcji liniowej. Do wyznaczenia funkcji przybliżającej owy wykres wykorzystaliśmy pierwszą i ostatnią wartość wykonanej symulacji. Dla takich wartości owa funkcja ma wzór

$$f(x) = 3.24x - 42.5$$



Rysunek 18. Wykres zawierający dopasowaną funkcję do wyniku symulacji przedstawionej na rysunku 17

Podsumowanie i wnioski z zadania 2

Proces Wienera, nazywany również ruchem Browna, zasymulowaliśmy metodą poznaną na laboratorium. Dla zadanych poleceń wykonaliśmy odpowiednie symulacje, z których wyciągnęliśmy wnioski.

Kształty wykresów dla średnich czasów wyjścia z zadanego przedziału $[a, b]$ z obu stron oraz tylko ze strony b są zgodne z intuicją.

W przypadku wyjścia z obu stron przedziału otrzymaliśmy parabolę. Wybierając x bliższe którejś z granic naszego zakresu zwiększamy prawdopodobieństwo, że uciekniemy z niego w mniejszej ilości kroków, dlatego też spodziewamy się, że na wykresie $\tau^x(x)$ otrzymamy symetryczny kształt wykresu, którego wartości będą rosły wraz ze zbliżaniem się do środka przedziału.

W przypadku wyjścia z przedziału tylko i wyłącznie od strony b otrzymaliśmy kształt funkcji liniowej $f(x)$. Ilość kroków czasowych t_j rośnie wprost proporcjonalnie wraz z przybliżaniem się wartości x do granicy b stąd owy kształt. Gdybyśmy chcieli wyrysować podobny wykres jednak liczony dla ilości wyjść od strony a , otrzymalibyśmy wykres funkcji liniowej $-f(x)$ ze zbliżonymi parametrami ponieważ proces Wienera nie jest deterministyczny.