Symulacje Komputerowe
Lista Zadań

1 Tydzień I. Zadania wprowadzające.

Funkcja `random()` w Pythonie generuje liczbę pseudolosową z rozkładu jednostajnego $U(0,1)$. Oznacza to, że liczbę z przedziału $[a,b]$, $0 \leqslant a < b \leqslant 1$ wylosujemy z prawdopodobieństwem $b-a$. Jeśli X jest zmienną losową z rozkładu $U(0,1)$, to $A+B\cdot X$ jest zmienną losową z rozkładu $U(A,B+A)$.

Zmienna losowa z rozkładu dyskretnego przyjmuje przeliczalną liczbę wartości $x_1,x_2,\ldots$. Każdej z tych wartości możemy przypisać określone prawdopodobieństwo: $P(X=x_i)=p_i$.

Zmienne losowe dyskretne można łatwo wygererować, jeśli mamy do dyspozycji zmienną losową z rozkładu jednostajnego. Przykładowo rzut symetryczną monetą możemy symulować w następujący sposób:

1. Losujemy zmienną losową X z rozkładu $U(0,1)$
2. jeśli $X > 0.5$, to zwracamy O. W przeciwnym wypadku zwracamy R.

Na pierwszych zajęciach zajmiemy się podstawami metody Monte-Carlo i implementacją podstawowych narzędzi.

Zadanie 1

- (2 pkt) Grześ i Jaś rzucają monetą do czasu, aż w trzech ostatnich rzutach pojawi się sekwencja OOR (wygrywa Jaś) lub ORR (wygrywa Grześ). Oszacuj prawdopodobieństwo zwycięstwa Grzesia.
- (2 pkt) Grześ ma 25 bitcoinów, a Jaś 5 bitcoinów (grupa środowa: 5). Przy każdej rozgrywce każdy z nich stawia jednego bitcoina. Chłopcy grają tak długo, aż jeden z nich zbankrutuje. Jakie jest prawdopodobieństwo bankructwa Grzesia?
- (w domu, 3 pkt) Do gry dołącza się Tosia z 50 bitcoinami i sekwencją RRR. Jakie teraz są prawdopodobieństwa zwycięstwa i bankructwa każdego z graczy? Czy da się znaleźć takie kapitały początkowe, aby każdy z graczy rozbijał bank (doprowadzał do bankructwa pozostałych graczy) z takim samym prawdopodobieństwem?

Zadanie 2 (6 pkt) Dana jest próba losowa $\mathbb{X}$ rozmiaru n. Dystrybuantą empiryczną nazywamy funkcję

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \chi_{(-\infty,x]}(X_i). \tag{1}$$

Zaprogramuj funkcję, która, przyjmując jako argumenty $\mathbb{X} \in \mathbb{R}^n$ oraz $x \in \mathbb{R}$ jako wynik zwróci wartość dystrybuanty empirycznej próby $\mathbb{X}$ w punkcie x. Następnie napisz funkcję, która rysuje wykres dystrybuanty empirycznej (na jakimś sensownym przedziale - jakie heurystyki wydają się sensowne?) próby $\mathbb{X}$. Z definicji dystrybuanty empirycznej wynika, że dystrybuanta empiryczna jest funkcją schodkową, niemalejącą, w której 'schodki' występują w punktach $x \in \mathbb{X}$ i każdy ze schodków ma 'wysokość' k/n.

2 Tydzień II. Całkowanie Monte-Carlo. Metoda odwracania dystrybuanty.

Zadanie 3 (5 pkt) Napisz funkcję, która dla zadanego przedziału $[a,b]$, funkcji f i liczby powtórzeń Monte-Carlo N szacuje wartość całki

$$\int_a^b f(x)dx.$$

Przetestuj funkcję dla $f = \sin(x)$ na przedziale $[0,2\pi]$, $f(x) = e^x$ na $[0,1]$ oraz $f(x) = \frac{1}{1+x^2}$ na przedziale $[-100,100]$ (na zajęciach wystarczy jedna z wymienionych funkcji, pozostałe do domu). Uwaga! Trzeba będzie przekazać funkcję jako argument.

Zadanie 4 (3 pkt) Napisz funkcję do oszacowywania liczby π metodą Monte-Carlo. Koło o promieniu 1 wpisujemy w kwadrat o boku 2. Wybieramy losowo, niezależnie kolejnych N punktów z kwadratu (współrzędna X z rozkładu jednostajnego $U(0,2)$ i współrzędna Y niezależnie też z $U(0,2)$). Liczbę punktów, które znalazły się w kole oznaczmy przez m. Teraz m/N będzie dążyć do $\pi/4$ (dlaczego?). Zbadaj, jak szybko ten stosunek zbiega do wartości granicznej: sporządź wykres oszacowanej średniej różnicy $r_N = |m/N - \pi/4|$ w zależności od N: w tym celu dla każdego $N = 10,20,50,100,200,500,1000$ wykonaj $M = 1000$ prób Monte-Carlo, aby oszacować r_N. Uzyskane wartości nanieś na wykres. Następnie sporządź ten sam wykres ze skalą logarytmiczną na wybranej osi.

Zadanie 5 (2 pkt, w domu) Czym różnią się algorytmy Monte Carlo od algorytmów Las Vegas? Zaimplementuj algorytm Bogosort (tzw. „zwariowane sortowanie”) i metodą Monte Carlo oszacuj średni czas potrzebny do posortowania tablicy (lub listy) 5-elementowej. W tym celu przyda się funkcja do 'tasowania' elementów: w Pythonie jest to

```
import random ;
random.shuffle(array)
```

Zadanie 6 (3 pkt, w domu) W podobny sposób jak w Zadaniu 4, oszacuj szybkość zbieżności do prawidłowego wyniku oszacowań funkcji $\sin x$ i e^x z Zadania 3, tzn. jak dobre (średnio) oszacowanie dostajemy dla różnych N i wybranych funkcji? Prawidłowy wynik potrafimy obliczyć analitycznie.

Zadanie 7 (2 pkt) Zaprogramuj za pomocą metody odwracania dystrybuanty funkcje generujące liczby z rozkładów:

- $Exp(\lambda)$
- $N(\mu, \sigma^2)$
- $Gamma(a, b)$

Wszędzie tam, gdzie potrafimy wyliczyć jawne wzory na dystrybuantę odwrotną, korzystamy z nich. Tam, gdzie się nie da, sięgamy po wskazówki do Internetu (np. funkcje norminv, gaminv w Matlabie).

(w domu, 2 pkt) Każdy napisany generator testujemy, rysując histogram, gęstość teoretyczną, dystrybuantę empiryczną, dystrybuantę teoretyczną, wykres kwantylowy (QQ-plot) a także licząc średnią i wariancję, i testy zgodności. Poczytać o najważniejszych testach zgodności. Dla rozkładu normalnego ciekawy artykuł można znaleźć tutaj: http://smarterpoland.pl/index.php/2013/04/wybrane-testy-normalnoci/

Podsumowanie Wszystkie rozwiązania wysłać po skończonych zajęciach na adres jacek.mucha@pwr.edu.pl. Na następnych zajęciach będziemy się zajmowali generowaniem liczb pseudolosowych z ustalonych rozkładów ciągłych i dyskretnych oraz ich implementacją. Dla zainteresowanych: przypomnijcie sobie, co to jest złożoność obliczeniowa. Ponadto przypomnijcie sobie relacje między najważniejszymi rozkładami prawdopodobieństwa i o ich występowaniu w nauce i technice. Źródła, np.

1. Wikipedia
2. John D. Cook
3. R. Magiera, *Modele i metody statystyki matematycznej*

3 Tydzień III. Uogólnienie metody akceptacji-odrzucenia.

Zadanie 8 (3 pkt) Napisać funkcję, która, korzystając z metody akceptacji-odrzucenia, generuje zmienną losową o rozkładzie $P(X=0)=\frac{1}{10}$, $P(X=1)=P(X=3)=\frac{2}{10}$, $P(X=2)=P(X=4)=\frac{1}{20}$, $P(X=5)=P(X=6)=P(X=7)=\frac{2}{15}$. Narysować histogram i wykres d.e. Funkcja powinna przyjmować jako argument słownik o kluczu: wartość zmiennej losowej i wartości: prawdopodobieństwo przyjęcia tej wartości.

Zadanie 9 (zadanie dodatkowe, ponadprogramowe, do domu, 7 pkt) Trzy rody rzymskie, Fabiusze, Emiliusze i Korneliusze po wygnaniu Tarkwiniusza Pysznego składają się wyłącznie z m_F, m_E, m_K mężczyzn. W każdym pokoleniu mężczyzna może mieć do n synów, którzy zapewniają przedłużenie rodu. Napisz funkcję, która jako parametr dostaje listę $P=(p_0, p_1, ...p_n)$, gdzie p_i oznacza prawdopodobieństwo posiadania i-synów przez jednego mężczyznę, i m (początkowa liczba mężczyzn) a która zwraca oszacowanie prawdopodobieństwa, że ród o płodności P i liczbie mężczyzn m wyginie w ciągu T pokoleń. Oszacuj prawdopodobieństwo wyginięcia wymienionych rodów dla $P_F = (0.4, 0.1, 0.2, 0.3)$, $m_F = 3$, $P_E = (0.3, 0.5, 0.05, 0.05, 0.05, 0.05)$, $m_E = 5$, $P_K = (0.5, 0.5)$, $m_K = 20$.

To zadanie stanowi przykład tzw. *branching process*. Podany tu przypadek da się wyliczyć analitycznie za pomocą tzw. funkcji tworzących.

Zadanie 10 (5 pkt, czyli 1 pkt za każdy podpunkt) Metoda akceptacji-odrzucenia dla rozkładów ciągłych o nośnikach zwartych.
Napisz funkcje generujące n zmiennych losowych z poniższych rozkładów, używając metody akceptacji-odrzucenia. Zadbaj o wykresy: dystrybuantę teoretyczną, empiryczną, histogram, gęstość teoretyczną. Policz także średnią i wariancję.

1. $f(x) = \frac{3}{2}(1 - x^2)$, $x \in [0, 1]$,
2. Rozkład trójkątny $f(x) = \frac{2(x-a)}{(b-a)(c-a)}$, $x \in [a, c]$, $f(x) = \frac{2(b-x)}{(b-a)(b-c)}$, $x \in [c, b]$,
3. $f(x) = \frac{\sin(x)}{2}$ na $[0, \pi]$,
4. $f(x) = \frac{3}{2} \sin(x) \cos^2(x)$ na $[0, \pi]$.
5. Rozkład Beta (z zadanymi parametrami)

4 Tydzień IV. Metoda Boxa-Müllera.

Zadanie 11 (4 pkt) Zaimplementuj metodę Boxa-Mullera generowania zmiennych losowych z rozkładu normalnego. Zbadaj zgodność rozkładu dla prób 10, 100, 10000-elementowych, używając testu Kołmogorowa-Smirnowa: w Matlabie jest to http://www.mathworks.com/help/stats/kstest w Pythonie zaś scipy.stats.kstest. Wygeneruj za pomocą funkcji wbudowanej 10-elementową próbę $\mathbb{X}_2$ z rozkładu $N(0.5, 2)$ i zbadaj, czy 10-elementowa próba $\mathbb{X}_1$ wygenerowana za pomocą metody B-M pochodzi z tego samego rozkładu, co $\mathbb{X}_2$. Jak często test K-S odrzuca hipotezę o zgodności tych rozkładów z rozkładem normalnym? Przeprowadzić 10000 prób dla obu metod.

(w domu, 2 pkt) O czym mówi wynik testu i jak go należy interpretować? Co to jest p-wartość? Na jakiej statystyce oparty jest ten test? Jakie są związki Testu K-S z Podstawowym Twierdzeniem Statystyki Matematycznej? Jak użyć tego testu do badania zgodności rozkładów dwóch prób losowych? Poczytać, co ASA sądzi na temat bezkrytycznego używania p-wartości (np. tutaj http://amstat.tandfonline.com/doi/pdf/10.1080/00031305.2016.1154108 lub tutaj: http://smarterpoland.pl/index.php/2016/03/publikacja-wyroku-w-sprawie/).

Uwaga. Pochodną można policzyć na kartce albo w Wolframie.

Zadanie 12 (3 pkt) Za pomocą metody odwracania dystrybuanty potrafimy generować zmienne losowe z rozkładu wykładniczego. Możemy więc zastosować metodę akceptacji-odrzucenia dla pewnej klasy rozkładów o nośniku $[0, \infty)$. Napisać funkcję do generowania zmiennych losowych z uogólnionego rozkładu normalnego o gęstości

$$f(y|\alpha, \beta) = \frac{2\beta^{\frac{\alpha}{2}}}{\Gamma\left(\frac{\alpha}{2}\right)} y^{\alpha-1} e^{-\beta y^2}, \; y, \alpha, \beta > 0. \tag{2}$$

(w domu, 1 pkt) Czy, używając rozkładu wykładniczego, moglibyśmy wygenerować zmienne losowe z rozkładu Weibulla z parametrem $\alpha > 1$? Dlaczego?

Zadanie 13 (3 pkt) Metoda Boxa-Mullera wymaga obliczania wartości funkcji $\sin x$ lub $\cos x$, co znacząco wydłuża obliczenia. Zaimplementuj metodę biegunową Boxa-Mullera.

```
success=false ;
while !success do
    V1=unif(-1,1), V2=unif(-1,1);
    R2=V1*V1+V2*V2;
    if R2<= 1 then
        success=true;
    end
end
X=V1*sqrt(-2*log(R2)/R2);
Y=V2*sqrt(-2*log(R2)/R2);
return (X,Y);
```

(w domu, 2 pkt) Przetestuj, która metoda działa szybciej (biegunowa, czy zwykła). W Matlabie możesz użyć np.

```
start=cputime;
//generowanie 1000000 zmiennych ;
end=cputime ;
end-start;
```

Uwaga! Szybkość działania obu metod może zależeń od szczegółów implementacyjnych, od wersji używanego języka, od rodzaju procesora (niektóre procesory mogą mieć wbudowane koprocesory matematyczne przyspieszające liczenie np. funkcji trygonometrycznych).

Narysuj wykres funkcji czasu działania (np. w mikrosekundach) obu algorytmów na Twoim procesorze w zależności od liczby zmiennych do wygenerowania.

Zadanie 14 (w domu, 4 pkt) Zaimplementuj funkcję, która rysuje wykres kwantylowy próby losowej w odniesieniu do zadanego rozkładu. **input:** $\mathbb{X}$ - próba losowa, F - rozkład, względem którego rysujemy QQ-plot, ... - parametry rozkładu F. Należy zdecydować, czy F będzie funkcją kwantylową, dystrybuantą, a może gęstością. Funkcja powinna działać dla następujących F: rozkład normalny, rozkład log-normalny, rozkład Gamma, rozkład Pareto, rozkład Poissona, rozkład jednostajny.

Zadanie 15 (dodatkowe, w domu, termin: przed pierwszym sprawozdaniem) Zaimplementuj testy normalności wymienione w http://smarterpoland.pl/index.php/2013/04/wybrane-testy-normalnoci/ (1,5 pkt za każdy test, +1 pkt za każdy test, który nie pojawił się w artykule). Wytłumacz, co to jest moc testu (0,5 pkt). Potrafimy generować zmienne losowe z rozkładów ciągłych o zadanej gęstości. Przetestuj i porównaj czułość swoich testów dla prób z rozkładów o różnego rodzaju odchyleniach od rozkładu normalnego (innymi słowy należy powtórzyć badanie prof. Biecka) - 1 pkt. za każde sensowne odstępstwo od normalności, +0.3 pkt za każde odstępstwo, które nie pojawiło się w artykule. Wskazówka: w artykule jest podana bibliografia i kody niektórych testów w R. R jest językiem otwartym, można zajrzeć do implementacji każdego z testów osobiście. Do celów edukacyjnych można się nimi inspirować.

5 Tydzień V. Generatory liczb pseudolosowych.

Zadanie 16 (2 pkt) Zaimplementuj generator liniowy kongruentny liczb pseudolosowych z zakresu $[0,M]$ (1 pkt). Podaj statystyki opisowe. Zbadaj zgodność liczb generowanych tą metodą z rozkładem równomiernym. (2 pkt)
Generator ten generuje (w sposób deterministyczny) rekurencyjnie liczby pseudolosowe z zakresu od 0 do $m-1$ za pomocą następującego wzoru:

$$x_i = (a \cdot x_{i-1} + 1) \mod m. \tag{3}$$

x_1 to tzw. ziarno. Aby zapewnić 'losowość', warto za x_1 przyjąć np. liczbę (*integer*!) zwracaną przez cputime. m powinno być bardzo duże, w przeciwnym razie istnieje duża szansa, że wpadniemy w cykl. a najlepiej wybrać tak, aby było o rząd mniejsze od m. Najczęściej zaleca się wybór $a = w21$, gdzie w jest jakąś liczbą parzystą.
Ciekawostka. W języku ANSI C generator liczb losowych wygląda następująco:

$$x_i = (1103515245 \cdot x_{i-1} + 12345) \mod 2^{32}. \tag{4}$$

Co się stanie, jeśli przyjmiemy $a = 19$, $m = 381$, $x_1 = 0$? (1 pkt)

Zadanie 17 (1 pkt) Zaimplementuj generator Fibonacciego,

$$x_n = (x_{n-1} + x_{n-2}) \mod m.$$

Podobnie jak w poprzednim zadaniu, zbadaj zgodność rozkładu z rozkładem równomiernym.

Zadanie 18 (2 pkt) Zaimplementuj za pomocą jednej z poznanych metod generator liczb pseudolosowych z rozkładu Poissona.

Zadanie 19 (3 pkt) Korzystając z generatorów liczb pseudolosowych całkowitych, zaimplementuj generator liczb pseudolosowych z rozkładu $U(a,b)$. Narysuj histogram i wykres kwantylowy. Wykonaj testy statystyczne. Oceniany jest także pomysł!

Motto: *Some programmers, when confronted with a problem, think "I know, I'll use floating point arithmetic." Now they have 1.999999999997 problems.* – @tomscott

Zadanie 20 (2 pkt.) Niech

$$b_i = a_1 b_{i-1} + \ldots + a_k b_{i-k} mod\ 2, \tag{5}$$

$i = k+1, k+2, ...$, gdzie $a_1, a_2, ...a_k, b_1, ..., b_k \in \{0,1\}$. Zaimplementuj generator TausworTha

$$U_i = \sum_{j=1}^{L} 2^{-j} b_{is+j}, \tag{6}$$

$i = 0, 1, 2...$, natomiast s jest ustaloną liczbą nieujemną.

6 Tydzień VI. Jednorodny proces Poissona.

Zadanie 21 (4 pkt.)

Zaimplementujemy metodę generowania czasów oczekiwania na kolejny skok. Oznaczenia:
I - liczba skoków N(t) na [0,T],
$S_1, ..., S_I$ - momenty skoków,
t - suma czasów oczekiwania T_i.

1. Podstaw I=0, t=0.
2. Generuj U z rozkładu $U(0,1)$.
3. Wstaw $t = t - \frac{1}{\lambda} \log U$. Jeśli $t > T$, to koniec, w przeciwnym przypadku przejdź do 4.
4. Wstaw I=I+1, $S_I = t$.
5. Wróć do 2.

Wygeneruj 5 trajektorii procesu Poissona powyższą metodą i nanieś wygenerowane trajektorie na wykres. Pamiętaj, że trajektoriami procesu Poissona są funkcje schodkowe.

Zadanie 22 (4 pkt.)

Druga metoda generowania procesu Poissona polega na generowaniu momentów skoku.

1. Generuj $n \sim P(\lambda T)$.
2. Jeśli n=0, to koniec (brak skoków).
3. Generuj $U_1, ..., U_n$ - i.i.d., $U_i \sim U(0,T)$.
4. Sortuj $(U_1, ..., U_n)$ aby otrzymać $(U_{1:n}, U_{2:n}, ..., U_{n:n})$
5. Wstaw $S_i = U_{i:n}$(statystyki pozycyjne), $i = 1, ..., n$

Wygeneruj 5 trajektorii procesu Poissona powyższą metodą i nanieś wygenerowane trajektorie na wykres.

Zadanie 23 (4 pkt.)

Porównaj metody z zadań 1. i 2. Wygeneruj po 100000 realizacji procesu Poissona każdą metodą dla ustalonych λ i T. Która metoda jest szybsza? Jakie są ich wady? Zaproponuj metodę symulowania procesu Poissona w czasie rzeczywistym, tj. stwórz wątek (thread; dodatkowe plusy za użycie wątku w sensie programistycznym), który będzie wydawał krótkie sygnały dźwiękowe w momentach skoku. Czy któraś z powyższych metod się do tego nadaje?

(w domu, dodatkowe, 3 pkt.) Stwórz animację,która pokazywać będzie ewolucję trajektorii procesu Poissona. Zadbaj o estetykę.

Zadanie 24 (6 pkt.) Korzystając z funkcji generujących momenty kolejnych skoków, napisz funkcję, która dokonuje transkrypcji danej trajektorii procesu Poissona na tekst odczytywany na podstawie Alfabetu Morse'a. Co będzie kropką, a co kreską?

Tutaj przydatny link (nie musi być aż tak profesjonalnie, chcę, abyście używali własnych, być może niedoskonałych, pomysłów, uczymy się być wynalazcami!) https://wwwhome.ewi.utwente.nl/ ptdeboer/ham/rscw/algorithm.html

(dodatkowy 1 pkt.) Ile średnio trzeba czekać, aż zostanie wygenerowany napis „NAVIGARE EST"? (co najmniej 10 prób Monte-Carlo) A w przypadku generatora, który generuje każdą literę z tym samym prawdopodobieństwem (to ostatnie policzyć analitycznie)?

7 Tydzień VII. Niejednorodny proces Poissona.

Za $\lambda(t)$ przyjąć $\lambda_1(t) = \sin^4 t$, $\lambda_2(t) = t^4$, $\lambda_3(t) = e^{-t^2}$, $\lambda_4(t) = t$, natomiast $\lambda_5(t)$ należy wymyślić samodzielnie (każda grupa powinna mieć inne).

Zadanie 25 (5 pkt) Niech $M \geqslant \max_{t\in[0,T]} \lambda(t)$. Pierwsza metoda polega na wygenerowaniu momentów skoków S_i.
Pseudokod:

1. Podstaw t=0, I=0.
2. Wygeneruj $U_1 \sim U(0,1)$.
3. Podstaw $t = t - \frac{1}{M}\log U_1 (\sim Exp(\lambda))$. Jeśli $t > T$, to koniec.
4. Generuj $U_2 \sim U(0,1)$ (U_2 i U_1 - niezależne)
5. Jeśli $U_2 \leqslant \frac{\lambda(t)}{M}$ to I=I+1, $S_I = t$
6. Wróć do punktu 2.

Zadanie polega na zaimplementowaniu powyższej metody i narysowaniu kilku przykładowych realizacji NPP.

Zadanie 26 (5 pkt) Generowanie kolejnych czasów oczekiwania na skok. Niech $N(t)$ będzie niejednorodnym procesem Poissona z funkcją intensywności $\lambda(t)$. Jeśli proces $N(t)$ miał skok w punkcie s to dystrybuanta czasu oczekiwania na kolejny skok jest równa $F_S(x) = 1 - e^{(-\int_s^{s+x}\lambda(u)du)}$

Pseudokod:
Założenie: Potrafimy wyznaczyć $F_s^{-1}(y)$ - funkcję odwrotną do $F_s(x)$ względem zmiennej x.

1. Wstaw $t = 0$, $I = 0$.
2. Generuj $U_1 \sim U(0,1)$.
3. Wstaw $\tau = F_t^{-1}(U), t = t + \tau$. Jeśli $t > T$, to kończymy
4. Wstaw $I = I + 1$, $S_I = t$
5. Wróć do punktu 2.

Zadanie polega na zaimplementowaniu powyższej metody z wybraną funkcją λ i narysowaniu kilku przykładowych realizacji NPP.

Zadanie 27 (5 pkt) Cel: wygenerowanie wektora momentów skoków $(S_1, ..., S_n)$ na [0,T].

Niech $N(t)$ będzie niejednorodnym procesem Poissona z funkcją intensywności $\lambda(t)$. Wtedy warunkowy rozkład wektora momentów skoków $(S_1, ..., S_n)$ pod warunkiem $N(t) = n$, jest równy rozkładowi wektora statystyk pozycyjnych $(V_{1:n}, ..., V_{n:n})$, gdzie zmienne losowe $V_1, ..., V_n$ są niezależnymi zmiennymi o jednakowym rozkładzie o gęstości $f(t) = \frac{\lambda(t)}{\int_0^T \lambda(u)du}, t \in [0,T]$. Niech n - liczba skoków $N(t)$ na $[0,T]$, $S_1, ..., S_n$ - momenty skoków.
Pseudokod:

1. Generuj $n \sim Poiss(\int_0^T \lambda(u)du)$.
2. Jeśli $n = 0$, to kończymy
3. Generuj niezależnie zmienne losowe $V_1, ..., V_n$ za pomocą metody akceptacji-odrzucenia
4. Sortuj $V_1, ..., V_n$, aby otrzymać $(V_{1:n}, ..., V_{n:n})$.
5. Wstaw $S_i = V_{i:n}$, $i = 1, ..., n$.

Zadanie polega na zaimplementowaniu powyższej metody i narysowaniu kilku przykładowych realizacji NPP.

Zadanie 28 (8 pkt.) Porównaj wszystkie trzy metody, nanosząc na wykres średni czas działania każdej z nich na wybranym horyzoncie czasowym.

8 Tydzień VIII. Mieszany proces Poissona. Ruch Browna.

Zadanie 29 (3 pkt) Zaimplementuj następujący algorytm:

1. Generuj jedną realizację λ zmiennej losowej Λ.
2. Ustaw $t = 0, I = 0$.

3. Generuj $U \sim U(0,1)$, U i Λ - niezależne.
4. Wstaw $t = t - \frac{1}{\lambda}\log(U)$. Jeśli $t > T$, to koniec.
5. Wstaw $I = I + 1$, $S_I = t$.
6. Wróć do 3.

Zadanie 30 (3 pkt) Zaimplementuj następujący algorytm:

1. Generuj jedną realizację λ zmiennej losowej Λ.
2. Generuj $n \sim Pois(\lambda T)$.
3. Jeśli $n = 0$, to koniec.
4. Generuj $U_1, ..., U_n$ - iid, $U_i \sim U(0,T)$, $U_i \| \Lambda$.
5. Sortuj $U_1, ..., U_n$ aby uzyskać $(U_{1:n}, ..., U_{n:n})$.
6. Wstaw $S_i = U_{1:n}$, $i = 1, ..., n$.

Zadanie 31 (3 pkt) Przetestuj różne rozkłady zmiennej losowej Λ, np. rozkład wykładniczy. Wygeneruj 200 realizacji MPP, $(N_t^{(i)})_{t\in[0,T]}$, $i = 1,2,...200$ i narysuj histogramy dla $(N_{\frac{T}{2}}^{(1)}, N_{\frac{T}{2}}^{(2)}, \dots, N_{\frac{T}{2}}^{(200)})$ oraz $(N_T^{(1)}, N_T^{(2)}, \dots, N_T^{(200)})$. Co można zaobserwować? Czy umiesz odgadnąć rozkład zmiennej N_T?

Zadanie 32 (2 pkt) Niech Λ_i będą niezależnymi zmiennymi losowymi z rozkładu wykładniczego z parametrem $\lambda = 1$ i niech τ_i będą iid (niezależne także z Λ_i) z rozkładu $U(0,T)$. Niech $(N_t^{(i)})_{t\in[0,\tau_i]}$ będzie rodziną mieszanych procesów Poissona z intensywnościami Λ_i; $i = 1, ...300$. Narysuj histogram zmiennej losowej z rozkładu N_τ.

Zadanie 33 (4 pkt) Zaimplementuj symulator trajektorii dwuwymiarowego ruchu Browna. Narysuj przykładowe trajektorie na przedziale $[0,T]$ z rozdzielczością μ. Uwaga na skalę! Skorzystaj z faktu, że

$$B_{t_{i+1}} - B_{t_i} =_d \sqrt{t_{i+1} - t_i} Z_i,$$

gdzie Z_i są niezależne o rozkładzie $N(0,1)$.

Zadanie 34 (1 pkt. za każdy podpunkt) Niech W_t, $t > 0$ będzie jednowymiarowym procesem Wienera. Zbadaj rozkład zmiennej losowej

$$X_n = |\{i \in \mathbb{N} : W_{t_i} \cdot W_{t_{i+1}} < 0,\ i \leqslant n\}|, \tag{7}$$

gdzie

1. $t_i = \frac{i}{10}$
2. $t_i = \frac{1}{i}$
3. $t_i = 1 - \frac{1}{i}$

4. $t_i = e^i$

5. $t_i = |\sin(i/10)|$

6. $t_i = i!$.

Jaka musi być rozdzielczość procesu, aby uzyskać sensowne rozkłady X_n?

9 Tydzień IX. Proces Coxa. Model Blacka-Scholesa.

Podwójnie stochastyczny proces Poissona.

Metoda 1

1. Generuj $\Lambda(t)$ na [0,T].
2. Generuj $\tilde{N}(t)$ na $[0, \Lambda(t)]$
3. Wstaw $N(t) = \tilde{N}(\Lambda(t)),\ t \in [0, T]$.

Metoda 2

1. Generuj $\lambda(t)$ na [0,T].
2. Wyznacz $\lambda \geqslant \max \lambda(t), t \in [0, T]$
3. Wstaw $t = 0,\ I = 0$.
4. Generuj $U_1 \sim U(0,1)$
5. Wstaw $t = t - \frac{1}{\lambda} \log U_1$, jeśli $t > T$ to koniec
6. Generuj $U_2 \sim U(0,1)$
7. Jeśli $U_2 \leqslant \frac{\lambda(t)}{\lambda}$, to $I = I + 1$, $S_I = t$
8. Wróć do 4.

Uwaga!

$$\Lambda(t) = \int_0^t \lambda(s)ds. \tag{8}$$

Zadanie 35 (3 pkt., dodatkowe) Zaimplementuj Metodę 1.

Zadanie 36 (3 pkt.) Zaimplementuj Metodę 2.

Zadanie 37 (6 pkt., po 1.5 za każdy podpunkt) Przetestuj powyższe metody dla następujących Λ_t

- $\lambda_1(t) = \max\{0, \ln X_t\}$, $X_t \sim\ N(2, \frac{1}{t})$, $t = 1, 2, 3 \ldots, 100$
- $\lambda_2(t) = \exp(-X_t)$, $X_t \sim\ N(0, t^2)$, $t = 1, 2, 3 \ldots, 100$
- $\lambda_3(t) = X_t \cdot \exp(-Y_t)$, X_t jest zmienną z rozkładu wykładniczego z parametrem $\lambda = \sin^2(t)$, Y_t jest niezależna od X_t i pochodzi z rozkładu Cauchy'ego, $t = 1, 2, 3 \ldots, 100$
- $\lambda_4(t) = |B_t|$, gdzie B_t jest ruchem Browna

Zadanie 38 (8 pkt) Model Blacka-Scholesa. Rozważmy proces stochastyczny (geometryczny ruch Browna)

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}, \tag{9}$$

gdzie W_t jest standardowym ruchem Browna, $\mu = 1$ to trend, $S_0 = 5$ to cena początkowa, a $\sigma = 0.8$ to zmienność (*volatility*). GRB stosowany jest do modelowania zmienności cen akcji. Oszacuj prawdopodobieństwo p, że w ciągu dwóch lat (za jednostkę czasu przyjmujemy 1 dzień) cena akcji spadnie poniżej $\frac{S_0}{n}$, $n = 5$. Następnie narysuj wykresy funkcji:

- $p(\mu)$
- $p(\sigma)$
- $p(n)$.

10 Tydzień X. Proces ryzyka.

Zadanie 39. Proces ryzyka (5 pkt.) Kapitał firmy ubezpieczeniowej jest więc dany wzorem:

$$R(t) = r_0 + p(t) - \sum_{i=1}^{N_t} X_i, \tag{10}$$

Gdzie N_t jest jednorodnym procesem Poissona o intensywności λ, X_i to wielkości szkód (np. z rozkładu gamma, Pareto, Weibulla), r_0 to kapitał początkowy, a $p(t)$ to funkcja zysków. Stwórz symulację takiego procesu, która zatrzymuje się przy uderzeniu w 0, z $p(t) = ct$ i $X_i \sim Exp(\kappa)$.

Zadanie 40 Zbadaj rozkład (histogramy, statystyki opisowe) czasu ruiny w zależności od r_0, c, κ, λ (po 1,5 pkt. za każdy parametr; weź pod uwagę różne poziomy pozostałych parametrów).

Zadanie 41. Ruina Paryska (2 pkt.) Uogólnij symulację z Zadania 39, aby firma nie bankrutowała automatycznie po uderzeniu w 0. Firma będzie teraz bankrutować po T jednostkach czasu przebywania poniżej 0 (po wyjściu ponad poziom 0 zegar restartujemy).

Zadanie 42 (2 pkt.) Dla takich parametrów r_0, c, κ, λ, dla których firma w klasycznym przypadku bankrutuje „często” na przedziale czasu od 10 do 100, zbadaj średni czas bankructwa paryskiego w zależności od T z Zadania 41.

Zadanie 43. Zabawa (5 pkt.)

$$R(t) = r_0 + p_n(t) - \sum_{i=1}^{N_t} X_i, \tag{11}$$

Zakładamy, że firma ubezpieczeniowa ma liniowe przychody ($p_0(t) = ct$, $c = 1/3$); wypłaty odszkodowań obserwujemy jako niejednorodny proces Poissona N_t, przy czym intensywność $\lambda(t)$ w dzień powszedni zwykły wynosi 0.8, w dzień powszedni w sezonie wakacyjnym 1.0, w weekendy zaś i święta 1.3. Wielkość wypłat pochodzi z rozkładu wykładniczego $\mathcal{E}(3)$. Kapitał początkowy firmy to $r_0 = 3$. Jeden dzień trwa jedną jednostkę czasu.

Jeśli trafimy w 0:

1. rozpoczynamy odliczanie paryskiego zegara: jeśli będziemy na minusie dłużej niż 4 dni, przegrywamy.
2. nowe $p_n(x) = 0.9 \cdot p_{n-1}(x)$ (zaciągamy kredyt, nasze dochody spadają)
3. jeśli spadniemy poniżej -10, to automatycznie przegrywamy

Niech k będzie liczbą dotychczasowych inwestycji (na początku $k = 1$). Jeśli majątek firmy wzrośnie powyżej $20k$, to:

1. $k++$
2. $p_n(t) = p_{n-1}(x) \cdot 1.05$ (inwestujemy, więc dochody wzrastają)
3. Inwestycje to koszty: od obecnego majątku odejmujemy $10k$.

Oszacuj prawdopodobieństwo ruiny w ciągu 5 lat.

Zadanie 44, dodatkowe 3 pkt. Dobierz stałą c przy $p_0(t)$ z poprzedniego zadania tak, aby prawdopodobieństwo ruiny wynosiło ok. 0.1.

11 Tydzień XI. Proces Odnowy.

Proces liczący $N(t)$ nazywamy procesem odnowy, jeżeli czasy oczekiwania na kolejny skok $\{T_i\}$ są niezależnymi dodatnimi zmiennymi losowymi o jednakowym rozkładzie. Oznaczmy $S_n = \sum_{i=1}^{n} T_i$. Wtedy $N(t) = \max\{n \in \mathbb{N} : S_n \leqslant t\}$.

Zadanie 45. Proces Odnowy (3 pkt.) Napisz funkcję generującą proces odnowy, która jako parametry będzie przyjmować T (chwila, do której chcemy generować proces), f - funkcję generującą zmienne losowe będące czasami oczekiwania na kolejny skok, *par* - listę parametrów dla f, N - liczbę trajektorii do wygenerowania.

Zadanie 46 (2 pkt.) Narysuj kilka przykładowych trajektorii procesu odnowy dla dwóch różnych f.

Zadanie 47 (2 pkt.) Niech $M_t = N_t - U_t$, gdzie U_t jest procesem odnowy, w którym czas oczekiwania pochodzi z rozkładu $U(2,4)$, natomiast N_t jest jednorodnym procesem Poissona o intensywności $\lambda = 3$. Jaka jest wartość oczekiwana $\mathbb{E}M_t$ w dowolnej chwili $t \geqslant 0$? Wygeneruj i nanieś na wykres kilka przykładowych trajektorii procesu M_t.

Zadanie 48 (2 pkt.) Oszacuj prawdopodobieństwo $p = P(\max_{t\in[0,100]} M_t > 10)$. Narysuj histogram rozkładu zmiennej losowej M_{100}.

Zadanie 49 (dodatkowe, 2 pkt.) Powtórz powyższe zadanie w przypadku, gdy N_t jest niejednorodnym procesem Poissona z funkcją intensywności $\lambda(t) = 3 + \sin(t/\lambda_0)$, $\lambda_0 > 0$. Dla jakich λ_0 prawdopodobieństwo p będzie największe w rozważanym przypadku?

Zadanie 50, odbity ruch Browna (2 pkt.) Zaimplementuj funkcję do symulowania jednowymiarowego procesu Wienera odbitego w $a \in \mathbb{R}$.

Zadanie 51 (3 pkt.) Niech $\{T_i\}$ będą iid czasami oczekiwania na kolejny skok, a $\{X_i\}$ ciągiem iid niezależnym od $\{T_i\}$. Zakładamy, że $\mathbb{E}T_1 < \infty$. Zbadaj symulacyjnie dla dwóch wybranych par rozkładów, że

$$\frac{\sum_{i=1}^{N(t)} X_i}{t} \longrightarrow \frac{\mathbb{E}X_1}{\mathbb{E}T_1} \tag{12}$$

prawie na pewno przy $t \to \infty$.

Zadanie 52 (dodatkowe 6 pkt.) Niech X_t, Y_t będą niezależnymi procesami Wienera. Nioech $Z_y = |X_t| \cdot e^{iY_t}$. Narysuj kilka trajektorii Z_t i porównaj rozkład czasu trafienia Z_t w barierę $A = \{(x,5), x \in \mathbb{R}\}$ z rozkładem czasu trafienia dwuwymiarowego ruchu Browna w A.

12 Niedogarki. Łańcuchy Markowa. Miara harmoniczna (kind of)

Niedogarki należy oddać do końca semestru.

Zadanie 39 Niech B_t będzie ruchem Browna w $\mathbb{R}^2$ o jednostkowej macierzy kowariancji. Badamy wartość procesu B_t w momencie wyjścia ze zbioru D, τ_D, czyli B_{τ_D}. W przypadku, gdy D jest kulą jednostkową, $D = B((0,0),1)$, rozkład zmiennej losowej B_{τ_D} jest rozkładem jednostajnym na sferze $S((0,0),1)$.

- Sprawdź to eksperymentalnie! Narysuj histogram punktów wyjścia z $B((0,0),1)$, parametryzując je kątem $\phi \in [0,2\pi)$ (5 pkt.)
- Powtórz powyższe zadanie w przypadku, gdy D jest kwadratem $[-1,1]\times[-1,1]$ oraz trójkątem równobocznym o środku w $(0,0)$. Ze względu na niezmienniczość ruchu Browna na obroty, histogram można narysować w jednym wymiarze. (dodatkowe, 8 pkt.)

Zadanie 40. Błądzenie losowe na grafie. Czas pokrycia.

13 Dodatkowe zadania uzupełniające

Na podstawie m.in.: Ross, Simulation. Do oddania najpóźniej na przedostatnich zajęciach.

Sumy zmiennych jednostajnych (1pkt.) Niech $U_1,\dots,U_n$ będą niezależnymi zmiennymi losowymi z rozkładu jednostajnego $\mathcal{U}(0,1)$. Niech $N = \min\{n \in \mathbb{N} : \sum_{i=1}^{n} U_i \geqslant 1$. Zbadaj $\mathbb{E}N$ poprzez wygenerowanie 100, 1000, 10000 zmiennych N. Jaką możesz postawić hipotezę?

Iloczyny zmiennych jednostajnych (1 pkt.) Niech $M = \max\{n : \prod_{i=1}^{n} U_i \geqslant e^{-3}\}$. Wyznacz symulacyjnie $\mathbb{E}M$ oraz $P(M = i)$ dla $i = 1,2,3,4,5,6,7$.

Korelacje, 1 pkt. Wyznacz symulacyjnie wielkości $Corr(U, \sqrt{1-U^2})$ oraz $Corr(U^2, \sqrt{1=U^2})$.

Permutacje losowe (3 pkt.) Napisz funkcję $shuffle(n)$, która zwraca losową permutację liczb $1,\dots,n$. Każda permutacja powinna występować z tym samym prawdopodobieństwem.

Warunkowa wartość oczekiwana (1 pkt.) Niech X będzie zmienną z rozkładu wykładniczego z parametrem $\lambda = 1$. Oszacuj $\mathbb{E}(X|X < 0.1)$.

Nierówność z artykułu Pruitta (3 pkt.) Niech Y będzie zmienną losową z dowolnego rozkładu o wartościach w $\mathbb{R}$ i niech $s > 0$. Uzasadnij (dla kilku wybranych rozkładów) symulacyjnie nierówność

$$P(|Y| \leqslant s) \leqslant \frac{s}{1-\cos 1} \int_0^{\frac{1}{s}} |\mathbb{E}e^{iwY}| dw, \tag{13}$$

rysując wykres lewej i prawej strony nierówności ww zależności od $s > 0$.

Kwadratura koła, 6 pkt. Na obwodzie kwadratu $[-1,1] \times [-1,1]$ rozmieszczone są równomiernie punkty $S_1,\dots,S_n$. Rozważmy n niezależnych ruchów Browna z dodaną siłą przyciągania („potencjałem”), $X_1,\dots X_n$, takich że X_i startuje z S_i oraz X_i jest ruchem Browna na prostej wyznaczonej przez wektor $\vec{0S_i}$, natomiast „potencjał” wynosi $\vec{0S_i}$ dla $|X_i| < |S_i|$ oraz $-\vec{0S_i}$ w przeciwnym wypadku. Wykonaj animację powyższego procesu. Oszacuj średni czas $\tau(\varepsilon)$ taki, że $|X_1(\tau)| \in [|S_1|-\varepsilon, |S_1|+\varepsilon], \dots, |X_n(\tau)| \in [|S_1|-\varepsilon, |S_1|+\varepsilon]$, czyli kiedy „kwadrat zamieni się w koło”.

Subordynacja, 4 pkt. Rozważmy dwuwymiarowy ruch Browna (Y_t, Z_t) w $\mathbb{R}^2$. Niech $\tau_t = \inf\{s : Y_s \geqslant t\}$. Narysować trajektorie procesu $X_t = Z_{\tau_t}$. Czy jest to proces Cauchy'ego?

ASEP, 5 pkt. Rozważmy graf o n wierzchołkach $V_1, \ldots V_n$ taki, że V_i jest połączony z V_{i-1} oraz V_{i+1}. W V_1 pojawiają się cząstki niezależnie z intensywnością α i zajmuje ten wierzchołek. Teraz intensywność skoku z wierzchołka V_i do V_{i+1} (o ile wierzchołek ów nie jest zajęty) wynosi 1, natomiast do V_{i-1} (o ile nie jest zajęty) wynosi $q \in [0,1)$. Prawdopodobieństwo opuszczenia układu przez cząstkę znajdującą się w V_n wynosi β. Proces można symulować jako losową liczbę zależnych cząstek błądzących po grafie albo jako proces Markowa o 2^n stanach. Stwórz program do symulowania takiego procesu. Zbadaj przypadki, gdy:

- $\alpha > 1-q$, $\beta > 1-q$
- $\alpha + \beta < 1-q$
- któryś z pozostałych (znak $=$ ma znaczenie)

Metoda generowania zmiennych normalnych metodą przybliżoną - badanie, 5 pkt. Napisz generator zmiennych losowych o rozkładzie normalnym, używając:

- Sumy N zmiennych losowych o rozkładzie jednostajnym
- Rozkładu dwumianowego $B(n,p)$
- Rozkładu Poissona z parametrem λ. Zbadaj, jak zmienia się szybkość generowania zmiennych losowych w zależności od N, n, p, λ oraz jak dobre przybliżenia dostajemy. Porównaj te generatory z wcześniej zaimplementowanymi generatorami zmiennych losowych z rozkładu normalnego.