Symulacje stochastyczne i teoria Monte Carlo

Tomasz Rolski

Instytut Matematyczny, Uniwersytet Wrocławski

Wrocław, 2017

Wersja v2.4.

Spis treści

Rozdział I

Wstęp

[1]

1 O skrypcie

Stochastyczne symulacje i *teoria Monte Carlo* są bardzo często są nazwami tej samej teorii. Jednakże w opinii autora tego skryptu warto je rozróżnić. Pierwszy termin będzie więc dotyczył teorii generatorów oraz metod generowania liczb losowych o zadanych rozkładach. Właściwie powininno się mówić o liczbach pseudo-losowych, bo tylko takie mogą być generowane na komputerze, ale w tekście często będziemy mówili o zmiennych lub liczbach losowych. Natomiast przez teorię Monte Carlo będziemy tutaj rozumieć teoretycznymi podstawami opracowania wyników, planowania symulacji, konstruowaniu metod pozwalających na rozwiązywanie konkretnych zadań, itp.

W przeciwieństwie do metod numerycznych, gdzie wypracowane algorytmy pozwalają kontrolować deterministycznie błędy, w przypadku obliczeń za pomocą metod stochastycznych dostajemy wynik losowy, i jest ważne aby zrozumieć jak taki wynik należy interpretować, co rozumiemy pod pojęciem błędu, itd. Do tego przydają się pojęcia i metody statystyki matematycznej.

Skrypt jest pisany z myślą o studentach matematyki Uniwersytetu Wrocławskiego, w szczególności studentach interesujących się informatyką oraz zastosowaniami rachunku prawdopodobieństwa. Dlatego, ideą przewodnią tego skryptu, jest przeprowadzanie symulacji rozmaitych zadań rachunku prawdopodobieństwa,

[1] Poprawiane 5.10.2017

przez co czytelnik poznaje przy okazji obszary, które trudno poznać bez specjalistycznej wiedzy z rachunku prawdopodobieństwa i procesów stochastycznych. Dlatego też nie zakłada się od czytelnika szczegółowej wiedzy z tych przedmiotów, a tylko zaliczenie podstawowego kursu uniwersyteckiego z rachunku prawdopodobieństwa i statystyki matematycznej. Natomiast po ukończeniu kursu, oprócz znajomości symulacji stochastycznej i teorii Monte Carlo, autor skryptu ma nadzieję, że czytelnik będzie zaznajomiony z wieloma klasycznymi zadaniami teorii prawdopodobieństwa, modelami stochastycznymi w badaniach operacyjnych i telekomunikacji, itd.

W dzisiejszej literaturze rozróżnia się generatory liczb losowych od generatorów liczb pseudoloswych. W tym wykładzie zajmiemy się liczbami pseudolosowymi, to jest otrzymanymi za pomocą pewnej rekurencji matematycznej. Natomiast ciągi liczb losowych możemy otrzymać na przykład przez rzuty monetą, rzuty kostką, obserwacji pewnych zjawisk atmosferycznych, zapisu rejestrów itp. Rozróżnia się hardwerowe generatory licz" losowych od sofwerowych genratorów liczb losowych. Jednakże tutaj nie będziemy sie tym zajmowali. Oprócz liczb pseudolosowych mamy też liczby quasi-losowe. Są to ciągi liczbowe które mają które spełniają wląsnośc 'low discrepacy".

Skrypt składa się z rozdziałów dotyczących:

- zarysu teorii generatorów liczb pseudolosowych,
- sposobów generowania zmiennych losowych o zadanych rozkładach,
- podstawowych pojęć dotyczących błedów, poziomu istotności, liczby replikacji i ich związków,
- przykładowych zadań rozwiązywanych metodami symulacji stochastycznej,
- metodami zmniejszenia liczby replikacji przy zadanym poziomie błędu i poziomie istotności.

Istotnym składnikiem tego skryptu jest dodatek, w którym znajdują się pojęcia z rachunku prawdodobieństwa, przegląd rozkładów i ich własności, potrzebne wiadomości ze statystyki matematyczne, zarys teorii kolejek, itd. Niektóre fragmenty znajdujące się w dodatku muszą zostać przerobione na wykładzie, bowiem trudno zakładać ich znajomość, a bez których wykład byłby niepełny.

Rozdział II

Teoria generatorów

1 Wstęp

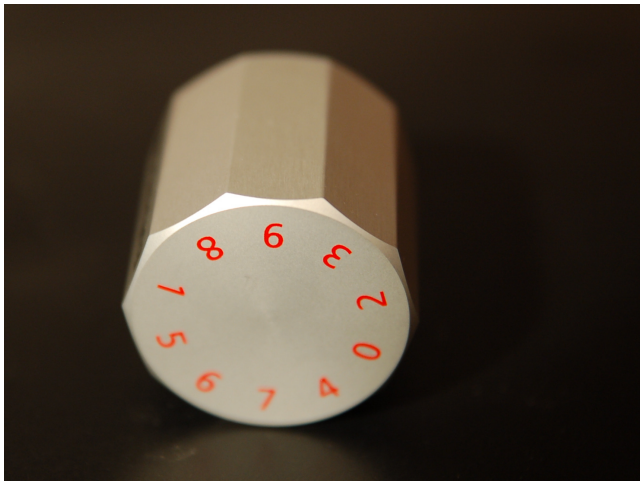
Podstawą symulacji stochastycznej jest pojęcie *liczby losowej*. Nie będziemy tutaj wdawać się szczegółowo w filozoficzne dysputy znaczenia słowa "losowy". Zacytujmy tylko co na ten temat pisze Knuth w swoim fundamentalnym dziele [23], str 2: "W pewnym sensie nie ma czegoś takiego jak liczba losowa; czy na przykład 2 jest liczbą losową. Można natomiast mówić o *ciągu niezależnych zmiennych losowych* o określonym *rozkładzie*, a to oznacza, z grubsza rzecz biorąc, że każda z liczb została wybrana zupełnie przypadkowo, bez żadnego związku z wyborem pozostałych liczb w ciągu, i że każda liczba mieści sie w dowolnym ustalonym zakresie wartości z określonym prawdopodobieństwem." W skrypcie terminy liczba losowa i zmienna losowa są używane wymiennie.

Podstawowym rozkładem w tej teorii jest rozkład jednostajny U(0,1). Jednakże na komputerze nie możemy otrzymać wszystkich liczb z odcinka $[0, 1]$. Dlatego tak naprawdę generujemy liczby przyjmujące wartości

$$0, 1/M, 2/M, \ldots, (M-1)/M$$

dla pewnego $M = 2^k$ i będziemy oczekiwać, że ona ma rozkład jednostajny na tym zbiorze. W tym wypadku będziemy mówili często o liczbie losowej bez wymieniania nazwy rozkładu , tj. zmiennej losowej U o rozkładzie jednostajnym U(0,1) na (0,1). Oznaczenia rozkładów, wraz definicjami, ich podstawowymi charakterystykami są podane w dodatku XI.1.1 oraz tablicy XI.4.

Przed nastaniem ery komputerowej, tworzono tablice liczb losowych. Najprostszą tablicę można zrobić samemu, wyciągając na przykład kule z urny, rzucając kostką, itd, następnie te wyniki zapisując. Można też używać "kostki" dającej wyniki od 0,1 do 9; patrz Rys. 1.1 i przez kolejne rzuty produkować



Rysunek 1.1: Kostka z Eurandom

ciąg *cyfr losowych*. Następnie grupując względem ustalonej liczby cyfr, poprzedzając je zerem i kropką [1] otrzymać żądane tablice liczb losowych. Takie tablice mają często "dobre własności" (co pod tym rozumiemy opowiemy później) ale gdy potrzebujemy bardzo dużo liczb losowych (na przykład setki tysięcy) opisana metoda jest bezużyteczna.

Dzisiaj do tworzenia ciągu liczb losowych (lub jak się mówi do ich generowania) stosuje się komputery. Chyba pierwszym, który to zrobił był John von Neumann około 1946 r. Zaproponował on aby tworzyć następną liczbę podnosząc do kwadratu poprzednią i wycięciu środkowych. "Metoda środka kwadratu" okazała się niezbyt dobra; bowiem ma tendencje do wpadania w koleinę – krótki cykl powtarzających się elementów. Na przykład ciąg dwucyfrowych liczb zaczynających się od 43 według tego algorytmu jest $84, 05, 00, 00, \ldots$. Mianowicie mamy $43^2 = 1849$, $84^2 = 7056$ i $05^2 = 25$. Powstała cała obszerna teoria dotycząca generatorów, tj. algorytmów produkujących ciągi "liczb losowych", analizy dobroci generatorów. Teoria ta wykorzystuje osiągnięcia współczesnej algebry i statystyki.

[1] W skrypcie używamy konwencję 0.1234 na zapis liczby z przedziału (0,1) a nie 0,1234.

Niestety komputer nie jest w stanie wygenerować idealnego ciągu niezależnych zmiennych losowych o rozkładzie jednostajnym. Dlatego czasem mówi się o liczbie pseudo-losowej generowanej przez komputer. Rozumiejąc te ograniczenia będziemy przeważnie dalej pisać o ciągu zmiennych losowych o rozkładzie U(0,1).

Ogólny schemat generowania liczb losowych na komputerze można scharakteryzować następująco przez piątkę: (S, s_0, f, U, g), gdzie

- S jest skończoną przestrzenią stanów,
- $s_0 \in S$ jest wartościa początkową rekurencji (ziarno),
- $s_{i+1} = f(s_i)$, gdzie $f : S \to S$,
- natomiast U skończoną przestrzenią wartości oraz
- $g : S \to U$, i $u_i = g(x_i)$ $(i = 0, 1, \ldots$ jest wygneroanym ciągiem liczb pseudolosowych.

Wtedy mówimy, że (S, s_0, f, U, g) jest generatorem liczb losowych (GLL).

Łatwo zauważyć, że ciąg generowany przez $X_{n+1} = f(X_n)$ zawsze wpada w pętlę: w końcu pojawia się cykl, który nieustannie się powtarza. Długość takiego cyklu nazywa sie *okresem*. Dobroć generatora zależy od okresu. Zależy nam na bardzo długich okresach. Zauważmy, że teoretycznie cykl nie może być dłuższy niż moc S.

1.1 Formalna definicja ciagu liczb losowych; liczby pseudolosowe

Przez ciąg liczb losowych rozumiemy ciąg U_j $(j = 1, \ldots)$ zmiennych losowych $(U_i : (\Omega, \mathcal{F}, \mathbb{P}) \to [0, 1])$ o jednakowycm rozkładzie jednostajnym $\mathcal{U}[0, 1]$, tzn.

(i) U_j ma rozkład jednostajny, tzn.

$$\mathbb{P}(U_j \leq t) = \begin{cases} 0 & t < 0, \\ t & 0 \leq t < 1 \\ 1 & t \geq 1, \end{cases}$$

(ii) $U_1, U_2, \ldots$ są niezależne o jednakowym rozkładzie, tj. dla $n = 1, 2, \ldots$

$$\mathbb{P}(U_1 \leq t_1, \ldots, U_n \leq t_n) = t_1 t_2 \cdots t_n,$$

dla $0 \leq t_1, \ldots, t_n \leq 1$.

Przez *losowy ciąg bitów* rozumiemy ciąg zmiennych losowych $\xi_1, \xi_2, \ldots$ $(\xi_j : (\Omega, \mathcal{F}, \mathbb{P}) \to \{0, 1\})$, taki, że

(i) $\mathbb{P}(\xi_j = 0) = \mathbb{P}(\xi_j = 1) = 0.5$,

(ii) zmienne losowe $\xi_1, \xi_2, \ldots$ są niezależne o jednakowym rozkładzie.

Zauważmy ważna i nietrywialną własność takich ciągow liczb (bitów) losowych. Jeśli $n_1 < n_2 < \ldots$ jest podciągiem liczb naturalnych, to ciąg $U_{n_1}, U_{n_2}, \ldots$ jest ciągiem lczb losowych. Podobnie jest dla losowego ciągu bitów.

Przedmiotem dalszych rozważań w tym wykładzie będą ciągi liczb (bitów) pseudo-losowych, tj. takich które imitują ciągi liczb (bitów) losowych. Takie ciągi można otrzymywać na różne sposoby, ale my zajmiemy się generowanymi przez odpowiednie algorytmy na komputerach. Mówi sie o generatorach liczb pseudolosowych (w języku angielskim *pseudo-random number generator* (PRNG). Ponieważ te dwa pojęcia sie zlewają, w dalszej część wykładu będziemy opuszczali słowo *pseudo*. W literaturze często pisze sie po prosu *random number generator* (RNG), który oczywiście musi być generatorem liczb pseudolosowych. Człowiek nie mjest bowiem w stanie generować idealny ciąg liczb które są realizacjami ciągu niezależnych zmiennych losowych o jednakowym rozkładzie. Kontynując rozważania terminologicze, zauważmy, że jesli jest mowa o R *replikcajach* liczby losowej U, to jeśli nie będzie powiedziane co innego, mamy ciąg $U_1, \ldots, U_R$ zmiennych losowych o jednakowym rozkładzie jednostajnym.

1.2 Rozkład jednostajny na $\{0, \ldots, M-1\}$

W niektórych zadaniach potrzebny jest ciąg liczb losowych $Y_1, Y_2, \ldots$ przyjmujących wartości od $0, \ldots, M-1$. Ciąg tem nazywamy losowy, jesli jest realizacją z próby niezależnych zmiennych losowych o jednakowym rozkładzie jednostajnym na $\bar{M} = \{0, \ldots, M-1\}$, tj.

$$\mathbb{P}(Y_j = i) = \frac{1}{M} \qquad i = 0, \ldots, M-1 \, .$$

Mając do dyspozycji generator, na przykład MATLABowski

```
rand
```

możemy takie liczby w prosty sposób otrzymać następująco: $Y_i = \lfloor MU_i \rfloor$. Jeśli więc chcemy otrzymać ciąg 100 elementowy z $M = 10$ to piszemy skrypt

```
%Generate uniform integers on the set 0:M-1:
n = 100; M=10;
Y = floor(M.*rand(n,1));
```

Polecamy zastanowić się, jak zmienić powyższy skrypt aby otrzymać cyfry losowe równomiernie rozłożone na $1, \ldots, M$.

2 Losowe permutacje

Niech $\mathcal{S}_n$ bedzie zbiorem wszystkich permutacji zbioru $[n] = \{1, \ldots, n\}$. Przez losową permutację rozumiemy taką operację która generuje każdą permutację z $\mathcal{S}_n$ z prawdopdobieństwem $1/n!$.

2.1 Algorytm na losową permutację

Podamy teraz algorytm na generowanie losowej permutacji liczb $1, \ldots, n$. A więc celem jest wygenerowanie 'losowej' permutacji tych liczb, tj. takiej, że każdy wynik można otrzymać z prawdopodobieństwem $1/n!$. Zakładamy, że mamy do dyspozycji generator liczb losowych $U_1, U_2, \ldots$, tj. ciąg niezależnych zmiennych losowych o jednakowym rozkładzie jednostajnym $\mathrm{U}(0, 1)$. Podajemy teraz algorytm na generowanie losowej permutacji ciągu $1, \ldots, n$

Losowa permutacja

```
1. podstaw  t = n oraz A[i] = i dla i = 1, . . . , n;
2. generuj liczbę losową u pomiędzy  0 i 1;
3. podstaw  k = 1 + \lfloor tu\rfloor  ; zamień  A[k] z A[t];
4. podstaw  t = t - 1;
   jeśli t > 1, to powrót do kroku 2;
   w przeciwnym razie stop i A[1], . . . , A[n ]
   podaje losową  permutację.
```

Złożoność tego algorytmu jest $\mathrm{O}(n)$ (dlaczego ?). Algorytm jest konsekwencją zadania 10.2.

W MATLAB-ie generujemy losową permutację za pomocą instrukcji

```
randperm
```

3 Generatory liczb losowych

3.1 Metody kongruencji liniowej

Najprostszym przykładem generatora liczb pseudo-losowych jest ciąg u_n zadany przez x_n/M, gdzie

$$x_{n+1} = (ax_n + c) \mod M. \tag{3.1}$$

Należy wybrać

M,	moduł;	$0 < M$
a,	mnożnik;	$0 \leq a < M$
c,	krok	$0 \leq c < M$
x_0,	wartość początkowa;	$0 \leq x_0 < M$.

Aby otrzymać ciąg z przedziału $(0, 1)$ musimy teraz wybrać $u_n = x_n/M$. Ciąg (x_n) ma okres nie dłuższy niż M.

Następujace twierdzenie (patrz twierdzenie A na stronie 17 swojej książki Knuth [23]) podaje warunki na to aby kongruencja liniowa miała okres M.

Twierdzenie 3.1 *Niech $b = a - 1$. Na to aby ciąg (x_n) zadany przez (3.1) mial okres równy M potrzeba i wystarcza aby*

- *c jest względnie pierwsze z M,*
- *b jest wielokrotnością p dla każdej liczby pierwszej p dzielącej M,*
- *jeśli 4 jest dzielnikiem M, to $a = 1 \mod 4$.*

Podamy teraz przykłady generatorów liczb pseudolosowych otrzymanych za pomocą metody konguencji liniowej.

Przykład 3.2 Można przyjąć $M = 2^{31} - 1 = 2147483647$, $a = 7^5 = 16867$, $c = 0$. Ten wybór był popularny na 32-bitowych komputerach. Innym przykładem proponowanego ciągu o niezłych własnościach jest wybór $M = 2147483563$ oraz $a = 40014$. Dla 36-bitowych komputerów wybór $M = 2^{35} - 31$, $a = 5^5$ okazuje się dobrym. [2]

[2] Podajemy za Asmussenem i Glynem oraz Rossem (str.37)

Przykład 3.3 [Resing [37]]. Jesli $(a, c, M) = (1, 5, 13)$ i $X_0 = 1$, to mamy ciąg (X_n): 1,6,11,3,8,0,5,10,2,7,12,4,9,1,..., który ma okres 13.
Jesli $(a, c, M) = (2, 5, 13)$, oraz $X_0 = 1$, to otrzymamy ciąg
$1, 7, 6, 4, 0, 5, 2, 9, 10, 12, 3, 11, 1, \ldots$ który ma okres 12. Jesli natomiast $X_0 = 8$, to wszytkie elementy tego ciągu są 8 i ciąg ma okres 1.

Uogólniona metoda kongruencji liniowej generuje ciągi według rekurencji

$$x_{n+1} = (a_1 x_{n-1} + a_2 x_{n-2} + \ldots + a_k x_{n-k}) \mod M$$

Jeśli M jest liczbą pierwszą, to odpowiedni wybór stałych a_j pozwala na otrzymanie okresu M^k. W praktyce wybiera się $M = 2^{31} - 1$ na komputerach 32 bitowych.

Za L'Ecuyer [29] podamy teraz kilka popularnych generatorów używanych w komercyjnych pakietach.

Przykład 3.4

Java

$$\begin{aligned} x_{i+1} &= (25214903917 x_i + 11) \mod 2^{48} \\ u_i &= (2^{27} \lfloor x_{2i}/2^{22} \rfloor + \lfloor x_{2i+1}/2^{21} \rfloor)/2^{53} . \end{aligned}$$

VB

$$\begin{aligned} x_i &= (1140671485 x_{i-1} + 12820163) \mod 2^{24} \\ u_i &= x_i/2^{24} . \end{aligned}$$

Excel

$$u_i = (0.9821 U_{i-1} + 0.211327) \mod 1 .$$

Jeśli $c = 0$, to generator nazywa się liniowy multiplikatywny kongruencyjny. W Lewis et al [3] zaproponowano następujące parametry takiego generatora: $M = 2^{31} - 1$, $a = 1680$.

[3] Lewis, P.A.P., Goodman, A.S. and Miller, J.M. (1969)

Spotyka sie też generatory bazujące na kongruencjach wyższych rzędów, np. kwadratowych czy sześciennych. Przykłady takich generatorów można znaleźć w w Rukhin et al 2010.[4]

Generator kwadratowej kongruencji I (QCG I). Ciągi zadan są rekurencją

$$x_{i+1} = x_i^2 \mod p$$

gdzie p jest 512 bitową liczbą pierwszą zadaną w układzie szenastkowym

$p = 987b6a6bf2c56a97291c445409920032499f9ee7ad128301b5d0254aa1a9633$
$fdbd378d40149f1e23a13849f3d45992f5c4c6b7104099bc301f6005f9d8115e1$

Generator kwadratowej kongruencji II (QCG II). Ciąg jest zadany rekurencją

$$x_{i+1} = 2x_i^2 + 3x_i + 1 \mod 2^{512}.$$

Generator sześciennej kongruencji (CCG). Ciąg jest zadany rekurencją

$$x_{i+1} = 2x_i^3 \mod 2^{512}.$$

Generator Fibonacciego. Generator Fibonnaciego zadany jest rekurencją

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

Modyfikacją dającą lepsze własności niezależności jest opóźniony generator Fibonnaciego (lagged Fibonnaci generator) zadany rekurencją

$$x_n = x_{n-k} + x_{n-l} \ mod\, M \ (n \geq \max(k, l)).$$

Inną modyfikacją jest *substractive lagged Fibonnaci generator* zadany rekurencją

$$x_n = x_{n-k} - x_{n-l} \ mod\, M \ (n \geq \max(k, l)).$$

Knuth zaproponował ten generator z $k = 100$, $l = 37$ oraz $M = 2^{30}$.

[4]Rukhin, A., Soto, J. Nechvatal, J, Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J. and Vo, S. 2010, A statitical test suite fro random number generators for cryptographic applications.National Institute of Standard and Technology, Special Publication 800-22, Rev 1a.

3.2 Generatory w MATLABie

W MATLABie[5] są zaimplementowane 3 generatory: 'state', 'seed' i 'twister'. Jeśli nic nie zostanie zadeklarowane to jest w użyciu 'state'. W przeciwnym razie musimy zadeklarować wpisując komendę rand(method,s), gdzie

- w 'method' należy wpisać jedną z metod 'state', 'seed' lub 'twister',
- wartość 's' zależy od zadeklarowanej metody:

  w przypadku 'state' or 'twister', s musi być albo liczbą całkowitą od 0 do $2^{32}-1$ lub wartościa wyjściową rand(method),

  w przypadku 'seed' wielkość s musi być albo liczbą całkowitą od 0 do $2^{31}-1$ lub wartościa wyjściową rand(method).

Przykład 3.5 Proponujemy na początek następujący eksperyment w MATLABie. Po nowym wywołaniu MATLAB napisać rozkaz

```
rand
```

a następnie

```
rand('state',0).
```

W obu przypadkach dostaniemy taką samą liczbę 0.9501. Natomiast

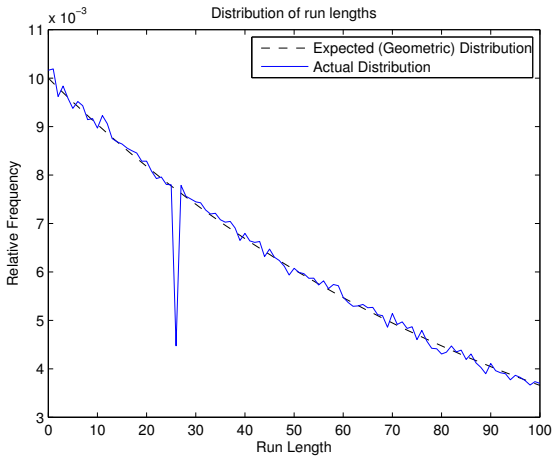
```
rand('twister',0)
```

Przykład 3.6 Dla generatora 'state' znaleziono przykłady na odchylenie od losowości, o których chcemy teraz opowiedzieć. Należy jednak zdawać sobie sprawę, że podobne przykłady można znaleźć dla wszystkich generatorów. Rekomendujemy czytelnikowi zrobienie następującego eksperymentu, w którym dla kolejnych liczb losowych $U_1, \ldots, U_n$ będziemy rejestrować odstępy $X_1, X_2, \ldots$ pomiędzy liczbami mniejszymi od δ. Oczywiście z definicji takie odstępy mają rozkład geometryczny z funkcją prawdopodobieństwa $\delta(1-\delta)^k$, $k = 0, 1, \ldots$. Ponadto kolejne odstępy $X_1, X_2, \ldots$ są niezależne o jednakowym rozkładzie geometrycznym. Jeśliby więc zrobić histogram pojawiania się poszczególnych długości to powinien dla dużej liczby wygenerowanych liczb losowych przypominać funkcję $\delta(1-\delta)^k$. Polecamy więc czytelnikowi przeprowadzenie następujących eksperymentów w MATLABIE; najpierw dla metody 'state' a potem 'twister', z róznymi n, δ według następującego algorytmu:

[5]MATLAB Version 7.2.0.294 (R2006a).

```
rand('state',0);
n = 5*10^7;
delta = .01;
runs = diff(find(rand(n,1)<delta))-1;
y = histc(runs, 0:100) ./ length(runs);
plot(0:100,(1-delta).^(0:100).*delta,'k--', 0:100,y,'b-');
title('Distribution of run lengths')
xlabel('Run Length'); ylabel('Relative Frequency');
legend({'Expected (Geometric) Distribution' 'Actual Distribution'})
```

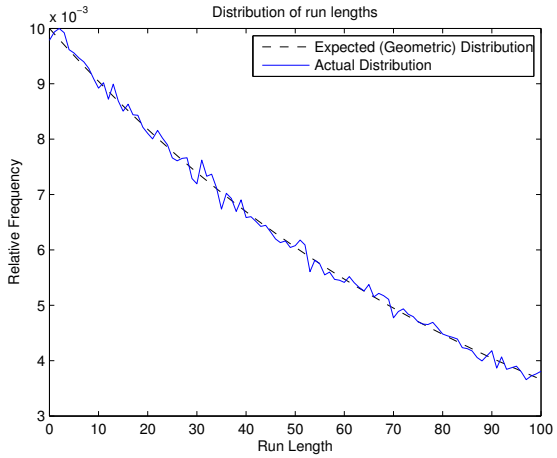
Dla pierwszej z tych metod możemy zaobserwować tajemniczą odchyłkę przy $k = 27$; patrz rys. 3.2. [6]



Rysunek 3.2: Histogram pojawiania się małych wielkości; generator 'state', $n = 5 * 10^7$, $\delta = 0.01$

Jest to spowodowane tym, ze generator niepoprawnie generuje małe wartości. Można się o tym przekonać eksperymentując z δ. Tak naprawdę, dobrze dopasowaną do zadanego $n = 5 * 10^7$ jest $\delta = 0.05$ jednakże wtedy rysunki się zlewają. Natomiast dla bardzo małych δ należałoby zwiększyć n. Wyjaśnieniem jak dobierać n zajmiemy się w następnych rozdziałach. Dla porównania na rys. 3.3 pokazujemy wynik dla podobnej symulacji ale z generatorem 'twister'. [7]

[6] plik przykladZlyGenState.m
[7] plik przykladZlyGenTwister.m

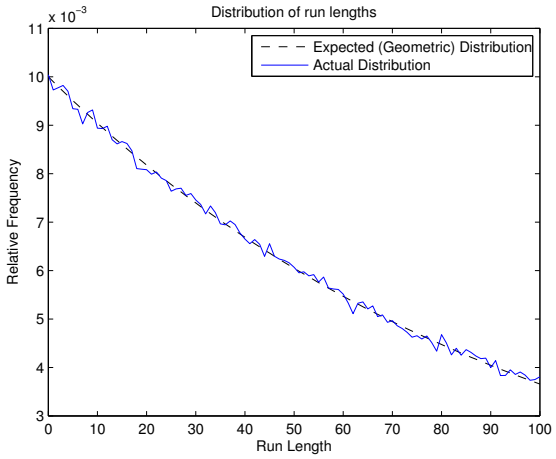


Rysunek 3.3: Histogram pojawiania się małych wielkości; generator 'twister', $n = 5 * 10^7$, $\delta = 0.01$

Dla kontrastu na rys. 3.4 podajemy wynik symulacji jeśli zamiast odstępów pomiędzy liczbami mniejszymi od $\delta = 0.01$ rozważamy odstępy pomiędzy liczbami większymi od $\delta = 0.99$. [8] W tym przypadku odstępy $X_1, X_2, \ldots$ mają ten sam rozkład geometryczny.

Generator Mersenne Twister Jednym z generatorów używanych przez MATLAB (ale nie tylko bo jest zaimplementowany w wielu innych pakietach) jest generator Mersenne Twister MT19937, który produkuje ciągi 32 bitowe liczb całkowitych. Charakteryzuje się on bardzo długim okresem $2^{19937} - 1$. Dla wielu zastosowań wystarczy już krótszy okres. Generator wymaga bardzo dużej przestrzeni CPU i jest przez to powolny. Problemem jest też wybór ziarna, które powinno być wysoce nieregularne, ponieważ w przypadku zbyt regularnego, aby otrzymać wyniki zbliżone do postulowanej imitacji losowego ciągu bitów, musi się wykonać wiele iteracji.

[8] plik przykladZlyGenStateAnty.m



Rysunek 3.4: Histogram pojawiania się dużych wielkości; generator 'state', $n = 5 * 10^7$, $\delta = 0.01$

3.3 Inne generatory

Przykład 3.7 [9] Wiele problemów przy rozwiązaniu wymaga bardzo długich okresów. Takie zapewniają tzw. generatory mieszane. Na przykład niech

1. $x_n = (A_1 x_{n-2} - A_2 x_{n-3}) \mod M_1$
2. $y_n = (B_1 y_{n-1} - B_2 y_{n-3}) \mod M_2$
3. $z_n = (x_n - y_n) \mod M_1$
4. jeśli $z_n > 0$, to $u_n = z_n/(M_1 + 1)$; w przeciwnym razie $u_n = M_1/(M_1 + 1)$.

z wyborem $M_1 = 4294967087$, $M_2 = 4294944443$, $A_1 = 1403580$, $A_2 = 810728$, $B_1 = 527612$, $B_2 = 1370589$. Jako warunki początkowe należy przyjąć pierwsze trzy x-sy oraz trzy y-ki. Algorytm został stworzony przez L'Ecuyera [28] przez komputerowe dopasowanie parametrów zapewniające optymalne własności generatora.

[9] Kelton et al (2004) W.D. Kelton, R.P.Sadowski and D.T. Sturrock (2004) Simulation with Arena (3rd ed.) McGraw Hill.

4 Testowanie dobrych własności generatora

Ciąg liczb pseudolosowych otrzymany za pomocą generatora "ma wyglądać" jak ciąg niezależnych zmiennych losowych o rozkładzie jednostajnym. Po wygenerowaniu próbki liczącej tysiące, trudno coś sprawdzić jak wygląda i dlatego trzeba się odwołać do metod statystycznych. Przygotowywaniu testów do sprawdzenia łosowościćiagu liczb zajmowano się od początku istnienia teorii generatorów. Jednakże obecnie teoria nabrała rumieńców gdy zauważono, jej użyteczność w kryptografii. Zauważono bowiem, że idealną sytuacja jest gdy zaszyfrowany ciąg wygląda zupełnie losowo, ponieważ odstępstwa od losowości moŋa zwiększyć szanse złamania protokułu kryptograficznego, odkrycia klucza prywatnego, itp. Ze względu na wspomniane zapotrzebowanie przygotowywane są zestawy testów. Najbardziej znane to

- Diehard tests. Jest to zestaw opracowany przez George'a Marsaglia z 1995 r.
- TestU01. Zestaw opracowany przez Pierre L'Ecuyer i Richard Simard z 2007 r.
- NIST Test Suite Zestaw opracowany przez organizację *Natonal Institute of Standards and Technology* Zestaw ten jest napisany pod kątem zastosowań w kryptografii.

Dla ilustracji zrobimy teraz przegląd kilku podstawowych metod. Musimy wpierw sprawdzić czy generowane liczby przypominają zmienne losowe o rozkładzie jednostajnym. Do tego służą w statystyce testy zgodności: χ^2 oraz test Kołmogorowa-Smirnowa.

Będziemy przyjmować hipotezę $\mathcal{H}_0$ mówiącą, że zmienne losowe $U_1, \ldots$ są niezależne o jednakowym rozkładzie U(0,1). Przypomnijmy, że dla zmiennych losowych $U_1, \ldots$ definiujemy *dystrybunatę empiryczną*

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(U_i \leq t), \qquad 0 \leq t \leq 1 \; .$$

W użyciu są, między innymi, następujące typy testów:

- testy równomierności – sprawdzający czy wygenerowane liczby są równomiernie rozmieszczone między zerem a jedynką,
- test serii – podobnie jak równomierności tylko dla kolejnych t-rek,

- test odstępów – notujemy odstępy pomiędzy liczbami losowymi, które nie należą do $[a, b]$, gdzie $0 \leq a < b \leq 1$, i sprawdzamy zgodność z rozkładem geometrycznym.
- testy oparte na klasycznych zadaniach kombinatorycznych,
- testy oparte na własnościach relizacji błądzenia przypadkowego.

Od generatora pseudolosowych ciągu bitów należy wymagać:

- Jednostajności: jakikolwiek badany podciąg powinien średnio zawierać połowę zer i połowę jedynek,
- Skalowalności: Test zastosowany do całego ciągu powinien dać te same rezultaty dla dowolnego wybranego podciągu.
- Zgodności: Dobroć generatora powinna być testowana dla dowolnych ziaren. Jest to szczególnie ważne w kontekście zastosowań kryptograficznych. Sa znane generatory (np. Mersenne twister), którego "dobre"własności zależą od łosowego"wybrania ziarna.

4.1 Funkcja p-wartość

Używając metod statystyki możemy weryfikować hipotezę. W naszym przypadku weryfikujemy hipotezę dotyczącą niezależności i jednakowości rozkładów (jednostajnych) sprawdzanego ciągu. Przyjmuje się poziom istotności (błąd pierwszego rodzaju α - prawdopodobieństwo odrzucenia hipotezy zerowej, jeśli jest ona prawdziwa) i następnie dla ustalonej statystyki testowej T znajduje sie obszar akceptacji. Najcześciej mamy do czynienia z testami jednostronnymi z obszarami akceptacji $T > x_0$ lub $T \leq x_0$ lub $|T| \leq t_0$. Wtedy x_0 jest dobierane tak aby odpowiednio $\mathbb{P}(T \notin$ obszaru akceptacji$) = \alpha$.

W przeciwieństwie do tego co się uczy na statystyce, że najpierw ustala się poziom istotności testu i obszar akceptacji, do testowania generatorów zalecana jest następująca procedura. Niech T ma dystrybantę $G(x)$ i w zagadnieniu obszar akceptacji testu jest postaci $(-\infty, a]$. Obliczamy statystykę testową T dla bardzo dużej liczby prób (na przykład $n = 2^{10}$) i następnie jeśli $T = t_0$ to obliczamy wartości tzw. p-wartości[10]:

$$p = 1 - G(t_0) \; . \tag{4.2}$$

[10] p–value

Czasami spotykamy testy z obszarem akceptacji postaci (a, ∞) i wtedy obliczamy lewą p wartość

$$p_l = G(t_0) \ . \tag{4.3}$$

Jeśli jedna z tych wartości jest bardzo bliska 0 to może wskazywać jakiś problem z generatorem. Przez małą liczbę L'Ecuyer poleca $p = 10^{-15}$–10^{-10}. Natomiast mamy przypadek "podejrzany" gdy p są bliskie 0.005. Jednakże wydaje sie, ze te zalecenia są zbyt ostre.

Ogólnie zalecane jest wygenerowanie $m = 1000 - -10000$ ciągów długości $n = 2^{20} - -2^{32}$. Można powtarzać eksperyment dla róznych $n = 2^k$ zwiększając k i obserwując czy nie ma załamania się wartości p. Po przeprowadzeniu eksperymentu otrzymujemy ciąg $p_1, \ldots, p_m$ p-wartości. Przyjmujemy wartość krytyczną, na przykład zalecane w raporcie NIST [41] $\alpha = 1/100 - 1/1000$. W przypadku $\alpha = 1/100$ to spodziewamy się że co najwyżej około 1% eksperymentów obleje test (to znaczy poda mała p-wartość). A więc w przeciwnym przypadku mamy podejrzenie, że generator jest kiepski.

Powtarzamy ten eksperyment dla $n = 2^k$ obserwując czy nie ma załamania się wartości p.

Zgodność z rozkładem jednostajnym; test λ Kołmogorowa

Teraz zajmiemy się testowaniem hipotezy, że rozkład brzegowy U_i jest U(0,1). Do tego służą testy zgodności, na przykład test Kołmogorowa-Smirnowa.

Generujemy liczby $U_1, \ldots, U_n$, i definiujemy odpowiadającą im

Pamiętając, że naszą hipotezą jest rozkład jednostajny $\mathrm{U}(0,1)$, tj. $F(t) = t$ dla $t \in (0,1)$, naszą statystyką testową jest

$$D_n = \sup_{0 \le t \le 1} |\hat{F}_n(t) - t| \ .$$

Z twierdzenia Gliwienko–Cantelli mamy, że jeśli U_i są rzeczywiście o rozkładzie jednostajnym, to z prawdopodobieństwem 1 ciąg D_n zmierza do zera. Natomiast unormowane zmienne $\sqrt{n}D_n$ zbiegają według rozkładu do rozkładu λ-Kołmogorowa, tj.

$$\lim_{n\to\infty} \mathbb{P}(\sqrt{n}D_n \le t) = K(t) = \sum_{j=-\infty}^{\infty} (-1)^j \exp(-2j^2t^2), \qquad t > 0 \ .$$

Rozkład ten jest stablicowany, patrz np tablice Zielińskich [52]. W szczególności mamy $\lambda_{0.1} = 1.224$, $\lambda_{0.05} = 1.358$ oraz $\lambda_{0.01} = 1.628$, gdzie

$$1 - K(\lambda_\alpha) = \alpha \ .$$

Do wyznaczenia D_n mamy następujący algorytm. Niech $U_{(1)}, \ldots, U_{(n)}$ będzie statystyką porządkową z próby $U_1, \ldots, U_n$, tj. ich niemalejące uporządkowanie oraz

$$\begin{aligned} D_n^+ &= \max_{1\le i\le n}\left(\frac{i}{n} - U_{(i)}\right), \\ D_n^- &= \max_{1\le i\le n}\left(U_{(i)} - \frac{i-1}{n}\right) . \end{aligned}$$

Wtedy

$$D_n = \max(D_n^+, D_n^-) .$$

Innym testem zgodności jest test χ^2. Pokażemy jego zastosowanie do sprawdzenia nie tylko zgodności z rozkładem jednostajnym ale również niezależności.

Test częstości par [11] Chcemy przetestować czy ciąg $u_1, u_2, \ldots \in [0,1]$ jest realizacją z próby prostej rozkładzie jednostajnym na $[0,1]$ $U_1, U_2, \ldots$. Zamiast ciągu łiczb rzeczywistych"przechodzimy do ciągu $y_1, y_2, \ldots$ stosując przekształcenie $y_j = \lfloor M u_j \rfloor$. W zależności od komputera musimy wybrać $M = 2^k$. Na przykład stosuje się $M = 64 = 2^6$ gdy komputer pracuje w systemi dwójkowym, i wtedy y_k reprezentuje 6 najbardziej znaczących bitów w dwójkowej reprezentacji u_j. Dla naszego test par rozpatrujemy ciąg n elementowy

$$(y_1, y_2), (y_3, y_4), \ldots, (y_{2n-1}, y_{2n}).$$

Wtedy liczymy X_{qr} liczbę powtórzeń $y_{2j-1} = q, y_{2j} = r$ dla każdej pary $0 \le q, r \le M-1$ i stosujemy test χ^2 z $M^2 - 1$ stopniami swobody z prawdopodobieństwem teoretycznym $1/M^2$ dla każdej pary (q, r).

Test odstępów Przykład pokazujący nieprawidłowości generatora 'state' w MATLABie (patrz podrozdział 3.2 wykorzystywał odstępy pomiędzy małymi wielkości. Ogólnie *test odstępów* bada odstępy pomiędzy pojawieniem się wielkości $u_j \in (\alpha, \beta]$, gdzie $0 \le \alpha < \beta \le 1$. Rozkład odstępu jest geometryczny z prawdopodobieństwem odstępu równym j wynoszącym $p(1-p)^{j-1}$ $(j = 1, 2, \ldots)$. Tutaj $p = \beta - \alpha$ i testujemy zgodność rozkładem chi kwadrat.

[11] Knuth v. ang. str. 60.

4.2 Testy oparte na schematach urnowych

Cała grupa testów jest oparta na schemacie rozmieszczenia n kul w k urnach. Zdefiniujemy najpierw kostkę w przestrzeni $[0,1)^t$. Dzielimy każdy z t boków na d (najlepiej $d = 2^\ell$) równych odcinków $[i/d, (i+1)/d)$, $i = 0, \ldots, d-1$, i następnie produktując je t razy otrzymujemy d^t małych kostek. Oznaczmy $k = d^t$, i ponumerujmy kostki od 1 do k. To będa nasze urny. Teraz generujemy nt liczb losowych $U_1, \ldots, U_{nt}$ i grupujemy je w n wektorów z $[0,1)^t$:

$$\boldsymbol{U}_1 = (U_1, \ldots, U_t), \boldsymbol{U}_2 = (U_{t+1}, \ldots, U_{2t}),$$
$$\ldots, \boldsymbol{U}_n = (U_{(n-1)t+1}, \ldots, U_{nt}) .$$

Każdy taki d wymiarowy wektor traktujemy jako kulę.

Test serii Niech X_j będzie liczba wektorów $\boldsymbol{U}_i$, które wpadną do kostki $j = 1, \ldots, k$ i zdefiniujmy statystykę

$$\chi^2 = \sum_{j=1}^{k} \frac{(X_j - n/k)^2}{n/k}.$$

Jeśli jest prawdziwa hipoteza $\mathcal{H}_0$ mówiąca, że zmienne losowe są niezależne o jednakowym rozkładzie U(0,1), to χ^2 ma średnią $\mathbb{E}\,\chi^2 = \mu = k-1$ oraz $\mathrm{Var}\,\chi^2 = \sigma^2 = 2(k-1)(n-1)$. W przypadku, gdy $t = 1$ mamy test zgodności z rozkładem brzegowym jednostajnym, bez testowania niezależności. Ze względu na duża liczbę kostek gdy t jest duże, używa się testu dla $t = 2$, tj. gdy testujemy jednostajność poszczególnych liczb i parami niezależność. Jesli chcemy sprawdzić niezależność dla większych t to stosujemy inne testy, jak np. test kolizji czy test urodzin. Na czym one polegają wyjaśnimy w następnym podrozdziale. Wiadomo, że

- jeśli $n \to \infty$ i k jest ustalone to χ^2 dąży do rozkładu χ^2 z $k-1$ stopniami swobody,

- jeśli $n \to \infty$ i $k \to \infty$ tak aby $n/k \to \gamma$, to $T = (\chi^2 - \mu)/\sigma$ dąży do rozkładu standardowego normalnego. Zalecane jest aby $0 \le \gamma < 5$, bo w przeciwnym razie χ^2 jest lepiej aproksymowany przez χ^2. [12]

[12] Sprawdzić w książkach statystycznych

Do przyjęcia lub odrzucenia hipotezy stosujemy test przy $k \gg n$, odrzucając hipotezą jesli χ^2 jest bardzo małe lub duże. [13]

Teraz przejdziemy do testów powstałych na bazie klasycznych zadań kombinatorycznych rachunku prawdopodobieństwa.

Test pokerowy Tak jak w teście częstotliwości par rozpatrujemy ciąg liczb $y_1, y_2, \ldots, y_{5n} \in \bar{M}$, który przy założeniu hipotezy zerowej, jest realizacą ciągu zmiennych losowych $Y_1, Y_2, \ldots$ o jednakowym rozkładzie jednostajnym $\mathrm{U}(\bar{M})$. Rozpatrzmy pierwszą piątkę $Y_1, \ldots, Y_5$. Wprowadzamy następujące wzorce, które są rozsądną modyfikacją klasyfikacji pokerowych. Chociaż ideą były wzorce z pokera (patrz zadanie 10.9), ale te wzorce są trudne w implementacji. Definiujemy więc za Knuth'em [23][14]

5 różnych = wszystkie różne,

4 różne = jedna para,

3 różne = dwie pary lub trójkę,

2 różne = full lub czwórkę

1 różna = wszystkie takie same.

Musimy policzyć prawdopodobieństwo teoretyczne p_r, że w piątce $Y_1, \ldots, Y_5$ będziemy mieli r różnych ($r = 1, \ldots, 5$). Niech $S_2(5, r)$ oznacza liczbę możliwych sposobów podziału zbioru pięcio elementowego na r niepustych podzbiorów (są to tzw. liczby Stirlinga 2-go rodzaju). Zostawiamy czytelnikowi do pokazania, że $S_2(5,1) = 1, S_2(5,2) = 15, S_2(5,3) = 25, S_2(5,4) = 10, S_2(5,5) = 1$. Ponieważ liczba różnych ciągów 5 elementowych wynosi $d(d-1)\cdots(d-r+1)$, więc szukane prawdopdobieństwo równa się

$$p_r = \frac{M(M-1)\cdots(M-r+1)}{M^5} S_2(5,r).$$

Dzielimy obserwacje na piątki $(y_{5(i-1)+1}, y_{5(i-2)+1}, \ldots, y_{5i})$ $(i = 1, \ldots, n)$ i następnie obliczamy próbkowe częstotliwości $\pi_1, \ldots, \pi_r$. Testujemy zgodność z prawdopdobieństwami teoretycznymi p_r testem chi kwadrat z 4 ma stopniami swobody.

[13] Referencje

[14] w ang wydaniu 1981 sekcja 3.3.2, p. 62

W następnych testach będziemy rozważać łosoweżozmieszczenie kul w komórkach. Za wzorzec przyjmiemy zadanie o urodzinach znane z wykładów rachunku prawdopodobieństwa.

Test kolizji Następujące zadanie jest podstawą tzw. *testu kolizji*, używanego do sprawdzania generatorów. Rozmieszczamy 'losowo' po kolei n kul w k (tak jak w zadaniu o urodzinach) komórkach (interesuje nas przypadek gdy $k >> n$). Rozważamy liczbę wszystkich kolizji C, gdzie *kolizję* definiujemy następująco. Po rozmieszczeniu kolejnej kuli, jeśli komórka do tej pory była pusta, to zapisujemy, że od tego momentu jest zajęta. Jeśli dla kolejnej wylosowanej kuli któraś komórka jest zajęta, to odnotowujemy *kolizję*. Podobnie jak w teście pokerowym można plokazać, że

$$\mathbb{P}(C = c) = \frac{k(k-1)\dots(k-n+c+1)}{k^n} S_2(n, n-c),$$

gdzie $S_2(n, d)$ jest liczbą Stirlinga drugiego rodzaju.

Polecamy czytelnikowi sprawdzenie czy jest prawdziwa aproksymacja $n^2/(2k)$ dla $\mathbb{E}\, C$ przy czym k jest dużo większe niż n(patrz Knuth []) . Można pokazać następujące prawdopodobieństwa $p = \mathbb{P}(C \leq c)$:

$c \leq$	101	108	119	126	134	145	153
p	.009	.043	.244	.476	.742	.946	.989

W zadanich o kulach i urnach rozpatruje się problem zajętości, to znaczy rozpatruje się zmienną Z niezajętych kul. Zostawiamy czytelnikowi do pokazania, że

$$Z = k - n + C$$

co pozwala wykorzystać znane fakty z klasycznego zadania o rozmieszczeniu. [15] Mianowicie wiadomo (patrz zadanie 10.10), że gdy gdy k, n rosną tak,że $\lambda = ke^{-\frac{n}{k}}$ pozostaje ograniczone, to dla każdego c

$$\mathbb{P}(Z = c) \to \frac{\lambda^c}{c!} e^{-\lambda}.$$

A więc

$$\mathbb{E}\, Z \sim ke^{-\frac{n}{k}}$$

[15] Uzupełnic z Fellera t.1

skąd mamy

$$\begin{aligned}\mathbb{E}\, C &= \mathbb{E}\, Z - k + n \\ &\sim ke^{-\frac{n}{k}} - k + n = k(1 - \tfrac{n}{k} + \tfrac{n^2}{2k^2} + o(\tfrac{n^2}{k^2})) - k + n \\ &= \tfrac{n^2}{2k} + o(\tfrac{n^2}{k}).\end{aligned}$$

Na tej podstawie zbudowano następujący test do zbadania losowości ciągu $y_1, \ldots, y_{20n}$. Na przykład niech $n = 2^{14}$ i $k = 2^{20}$. Wtedy możemy użyć testu do sprawdzenie generatora dla 20-sto wymiarowych wektorów: $v_j = (y_{20(j-1)+1}, \ldots, y_{20j})$ dla $1 \leq j \leq n$. [16]

Dni urodzin W rozdziale 2-gim książki Fellera [13] mamy zadanie o dniach urodzin. Rozważamy grupę (na przykład studencką) n osób i notujemy ich dzień urodzin. Przyjmujemy, że jest $k = 365$ dni w roku. Zakładając stałą częstośc urodzin w ciągu roku, zadanie to jest zadaniem o rozmieszczeniu łosowym"n kul w k komórkach. Klasycznym pytaniem jest o prawdopodobieństwo przynajmniej jednej sytuacji wielokrotnych urodzin, tj. o prawdopodobieństwo, że przynajmniej w jednej komórce mamy więcej niż jedną kulę. To prawdopodobieństwo łatwo da się obliczyć

$$p = 1 - \frac{365!}{(365-n)!(365)^n}\ .$$

Inne możliwe pytanie, na które można łatwo odpowiedzieć przez symulacje są następujące.

1. Jakie jest prawdopodobieństwo, że wśród n losowo wybranych osób przynajmniej dwie mają urodziny w przeciągu r dni jeden od drugiego?

2. Przypuśćmy, że osoby pojawiają się kolejno. Jak długo trzeba czekać aby pojawiły sie dwie mające wspólne urodziny?

Test odstępów dni urodzin Innym znanym testem sprawdzającym dobroć generatorów jest tzw. *test odstępów dni urodzin.* [17] Test ten jest oparty na statystyce K zdefiniowanej następująco. Zapisujemy dni urodzin kolejnych osób $Y_1, \ldots, Y_n \in \overline{k} = \{1, \ldots, k\}$ i ustawiamy w porządku niemalejącym $Y_{(1)} \leq \ldots \leq Y_{(n)}$. Następnie definiujmy, tzw. *odstępy*

$$S_1 = Y_{(2)} - Y_{(1)}, \ldots, S_{n-1} = Y_{(n)} - Y_{(n-1)}$$

[16] Knuth str. 73–75. Praca Tsang et al

[17] birthday spacing test

i K będzie liczbą równych odstępów Tj. ile razy mamy równość pomiędzy S-ami. Okazuje się, że jeśli n jest duże i $\lambda = n^3/(4k)$ jest małe, to przy założeniu naszej hipotezy $\mathcal{H}_0$, zmienna K ma w przybliżeniu rozkład Poissona z parametrem λ. Jednym z wyborów jest $n = 2^{10}$ i $k = 2^{24}$. Wtedy $\lambda = 16$.

[18] A więc jeśli $K = y$, to p-wartość wynosi

$$\mathbb{P}(K \geq y|\mathcal{H}_0) \approx 1 - \sum_{j=0}^{y-1} \frac{\lambda^j}{j!} e^{-\lambda}.$$

Okazuje się, że wtedy błąd przybliżenia jest rzędu $O(\lambda^2/n)$. W pracy L'Ecuyera i Simarda [30] zalecane jest aby $n^3 \leq k^{5/4}$.

5 Ciągi pseudolosowe bitów

Przez ciągi pseudolosowe bitów rozumiemy ciągi zero-jedynkowe. Wiele generatorów liczb pseudo-losowych jest generowany przez losowe ciągi bitów, które są dzielone na bloki zadanej długości, i następnie zamieniane na odpowiedni format. Na przykład na liczby z przedziału $[0, 1]$. Pytanie odwrotne jak z ciągu niezależnych zmiennych losowych o jednakowym rozkładzie $\mathrm{U}(\bar{M})$ otrzymać o wartościach w $\bar{M}$ otrzymąc losowy ciąg bitów. Można to zrobić jeśli $M = 2^k$. Szczegóły można znaleźć w zadaniu 10.1.

Innym powodem dla którego potrzeba rozpatrywać (pseudo)losowe ciągi bitów jest kryptografia.

5.1 Testy dla ciągów bitów

Niech $e_1, e_2, \ldots$ będzie ciągiem zero-jedynkowym, który przy założeniu hipotezy zerowej jest realizacjąlosowego ciągu bitów $\xi_1, \xi_2, \ldots$. Oznaczmy zmienne losowe $\eta_j = 2\xi_j - 1$, które przyjmują wartości $-1, 1$ z prawdopodobieństwem $1/2$ oraz przez *symetryczne błądzenie przypadkowe* rozumiemy ciąg

$$S_0, S_1 = \eta_1, \ldots, S_n = \eta_1 + \cdots + \eta_n.$$

W dalszej częsci oznaczamy $f_i = 2e_i - 1$. $s_0 = 0, s_1 = f_1, s_2 = f_1 + f_2, \ldots$.

[18] Trzeba poszukać gdzie jest dowód; Knuth z roku 1997; Exercises 3.3.2-28-30

Test częstotliwości Przy założeniu hipotezy zerowej, że ciągi pochodzą z obserwacji losowego ciągu bitów, z centralnego twierdzenia granicznego mamy, że zmienna losowa $S_n/\sqrt{n}$ ma w przybliżeniu rozkład normalny $\mathcal{N}(0,1)$, natomiast $T = |S_n/\sqrt{n}|$ ma rozkład normalny połówkowy. Mianowicie zauważamy, że $\mathbb{E}\,\eta_j = 0$ oraz $\operatorname{Var}\eta_j = 1$. Niech $F(t) = \mathbb{P}(T \leq t)$. Dla danego ciągu $e_1, \ldots$ weryfikujemy hipotezę o jego losowości obliczając statystykę $T = t_0$ i następnie obliczamy p-wartość z wzoru $1 - F(t_0)$.

Test częstotliwości w blokach Niech $n = bm$. Testowany ciąg $\xi_1, \ldots, \xi_n$, który przy założeniu hipotezy $\mathcal{H}_0$ jest losowym ciągiem bitów, długości n dzielimy na m bloków długości b. W każdym kolejnym bloku obliczamy frakcję jedynek:

$$\Pi_i = \frac{1}{b}\sum_{j=1}^{b} \xi_{(i-1)b+j}, \qquad i = 1, \ldots, m.$$

Mamy

$$\mathbb{E}\,\Pi_i = \frac{1}{2}, \qquad \operatorname{Var}\Pi_i = \frac{1}{4b}.$$

Dla odpowiednio dużych m, zmienna

$$\frac{1}{\sqrt{b}/2}\sum_{j=1}^{b}(\xi_{(i-1)b+j} - (b/2)),$$

ma w przyblizeniu standardowy rozkład normalny $\mathcal{N}(0,1)$. Niech $Z_1, \ldots, Z_m$ będzie ciągiem niezależnych zmiennych losowych o rozkładzie standarowym normalnym, i jeśli oznaczymy przez $\sim_{\mathrm{d}}$ - rozkład w przybliżeniu, to

$$\Pi_i - (b/2) \sim_{\mathrm{d}} \frac{1}{2\sqrt{b}} Z_i, \qquad i = 1, \ldots, m.$$

A więc

$$\sum_{i=1}^{m} Z_i^2 \sim_{\mathrm{d}} 4b\sum_{i=1}^{m}(\Pi_i - 1/2)^2,$$

i stąd mamy, że przy założeniu hipotezy $\mathcal{H}_0$ statystyka

$$T = 4b\sum_{i=1}^{m}(\Pi_i - 1/2)^2$$

ma rozkład chi-kwadrat z m stopniami swoboty. Jak zwykle teraz obliczamy p-wartość.

Test absolutnego maksimum błądzenia przypadkowego Dla prostego symetrycznego błądzenia przypadkowego S_n następujące twiedzenia Erdösa – Kaca jest prawdziwe:

$$\lim_{n\to\infty} \mathbb{P}(\max_{1\le j\le n} S_n/\sqrt{n} \le x) = \frac{4}{\pi}\sum_{j=0}^{\infty} \frac{(-1)^j}{2j+1} \exp(-(2j+1)^2\pi^2/8x^2), \qquad x \ge 0$$

[19]

Test oparty na wycieczkach błądzenia przypadkowego. Dla ciągu $s_0 = 0, s_1, s_2, \ldots, s_n$ niech $j_1 < j_2 < j_3 < \ldots$ będą indeksy kolejnych powrotów (s_n) do zera; tj. $s_i \neq 0$ dla $j_l < i < j_{l+1}$ dla $l = 1, 2, \ldots$. Przez wycieczkę rozumie się $(s_0, s_1, \ldots, s_{j_1})$, $(s_{j_1}, \ldots, s_{j_2})$, ... Niech dla prostego symetrycznego błądzenia przypadkowego $(S_j)_{1\le j\le n}$ niech J będzie liczbą wycieczek. Test budujemy korzystając z twierdzenia granicznego

$$\lim_{n\to\infty} \mathbb{P}(\frac{J}{\sqrt{n}} \le y) = \sqrt{\frac{2}{\pi}} \int_0^y e^{-x^2/2}\, dx. \tag{5.4}$$

Rozważmy teraz liczbę wizyt $L(x)$ do stanu x podczas jednej wycieczki. Wtedy

$$\mathbb{P}(L(x) = k) = \begin{cases} 1 - \frac{1}{2|x|} & \text{dla } k = 0 \\ \frac{1}{x^2}(1 - \frac{1}{2|x|})^{k-1} & \text{dla } k \ge 1 \end{cases} \tag{5.5}$$

Testujemy generator na bazie twierdzenia granicznego (5.4) oraz korzystając z (5.4) stosujemy test chi-kwadrat.

Przykład 5.1 [Zagadnienie komiwojażera] Komiwojażer opuszcza miasto 0 i musi odwiedzić wszystkie miasta $1, \ldots, n$ zanim powróci do miasta 0. Odległości pomiędzy miastem i a miastem j są zadane macierzą $(c_{ij})_{i,j=0,\ldots,n}$. Jeśli nie ma drogi z i do j to $c_{ij} = \infty$. Należy znaleźć drogę, która minimizuje odległośc $H(\pi)$, gdzie dla zadanej permutacji π liczb $1, \ldots, n$

$$H(\pi) = c_{0\pi_1} + c_{\pi_1\pi_2} + \ldots + c_{\pi_{n-1}\pi_n} + c_{\pi_n,0}$$

Zadanie to dla dużych n jest trudne do wykonania metodą deterministyczną ze względu na dużą liczbę obliczeń. Natomiast okazuje się, że zrandomizowane algorytmy dają niespodziewanie dobre rezultaty. Oczywiście na początku można poszukać 'dobrej' drogi przez symulacje.

[19] Bull. Amer. Math. Soc. Volume 52, Number 4 (1946), 292-302. On certain limit theorems of the theory of probability P. Erdös and M. Kac

5.2 Tasowanie kart

W ostatnich latach napisano sporo prac o teoretycznych własnościach tasowania kart. Talię kart można utożsamić z ciągiem liczb $[m] = (1, \ldots, m)$, które można uporządkować na $m!$ sposobów, noszących nazwy permutacji. Rozpatrzmy zbiór $\mathcal{S}_m$ wszystkich permutacji ciagu $[m]$. Permutacją losową będziemy nazywali odwzorowanie $X : \Omega \to \mathcal{S}_m$ z przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$ takie, że $\mathbb{P}(X = \pi) = 1/m!$ dla dowolnej permutacji $\pi \in \mathcal{S}_m$. Takie odwzorowanie możemy otrzymać przez algorytm na permutację losową z 2.1 na podstawie zadania 10.2. Możemy też wprowadzać różne sposoby tasowania kart, które naśladują tasowania prawdziwych kart. Każde tasowanie jest losowe, a więc zaczynając od $X_0 = \pi$, po n tasowaniach mamy permutacje X_n. Pytaniem jest jak daleko X_n jest od permutacji losowej. Możemy na przykład zmierzyc odległość wahania

$$d_n = \frac{1}{2} \sum_{\pi \in \mathcal{S}_m} |\mathbb{P}(X_n = \pi) - \frac{1}{m!}|.$$

Można wymyślać różne schematy tasowania kart. Niech $\pi_1, \pi_2, \ldots, \pi_m$ będzie talią m kart.

Proste tasowanie kart [20] W każdym kroku bierzemy kartę z góry i wkładamy ją losowo do talii. Możemy to sformalizować następująco. Oznaczmy miejsca między kartami $k_1 \sqcup k_2 \sqcup \cdots k_m \sqcup$. W każdym kroku wybieramy losowo jedno z tych miejsc i następnie wkładamy w to miejsce górna kartę.

Losowe inwersje [21] Przez losową inwersję permutacji $\pi = (\pi_1, \ldots, \pi_m)$ nazywamy wylosowanie pary π_i, π_j. Możemy to robić albo wyciągając dwie jednocześnie (parę losową), lub pojedynczo, i wtedy gdy dwie są takie same to nic nie robimy. Powtarzamy losowania n razy. Ponieważ na każdym etapie wynik jest losowy, a więc po n krokach mamy permutację X_n.

Riffle shuffle Generujemy odwrotne tasowanie. Oznaczamy tył każdej karty losowo i niezależnie liczbą 0 lub 1. Następnie wyciagamy z talii wszystkie oznaczone zerem, zachowując ich uporządkowanie i kładziemy je na pozostałych kartach.

[20] top-to-random
[21] random exchange

Pytaniem jest ile należy zrobić tasowań aby być "blisko"permutacji losowej.

Dla *riffle shufle* z $m = 52$ tak jak jest to w prawdziwej talii dokladne wielkości d_n są następujące:

n	1	2	3	4	5	6	7	8	9	10
d_n	1.000	1.000	1.000	1.000	.924	.614	.334	.167	.085	.043

(wg. pracy Diaconisa [11]). Te wyniki sugerują, że $n = 7$ powinno wystarczyć do otrzymania w miarę bliskiego permutacji losowej (czyli dobrze potasowanej talii). To było treścią artykułu w New York Times ([24] pod takim tytułem. Tabela powyższa pokazuje pewien problem z którym często spotykamy się gdy symulujemy charakterystykę procesu, który się stabilizuje, i chcemy obliczyć wielkość w warunkach stabilności. Jest więc pytanie kiedy osiągneliśmy warunki stabilności. Do tego powrócimy później. Symulacja liczb w tabelce byłaby trudna dla $m = 52$, ale możemy spróbować policzyć d_n przez symulacje dla mniejszych m.

6 Proste zadania probabilistyczne

6.1 Własności symetrycznego błądzenia przypadkowego

W rozdziale 3-cim książki Fellera [13] rozważa się zadania związane z ciągiem rzutów symetryczną monetą. Jest dwóch graczy A i B. Jeden z nich obstawia orła a drugi reszkę. Jeśli gracz A odgadnie wynik, to wygrywa +1 od gracza B, w przeciwnym razie gracz B wygrywa 1 od gracza A. Zakładamy, że wyniki kolejnych rzutów są niezalezne. Przypuśćmy, że gracze zaczynają grać z pustym kontem; tj $S_0 = 0$. Niech S_n będzie wygraną gracza A po n rzutach (tj. przegrana gracza B po n rzutach. Zauważmy, że wygodnie będzie rozpatrywanie $2N$ rzutów, i remisy mogą być jedynie w momentach $2, 4, \ldots, 2N$. Możemy postawić różne pytania dotyczące przebiegu gry, tj. przebiegu $(S_n)_{n=1,\ldots,2N}$. Możemy też przedsatwić błądzenie w postaci łamanej, łącząc prostymi punkty $(i-1, S_{i-1}), (i, S_i)$. Wtedy będziemy mieli segment powyżej osi x-w, tj. A prowadzi na segmencie $(i-1, S_{i-1} \to (i, S_i)$ jeśli $S_{i-1} \geq 0$ lub $S_i \geq 0$.[22].

1. Niech $R_{2N} = \max\{i : S_i = 0\}$ będzie ostatnim momentem remisu. Co można powiedzieć o rozkładzie R_{2N}?

[22] Feller t.1 str 79

2. Jakie jest prawdopodobieństwo $P(\alpha, \beta)$, że przy $2N$ rzutach gracz A będzie prowadził pomiędzy 100α a 100β procent czasu?

3. Jak wyglądają oscylacje $(S_n)_{0\leq n\leq 2N}$.

4. Jakie jest prawdopodobieństwo, że jeden z graczy będzie cały czas prowadził.

Tego typu zadania formalizujemy następująco, co następnie pozwoli na symulacje. Niech $\xi_1, \xi_2, \ldots,$ będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie $\mathbb{P}(\xi_j = 1) = \mathbb{P}(\xi_j = -1) = 1/2$ i definiujemy

$$S_0 = 0, \qquad S_n = \sum_{j=1}^{n} \xi_j, \qquad n = 1, \ldots$$

Ciąg $S_0 = 0, S_1, \ldots$ nazywamy *prostym symetrycznym błądzeniem przypadkowym.*

Zauważmy, że formalnie definiujemy

$$L_{2n} = \sup\{m \leq 2n : S_m = 0\}.$$

Przez

$$L^+_{2n} = \#\{j = 1, \ldots, 2n : S_{j-1} \geq 0 \quad \text{lub} \quad S_j \geq 0\}.$$

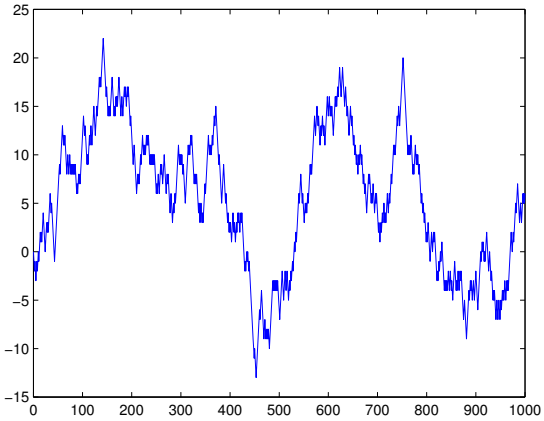
Inna charakterystyką może być, że gracz A ostro prowadzi w momencie m jeśli $S_m > 0$. Łączny czas ostrego prowadzenia A jest

$$\#\{1 \leq k \leq 2N : S_k > 0\}.$$

Przykładową symulację prostego symetrycznego błądzenia przypadkowego $S_0 = 0, S_1, \ldots, S_N$ można otrzymać za pomocą następującego programu. [23]

Proste symetryczne błądzenia przypadkowego

```
% N number of steps;
xi=2*floor(2*rand(1,N))-1;
A=triu(ones(N));
y=xi*A;
s=[0,y];
x=0:N;
plot(x,s)
```



Rysunek 6.5: Przykładowa realizacja prostego błądzenia przypadkowego; $N = 1000$.

Przykładowy wykres błądzenia przypadkowego podany jest na rys. 6.5

Na wszystkie pytania znane są odpowiedzi analityczne w postaci twierdzeń granicznych gdy $N \to \infty$; patrz Feller [13]. Na przykład wiadomo, że (patrz książka Fellera [13] lub Durreta [12])

$$\mathbb{P}(\alpha \leq L_{2N}^{+}/2N \leq \beta) \to \int_{\alpha}^{\beta} \pi^{-1}(x(1-x))^{-1/2}\, dx,$$

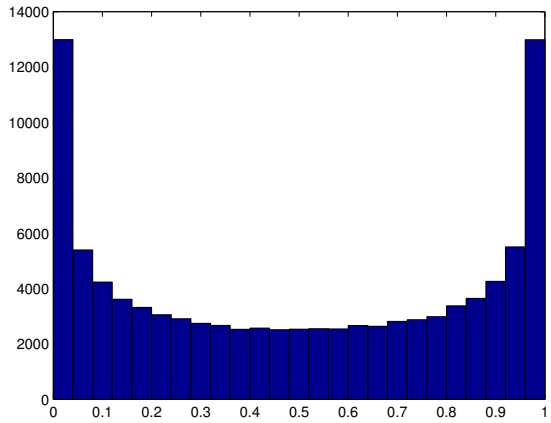
dla $0 \leq \alpha < \beta \leq 1$. Jest to tak zwane *prawo arcusa sinusa*. Nazwa bierze się stąd, że

$$\int_{0}^{t} \pi^{-1}(x(1-x))^{-1/2}\, dx = \frac{2}{\pi}(\arcsin(\sqrt{t}))\ .$$

Podobny rezultat jest dla L_{2N}; patrz Durret str. 197. My spróbujemy znaleźć odpowiedzi przez symulacje. Mając algorytm na symulacje błądzenia przypadkowego proponujemy zrobic histogramy dla powtórzeń L_{2N}^{+} – łącznego czasu prowadzenia przez A, L_{2N}. Jest jasne też, że gdy $N \to \infty$, to wielkości L_{2N}^{+}, $L2N$ rosną do nieskończoności. Ale jak się okazuje należy rozpatrywać frakcje czasu, czyli te wielkości podzielić przez $2N$. Metoda Monte Carlo polega na wielokrotnym powtórzeniu eksperymentu, tj. w naszym wypadku obliczenia powyższych wielkości. Niech R będzie liczbą przeprowadzonych

[23] dem2-1.m

eksperymentów, zwanych w teorii Monte Carlo *replikacjami* . Na rys 6.6 mamy histogram dla L^+_{1000} przy $R = 1000000$ replikacjach.
[24]



Rysunek 6.6: Histogram dla L^+_{1000};R=100000 replikacji, 25 klas

7 Zagadnienie sekretarki

Jest to problem z dziedziny optymalnego zatrzymania. Rozwiązanie jego jest czasami zwane regułą 37%. Można go wysłowić następująco:

- Jest jeden etat do wypełnienia.
- Na ten 1 etat jest n kandydatów, gdzie n jest znane.
- Komisja oceniająca jest w stanie zrobić ranking kandydatów w ściślym sensie (bez remisów).
- Kandydaci zgłaszają się w losowym porządku. Zakłada się, że każde uporządkowanie jest jednakowo prawdopodobne.

[24]Plik arcsinNew3.m Oblicza prawdopodobieństwo, że prowadzenie będzie krótsze niż 100*x % czasu, plik arcsinNew.m podaje histogram dla Lplus/$2N$ z $R = 100000$, $D = 2N = 1000$, i $n = 25$ jednakowej długości klas; obrazek arcsin.eps

- Po rozmowie kwalifikacyjnej kandydat jest albo zatrudniony albo odrzucony. Nie można potem tego cofnąć.
- Decyzja jest podjęta na podstawie względnego rankingu kandydatów spotkanych do tej pory.
- Celem jest wybranie najlepszego zgłoszenia.

Okazuje się, że optymalna strategia, to jest maksymizująca prawdopodobieństwo wybrania najlepszego kandydata jest, należy szukać wśród następujących strategi: opuścić k kandydatów i następnie przyjąć pierwszego lepszego od dotychczas przeglądanych kandydatów. Jesli takiego nie ma to bierzemy ostatniego. Dla dużych n mamy $k \sim n/e$ kandydatów. Ponieważ $1/e \approx 0.37$ więc stąd jest nazwa 37 %. Dla małych n mamy

n	1	2	3	4	5	6	7	8	9
k	0	0	1	1	2	2	2	3	3
P	1.000	0.500	0.500	0.458	0.433	0.428	0.414	0.410	0.406

To zadanie można prosto symulować jeśli tylko umiemy generować losową permutację. Możemy też zadać inne pytanie. Przypuśćmy, że kandydaci mają rangi od 1 do n, jednakże nie są one znane komisji zatrudniającej. Tak jak poprzednio komisja potrafii jedynie uporządkować kandydatów do tej pory starających sie o pracę. Możemy zadać pytanie czy optymalny wybór maksymizujący prawdopodobieństwo wybrania najlepszego kandydata jest również maksymizujący wartość oczekiwaną rangi zatrudnianej osoby. Na to pytanie można sobie odpowiedzieć przez symulację.

8 Metody quasi Monte Carlo

Alternatywną metodą do Monte Carlo (MC) używającą liczb pseudo-losowych są metody quasi-MC używającą liczb quasi-losowych. Jest to metoda szczególnie użyteczna przy obliczeniu całek $\int_{[0,1)^d} f(\boldsymbol{u})\ \mathrm{d}\boldsymbol{u}$. Korzystając z metody MC do obliczania tej całki, korzystamy z tego, ze jeśli $\boldsymbol{U}$ jest liczbą losową z $[0,1)^d$ (tzn. składowe wektora $\boldsymbol{U}$ są niezależne o jednakowym rozkładzie jednostajnym na $[0,1)$) i wtedy z mocnego prawa wielkich liczb, z prawdopodobieństwem 1

$$\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{U}_i) \to \mathbb{E}\, f(\boldsymbol{U}) = \int_{[0,1)^d} f(\boldsymbol{u})\ \mathrm{d}\boldsymbol{u}.$$

Szybkość tej metody jest rzędu $O(\frac{1}{\sqrt{n}}$ niezależnie od wymiaru d. Teraz przy użyciu metody quasi MC dany jest ciąg deterministyczny $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$, który posiada pewne własności o których za chwilę, i obliczmy szukaną całkę korzystając z tego, że

$$\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i) \to \mathbb{E}\, f(\boldsymbol{U}) = \int_{[0,1)^d} f(\boldsymbol{u})\ \mathrm{d}\boldsymbol{u}.$$

Jesli ciąg $(\boldsymbol{x}_i)$ jest ciągiem quasi-losowym, to wtedy szybkość zbieżności jest rzędu $O(n^{-(1-\epsilon)})$, gdzie $0 < \epsilon < 1$ jest dowolne. Dokładniejsza postać rzędu zbieżności będzie podana za chwilę. Oczywiście potrzebne są też pewne założenia z funkcji f, które w praktyce są spełnione. Metody quasi MC są szeroko stosowane do wycen produków finansowych.

Teoria liczb quasi-losowych opiera się na pojęciu *dyskrepancji* ciągu $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \in [0,1)^d$. Zacznijmy od dowolnego ciągu $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in [0,1)^d$ i niech $\mathcal{A}$ będzie pewną rodziną podzbiorów z $[0,1)^d$. Wtedy przez dyskrepancję rozumie sie wielkość

$$D(\boldsymbol{x}; \mathcal{A}) = \sup_{A \in \mathcal{A}} \left| \frac{\#\{i : \boldsymbol{x}_i \in A\}}{n} - \mathrm{vol}(A) \right|$$

gdzie $\mathrm{vol}(A)$ jest objętością zbioru A. W tym szkicu będziemy rozpatrywać rodzinę $\mathcal{A}^*$ podzbiorów postaci $\prod_{j=1}^{d}[0, u_j)$, gdzie $0 < u_j < 1$. Wtedy mówimy o gwiazdkowej dyskrepancji i oznaczamy $D^*(\boldsymbol{x}) = D(\boldsymbol{x}; \mathcal{A}^*)$. Dla każdego ustalonej długości ciągu n, możemy to zadanie wariacyjne rozwiązać i mamy $D^*(\boldsymbol{x}) \geq \frac{1}{2n}$, i równość jest osiągnięta dla $x_i = (2i-1)/(2n)$, $i8 = 1, \ldots, n$. Rozważmy teraz dowolny ciąg $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$. Okazuje się, że teraz problem jest dużo trudniejszy niż w przypadku skończonego ciągu. Główny konceptem tej teorii jest natepujące pojęcie. Mówimy, że ciąg $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \in [0,1)^d$ jest *niskiej dyskrepancji* jeśli $D^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \sim O\left(\frac{(\log n)^d}{n}\right)$. Okazuje się, że istnieją takie ciągi. Na przykład ciąg Haltona, ciąg Sobola, ciąg Faure. Dla tych ciągów można pokazać, że

$$D^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \leq C_d \frac{(\log n)^d}{n} + O(\frac{(\log n)^{d-1}}{n}.$$

W MATLABie w toolboxie Statistics and machne learning są zaimplementwane procedury generujące ciągi haltona i sobola (haltonset i sobolset).

9 Uwagi bibliograficzne

Na rynku jest sporo dobrych podręczników o metodach stochastycznej symulacji i teorii Monte Carlo. Zacznijmy od najbardziej elementarnych, jak na przykład książka Rossa [40]. Z pośród bardzo dobrych książek wykorzystujących w pełni nowoczesny aparat probabilistyczny i teorii procesów stochastycznych można wymienić książkę Asmussena i Glynna [4], Madrasa [33], Fishmana , Ripleya [38]. Wśród książek o symulacjach w szczególnych dziedzinach można wymienić książkę Glassermana [16] (z inżynierii finansowej), Fishmana [15] (symulację typu *discrete event simulation*. W języku polskim można znaleźć książkę Zielińskiego [50] oraz bardziej współczesną Wieczorkowskiego i Zielińskiego [48]. Wykłady prezentowane w obecnym skrypcie były stymulowane przez notatki do wykładu prowadzonego przez J. Resinga [37] na Technische Universiteit Eindhoven w Holandii. W szczegolności autor skryptu dziękuje Johanowi van Leeuwaardenowi za udostępnienie materiałów. W szczególności w niektórych miejscach autor tego skryptu wzoruje się na wykładach Johana van Leeuwardena i Marko Bono z Uniwersytetu Technicznego w Eindhoven.

Do podrozdziału II podstawowym tekstem jest monografia Knutha [22]; wydanie polskie [23]. W literaturze polskiej są dostępne książki Zielińskiego [51, 48], czy artykuł przeglądowy Kotulskiego [26]. Więcej na temat konstrukcji generatorów można znaleźć w artykułach przeglądowym L'Ecuyera [27, 30]. Omówienie przypadków efektu 'nielosowości' dla generatora *state* w MATLABie można znaleźć w pracy Savicky'ego [42] Natomiast przykład 3.6 jest zaczerpnięty z
http://www.mathworks.com/support/solutions/en/data/1-10HYAS /index.html?solution=1
10HYAS. Testowanie dobrych własności generatora jest bardzo obszerną dziś dziedziną. Podstawowym źródłem informacji jest monografia Knutha [23], rozdział III, również prace L'Ecuyera [27, 29, 30]. Podrozdział 5.1 jest oparty na raporcie [41]. Podrozdział 5.2 jest oparty na pracy Diaconisa [11]. Artykuł w New York Times był autorstwa Giny Kolaty [24]. Temu tematowi, i w ogólności szybkości zbieżności do rozkładu stacjonarnego łańcucha Markowa poświęcona jest książka [31].

Metoda quasi MC jest opisana w książce Niederreitera [35] i Glasserman [**?**].

10 Zadania

[25]

Zadania teoretyczne

10.1 Niech $X_1, X_2, \ldots$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie $\mathcal{U}(\bar{M})$, gdzie M jest liczbą naturalną oraz $\bar{M} = \{0, 1, \ldots, M-1\}$. Niech $M = 2^k$ oraz $b(x)$ dla $x \in \bar{M}$ będzie rozwinięciem dwójkowym przedstawionym w reprezentacji k cyfr zero lub jeden. Pokazać, że ciąg zmiennych losowych zero-jedynkowych zdefiniowany przez

$$\xi_1, \xi_2, \ldots = b(X_1), b(X_2) \ldots$$

jest ciągiem prób Bernoulliego. Pokazać na przykładzie, ze tak nie musi być jeśli M nie jest potęgą dwójki.

10.2 Uzasadnić, że procedura generowania losowej permutacji z wykładu daje losową permutację po rozwiązaniu następującego zadania z rachunku prawdopodobieństwa, w którym indukcyjnie definiujemy ciąg n permutacji zbioru $[n]$. Zerową permutacją π_0 jest permutacja identycznościowa. Dla $k = 1, 2, ..., n-1$ definiujemy indukcyjnie π_k-tą z π_{k-1}-szej następująco: zamieniamy element k-ty z J_k-tym gdzie J_k jest losowe (tj. o rozkladzie jednostajnym) na $\{k, \ldots, n\}$, niezależnie od $J_1, \ldots, J_{k-1}$. Pokazać, że π_{n-1} jest permutacją losową.

10.3 Dla przykładu 4.2, przy $n = 2^{14}$ i $k = 2^{20}$ uzupełnić tabelkę:

liczba kolizji $\leq$	100	110	120	130	140	150	160	170
z prawdopodobiestwem	?	?	?	?	?	?	?	?

10.4 Dany jest następujący ciąg zero jedynkowy : 10101101111100001010100
01011000101000101011110110
10010101001101010111111
01110001010110001000000001
00111001111010011111001111
00011101100010111001110000
10110000011110011100011011

[25] Poprawiane 13.10.2017

01001101101001010110101000
11110000100110110010000001
00010100011001000011001111
11110111100110010010011110

Przeprowadzić testy sprawdzające zgodność z hipotezą, że pochodzi z ciągu symetrycznych prób Bernoulliego.

10.5 Przez *funkcję błedu* $\mathrm{erf}(x)$ rozumie się

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt,$$

oraz przez *uzupełniającą funkcję błedu* $\mathrm{erfc}(x)$

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\, dt.$$

Pokazać, że jeśli $\Phi(x)$ jest dystrybuantą standardowego rozkładu normalnego, to

$$\Phi(x) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}(x/\sqrt{2}) = \frac{1}{2}\mathrm{erfc}(-x/\sqrt{2}).$$

10.6 Udowodnić nierówność Czernowa. Niech $S_n = \xi_1 + \ldots + \xi_n$, gdzie (ξ_i) są próbami Bernoulliego $\mathrm{Ber}(p)$. Pokazać, że dla $\delta < 1$

$$\begin{aligned} \mathbb{P}(S_n > (1+\delta)np) &\leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{np}, \\ \mathbb{P}(S_n < (1-\delta)np) &\leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{np}. \end{aligned}$$

10.7 Obliczamy całkę $\bar{f} = \int_{[0,1)^d} f(\boldsymbol{u})\, d\boldsymbol{u}$ przy założeniu, że $\sigma_f^2 = \mathrm{Var}\, f(\boldsymbol{U}) < \infty$. Korzystając z nierówności Czebyszewa pokazać, że z prawdopodobieństwm co najmniej δ mamy

$$|\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{U}_i) - \bar{f}| < \frac{\sigma_f}{\sqrt{\delta n}}.$$

10.8 Rzucono parę kostek $n = 144$ razy i otrzymano następujące wyniki. Przez E_s ozanaczamy liczbę oczekowanych obserwacji a przez O_s liczbę zaobserwowanych.

s	2	3	4	5	6	7	8	9	10	11	12
E_s	4	8	12	16	20	24	20	16	12	8	4
O_s	2	4	10	12	22	29	21	15	14	9	6

Zweryfikować testem chi-kwadrat hipotezę, że rzuty są niezależne i na każdej kostce wyniki są jednakowo prawdopodobne.

10.9 Pokazać, że dla układów pokerowych mamy następujące prawdopodobieństwa: Losujemy 5 kart z 52 i liczymy średnią liczbę wystąpień określonych sekwencji. Liczba wszystkich możliwych sekwencji wynosi $W = \binom{52}{5} = 2\ 598\ 960$ Prawdopodobieństwa wystąpień sekwencji kart w pokerze: - Jedna para: $1098240/W = 0.422569 = 1/2{,}4$ - Dwie Pary: $123552/W = 0.047539 = 1/21$ - Trójka: $54912/W = 0.0211 = 1/44$ - Full: $3744/W = 0.00144 = 1/694$ - Kareta: $624/W = 0.00024 = 1/4165$ - Kolor: $5148/W = 0.00198 = 1/50$ - Strit: $9 * 4^5/W$=9216/W = $0.003546 = 1/282$ - Poker: $9 * 4/W = 36/W = 0.0000138 = 1/72193$

10.10 [Feller, t. 1, p.94] Klasyczne zadanie o rozmieszczeniu n kul w m komórkach zakłada ich "losowość", tzn. każde rozmieszczenie ma prawdopodobieństwo m^{-n}. Pokazać, że prawdopodobieństwo $p_c(n, m)$ znalezienia dokładnie c pustych wynosi

$$\binom{m}{c} \sum_{j=0}^{m-c} (-1)^j \left(\frac{m-c}{j}\right) \left(1 - \frac{n+j}{m}\right).$$

Pokazać, że gdy m, n rosną tak,że $\lambda = me^{-\frac{n}{m}}$ pozostaje ograniczone, to dla każdego c

$$p_c(n, m) \to \frac{\lambda^c}{c!} e^{-\lambda}.$$

Zadania laboratoryjne

Zadania na wykorzystanie instrukcji MATLAB-owskich: rand, rand(m,n), rand('state', 0), randperm, randn.

10.11 Przeprowadzić test szybkości generatora: rand, rand('state'),rand('seed'), rand('twister') przy użyciu następującego skryptu:

```
clc; clear;
tic
for j=1:1000000
rand;
end
toc
```

10.12 Napisać program na sprawdzenie generatorów MATLABowskich przy użyciu
a. testu Kołmogorowa–Smirnowa,
b. testu serii z $t = 3$. Obliczyć p-wartość dla $d = 4$, $n = 2^m$; $(m = 10, 11, \ldots, 20)$.

10.13 [Savicky] Uruchomić procedurę

```
Z=rand(28,100000);
condition = Z(1,:)<1/16;
scatter(Z(16,condition),Z(28,condition),'.');
```

10.14 Uzasadnić, że w przypadku teoretycznym gdy Z jest macierzą składającą się z zmiennych losowych niezależnych o rozkładzie jednostajnym to na rysunku powinniśmy otrzymać chmurę par (U_{1i}, U_{2i}) $(i = 1, \ldots, 100000)$, gdzie U_{ij} są niezaleznymi zmiennymi losowymi o jednakowym rozkładzie jednostajnym U(0,1). Rysunek powinien więc wyglądać jak

```
Z=rand(2,100000);
condition = Z(1,:)<1/16;
scatter(Z(1,:),Z(2,:),'.');
```

Różnicę łatwo zauważyć. A co się dzieje gdyby użyć metody 'twister'. Wyjaśnić jak powinna wyglądać chmura w wypadku 'idealnym'.

10.15 Napisać procedurę na losowanie
1. z n obiektów wybieramy losowo k ze zwracaniem,
2. z n obiektów wybieramy k bez zwracania.
3. wybierającą losowy podzbiór k elementowyz $\{1, \ldots, n\}$.

10.16 Napisać procedurę generowania losowej permutacji `randpermmy` na podstawie algorytmu z podrozdziału 2.1. Porównać szybkość swojej procedury z matlabowska `randperm`.

10.17 Obliczyć przez symulacje prawdopodobieństwo p_n tego, że w permutacji losowej liczb $1, \ldots, n$, żadna liczba nie jest na swoim miejscu. Zrobić obliczenia dla $n = 1, \ldots, 10$. Do czego może zdążać p_n gdy $n \to \infty$.

10.18 Napisać procedurę n rzutów monetą $(\xi_1, \ldots, \xi_n)$.
Jeśli orzeł w i-tym rzucie to niech $\xi_i = 1$ jeśli reszka to $\xi_i = -1$. Niech $S_0 = 0$ oraz $S_k = \xi_1 + \ldots + \xi_k$. Obliczyć pierwszy moment i wariancję ξ_1. Zrobić wykres $S_0, S_1, \ldots, S_{1000}$. Na wykresie zaznaczyc linie: $\pm\mathrm{Var}(\xi)\sqrt{n}$, $\pm 2\mathrm{Var}(\eta)\sqrt{n}$, $\pm 3\mathrm{Var}(\xi)\sqrt{n}$.
Zrobić zbiorczy wykres 10 replikacji tego eksperymentu.

10.19 Rzucamy N razy symetryczną monetą i generujemy błądzenie przypadkowe $(S_n)_{0 \le n \le N}$. Napisać procedury do obliczenia:
1. Jak wygląda absolutna różnica pomiędzy maximum i minimum błądzenia przypadkowego gdy N rośnie?

10.20 Rozważamy ciąg $(S_n)_{0 \le n \le 2N}$. Przez prowadzenie w odcinku $(i, i+1)$ rozumiemy, że $S_i \ge 0$ oraz $S_{i+1} \ge 0$. Niech X będzie łączną długością prowadzeń dla N rzutów.
1. Jakie jest prawdopodobieństwo, że jeden gracz będzie przeważał pomiędzy 50% a 55% czasu? Lub więcej niż 95% czasu? Zrobić obliczenia symulacją. Eksperymentalnie sprawdzić ile powtórzeń należy zrobić aby osiągnąc stabilny wynik.
3. Niech $N = 200$. Zrobić histogram wyników dla 10000 powtórzeń tego eksperymentu.

10.21 Napisać procedurę do zadania dni urodzin.
1. Jakie jest prawdopodobieństwo, że wśród n losowo wybranych osób przynajmniej dwie mają ten sam dzień urodzin?
2. Jakie jest prawdopodobieństwo, że wśród n losowo wybranych osób przynajmniej dwie mają urodziny w przeciągu r dni jeden od drugiego?
3. Przypuśćmy, że osoby pojawiają się kolejno. Jak długo trzeba czekać aby pojawiły sie dwie mające wspólne urodziny? Zrobić histogram dla 1000 powtórzeń.

10.22 [Test odstępów dni urodzin] Używając symulacji znaleźć funkcję prawdopobieństwa statystyki R. Sprawdzić czy wyniki się pokrywają jeśli do obliczeń używa sie różnych generatorów.

10.23 [Zagadnienie komiwojażera] Komiwojażer opuszcza miasto 0 i musi odwiedzic wszystkie miasta $1, \ldots, n$ przed powrotem do punktu wyjściowego. Odległości pomiędzy miastem i a miastem j jest c_{ij}. Jeśli nie ma drogi do kładziemy $c_{ij} = \infty$. Jest $n = 100$ miast do odwiedzenia, których odległości są zadane w następujący sposób. Zacząć od rand('state', 0). Przydzielić odległości kolejno wierszami c_{ij} gdzie $i = 0, \ldots, 100$ oraz $j = 0, 1, \ldots, 100$ - odległość między adresem i oraz j. (W ten sposób wszyscy będą mieli taką samą macierz odległości). Znaleźć jak najlepszą drogę $0, \pi_1, \ldots, \pi_{100}, 0$ minimizującą przejechaną trasę. Czy uda się odnaleźć najkrótszą trasę. Znaleźć najkrótszą po 100, 200, itd. losowaniach.

Projekt

Projekt 1 Bob szacuje, że w życiu spotka 3 partnerki, i z jedną się ożeni. Zakłada, że jest w stanie je porównywać. Jednakże, będąc osobą honorową, (i) jeśli zdecyduje się umawiać z nową partnerką, nie może wrócić do już odrzuconej, (ii) w momencie decyzji o ślubie randki z pzostałymi są niemożliwe, (iii) musi się zdecydować na którąś z spotkanych partnerek. Bob przyjmuje, że można zrobić między partnerkami ranking: 1=dobra,2=lepsza, 3=najlepsza, ale ten ranking nie jest mu wiadomy. Bob spotyka kolejne partnerki w losowym porządku.

Napisz raport jaką stategię powinien przyjąc Bob. W szczególności w raporcie powinny się znaleźć odpowiedzi na następujące pytania. Przy sporządzaniu raport przeprowadzić analizę błedu, korzystając z fundamentalnego wzoru

$$b = \frac{1.96\sigma}{\sqrt{n}}$$

(patrz podsekcja IV.1.2, gdzie b - błąd, n - liczba replikacji, oraz σ wariancja estymatora). A więc jeśli chcemy wyestymaować r z błędem mniejszym niż $b = 0.02$ możemy skorzystać z fundamentalnego wzoru z wariancją uzyskaną wcześniej z pilotażowej symulacji 1000 repliacji.

(a) Bob decyduje się na ślub z pierwszą spotkaną partnerką. Oblicz prawdopodobieństwo P (teoretyczne), że ożeni się z najlepszą. Oblicz też oczekiwany rangę r

(b) Użyj symulacji aby określić P i r dla następującej strategii. Bob nigdy nie żeni się z pierwszą, żeni się z drugą jesli jest lepsza od pierwszej, w przeciwnym razie żeni się z trzecią.

(c) Rozważyć zadanie z 10-cioma partnerkami, które mają rangi $1, 2, \ldots, 10$. Obliczyć teoretycznie prawdopodobieństwo P i oczekiwaną rangę r dla strategii jak w zadaniu (a).

(d) Zdefiniować zbiór strategii, które są adaptacją strategii z zadania (b) na 10 partnerek. Podać jeszcze inną strategię dla Boba. Dla każdej stategii obliczyć P i r. Wsk. Należy pamiętać, że jak się nie zdecydujemy na jakąś kandydatkę, to nie ma potem do niej powrotu.

Projekt 2 Rozpatrzyć następujące generatory liczb losowych: "state", "twister" oraz excellowski $u_i = (0.9821u_{i-1} + 0.211327) \mod 1$. Ponadto rozpatrzyć pseudo-losowe ciągi bitów date.e, date.sqrt2 (z strony www.math.uni.wroc.pl/řolski).

(i) Dla tych ciągów przeprowadzić analizę ich "losowości", korzystając z przynajmniej 3 testów (odpowiednio dla każdego ciągu). Na przykład Kołmogorowa-Smirnow, test serii i test odstępów dni urodzin. Zamiast tego ostentiego można skorzystać z tesu kolizji.
(ii) Dla generatora "twister" zbadać dwa przypadki. Pierwszy gdy generator startuje z ziarna $u_o = 0$, a drugi gdy z ziarna $u_0 = 1812433253$.

Rozdział III

Symulacja zmiennych losowych

[1]

1 Metoda ITM; rozkład wykładniczy i Pareto

Naszym celem jest przeprowadzenie losowania w wyniku którego otrzymamy liczbę losową Z z dystrybuantą F mając do dyspozycji zmienną losową U o rozkładzie jednostajnym U(0,1). W dalszym ciągu będziemy często bez wyjaśnienia pisali $U_1, U_2, \ldots$ dla ciągu niezależnych zmiennych losowych o jednakowym rozkładzie $\mathcal{U}(0,1)$.

Niech

$$F^{\leftarrow}(t) = \inf\{x : \ t \leq F(x)\} \ .$$

W przypadku gdy $F(x)$ jest funkcją ściśle rosnącą, $F^{\leftarrow}(t)$ jest funkcją odwrotną $F^{-1}(t)$ do $F(x)$. Dlatego $F^{\leftarrow}(t)$ będziemy nazywać się *uogólnioną funkcją odwrotną* Zbadamy teraz niektóre własności uogólnionej funkcji odwrotnej $F^{\leftarrow}(x)$. Jeśli F jest ściśle rosnąca to jest oczywiście prawdziwa następująca równoważność:

- $t = F(x)$ wtedy i tylko wtedy gdy $F^{\leftarrow}(t) = x$.

Jednakże można podać przykład taki, że $F^{\leftarrow}(t) = x$, ale $t < F(x)$ (patrz rysunek **[??2:rys1??]**). Natomiast jest prawdą, że zawsze $F^{\leftarrow}(F(x)) = x$, chociaż łatwo podać przykład, że nie musi być $F(F^{\leftarrow}(t)) \neq t$ (zostawiamy czytelnikowi znalezienie takiego przykładu w formie rysunku takiego jak w **[??2:rys1??]**).

[1] Poprawiane 1.2.2016

W skrypcie będziemy oznaczać *ogon dystrybuanty* $F(x)$ przez

$$g(x) = \overline{F}(x) = 1 - F(x).$$

Lemat 1.1 (a) $u \leq F(x)$ *wtedy i tylko wtedy gdy* $F^{\leftarrow}(u) \leq x$.
(b) *Jeśli* $F(x)$ *jest funkcją ciągłą, to* $F(F^{\leftarrow}(x)) = x$.
(c) $\overline{F}^{\leftarrow}(x) = F^{\leftarrow}(1-x)$.

Fakt 1.2 *(a) Jeśli* U *jest zmienną losową, to* $F^{-1}(U)$ *jest zmienną losową z dystrybuntą* F.
(b) Jesli F *jest ciągłą i zmienna* $X \sim F$, *to* $F(X)$ *ma rozkład jednostajny U(0,1)*

Dowód (a) Korzystając z lemat 1.1 piszemy

$$\mathbb{P}(F^{\leftarrow}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x) \ .$$

(b) Mamy

$$\mathbb{P}(F(X) \geq u) = \mathbb{P}(X \geq F^{\leftarrow}(u)) \ .$$

Teraz korzystając z tego, że zmienna losowa X ma dystrybuantę ciągłą, a więc jest bezatomowa, piszemy

$$\mathbb{P}(X \geq F^{\leftarrow}(u)) = \mathbb{P}(X > F^{\leftarrow}(u)) = 1 - F(F^{\leftarrow}(u)) \ ,$$

które na mocy lematu 1.1 jest równe $1-u$. □

Oznaczmy przez

$$\begin{aligned} g(t) &= F^{\leftarrow}(t) \\ \overline{g}(t) &= g(1-t) \end{aligned}$$

Zauważmy, że

$$\overline{g}(t) = \inf\{u : \bar{F}(u) \leq 1 - t\} \ .$$

Korzystając z powyższgo faktu mamy następujący algorytm na generowanie zmiennej losowej o zadanym rozkładzie F jesli mamy daną funkcję $g(t)$.

Algorytm ITM

```
Z=g(rand);
```

Zauważmy jeszcze coś co będzie przydatne w dalszej części, że jeśli $U_1, U_2, \ldots$ jest ciągiem zmiennych losowych o jednakowym rozkładzie jednostajnym,, to $U_1', U_2' \ldots$, gdzie $U_j' = 1 - U_j$, jest też ciągiem niezależnych zmiennych losowych o rozkładzie jednostajnym.

Niestety jawne wzory na funkcję $g(x)$ czy $\bar{g}(x)$ znamy w niewielu przypadkach, które omówimy poniżej.

Rozkład wykładniczy Exp(λ). W tym przypadku

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - \exp(-\lambda x), & x \geq 0. \end{cases} \tag{1.1}$$

Jeśli $X \sim \mathrm{Exp}(\lambda)$, to $\mathbb{E}\, X = 1/\lambda$ oraz $\mathrm{Var}\, X = 1/(\lambda^2)$. Jest to zmienna z lekkim ogonem ponieważ

$$\mathbb{E}\, e^{sX} = \frac{\lambda}{\lambda - s}$$

jest określona dla $0 \leq s < \lambda$. W tym przypadku funkcja odwrotna

$$\bar{g}(x) = -\frac{1}{\lambda} \log(x) \ .$$

Zmienną losową $\mathrm{Exp}(\lambda)$ o rozkładzie wykładniczym generujemy następującym algorytmem.

Algorytm ITM-Exp.

```
% Dane lambda. Otrzymujemy Z o rozkładzie wykładniczym z par. lambda
Z=-log(rand)/lambda;
```

Zajmiemy się teraz przykładem zmiennej ciężko-ogonowej.

Rozkład Pareto Par(α). Rozkład Pareto definiujemy przez

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - \frac{1}{(1+x)^\alpha}, & x \geq 0\,. \end{cases} \tag{1.2}$$

Jeśli $X \sim \mathrm{Par}(\alpha)$, to $\mathbb{E}\,X = 1/(\alpha-1)$ pod warunkiem, że $\alpha > 1$ oraz $\mathbb{E}\,X^2 = 2/((\alpha-1)(\alpha-2))$ pod warunkiem, że $\alpha > 2$. Jest to zmienna z cieżkim ogonem ponieważ

$$\mathbb{E}\,e^{sX} = \infty$$

dla dowolnego $0 < s$. Funkcja odwrotna

$$\bar{g}(x) = x^{-1/\alpha} - 1.$$

Zauważmy, że rozkład Pareto jest tylko z jednym parametrem kształtu α. Dlatego jeśli chcemy miec szerszą klasę dopuszczającą dla tego samego parametru kształtu zadaną średnią m skorzystamy z tego, że zmienna $m(\alpha-1)X$ ma średnią m (oczywiście zakładając, że $\alpha > 1$. Można więc wprowadzić dwuparametrową rodzinę rozkładów Pareto $\mathrm{Par}(\alpha, c)$. Zmienna Y ma rozkład $\mathrm{Par}(\alpha, c)$ jeśli jej dystrybunata jest

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - \frac{c^\alpha}{(c+x)^\alpha}, & x \geq 0\,. \end{cases} \tag{1.3}$$

W literaturze nie ma jednoznacznej definicji rozkładu Pareto, i dlatego w poręcznikach można znaleźć inne warianty podanej definicji.

Algorytm ITM-Par.

```
% Dane alpha. Otrzymujemy Z o rozkładzie Par(alpha).
Z=rand^{(-1/\alpha)}-1.
```

2 Metoda ITM oraz ITR dla rozkładów kratowych; rozkład geometryczny

Przypuśćmy, że rozkład jest kratowy skoncentrowany na $\{0, 1, \ldots\}$. Wtedy do zadania rozkładu wystarczy znać funkcję prawdopodobieństwa $\{p_k,\ k = 0, 1, \ldots\}$.

Fakt 2.1 *Niech U będzie liczbą losową. Zmienna*

$$Z = \min\{k : U \leq \sum_{i=0}^{k} p_i\}$$

ma funkcję prawdopodobieństwa $\{p_k\}$.

Dowód Zauważmy, że

$$\mathbb{P}(Z = l) = \mathbb{P}(\sum_{i=0}^{l-1} p_i < U \leq \sum_{i=0}^{l} p_i) \ ,$$

i odcinek $(\sum_{i=0}^{l-1} p_i, \sum_{i=0}^{l} p_i]$ ma długość p_l. □

Stąd mamy następujący algorytm na generowanie kratowych zmiennych losowych Z.

Algorytm ITM-d. z

```
% Dane p_k=p(k). Otrzymujemy Z z funkcją prawd. {p_k}.
U=rand;
Z=0;
S=p(0);
while S<U
  Z=Z+1;
  S=S+p(Z);
end
```

Fakt 2.2 *Niech N będzie liczbą wywołań w pętli* while *w powyższym algorytmie. Wtedy N ma funkcję prawdopodobieństwa*

$$\mathbb{P}(N = k) = p_{k-1},$$

dla $k = 1, 2, \ldots$, skąd średnia liczba wywołań wynosi $\mathbb{E}\, N = \mathbb{E}\, Z + 1$.

W powyższego faktu widzimy więc, że jeśli zmienna Z ma średnią nieskończoną, to liczba wywołań ma również średnią nieskończona i algorytm nie jest dobry.

Rozkład geometryczny Geo(p) ma funkcję prawdopodobieństwa

$$p_k = (1-p)p^k, \qquad k = 0, 1, \ldots$$

Zmienna $X \sim \text{Geo}(p)$ ma średnią $\mathbb{E}\, X = p/q$. Zmienna

$$X = \lfloor \log U / \log p \rfloor$$

ma rozkład Geo(p). ponieważ

$$\begin{aligned} \mathbb{P}(X \geq k) &= \mathbb{P}(\lfloor \log U / \log p \rfloor \geq k) \\ &= \mathbb{P}(\log U / \log p \geq k) \\ &= \mathbb{P}(\log U) \leq (\log p)k) \\ &= \mathbb{P}(U \leq p^k) = p^k \,. \end{aligned}$$

W niektórych przypadkach następująca funkcja

$$c_{k+1} = \frac{p_{k+1}}{p_k}$$

jest prostym wyrażeniem. Podamy teraz kilka takich przykładów.

- $X \sim \text{B}(n, p)$. Wtedy

$$p_0 = (1-p)^n, \quad c_{k+1} = \frac{p(n-k)}{(1-p)(k+1)}, \qquad k = 0, \ldots, n-1.$$

- $X \sim \text{Poi}(\lambda)$. Wtedy

$$p_0 = e^{-\lambda}, \quad c_{k+1} = \frac{\lambda}{k+1}, \qquad k = 0, 1, \ldots$$

- $X \sim \text{NB}(r, p)$. Wtedy

$$p_0 = (1-p)^r, \quad c_{k+1} = \frac{(r+k)p}{k+1}, \qquad k = 0, 1, \ldots$$

W przypadki gdy znana jest jawna i zadana w prostej postaci funkcja c_{k+1} zamiast algorytmu ITM-d możemy używać następującego ITR. Oznaczamy c_{k+1} =c(k+1) i p_0 =p(0).

Algorytm ITR.

```
% Dane p(0) oraz c(k+1), k=0,1,... Otrzymujemy X z f. prawd. {p_k}.
S=p(0);
P=p(0);
X=0;
U=rand;
while (U>S)
  X=X+1;
  P=P*c(X);
  S=S+P;
end
```

3 Metody ad hoc.

Są to metody specjalne, które wykorzystują własności konkretnych rozkładów.

3.1 B(n,p)

Niech $X \sim \mathrm{B}(n,p)$. Wtedy

$$X =_{\mathrm{d}} \sum_{j=1}^{n} \xi_j,$$

gdzie $\xi_1, \ldots, \xi_n$ jest ciągiem prób Bernoulliego, tj. ciąg niezależnych zmiennych losowych $\mathbb{P}(\xi_j = 1) = p$, $\mathbb{P}(\xi_j = 0) = 1 - p$.

Algorytm DB

```
% Dane n,p. Otrzymujemy X z f. prawd. B(n,p)
X=0;
for i=1:n
  if (rand<= p)
    X=X+1;
  end
end
```

Polecamy czytelnikowi porównać tą metodę z metodą wykorzystującą ITR.

3.2 Erl(n, λ)

Jeśli $\xi_1, \ldots, \xi_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie Exp(λ), to

$$X = \xi_1 + \ldots + \xi_n$$

ma rozkład Erlanga Erl(n, λ). Wiadomo, że rozkład Erlanga jest szczególnym przypadkiem rozkładu Gamma(n, λ) z gęstością

$$f(x) = \frac{1}{\Gamma(n)} \lambda^n x^{n-1} e^{-\lambda x}, \qquad x \geq 0 \; .$$

Przez proste różniczkowanie można się przekonać, że ogon dystrybuanty rozkładu Erlanga Erl$(k+1,1)$ wynosi

$$e^{-x}(1 + x + \frac{x^2}{2!} + \ldots + \frac{x^k}{k!}) \; .$$

Będziemy tego faktu potrzebować poniżej.

Ponieważ liczbę losową o rozkładzie wykładniczym Exp(λ) generujemy przez $-\log(\text{rand})/\lambda$, więc $X \sim \text{Erl}(n, \lambda)$ możemy generować przez

$$-\log(U_1 \cdot \ldots \cdot U_n)/\lambda \; .$$

Należy zauważyć, że w przypadku dużych n istnieje ryzyko wystąpienia błędu zmiennnopozycyjnego niedomiaru przy obliczaniu iloczynu $U_1 \cdot \ldots \cdot U_n$.

3.3 Poi(λ).

Teraz podamy algorytm na generowanie zmiennej $X \sim \text{Poi}(\lambda)$ o rozkładzie Poissona. Niech $\tau_1, \tau_2, \ldots$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie Exp(1). Niech

$$N = \#\{i = 1, 2, \ldots : \tau_1 + \ldots + \tau_i \leq \lambda\}. \tag{3.4}$$

Fakt 3.1 *Mamy*

$$N =_{\text{d}} \text{Poi}(\lambda).$$

Dowód Zauważmy, że

$$\{N \leq k\} = \{\sum_{j=1}^{k+1} \tau_j > \lambda\}.$$

Ponieważ

$$\mathbb{P}(\sum_{j=1}^{k} \tau_j > x) = e^{-x}(1 + x + \frac{x^2}{2!} + \ldots + \frac{x^k}{k!}),$$

i pamiętając o konwencji $\sum_{j=1}^{0} = 0$, mamy więc

$$\mathbb{P}(N = 0) = \mathbb{P}(N \leq 0) = e^{-x},$$

oraz

$$\begin{aligned} \mathbb{P}(N = k) &= \mathbb{P}(\{N \leq k\} \cap \{N \leq k-1\}^c) \\ &= \mathbb{P}(\{N \leq k\}) - \mathbb{P}(\{N \leq k-1\}) \\ &= \frac{\lambda^k}{k!} e^{-\lambda}. \end{aligned}$$

□

Wzór (3.4) można przepisać następująco:

$$N = \begin{cases} 0, & \tau_1 > \lambda \\ \max\{i \geq 1 : \tau_1 + \ldots + \tau_i \leq \lambda\}, & \text{w przeciwnym razie.} \end{cases}$$

Na tej podstawie możemy napisać następujący algorytm.

Algorytm DP

```
% Dane lambda. Otrzymujemy X o rozkładzie Poi(lambda).
X=0;
S=-log(rand);
while (S<=lambda)
  Y=-log(rand);
  S=S+Y;
  X=X+1;
end
```

Algorytm DP możne jeszcze uprościć korzystając z następujących rachunków. Jeśli $\tau_1 \leq \lambda$ co jest równowazne $-\log U_1 \leq \lambda$ co jest równoważne $U_1 \geq$

$\exp(-\lambda)$, mamy

$$\begin{aligned} X &= \max\{n \geq 0 : \tau_1 + \ldots + \tau_n \leq \lambda\} \\ &= \max\{n \geq 0 : (-\log U_1) + \ldots + (-\log U_n \leq \lambda\} \\ &= \max\{n \geq 0 : \log \prod_{i=1}^{n} U_i \geq -\lambda\} \\ &= \max\{n \geq 0 : \prod_{i=1}^{n} U_i \geq e^{-\lambda}\} \ . \end{aligned}$$

Powyżej przyjmujemy konwencję, że jeśli $n = 0$, to $\tau_1 + \ldots + \tau_n = 0$.

3.4 Rozkład chi kwadrat

Jeśli ξ ma rozkład normalny N(0,1) to prosty argument pokazuje, ze ξ^2 ma rozkład Gamma$(1/2, 1/2)$. Jesli więc zmienne $\xi_1, \ldots, \xi_n$ są niezależne o jednakowym rozkładzie N$(0, 1)$ to

$$\xi_1^2 + \ldots + \xi_n^2 \tag{3.5}$$

ma rozkład Gamma$(n/2, 1/2)$. Taki rozkład nosi nazwę *chi kwadrat* z n stopniami swobody, który jest jednym z fundamentalnych rozkładów statystyki matematycznej. Jak generować zmienne losowe o rozkładzie N(0,1) będziemy się uczyć dalej. W MATLABie generuje się takie liczby losowe za pomocą instrukcji

```
randn
```

4 Rozkład jednostajny

Mówimy, że wektor losowy X ma rozkład jednostajny U(A) w obszarze $A \subset \mathbb{R}^n$, jeśli

$$\mathbb{P}(X \in D) = \frac{|D|}{|A|} \ .$$

Naszą podstawową zmienną jest jednostajna U(0,1) na odcinku $(0, 1)$. Jeśli więc potrzebujemy zmiennej U(a, b) jednostajnej na odcinku (a, b) to musimy przekształcić zmienną U liniowo:

$$V = (b - a)U + a \ .$$

Zostawiamy uzasadnienie tego faktu czytelnikowi. Aby generować rozkład jednostajny w dowolnym obszarze A, np. kuli lub elipsoidy, zrobimy następującą obserwację. Przypuśćy, że A i B będą pozdbiorami $\mathbb{R}^n$ dla których istnieje objętość $\int_B d\boldsymbol{x} = |B|$ i $\int_A d\boldsymbol{x} = |A|$ oraz $A \subset B$. Jesli $\boldsymbol{X}$ ma rozkład jednostajny $\mathrm{U}(B)$, to $(X|X \in A)$ ma rozkład jednostajny $\mathrm{U}(A)$. Mianowicie trzeba skorzystać, że dla $D \subset A$

$$\mathbb{P}(X \in D|X \in A) = \frac{|D \cap |/|B|}{|A|/|B|} = \frac{|D|}{|A|}.$$

Stąd mamy następujący algorytm na generowanie wektora losowego $\boldsymbol{X} = (X_1, \ldots, X_n)$ o rozkładzie jednostajnym $\mathrm{U}(A)$, gdzie A jest ograniczonym podzbiorem $\mathbb{R}^n$. Mianowicie taki zbiór A jest zawarty w pewnym prostokącie $B = [a_1, b_1] \times \ldots \times [a_n, b_n]$. Łatwo jest teraz wygenerować wektor losowy $(V_1, \ldots, V_n)$ o rozkładzie jednostajnym $\mathrm{U}(B)$ ponieważ wtedy zmienne V_i są niezależne o rozkładzie odpowiednio $\mathrm{U}[a_i, b_i]$ $(i = 1, \ldots, n)$.

Algorytm U(A)

```
%na wyjściu X o rozkładzie jednostajnym w A.
1. for i=1:n
     V(i)=(b(i)-a(i))*U(i)+a(i)
   end
2. jeśli V=(V(1),...,V(n) należy do A to X:= V,
3. w przeciwnym razie idź do 1.
```

5 Metoda superpozycji

Jest to metoda bazująca na wyrażeniu rozkładu poszukiwanej liczby losowej X jako mieszanki rozkładów. W przypadku gdy X ma gęstość $f_X(x)$ zakładamy, że

$$f(x) = \sum_{j=1}^{n} p_j g_j(x) \ ,$$

gdzie $g_j(x)$ są pewnymi gęstościami rozkładów (n może być ∞). Przypuśćmy dalej, że potrafimy symulować zmienne losowe o tych gęstościach. Aby wygenerować liczbę losową z gęstością $f(x)$, losujemy najpierw J z funkcją prawdopdobieństwa (p_j) i jeśli $J = j$ to losujemy liczbę losową z

gęstością $g_j(x)$ i tę wielkość przyjmujemy jako szukaną wartość X. Metoda ta ma w symulacji zastosowanie szersze niż by się wydawało. Bardzo ważnym przypadkiem jest gdy

$$f(x) = \sum_{n=0}^{\infty} p_n(\lambda) g_n(x) \tag{5.6}$$

gdzie $p_n(\gamma) = \frac{\lambda}{n!} e^{-\lambda}$ jest rozkładem Poissona a $g_n(x)$ jest gęstościa gamma $\Gamma(\alpha_n, \beta)$.

Przykład 5.1 [Niecentralny rozkład chi-kwadrat] W statystyce, matematyce finansowej i innych teoriach pojawia się rozkład chi-kwadrat. Najbardziej ogólna wersja *niecentalnego rozkładu chi-kwadrat z d stopniami swobody i parametrem niecentralności* λ (tutaj zakładamy jedynie $d > 0$ i $\lambda > 0$) ma gęstość postaci (5.6) z

$$p_n = \frac{(\frac{\lambda}{2})^n}{n!} e^{-\lambda/2}, \qquad g_n \sim \Gamma(j + \frac{1}{2}, \frac{1}{2}).$$

W przypadku gdy d jest naturalne, jest to gęstość rozkładu $X = \sum_{j=1}^{d} (Z_j + a_j)^2$, gdzie $\lambda = \sum_{j=1}^{d} a_j^2$ oraz $Z_1, \ldots, Z_d$ są niezależne o jednakowym rozkładzie $\mathcal{N}(0, 1)$. Można pokazać, że

$$X =_{\mathrm{d}} Y + W, \tag{5.7}$$

gdzie Y, W są niezależne, Y ma rozkład chi-kwadrat z $d - 1$ stopniami swobody oraz $W =_{\mathrm{d}} (Z + \sqrt{\lambda})^2$.

Przykład 5.2 [Dagpunar [9]] W fizyce atomowej rozpatruje się rozkład z gęstością

$$f(x) = C \sinh((xc)^{1/2}) e^{-bx}, \qquad x \geq 0$$

gdzie

$$C = \frac{2b^{3/2} e^{-c/(4b)}}{\sqrt{\pi c}} \ .$$

Po rozwinięciu w szereg $\sinh((xc)^{1/2})$ mamy

$$\begin{aligned} f(x) &= C\sum_{j=0}^{\infty}\frac{(xc)^{(2j+1)/2}e^{-bx}}{(2j+1)!} \\ &= C\sum_{j=0}^{\infty}\frac{\Gamma(j+\frac{2}{3})c^{j+\frac{2}{2}}}{b^{j+\frac{3}{2}}(2j+1)!}\frac{b^{j+3/2}x^{j+\frac{1}{2}}e^{-bx}}{\Gamma(j+\frac{2}{3})} \\ &= \sum_{j=0}^{\infty}C\frac{\Gamma(j+\frac{2}{3})c^{j+\frac{1}{2}}}{b^{j+\frac{2}{3}}(2j+1)!}g_j(x)\ , \end{aligned}$$

gdzie $g_j(x)$ jest gęstością rozkładu $\Gamma(j+\frac{3}{2},b)$. Teraz korzystając z tego, że

$$\Gamma(j+\frac{3}{2}) = (j-1+\frac{3}{2})(j-2+\frac{3}{2})\ldots(1+\frac{3}{2})(\frac{3}{2})\Gamma(\frac{3}{2})$$

(patrz Abramowitz & Stegun [2] wzór 6.1.16) oraz tego, że $\Gamma(\frac{3}{2}) = \pi^{\frac{1}{2}}/2$ (tamże, wzór 6.1.9) mamy

$$\begin{aligned} f(x) &= C\sum_{j=0}^{\infty}\frac{\pi^{1/2}c^{j+\frac{1}{2}}}{2^{2j+1}j!b^{j+\frac{3}{2}}}g_j(x) \\ &= \sum_{j=0}^{\infty}\left(\frac{c}{4b}\right)^j/j!e^{-c/(4b)}g_j(x)\ , \end{aligned}$$

czyli jest to mieszanka gęstości $\Gamma(j+\frac{3}{2},b)$, (p_j) będące rozkładem Poissona z parametrem $c/(4b)$. Teraz jeszcze warto zauważyć, że jesli Y jest liczbą losową o rozkładzie Gamma$(j+\frac{3}{2},\frac{1}{2})$ to $Y/(2b)$ liczbą losową o szukany rozkładzie Gamma$(j+\frac{3}{2},b)$. Zauważmy na koniec, że Gamma$(j+\frac{3}{2},\frac{1}{2})$ jest rozkładem chi kwadrat z $2j+1$ stopniami swobody. Jak pokazaliśmy wcześniej taki rozkład prosto się symuluje.

6 Metoda eliminacji

Podamy teraz pewną ogólną metodę na generowanie zmiennych losowych o pewnych rozkładach, zwaną metodą eliminacji. Metoda eliminacji pochodzi od von Neumanna. Zaczniemy od zaprezentowania jej dyskretnej wersji. Przypuśćmy, ze chcemy generować zmienną losową X z funkcją prawdopdobieństwa $(p_j, j\in E)$, gdzie E jest przestrzenią wartości X (na przykład $\mathbb{Z}$,

$\mathbb{Z}_+$,$\mathbb{Z}^2$, itd), podczas gdy znamy algorytm generowania liczby losowej Y z funkcją prawdopdobieństwa $(p_j^{'}, j \in E)$ oraz gdy istnieje $c > 0$ dla którego

$$p_j \leq cp_j^{'}, \qquad j \in E.$$

Oczywiście takie c zawsze znajdziemy gdy E jest skończona, ale tak nie musi być gdy E jest przeliczalna; polecamy czytelnikowi podać przykład. Oczywiście $c > 1$, jeśli funkcje prawdopodobieństwa są różne (dlaczego ?). Generujemy niezależne $U \sim \mathrm{U}(0,1)$ i Y z funkcją prawdopodobieństwa $(p_j^{'}, j \in E)$ i definiujemy *obszar akceptacji* przez

$$A = \{c\, U\, p_Y^{'} \leq p_Y\}.$$

Podstawą naszego algorytmu będzie następujący fakt.

Fakt 6.1

$$\mathbb{P}(Y = j|A) = p_j, \qquad j \in E$$

oraz $\mathbb{P}(A) = 1/c$

Dowód Bez straty ogólności możemy założyć, że $p_j^{'} > 0$. Jeśli by dla jakiegoś j mielibyśmy $p_j^{'} = 0$, to wtedy również $p_j = 0$ i ten stan można wyeliminować. Dalej mamy

$$\begin{aligned}
\mathbb{P}(Y = j|A) &= \frac{\mathbb{P}(\{Y = j\} \cap \{cUp_Y^{'} \leq p_Y\})}{\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(Y = j, U \leq p_j/(cp_j^{'}))}{\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(Y = j)\mathbb{P}(U \leq p_j/(cp_j^{'}))}{\mathbb{P}(A)} \\
&= p_j^{'}\frac{p_j/(cp_j^{'})}{\mathbb{P}(A)} = p_j\frac{c}{\mathbb{P}(A)}\ .
\end{aligned}$$

Teraz musimy skorzystać z tego, że

$$1 = \sum_{j \in E} \mathbb{P}(Y = j|A) = \frac{1}{c\mathbb{P}(A)}.$$

□

Następujący algorytm podaje przepis na generowanie liczby losowej X z funkcją prawdopdobieństwa $(p_j, j \in E)$.

Algorytm RM-d

```
% dane f. prawd. (p(j)) i  (p'(j))
1. generuj Y z funkcją prawdopodobieństwa (p'(j));
2. jeśli c rand p'(Y)<= p(Y), to podstaw X:=Y;
3. w przeciwnym razie idź do 1.
```

Widzimy, że liczba powtórzeń L w pętli ma rozkład obcięty geometryczny TGeo($\mathbb{P}(A)$)=TGeo($1-\frac{1}{c}$); $\mathbb{P}(L=k)=\frac{1}{c}(1-\frac{1}{c})^{k-1}$, $k=1,2,\ldots$. Oczekiwana liczba powtórzeń wynosi c, a więc należy dążyć aby c było jak najmniejsze.

Podamy teraz wersję ciągłą metody eliminacji. Przypuśćmy, ze chcemy generować liczbę losową X z gęstością $f(x)$ na E (możemy rozważać $E=\mathbb{R}_+,\mathbb{R},\mathbb{R}^d$, itd.), podczas gdy znamy algorytm generowania liczby losowej Y o wartościa w E z gęstością $g(x)$. Niech $c>0$ będzie takie, że

$$f(x)\le cg(x), \qquad x\in E.$$

Oczywiście $c>1$, jeśli gęstości są różne. Niech $U\sim \mathrm{U}(0,1)$ i definiujemy *obszar akceptacji* przez

$$A=\{cUg(Y)\le f(Y)\}.$$

Zakładamy, że U i Y są niezależne. Podstawą naszego algorytmu będzie następujący fakt, który wysłowimy dla $E\subset\mathbb{R}^n$.

Fakt 6.2

$$\mathbb{P}(Y\in B|A)=\int_B f(y)\,dy, \qquad dla \quad B\subset E$$

oraz $\mathbb{P}(A)=1/c$

Dowód Mamy

$$\begin{aligned}\mathbb{P}(Y\in B|A) &= \frac{\mathbb{P}(Y\in B, cUg(Y)\le f(Y))}{\mathbb{P}(A)}\\ &= \frac{\int_B \frac{f(y)}{cg(y)}g(y)\,dy}{\mathbb{P}(A)}\\ &= \int_B dy\frac{1}{c\mathbb{P}(A)}.\end{aligned}$$

Podstawiając $B = E$ widzimy, że $\mathbb{P}(A) = 1/c$. Ponieważ równość jest dla dowolnych $B \subset E$, więc f jest szukaną gęstością, co kończy dowód. □

Następujący algorytm podaje przepis na generowanie liczby losowej X z gęstością $f(x)$.

Algorytm RM

```
%zadana gęstość f(x)
1. generuj Y z gęstością g(y);
2. jeśli c* rand * g(Y)<= f(Y) to podstaw X:=Y
3. w przeciwnym razie idź do 1.
```

Przykład 6.3 [Normalna zmienna losowa] Pokażemy, jak skorzystać z metody eliminacji aby wygenerować zmienną losową X o rozkładzie normalnym $\mathrm{N}(0,1)$. Najpierw wygenerujemy zmienną losową Z z gęstością

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad 0 < x < \infty \,. \tag{6.8}$$

Niech $g(x) = e^{-x}, x > 0$ będzie gęstością rozkładu wykładniczego $\mathrm{Exp}(1)$. Należy znaleźć maksimum funkcji

$$\frac{f(x)}{g(x)} = \frac{2}{\sqrt{2\pi}} e^{x-\frac{x^2}{2}}, \qquad 0 < x < \infty \,,$$

co się sprowadza do szukania wartości dla której $x - x^2/2$ jest największa w obszarze $x > 0$. Maksimum jest osiągnięte w punkcie 1 co pociąga

$$c = \max \frac{f(x)}{g(x)} = \sqrt{\frac{2e}{\pi}} \approx 1.32 \,.$$

Akceptacja jest przy $U \leq f(Y)/(cg(Y)$. Zauważmy, że przy symulacjach możemy wykorzystać

$$\frac{f(x)}{cg(x)} = e^{x-x^2/2-1/2} = e^{-(x-1)^2/2} \,.$$

Teraz aby wygenerować zmienną losową X o rozkładzie $\mathcal{N}(0,1)$ musimy zauważyć, że należy wpierw wygenerować Z z gęstoścla $f(x)$ i następnie wylosować znak plus z prawdopodobieństwem $1/2$ i minus z prawdopodobieństwem $1/2$ (patrz Zadanie 9.14). To prowadzi do następującego algorytmu na generowanie liczby losowej o rozkładzie $\mathcal{N}(0,1)$.

Algorytm RM-N

```
%  na wyjściu X o rozkladzie N(0,1)
1. generuj Y=-\log rand;
2. jeśli U<= exp(-(Y-1)^2/2) to podstaw Z:=Y;
3. w przeciwnym razie idź do 1.
4. generuj  I=2*floor(2*rand)-1;
5. Podstaw X:=I*Z;
```

Przykład 6.4 [Rozkład Gamma(α, λ); Madras [33]]

Przypadek $\alpha < 1$ Bez straty ogólności można założyć $\lambda = 1$ (Dlaczego? Jak otrzymać z X z gęstością $f(x)$ o rozkładzie Gamma$(\alpha, 1)$ zmienną Y o rozkładzie Gamma(α, β)? Patrz Zadanie 9.14). Teraz za $g(x)$ weźmiemy

$$g(x) = \begin{cases} Cx^{\alpha-1}, & 0 < x < 1 \\ Ce^{-x}, & x \geq 1 \end{cases}$$

Z warunku, że $\int_0^\infty g(x)\,dx = 1$ wyznaczamy

$$C = \frac{\alpha e}{\alpha + e}\ .$$

Jeśli więc położymy $g_1(x) = \alpha x^{\alpha-1}\mathbf{1}(0 < x < 1)$ i $g_2(x) = ee^{-x}\mathbf{1}(x \geq 1)$ to

$$g(x) = p_1 g_1(x) + p_2 g_1(x)\ ,$$

gdzie $p_1 = C/\alpha$ oraz $p_2 = C/e$. A więc możemy symulować liczbę losową Y z gęstością g używając metody superpozycji. Zostawiamy czytelnikowi napisanie procedury generowania $g_1(x)$ oraz $g_2(x)$. Teraz należy zauważyć, że istnieje c takie, że

$$f(x) \leq cg(x), \qquad x \in \mathbb{R}_+,$$

co pozwala wykorzystać metodę eliminacji do symulacji zmiennej X. Zostawiamy czytelnikowi napisanie całej procedury.

Przypadek $\alpha > 1$**.** Oczywiście przypadek $\alpha = 1$ jest nam dobrze znany, ponieważ wtedy mamy do czynienia z liczbą losową o rozkładzie wykładniczym. Jeśli $\alpha > 1$, to przyjmujemy $g(x) = \lambda e^{-\lambda x}\mathbf{1}(0 \le x < \infty)$ z $\lambda < 1$. Oczywiście w tym przypadku

$$c = \sup_{x \ge 0} \frac{f(x)}{g(x)} = \sup_{x \ge 0} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-(1-\lambda)x} < \infty \ .$$

Stałą λ powinno się dobrać tak aby c było minimalne. [2]

7 Generowanie zmiennych gaussowskich

7.1 Metody ad hoc dla N(0,1).

Przedstawimy dwa warianty metody, zwanej Boxa-Mullera, na generowanie pary niezależnych zmiennych losowych Z_1, Z_2 o jednakowym rozkładzie normalnym N(0,1). Pomysł polega na wykorzystaniu współrzędnych biegunowych. Zacznijmy od udowodnienia lematu.

Lemat 7.1 *Niech Z_1, Z_2 będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie normalnym N(0,1). Niech (R, Θ) będzie przedstawieniem punktu (Z_1, Z_2) we współrzędnych biegunowych. Wtedy R i Θ są niezależne, R ma gęstośc*

$$f(x) = xe^{-x^2/2}1(x > 0) \tag{7.9}$$

oraz Θ ma rozkład jednostajny $U(0, 2\pi)$.

Dowód Oczywiście, (Z_1, Z_2) można wyrazić z (R, Θ) przez

$$\begin{aligned} Z_1 &= R\cos\Theta, \\ Z_2 &= R\sin\Theta \end{aligned}$$

a wektor zmiennych losowych Z_1, Z_2 ma łączną gęstość $\frac{1}{2\pi}\exp(-(x_1^2+x_2^2)/2)$. Skorzystamy z wzoru (XI.1.1). Przejście do współrzędnych biegunowych ma

[2] Proponujemy czytelnikowi znalezienie optymalnej λ (dla której c jest minimalne). Można pokazać, że

$$c = \frac{(\alpha-1)^{\alpha-1}}{(1-\lambda)^{\alpha-1}} e^{-(\alpha-1)}.$$

jakobian

$$\begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

a więc (R, Θ) ma gęstość

$$f_{(R,\Theta)}(r,\theta) = \frac{1}{2\pi} e^{-\frac{r^2}{2}} r\, dr d\theta, \qquad 0 < \theta \leq 2\pi, \quad r > 0.$$

□

Zauważmy, że rozkład zadany wzorem (7.9) nazywa się *rozkładem Raleigha.*

W następującym lemacie pokażemy jak generować zmienne o rozkładzie Raleigha.

Lemat 7.2 *Jeśli R ma rozkład Raleigha, to $Y = R^2$ ma rozkład wykładniczy Exp(1/2). W konsekwencji, jeśli U ma rozkład jednostajny U(0,1), to zmienna losowa $(-2\log U)^{1/2}$ ma rozkład Raleigha.*

Dowód W tym celu zauważmy, że $Y = R^2$ ma rozkład wykładniczy $\mathrm{Exp}(1/2)$ bo

$$\begin{aligned} \mathbb{P}(R^2 > x) &= \mathbb{P}(R > \sqrt{x}) \\ &= \int_{\sqrt{x}}^{\infty} r e^{-r^2/2}\, dr = e^{-x/2}, \qquad x > 0. \end{aligned}$$

Ponieważ $(-2\log U)$ ma rozkład wykładniczy $\mathrm{Exp}(1/2)$, więc jak wcześniej zauważylismy, pierwiastek z tej zmiennej losowej ma rozkład Raleigha. □

Fakt 7.3 *Niech U_1 i U_2 będą niezależnymi zmiennymi loswymi o jednakowym rozkładzie jednostanym $U(0,1)$. Niech*

$$\begin{aligned} Z_1 &= (-2\log U_1)^{1/2}\cos(2\pi U_2), \\ Z_2 &= (-2\log U_1)^{1/2}\sin(2\pi U_2)\ . \end{aligned}$$

Wtedy Z_1, Z_2 są niezależnymi zmiennymi losowymi o jednakowym rozkładzie normalnym N(0,1).

Algorytm BM

```
%  na wyjściu niezalezne Z1, Z2 o rozkladzie N(0,1)
   Generuj U1,U2
   R<- -2*log U1
   Z1 <- sqrt(R)* cos V
   Z2 <- sqrt(R)* sin V
   zwrot Z1,Z2
```

Dowód Mamy $Z_1 = R\cos\Theta$, $Z_2 = R\sin\Theta$, gdzie R, Θ są niezależne, R ma rozkład Raleigha a Θ jednostajny na $(0, 2\pi]$. Trzeba wygenerować zmienną Θ o rozkładzie jednostajnym na $(0, 2\pi]$ (proste) oraz R z gęstościa $re^{-\frac{r^2}{2}}$, $r > 0$ co wiemy jak robić z lematu 7.2

Druga metoda jest modyfikacją Box-Mullera i pochodzi od Marsaglia-Breya. Opiera się ona na następującym fakcie.

Fakt 7.4 *Niech*

$$\begin{aligned} Y_1 &= \{-2\log(V_1^2 + V_2^2)\}^{1/2} \frac{V_1}{(V_1^2 + V_2^2)^{1/2}}, \\ Y_2 &= \{-2\log(V_1^2 + V_2^2)\}^{1/2} \frac{V_2}{(V_1^2 + V_2^2)^{1/2}}. \end{aligned}$$

i V_1, V_2 są niezależnymi zmiennymi losowymi o jednakowym rozkładzie jednostajnym U(-1,1). Wtedy

$$(Z_1, Z_2) =_{\mathrm{d}} ((Y_1, Y_2)|V_1^2 + V_2^2 \leq 1)$$

jest parą niezależnych zmiennych losowych o jednakowym rozkładzie normalnym N(0,1).

Dowód powyższego faktu wynika z następującego lematu. Niech K będzie kołem o promieniu 1 i środku w początku układu $\mathbb{R}^2$.

Lemat 7.5 *Jeśli $(W_1, W_2) \sim U(K)$, i (R, Θ) jest przedstawieniem punktu (W_1, W_2) w współrzędnych biegunowych, to R i Θ są niezależne, R^2 ma rozkład jednostajny U(0,1) oraz Θ ma rozład jednostajny $U(0, 2\pi)$.*

Dowód Mamy

$$W_1 = R\cos\Theta, \qquad W_2 = R\sin\Theta\ .$$

Niech $B = \{a < r \le b, t_1 < \theta \le t_2\}$. Wtedy

$$\begin{aligned}
\mathbb{P}(t_1 < \Theta \le t_2, a < R \le b) &= \mathbb{P}((W_1, W_2) \in B) \\
&= \frac{1}{\pi}\int_B 1((x,y) \in K)\, dx\, dy \\
&= \frac{1}{\pi}\int_{t_1}^{t_2}\int_a^b r\, dr\, d\theta \\
&= \frac{1}{2\pi}(t_2 - t_1) \times (b^2 - a^2) \\
&= \mathbb{P}(t_1 < \Theta \le t_2)\mathbb{P}(a < R \le b).
\end{aligned}$$

Stąd mamy, ze R i Θ są niezależne, Θ o rozkładzie jednostajnym $\mathrm{U}(0, 2\pi)$. Ponadto $\mathbb{P}(R \le x) = x^2 (0 \le x \le 1)$ czyli R^2 ma rozkład jednostajnym U(0,1).

Algorytm MB

```
%  na wyjściu niezalezne Z1, Z2 o rozkladzie N(0,1)
   while (X>1)
   generate U1,U2
   U1 <- 2*U1-1, U2 <- 2*U2-1
   X = U1^2+U2^2
   end
   Y <- sqrt((-2* log X)/X)
   Z1 = U1*Y, Z2=U2*Y
   return Z1,Z2
```

7.2 Metoda ITM dla zmiennych normalnych

[3] Aby generować standardowe zmienne normalne $\mathcal{N}(0,1)$ metodą ITM musimy umieć odwrócić dystrybuantę $\Phi(x)$. Niestety w typowych pakietach nie ma zaszytej funkcji $\Phi^{-1}(u)$ (tutaj mamy klasyczną funkcję odwrotną). Istnieją rozmaite algorytmy obliczające $\Phi^{-1}(u)$. Jak wszystkie algorytmy zaszyte

[3] Glasserman [16] str. 67.

w pakietach lub kalkulatorach podają przybliżone wartości z pewnym błędem, który powinien być podany. Zagadnieniem jest rozwiązanie równania $\Phi(x) = u$ dla zadanego $u \in (0,1)$. Punktem wyjścia może być standardowa metoda Newtona szukania zera funkcji $x \to \Phi(x) - u$. Zauważmy najpierw, że ponieważ

$$\Phi^{-1}(1-u) = -\Phi(u)$$

więc wystarczy szukać rozwiązania albo dla $u \in [1/2, 1)$ lub $(0, 1/2)$. Metoda Newtona jest metodą iteracyjną zadaną rekurencją

$$\begin{aligned} x_{n+1} &= x_n - \frac{\Phi(x_n) - u}{\phi(x_n)} \\ &= x_n + (u - \Phi(x_n))e^{0.5x_n^2 + c}, \end{aligned} \tag{7.10}$$

gdzie $c = \log\sqrt{2\pi}$. Sugeruje się aby zacząć z punktu

$$x_0 = \pm\sqrt{|-1.6\log(1.0004 - (1-2u)^2)|}.$$

Powyzej znak zależy od tego czy $u \geq 1/2$ (plus) czy $u < 1/2$ (minus). Powyższa procedura wymaga znajomości funkcji Φ i oczywiście wykładniczej. Poniżej przedstawiamy inny algorytm, tzw. algorytm Beasley'a-Zamana-Moro (za książką Glassermana [16]).

function x=InvNorm(u)

```
a0=2.50662823884;
a1=-18.61500062529;
a2=41.39119773534;
a3=-25.44106049637;

b0=-8.47351093090;
b1=23.08336743743;
b2=-21.06224101826;
b3=3.13082909833;

c0=0.3374754822726147;
c1=0.9761690190917186;
c2=0.1607979714918209;
c3=0.0276438810333863;
c4=0.0038405729373609;
```

```
c5=0.0003951896511919;
c6=0.0000321767881768;
c7=0.0000002888167364;
c8=0.0000003960315187;

y<-u-0.5;
if |y|<0.42;
r <- y*y
x <- (((a3*r+a2)*r+a1)*r+a0)/((((b3*r+b2)*r+b1)*r+b0)+1)
else
r <- u
if (y>0) r <- 1-u
r <- log (-log)(r)
x <- c0+r*(c1+r*(c2+r*(c3+r*(c4+r*(c5+r*(c6+r*(c7+r*c8)))))))
if (y<0) x<- -x
return x
```

Aby to jeszcze polepszyć mozna zastosować do wyniku jeszcze jeden krok według rekurencji (7.10) wychodząc z otrzymanego $x = \Phi^{-1}(u)$. Glasserman [16] podaje, ze wtedy błąd nie przekracza 10^{-15}.

7.3 Generowanie wektorów losowych $\mathrm{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$.

Wektor losowy $\boldsymbol{X}$ o rozkładzie wielowymiarowym normalnym $\mathrm{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$ ma średnią $\boldsymbol{m}$ oraz macierz kowariancji $\boldsymbol{\Sigma}$ oraz dowolna kombinacja $\sum_{j=1}^{n} a_j X_j$ ma rozkład normalny. Oczywiście gdy mówimy o n-wymiarowym wektorze normalnym $\boldsymbol{X}$, to $\boldsymbol{m}$ jest n wymiarowym wektorem, $\boldsymbol{\Sigma}$ jest macierzą $n \times n$. Przypomnijmy też, że macierz kowariancji musi być symetryczna i nieujemnie określona. W praktyce tego wykładu będziemy dodatkowo zakładać, że macierz $\boldsymbol{\Sigma}$ jest niesingularna. Wtedy rozkład $\boldsymbol{X}$ ma gęstość zadaną wzorem

$$f(x_1, \ldots, x_n) = \frac{1}{(2\pi \det(\boldsymbol{\Sigma}))^{n/2}} e^{-\frac{\sum_{j,k=1}^{n} c_{ij} x_j x_k}{2}}.$$

gdzie

$$\boldsymbol{C} = (c_{jk})_{j,k=1}^{n} = \boldsymbol{\Sigma}^{-1}$$

Przykład 7.6 Dla $n = 2$ macierz kowariancji można przedstawić jako

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

gdzie $\rho = \mathrm{Corr}(X_1, X_2)$, $\sigma_i^2 = \mathrm{Var}\, X_i$ $(i = 1, 2)$. Wtedy

$$\boldsymbol{C} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} .$$

Aby wygerować taki dwuwymiarowy wektor normalny $\boldsymbol{X} \sim \mathrm{N}(0, 1)$, niech

$$Y_1 = \sigma_1(\sqrt{1-|\rho|}Z_1 + \sqrt{|\rho|}Z_3), \quad Y_2 = \sigma_2(\sqrt{1-|\rho|}Z_2 + \mathrm{sign}(\rho)\sqrt{|\rho|}Z_3) ,$$

gdzie Z_1, Z_2, Z_3 są niezależnymi zmiennymi losowymi o rozkładzie $N(0, 1)$. Teraz dostaniemy szukany wektor losowy $(X_1, X_2) = (Y_1 + m_1, Y_2 + m_2)$. To że (Y_1, Y_2) ma dwuwymiarowy rozkład normalny jest oczywiste. Należy pokazać, że macierz kowariancji tego wektora jest taka jak trzeba.

Dla dowolnego n, jeśli nie ma odpowiedniej struktury macierzy kowariancji $\boldsymbol{\Sigma}$ pozwalającej na znalezienie metody ad hoc polecana jest nastepująca metoda generowania wektora losowego $\mathrm{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$. Niech $\boldsymbol{X} = (X_1, \ldots, X_n)$ oraz macierz $\boldsymbol{A}$ będzie taka, że

$$\boldsymbol{\Sigma} = (\boldsymbol{A})^T \boldsymbol{A}. \tag{7.11}$$

Macierz $\boldsymbol{A}$ możemy nazywać pierwiastkiem z macierzy $\boldsymbol{\Sigma}$. Każda macierz kowariancji jest symetryczną macierzą nieujemnie określoną, a wiadomo, że dla takich macierzy istnieje pierwiastek.

Fakt 7.7 *Jeśli $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie normalnym $N(0, 1)$ to*

$$\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{A} + \boldsymbol{m}$$

ma rozkład $N(\boldsymbol{m}, \boldsymbol{\Sigma})$.

Aby obliczyć 'pierwiastek" $\boldsymbol{A}$ z macierzy kowariancji $\boldsymbol{\Sigma}$ można skorzystać z gotowych metod numerycznych. Na przykład w MATLAB-ie jest instrukcja chol($\boldsymbol{\Sigma}$), która używa metody faktoryzacji Choleskiego. W wyniku otrzymujemy macierz górnotrójkątną. Zauważmy, że jeśli Σ jest nieosobliwa, to istnieje jedyna A dla której faktoryzacja (7.11) zachodzi i na przekątnej macierz A ma dodatnie elementy. W przypadku gdy Σ jest singularna to taki rozkład nie jest jednoznaczny.

Proponujemy następujący eksperyment w MATLABIE.

```
>> Sigma=[1,1,1;1,4,1;1,1,8]

Sigma =

     1     1     1
     1     4     1
     1     1     8

>> A=chol(Sigma)

A =

    1.0000    1.0000    1.0000
         0    1.7321         0
         0         0    2.6458

>> A'*A

ans =

    1.0000    1.0000    1.0000
    1.0000    4.0000    1.0000
    1.0000    1.0000    8.0000
```

8 Przykładowe symulacje

W dwóch wykresach na rys. 8.2 i 8.3 podajemy wyniki S_n/n, $n = 1, \ldots, 5000$ dla odpowiednio rozkład wykładniczego i Pareto ze średnią 1. Zmienne Pareto są w tym celu pomnożone przez 0.2. Warto zwrócić uwagę na te symulacje. Widać, że dla zmiennych Pareto zbieżność jest dramatycznie wolniejsza. Jest to efekt ciężko-ogonowości.

Na rys. 8.4 podane sa wyniki symulacji, kolejno 5000 replikacji rozkładu normalnego N(1,1), 5000 replikacji rozkładu wykładniczego Exp(1), oraz 5000 replikacji rozkładu Pareto(1.2) przemnożonego przez 0.2. Polecamy zwrócić uwagę na charakter poszczególnych fragmentów na rys. 8.4. Chociaż wszystkie zmienne mają średnią 1, to w pierwszym fragmencie (dla zmiennych o rozkładzie normalnym) zmienne są mało rozrzucone, w drugiej części

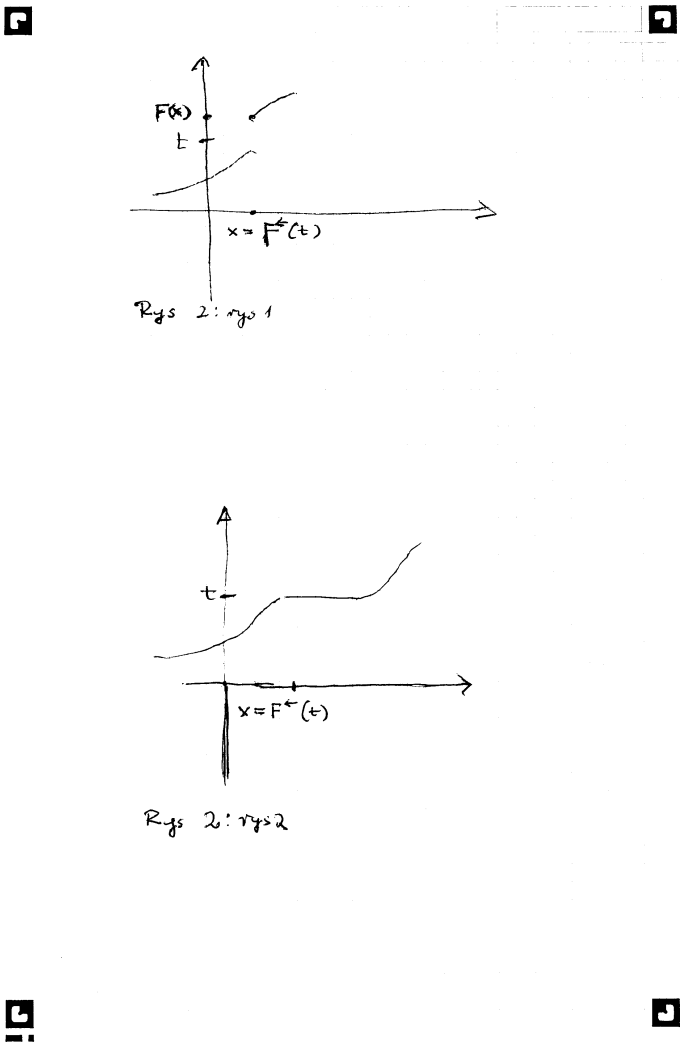
trochę bardziej, natomiast duży rozrzut można zaobserwować w części ostatniej (tj. dla zmiennych Pareto). Jest to spowodowane coraz cięższym ogonem rozkładu. Do symulacji zmiennych losowych o rozkładzie normalnym N(0,1) użyto MATLABowskiej komendy randn. Aby uzyskać zmienne N(1,1) trzeba wziąć randn+1.

Dla ilustracji na rysunkach 8.5, 8.6 i 8.7 podane są przykładowe symulacje dwuwymiarowego wektora normalnego z zerowym wektorem średniej, $\mathrm{Var}\,(X_1) = 3$, $\mathrm{Var}\,(X_2) = 1$ i współczyniku korelacji $\rho = 0.1, 0.6, 0.9$.[4]

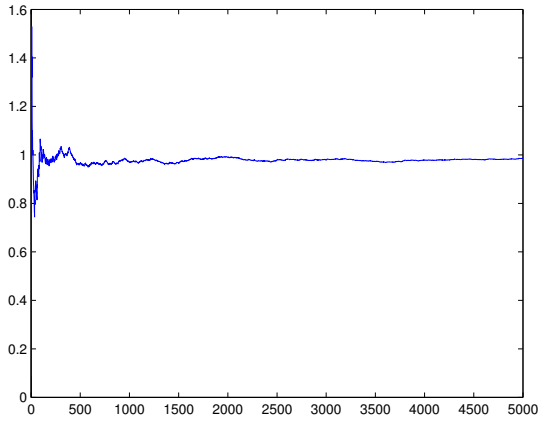
8.1 Uwagi bibliograficzne

Na temat generowania zmiennych losowych o zadanych rozkładach istnieje obszerna literatura książkowa: patrz e.g. elementarny podręcznik Rossa [40], fundamentalna monografia Asmussena i Glynna [4], obszerna monografia Fishmana [14], [9], bardzo ciekawy nieduży podręcznik Madrasa [33].

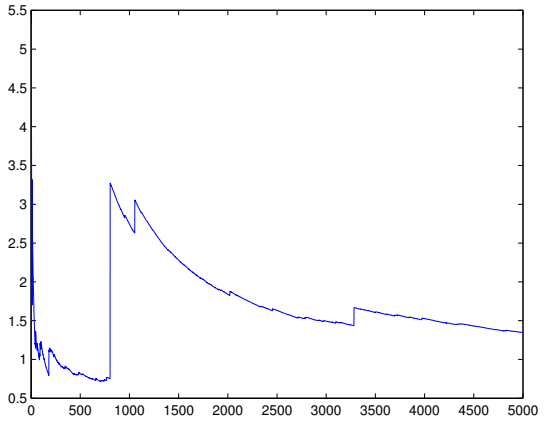
[4] skrypt: normalnyRho.m



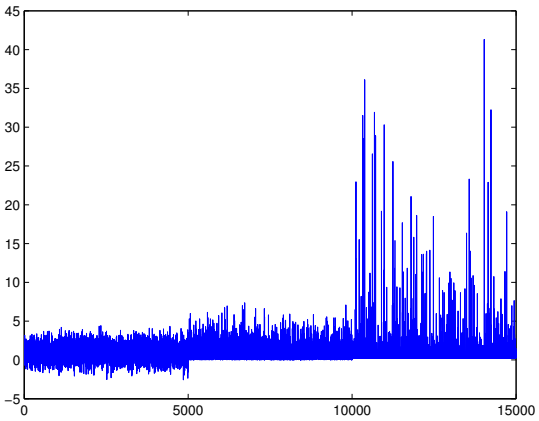
Rysunek 1.1: Rys 2-1



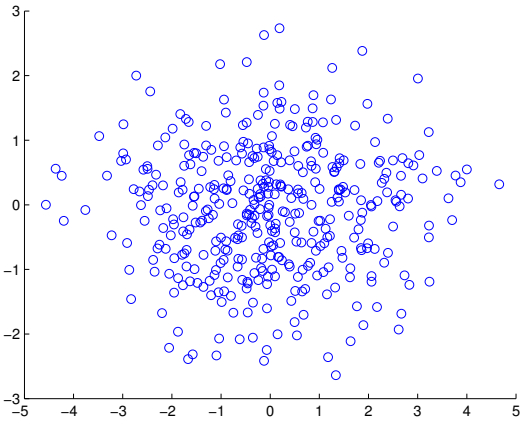
Rysunek 8.2: Rozkład wykladniczy Exp(1), 5000 replikacji



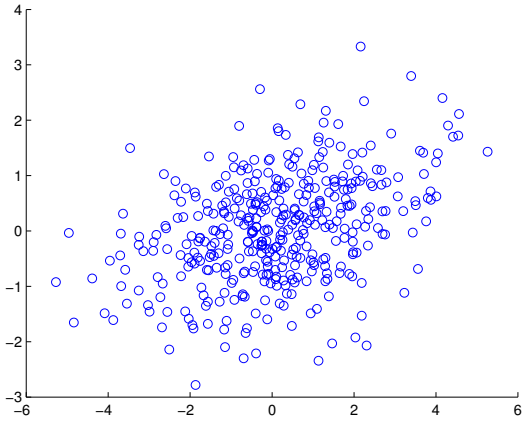
Rysunek 8.3: Rozkład Pareto(1.2), 5000 replikacji



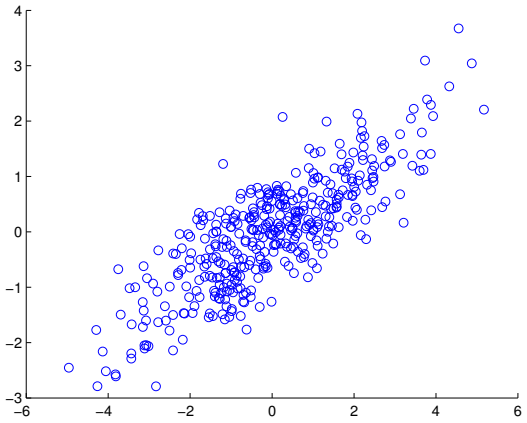
Rysunek 8.4: Rozkład N(1,1) (5000 repl.), Exp(1) (5000 repl.), Pareto(1.2), (5000 repl.)



Rysunek 8.5: Wektor normalny, $\rho = 0.1$, $\sigma_1^2 = 3$, $\sigma_1^2 = 1$, 400 replikacji



Rysunek 8.6: Wektor normalny, $\rho = 0.6$, $\sigma_2^2 = 3$, $\sigma_1^2 = 1$, 400 replikacji



Rysunek 8.7: Wektor normalny, $\rho = 0.9$, $\sigma_1^2 = 3$, $\sigma_2^2 = 1$, 400 replikacji

9 Zadania

[5]

Zadania teoretyczne

9.1 Niech X ma dystrybuantę F i chcemy wygenerować zmienna losową o warunkowym rozkładzie $X|X \in (a,b)$ gdzie $\mathbb{P}(X \in (a,b)) > 0$. Niech

$$V = F(a) + (F(b) - F(a))U.$$

Jaki rozkład ma V. Pokazać, że $Y = F^{-1}(V)$ ma żądaną warunkową dystrybuantę G, gdzie

$$G(x) = \begin{cases} 0 & x < a, \\ \frac{F(x)-F(a)}{F(b)-F(a)}, & a \le x < b, \\ 1 & x \ge b \end{cases}$$

9.2 Podać procedurę generowania liczb losowych o rozkładzie *trójkątnym* Tri(a,b), z gęstością

$$f(x) = \begin{cases} c(x-a), & a < x < (a+b)/2, \\ c((b-x), & (a+b)/2 < x < b, \end{cases}$$

gdzie stała normujaca $c = 4/(b-a)^2$. Wsk. Rozważyć najpierw gęstość $U_1 + U_2$.

9.3 Pokazać, że

$$X = \log(U/(1-U))$$

ma rozkład *logistyczny* o dystrybuancie $F(x) = 1/(1+e^{-x})$.

9.4 Podać dwie procedury na generowanie liczby losowej z gęstością postaci

$$f(x) = \sum_{j=1}^{N} a_j x^j, \qquad 0 \le x \le 1,$$

gdzie $a_j > 0$. Wsk. Obliczyć gęstość $\max(U_1, \ldots, U_n)$ i następnie użyć metody superpozycji. Jaki warunek muszą spełniać (a_i)?

[5] Poprawiane 1.2.2016

9.5 Podać procedurę generowania zmiennej losowej X z gęstością $f(x) = n(1-x)^{n-1}$.

9.6 Podać dwie procedurę generowania liczby losowej o rozkładzie Cauchego.

9.7 Niech F jest dystrybuantą rozkładu Weibulla $\mathrm{W}(\alpha, c)$,

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \exp(-cx^{\alpha}), & x \geq 0\,. \end{cases}$$

Zakładamy $c > 0, \alpha > 0$. Pokazać, że

$$X = \left(\frac{-\log U}{c} \right)^{1/\alpha}$$

ma rozkład $\mathrm{W}(\alpha, c)$.

9.8 Podać przykład pokazujący, że c w metodzie eliminacji (zarówno dla przypadku dyskretnego jak i ciągłego) nie musi istnieć skończone.

9.9 a) Udowodnić, że $\mathbb{P}(\lfloor U^{-1} \rfloor = i) = \frac{1}{i(i+1)}$, dla $i = 1, 2, \ldots$.
b) Podać algorytm i obliczyć prawdopodobieństwo akceptacji na wygenerowanie metodą eliminacji dyskretnej liczby losowej X z funkcją prawdopodobieństwa $\{p_k,\ k = 1, 2, \ldots\}$, gdzie

$$p_k = \frac{6}{\pi^2} \frac{1}{k^2}, \qquad k = 1, 2, \ldots.$$

c) Podać algorytm i obliczyć prawdopodobieństwo akceptacji na wygenerowanie metodą eliminacji dyskretnej liczby losowej X z funkcją prawdopodobieństwa $\{p_k,\ k = 1, 2, \ldots\}$, gdzie $p_k = \frac{90}{\pi^4} \frac{1}{n^4}$.

9.10 Podaj metodę generowania liczby losowej z gęstością:

$$f(x) = \begin{cases} e^{2x}, & -\infty < x < 0 \\ e^{-2x}, & 0 \geq x < \infty \end{cases}$$

Taki rozkład nazywa się *podwójny wykładniczyczy.*

9.11 *Rozkład beta* Beta(α, β) ma gęstość

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \qquad 0 < x < 1.$$

gdzie beta funkcja jest zdefiniowana przez

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx\ .$$

Przy założeniu $\alpha, \beta > 1$ podać algorytm wraz z usadnieniem generowania liczb losowych Beta(α, β) metodą eliminacji. Wsk. Przyjąć $g(x) = \mathbf{1}(0 < x < 1)$.

9.12 Gęstość rozkładu zadana jest wzorem

$$f(x) = Cxe^{-x^4}, \qquad x \geq 0,$$

gdzie $C = 4/\sqrt{\pi}$. Podać procedurę generowania zmiennych losowych o rozkładzie f.

9.13 Gęstość dwuwymiarowego wektora losowego $\boldsymbol{Y} = (Y_1, Y_2)$ jest dana

$$f(x_1, x_2) = ce^{-|\boldsymbol{x}|^4} = ce^{-(x_1^2+x_2^2)^2}, \qquad \boldsymbol{x} \in \mathbb{R}^2,$$

gdzie c jest znaną stała normującą. Zaproponować algorytm na sumulację wektora losowego $\boldsymbol{Y}$. (Wsk. Przedyskutować dwie metody: eliminacji oraz przejście do współrzędnych biegunowych oraz zastanowić się którą lepiej używać).

9.14 a) Niech η będzie zmienną losową niezależną od X, $\mathbb{P}(\eta = -1) = \mathbb{P}(\eta = -1) = 1/2$, oraz X ma rozkład z gęstością daną w (6.8). Pokazać, że ηX ma rozkład standardowy normalny N(0,1).
b) Pokazać, że jeśli X ma rozkład Gamma(α,1) to X/β ma rozkład Gamma(α,β).

9.15 Podać dwa algorytmy na generowanie punktu losowego o rozkładzie U(B), gdzie $B = \{x \in \mathbb{R}^2 : |x| \leq 1\}$. Wsk. Do jednego z algorytmów wykorzystać lemat 7.5.

9.16 Niech (X, Y) będzie wektorem losowym z gęstością $f(x, y)$ i niech $B \subset \mathbb{R}^2$ będzię obszarem takim, że

$$\int_B f(x, y)\, dx\, dy < \infty.$$

Pokazać, że $X|(X, Y) \in B$ ma gęstość

$$\frac{\int_{y:(x,y)\in B} f(x, y)\, dy}{\int_B f(x, y)\, dx\, dy}.$$

9.17 [Marsaglia [34]] Niech J ma funkcję prawdopodobieństwa $(p_m)_{m=0}^{\infty}$ z $p_m = 1/(ce^{m+1})$ i $c = 1/(e-1)$. Niech I ma funkcję prawdopodobieństwa $(q_n)_{n=1}^{\infty}$, gdzie $q_n = c/n!$. Zakładamy, że I, J oraz $U_1, \ldots$ są niezależne. Pokazać, że $X = J + \min(U_1, \ldots, U_I)$ ma rozkład wykładniczy Exp(1). Powyższa własność rozkładu wykładniczego daje możliwość napisania algorytmu na generowanie wykładniczych zmiennych losowych bez użycia kosztownej funkcji log. Zastanowić się jak można szybko generować liczby losowe I, J. Wsk. Zauważyć, że I ma rozkład Poissona Poi(e) przesunięty o 1, natomiast J ma rozkład geometryczny Geo(e^{-1}). Wsk. Zauważyć, że $J \sim$Geo(e^{-1}) oraz $c_{k+1} = 1/(k+1)$ [6]

9.18 Uzasadnić dwie metody generowania rozkładu wielomianowego M$(n, \boldsymbol{a})$. Pierwsza metoda jest uogólnieniem metody ad hoc dla rozkłady dwumianowego (patrz podrozdział 3). Polega na generowaniu niezależnych wektorów losowych przyjmujących wartości $(0, \ldots, 1, \ldots 0)$ (jedynka na i-tym miejscu) z prawdopodobieństwem a_i i następnie zsumowanie liczby jedynek na każdej koordynacie. Natomiast druga metoda wykorzystuje unikalną własność rozkładu wielomianowego wektora $\boldsymbol{X} = (X_1, \ldots, X_d)$: rozkład X_{j+1} pod warunkiem $X_1 = k_1, \ldots, X_j = k_j)$ jest dwumianowy B$(n - k_1 - \ldots - k_j, a_{j+1}/(a_1 + \ldots + a_j))$ i X_1 ma rozkład B(n, a_1).

9.19 a) Pokazać, że jeśli $Z_1, \ldots, Z_n$ są niezależnymi zmiennymi o rozkładzie normalnym N(0,1) i $||Z|| = \sum_{j=1}^n Z_j^2$, to punkt $X = (Z_1/||Z||, \ldots, Z_2/||Z||)$ ma rozkład jednostajny na sferze $S_{n-1} = \{x = (x_1, \ldots, x_n) : \sum_{j=1}^n x_j^2 = 1$.

[6] A jak szybko generować liczby losowe I, J?

b) Jeśli ponadto punkt ze sfery X z punktu a) przemnożymy przez $U^{1/n}$, gdzie U i X są niezależne, to otrzymamy punkt losowy z kuli $B_n = \{x : \sum_{j=1}^{n} x_j^2 \leq 1\}$.

Zadania laboratoryjne

9.20 Podać wraz z uzasadnieniem jak generować liczbę losową X o rozkładzie U[a,b]. Napisać MATLABowska funkcję Unif(a,b) na generowanie liczby losowej X.

9.21 W literaturze aktuarialnej rozważa się *rozkład Makehama* przyszłego czasu życia T_x dla x-latka z ogonem rozkładu

$$\mathbb{P}(T_x > t) = e^{-At - m(c^{t+x} - c^x)}.$$

Napisać procedurę Makeham(A,m,c,x) generowania liczb losowych o takim rozkładzie. Przyjąć $A = 5 \cdot 10^{-4}$, $m = 7.5858 \cdot 10^{-5}$ oraz $c = \log 1.09144$(tzw. G82). Wsk. Zinterpretować probabilistycznie jaka operacja prowadzi do faktoryzacji $e^{-At-m(c^{t+x}-c^x)} = e^{-At}e^{-m(c^{t+x}-c^x)}$. Zastanowić się czy rozkład z ogonem dystrybuanty $e^{-m(c^{t+x}-c^x)}$ można przedstawić jako warunkowy.

9.22 Napisać procedure Poi(lambda) generowania liczb losowych o rozkładzie Poi(λ). Uzasadnić jej poprawność.

9.23 Napisac funkcję Gamma(a,b) na generowanie liczby losowej o rozkładzie Gamma(a, b). Wsk. Algorytm musi się składać z dwóch częśći: $a < 1$ oraz $a \geq 1$.

9.24 Przeprowadzić następujący eksperyment z MATLABowskimi generatorami. Obliczyć objętośc kuli $k = 30$ wymiarowej o promieniu 1 i porównąc z wynikiem teoretycznym na objętośc takiej kuli:

$$V_k = \frac{2}{k} \frac{\pi^{k/2}}{\Gamma(k/2)} .$$

W zadaniu należy przeprowadzić symulacje punktu z $(0, 1)^k$ i następnie sprawdzić czy leży on w kuli. Zastanowić się ile należy wziąć replikacji.

9.25 Porównać dwie procedury generowania rozkładu normalnego N(0,1) z podsekcji 7.1. Porównać szybkość tych procedur mierząc czas wykonania 100 000 replikacji dla każdej z nich.

9.26 Korzystając z zadania 9.18 porównać dwie procedury generowania rozkładu wielomianowego. Do generowania rozkładu dwumianowego wykorzystać procedurę ITR. Przeprowadzić eksperyment z $n = 10$ i $a_1 = 0.55, a_2 = \ldots = a_{10} = 0.05$. Porównać szybkość tych procedur mierząc wykonanie 100 000 replikacji dla każdej z nich.

9.27 Napisać procedurę generowania liczb losowych z gęstością

$$f(x) = pe^{-x} + \frac{(1-p)}{0.2}\left(\frac{1}{1+(x/0.2)}\right)^{1.2}.$$

Zrobić wykładniczy i paretowski wykres kwantylowy dla 300 zmiennych losowych wygenerowanych z $p = 0.05$ i $p = 0.95$.

Projekt

9.28 Napisać procedurę MATLAB-owską generowania liczby losowej N(0,1) (tablicy liczb losowych N(0,1))
(i) metodą eliminacji (randnrej),
(ii) 1-szą metodą Boxa-Mullera (randnbmfirst),
(iii) 2-gą metodą Boxa-Mullera (randnbmsec). Przeprowadzić test z pomiarem czasu 100 000 replikacji dla każdej z tych metod oraz MATLAB-owskiej randn. Użyć instrukcji TIC i TOC. Każdy eksperyment zacząć od rand('state',0). Sporządzić raport z eksperymentów. Zrobić wykresy kwantylowe dla każdej z procedur z 1000 replikacji.

9.29 Używając rozważania z ćwiczenia 9.17 napisać algorytm na generowanie liczb losowych o rozkładzie wykładniczym Exp(1). Zastanowić się jak optymalnie generować liczby losowe I, J z zadania 9.17. Przeprowadzić test z pomiarem czasu 100 000 replikacji dla tej metody i metody ITM $(-\log U)$. Każdy eksperyment zacząć od rand('state',0). Sporządzić raport z eksperymentów.

Rozdział IV

Analiza wyników; niezależne replikacje

[1] Przypuśćmy, że chcemy obliczyć wartość I o której wiemy, że można przedstawić w postaci

$$I = \mathbb{E}\, Y$$

dla pewnej zmiennej losowej Y, takiej, że $\mathbb{E}\, Y^2 < \infty$. Ponadto umiemy również generować zmienną losową Y. Zauważmy, że dla zadanego I istnieje wiele zmiennych losowych, dla których $I = \mathbb{E}\, Y$. Celem tego rozdziału będzie zrozumienie, jak najlepiej dobierać Y do konkretnego problemu. Przy okazji zobaczymy jak dobierać liczbę replikacji potrzebną do zagwarantowania żądanej dokładności.

1 Przypadek estymacji nieobciązonej

Przypuśćmy, że chcemy obliczyć wartość I o której wiemy, że można przedstawić w postaci

$$I = \mathbb{E}\, Y$$

dla pewnej liczby losowej Y, takiej, że $\mathbb{E}\, Y^2 < \infty$. W tym podrozdziale będziemy przyjmować, że $Y_1, \ldots, Y_n$ będą niezależnymi replikacjami liczby losowej Y i będziemy rozważać *estymatory* szukanej wielkości I postaci

$$\hat{Y}_n = \frac{1}{n} \sum_{j=1}^{n} Y_j.$$

[1] Poprawiane 1.2.2016

Korzystając z elementarnych faktów teorii prawdopodobieństwa, estymator $\hat{I}_n$ jest *nieobiążony* t.j.

$$\mathbb{E}\,\hat{Y}_n = I\ ,$$

oraz *mocno zgodny* to znaczy z prawdopodobieństwem 1 zachodzi zbieżność dla $n \to \infty$

$$\hat{Y}_n \to I,$$

co pociąga ważną dla nas słabą zgodność, to znaczy, że dla każdego $b > 0$ mamy

$$\mathbb{P}(|\hat{Y}_n - I| > b) \to 0.$$

Zauważmy ponadto, że warunek $|\hat{Y}_n - Y| \leq b$ oznacza, że błąd bezwzględny nie przekracza b. Stąd widzimy, że dla każdego b i liczby $0 < \alpha < 1$ istnieje n_0 (nas interesuje najmniejsze naturalne) takie, że

$$\mathbb{P}(\hat{Y}_n - b \leq I \leq \hat{Y}_n + b) \geq 1 - \alpha, \qquad n \geq n_0.$$

A więc z prawdopodobieństwem większym niż $1-\alpha$ szukana wielkość I należy do przedziału losowego $[\hat{Y}_n - b, \hat{Y}_n + b]$ i dlatego α nazywa się *poziomem istotności* lub *poziomem ufności*. Równoważnie możemy pytać się o błąd b dla zadanej liczby replikacji n oraz poziomu ufności α, lub przy zadanym n i b wyznaczyć poziom ufności α. Zgrubną odpowiedź dostaniemy z nierówności Czebyszewa.

Uwaga Niech b_n będzie błędem przy n replikacjach. Wtedy

$$\mathbb{P}(|\hat{Y}_n - I| \geq b_n) \leq \frac{\operatorname{Var} \hat{Y}_n}{b_n^2}$$

Jeśli więc chcemy aby błąd był na poziomie α, to

$$\alpha = \frac{\operatorname{Var} \hat{Y}_n}{b_n^2} = \frac{\sigma_Y^2}{n b_n^2}$$

i stąd rozwiązując równanie mamy

$$b_n = \frac{\sigma_Y / \alpha^{1/2}}{n^{1/2}}$$

czyli $b_n = O(n^{-1/2})$.

Wyprowadzimy teraz wzór bardziej dokładny wzór na związek pomiędzy n, b, α. Ponieważ w metodach Monte Carlo n jest zwykle duże, możemy skorzystać z *centralnego twierdzenia granicznego* (CTG), które mówi, że jeśli tylko $Y_1, \ldots, Y_n$ są niezależne i o jednakowym rozkładzie oraz $\sigma_Y^2 = \operatorname{Var} Y_j < \infty$, to dla $n \to \infty$

$$\mathbb{P}\left(\sum_{j=1}^{n} \frac{Y_j - I}{\sigma_Y \sqrt{n}} \leq x\right) \to \Phi(x) \ , \tag{1.1}$$

gdzie $\sigma_Y = \sqrt{\operatorname{Var} Y_1}$ oraz $\Phi(x)$ jest dystrybuantą standardowego rozkładu normalnego. Niech $z_{1-\alpha/2}$ będzie $\alpha/2$–kwantylem rozkładu normalnego tj.

$$z_{1-\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2}).$$

Stąd mamy, że

$$\lim_{n\to\infty} \mathbb{P}(-z_{1-\alpha/2} < \sum_{j=1}^{n} \frac{Y_j - I}{\sigma_Y \sqrt{n}} \leq z_{1-\alpha/2}) = 1 - \alpha.$$

A więc, pamiętając, że $I = \mathbb{E}\, Y_1$,

$$\mathbb{P}(-z_{1-\alpha/2} < \sum_{j=1}^{n} \frac{Y_j - I}{\sigma_Y \sqrt{n}} \leq z_{1-\alpha/2}) \approx 1 - \alpha \ .$$

Przekształcając otrzymujemy równoważną postać

$$\mathbb{P}(\hat{Y}_n - z_{1-\alpha/2}\frac{\sigma_Y}{\sqrt{n}} \leq I \leq \hat{Y}_n + z_{1-\alpha/2}\frac{\sigma_Y}{\sqrt{n}}) \approx 1 - \alpha \ ,$$

skąd mamy *fundamentalny związek* wiążący n, α, σ_Y^2:

$$b = z_{1-\alpha/2}\frac{\sigma_Y}{\sqrt{n}} \ . \tag{1.2}$$

Uwaga Zauważmy, że CTG można wyrazić następująco:

$$\sqrt{n}\hat{Y}_n \to_{\mathrm{d}} N(0, \sigma_Y^2).$$

Jeśli nie znamy σ_Y^2, a to jest raczej typowa sytuacja, to możemy zastąpić tę wartość przez wariancję próbkową

$$\hat{S}_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_i - \hat{Y}_n)^2,$$

ponieważ z teorii prawdopodobieństwa wiemy, że jest również prawdą

$$\mathbb{P}\left(\sum_{j=1}^{n}\frac{Y_j - I}{\hat{S}_n\sqrt{n}} \leq x\right) \to \Phi(x)\ . \tag{1.3}$$

Estymator $\hat{S}_n^2$ jest nieobciążony, tj. $\mathbb{E}\,\hat{S}_n^2 = \sigma_Y^2$ oraz zgodny. Przekształcając (1.3)

$$\mathbb{P}(-z_{1-\alpha/2} < \sum_{j=1}^{n}\frac{Y_j - I}{\hat{S}_n\sqrt{n}} \leq z_{1-\alpha/2}) \approx 1-\alpha\ ,$$

i po dalszych manipulacjach otrzymujemy równoważną postać

$$\mathbb{P}(\hat{Y}_n - z_{1-\alpha/2}\frac{\hat{S}_n}{\sqrt{n}} \leq I \leq \hat{Y}_n + z_{1-\alpha/2}\frac{\hat{S}_n}{\sqrt{n}}) \approx 1-\alpha\ .$$

Stąd możemy napisać inną wersję fundamentalnego związku

$$b = z_{1-\alpha/2}\frac{\hat{S}_n}{\sqrt{n}}$$

W teorii Monte Carlo przyjęło rozważać się $\alpha = 0.05$, i z tablic możemy odczytać, że $z_{1-0.025} = 1.96$ i tak będziemy robić w dalszej części wykładu. Jednakże zauważmy, że istnieje mozliwość wnioskowania na poziomie ufności, na przykład, 99%. Wtedy odpowiadająca $z_{1-0.005} = 2.58$. Polecamy czytelnikowi obliczenie odpowiednich kwantyli dla innych poziomów istotności wykorzystając tablicę rozkładu normalnego. A więc na poziomie istotności 0.05 fundamentalny związek jest

$$b = \frac{1.96\sigma_Y}{\sqrt{n}}. \tag{1.4}$$

A więc aby osiągnąć dokładność b należy przeprowadzić liczbę $n = 1.96^2\sigma_Y^2/b^2$ prób.

Jak już wspomnieliśmy, niestety w praktyce niedysponujemy znajomością σ_Y^2. Estymujemy więc σ_Y^2 estymatorem $\hat{S}_n^2$. Można to zrobić podczas pilotażowej symulacji, przed zasadniczą symulacją mającą na celu obliczenie I. Wtedy zapisujemy wynik symulacji w postaci przedziału ufności

$$(\hat{Y}_n - \frac{z_{1-\alpha/2}\hat{S}_n}{\sqrt{n}}, \hat{Y}_n + \frac{z_{1-\alpha/2}\hat{S}_n}{\sqrt{n}})$$

lub w formie $(\hat{Y}_n \pm \frac{z_{1-\alpha/2}\hat{S}_n}{\sqrt{n}})$.

W MATLABie srednią i odchylenie standardowe liczymy za pomocą instrukcji mean(X) oraz std(X), gdzie X jest wektorem. Jeśli potrzebujemy liczyć odchylenie standardowe samemu to przydatne są następujące wzory rekurencyjne Podamy teraz wzór rekurencyjny na obliczanie $\hat{Y}_n$ oraz $\hat{S}_n^2$.

Algorytm

$$\hat{Y}_n = \frac{n-1}{n}\hat{Y}_{n-1} + \frac{1}{n}Y_n$$

i

$$\hat{S}_n^2 = \frac{n-2}{n-1}\hat{S}_{n-1}^2 + \frac{1}{n}(Y_n - \hat{Y}_{n-1})^2$$

dla $n = 2, 3, \ldots$ i

$$\hat{Y}_1 = Y_1 \qquad \hat{S}_1^2 = 0\ .$$

Przykład 1.1 [Asmussen & Glynn [4]] Przypuśćmy, że chcemy obliczyć $I = \pi = 3.1415\ldots$ metodą Monte Carlo. Niech $Y = 4\mathbf{1}(U_1^2 + U_2^2 \leq 1)$, gdzie U_1, U_2 są niezależnymi jednostajnymi U(0,1). Wtedy $\mathbf{1}(U_1^2 + U_2 \leq 1)$ jest zmienną Bernoulliego z średnią $\pi/4$ oraz wariancją $\pi/4(1-\pi/4)$. Tak więc

$$\mathbb{E}\, Y = \pi$$

oraz wariancja Y równa się

$$\mathrm{Var}\, Y = 16\pi/4(1-\pi/4) = 2.6968.$$

Aby więc, przy zadanym standardowym poziomie istotności, dostać jedną cyfrę wiodącą, tj. 3 należy zrobić $n \approx 1.96^2 \cdot 2.7^2/0.5^2 = 112$ replikacji, podczas gdy aby dostać drugą cyfrę wiodąca to musimy zrobić $n \approx 1.96^2 \cdot 2.7^2/0.05^2 = 11\ 200$ replikacji, itd.

Wyobraźmy sobie, że chcemy porównać kilka eksperymentów. Poprzednio nie braliśmy pod uwagę czasu potrzebnego na wygenerowanie jednej replikacji Y_j. Przyjmijmy, że czas można zdyskretyzować ($n = 1, 2, \ldots$), i że dany jest łączny czas m (zwany *budżetem*) którym dysponujemy dla wykonania całego eksperymentu. Niech średnio jedna replikacja trwa czas t. A więc średnio można wykonąc w ramach danego budżetu jedynie m/t replikacji. A więc mamy

$$\hat{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j \approx \frac{1}{m/t}\sum_{j=1}^{m/t} Y_j.$$

Aby porównać ten eksperyment z innym, w którym posługujemy się liczbami losowymi $Y_1^{'}, \ldots$, takimi, że $Y = \mathbb{E}\, Y^{'}$, gdzie na wykonanie jednej replikacji $Y_j^{'}$ potrzebujemy średnio $t^{'}$ jednostek czasu, musimy porównać wariancje

$$\operatorname{Var} \hat{Y}_n, \qquad \operatorname{Var} \hat{Y}_n^{'}.$$

Rozpisując mamy odpowiedno $(t\operatorname{Var} Y)/m$ i $(t^{'}\operatorname{Var} Y^{'})/m$. W celu wybrania lepszego estymatora musimy więc porównać powyższe wielkości, i mniejszą z wielkość nazwiemy *efektywnością* estymatora.

1.1 Obliczanie całek metodą Monte Carlo

Przypuśćmy, że szukaną wielkość I możemy zapisać jako

$$I = \int_B k(x)\, dx,$$

gdzie $|B| < \infty$. Będziemy dalej zakładać, że $|B| = 1$ (jeśli ten warunek nie jest spełniony, to zawsze możemy zadanie przeskalować). Przykładem może być $B = [0,1]$ lub $[0,1]^d$.

Zgrubna metoda Monte Carlo (CMC) Połóżmy $Y = k(V)$, gdzie V ma rozkład jednostajny na B mamy $\mathbb{E}\, Y = I$. Niech $V_1, \ldots, V_n$ będą niezależnymi kopiami V o rozkładzie jednostajnym $\mathrm{U}(B)$. Wtedy

$$\hat{Y}^{\mathrm{CMC}} = \hat{Y}_n^{\mathrm{CMC}} = \frac{1}{n} \sum_{j=1}^{n} k(V_j)$$

jest *zgrubnym estymatorem Monte Carlo*(crude Monte Carlo) i stąd skrót CMC. Będziemy dalej mówili o estymatorach CMC.

Przykład 1.2 [Madras [33]] Niech $k : [0,1] \to \mathbb{R}$ będzie funkcją całkowalną z kwadratem $\int_0^1 k^2(x)\, dx < \infty$. Niech $Y = k(U)$. Wtedy $\mathbb{E}\, Y = \int_0^1 k(x)\, dx$ oraz $\mathbb{E}\, Y^2 = \int_0^1 k^2(x)\, dx$. Stąd

$$\sigma_Y = \sqrt{\int_0^1 k^2(x)\, dx - (\int_0^1 k(x)\, dx)^2}.$$

W szczególności rozważmy

$$k(x) = \frac{e^x - 1}{e - 1}.$$

Wtedy oczywiście mamy

$$\begin{aligned} \int_0^1 k(x)\,dx &= \frac{1}{e-1}\Big(\int_0^1 e^x\,dx - 1\Big) = \frac{e-2}{e-1} = 0.418 \\ \int_0^1 k^2(x)\,dx &= 0.2567. \end{aligned}$$

Stąd $\sigma_Y^2 = 0.0820$ a więc $n = 0.3280/b^2$.

Metoda chybił–trafił. Do estymacji $I = \int_B k(x)\,dx$ możemy używać innych estymatorów. Załóżmy dodatkowo, że $B = [0,1]$ i $0 \le k(x) \le 1$. Umieścmy teraz krzywą $k(x)$ w kwadracie $[0,1] \times [0,1]$ rozbijając kwadrat na dwie części: nad i pod krzywą. Część do której celujemy, rzucając "losowo" punkt (U_1, U_2) na kwadrat, jest pod krzywą. Ten pomysł prowadzi do tzw. estymatora *chybił-trafił* (Hit 0r Miss) (HoM). Formalnie definiujemy

$$Y = \mathbf{1}(k(U_1) > U_2).$$

Mamy bowiem

$$I = \mathbb{E}\,Y = \int_0^1 \int_0^1 \mathbf{1}_{k(u_2)>u_1}\,du_1\,du_2 = \int_0^1 k(u_2)\,du_2.$$

Wtedy dla $U_1, \ldots, U_{2n}$ obliczamy replikacje $Y_1, \ldots, Y_n$, skąd obliczamy estymator

$$\hat{Y}_n^{\mathrm{HoM}} = \frac{1}{n}\sum_{j=1}^{n} Y_j.$$

zwanym estymatorem *chybił trafił*. Wracając do przykładu 1.2 możemy łatwo policzyć wariancję $\mathrm{Var}\,Y = I(1-I) = 0.2433$. Stąd mamy, że $n = 0.9346/b^2$, czyli dla osiągnięcia tej samej dokładności trzeba około 3 razy więcej replikacji Y-ków aniżeli przy użyciu metody CMC. W praktyce też na wylosowanie Z w metodzie chybił trafił potrzeba dwa razy więcej liczb losowych.

Jak widać celem jest szukanie estymatorów z bardzo małą wariancją.

1.2 Symulacja funkcji od wartości oczekiwanej

Chcemy obliczyć $z = f(I)$, gdy wiadomo jak symulować liczbę losową Y i $I = \mathbb{E}\, Y$. Zakładamy, że f jest funkcją ciągłą w otoczneiu punktu I. Wtedy naturalnym estymatorem jest

$$\hat{Z}_n = f(\hat{Y}_n) \ ,$$

gdzie jak zwykle $\hat{Y}_n = (1/n) \sum_{j=1}^{n} Y_j$. Niestety estymator $\hat{Z}_n$ jest przeważnie obciążony dla $f(I)$. Przy założeniu, że $\sigma^2 = \mathrm{Var}\, Y < \infty$ mamy

$$\sqrt{n}(\hat{Y}_n - I) \xrightarrow{\mathcal{D}} \mathrm{N}(0, \sigma^2) \ .$$

Zakładając ponadto, że $f(x)$ jest ciągle różniczkowalna w otoczeniu I, ze wzoru Taylora $f(x) - f(I) = f^{'}(I)(x - I) + o(x - I)$, mamy

$$\sqrt{n}(f(\hat{Y}_n) - f(I)) = f^{'}(I)\sqrt{n}(\hat{Y}_n - I) + \sqrt{n}\ o((\hat{Y}_n - I)).$$

Ponieważ

$$\sqrt{n} o((\hat{Y}_n - I)) = \sqrt{n}(\hat{Y}_n - I))o(1)) \xrightarrow{\mathrm{Pr}} 0$$

(detale można znaleźć w książce van der Vaarta [44]), więc

$$\sqrt{n}(f(\hat{Y}_n) - f(I)) \xrightarrow{\mathcal{D}} \mathrm{N}(0, f^{'}(I)^2\sigma^2).$$

Technika dowodowa polegająca na rozwinięciu Taylorowskim funkcji nazywana jest *metodą delty*.

Rozpatrzymy teraz ogólną sytuację gdy mamy estymator $\hat{Y}_n$ wielkości I jest obciążony z *obciążeniem* $\delta_n = \mathbb{E}\, \hat{Y}_n - I$. Zakładamy, że

B.1 $\sqrt{n}(\hat{Y}_n - \mathbb{E}\, \hat{Y}_n) \to_{\mathrm{d}} \mathrm{N}(0, \sigma^2)$, gdzie $0 < \sigma^2 < \infty$ gdy $n \to \infty$,

B.2 $\sqrt{n}\delta_n \to 0$, gdy $n \to \infty$.

Wtedy korzystając z własności zbieżności według rozkładu (patrz (XI.1.15)) własności B.1 i B.2 pociągają

$$\sqrt{n}(\hat{Y}_n - I) \xrightarrow{\mathcal{D}} \mathrm{N}(0, \sigma^2) \ .$$

A więc jak poprzednio metodą delty mozna udowodnić następujący fakt.

Fakt 1.3 *Jesli funkcja f jest ciągle różniczkowalna w otoczeniu I oraz spełnione są założenia B.1-B.2, to*

$$\sqrt{n}(f(\hat{Y}_n) - f(I)) \xrightarrow{\mathcal{D}} \mathrm{N}(0, w^2) \ ,$$

gdzie $w^2 = (f^{'}(I))^2\sigma^2$.

Teraz heurystycznie pokażemy, że jeśli funkcja f jest dwukrotnie ciągle różniczkowalna w I, oraz

B.2' dla pewnego $0 < \delta < \infty$

$$n\delta_n \to \delta$$

(zauważmy, że B.2' jest mocniejszy niż B.2), to

$$\mathbb{E}\,(f(\hat{Y}_n) - f(I)) \sim \frac{1}{n}\left(\delta f^{'}(I) + \frac{\sigma^2 f^{''}(I)}{2}\right) \ .$$

Mianowicie z twierdzenia Taylora mamy

$$f(\hat{Y}_n) - f(I) = f^{'}(I)(\hat{Y}_n - f(I)) + \frac{f^{''}(I)}{2}(\hat{Y}_n - I)^2 + o((\hat{Y}_n - I)^2) \ . \quad (1.5)$$

i dalej mnożymy (1.5) przez n oraz korzystamy z tego, że

- $nf^{'}(I)(\mathbb{E}\,\hat{Y}_n - I) \to \delta f^{'}(I))$,
- $n\frac{f^{''}(I)}{2}\mathbb{E}\,(\hat{Y}_n - I)^2 \to \frac{f^{''}(I)}{2}\sigma^2$,
- $\mathrm{n}\mathbb{E}\,[o((\hat{Y}_n - I)^2)] \to 0$.

Niestety tym razem aby przeprowadzić formalny dowód musimy skorzystać z własności jednostajnej całkowalności co jest raczej niemożliwe bez dodatkowych założeń.

Omówimy teraz rozszerzenie faktu 1.3 na przypadek wielowymiarowy. Mianowicie chcemy wyestymować $z = f(I^1, \ldots, I^d)$ jeśli wiemy, ze $(I^1, \ldots, I^d)$ $= \mathbb{E}\,(\hat{Y}^1, \ldots, \mathbb{E}\,\hat{Y}^d)$ oraz potrafimy symulować
$(\hat{Y}^1, \ldots, \hat{Y}^d)$. Wygodnie jest wprowadzić oznaczenia wektorowe

$$\hat{\boldsymbol{Y}} = (\hat{Y}^1, \ldots, \hat{Y}^d), \qquad \boldsymbol{I} = (I^1, \ldots, I^d).$$

Przez $(\nabla f)(\boldsymbol{I})$ oznaczamy wektor $(f^{(1)}(\boldsymbol{I}), \ldots, f^{(d)}(\boldsymbol{I}))$, gdzie $f^{(i)}(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial x_i}$. Dowód następującego faktu przy użyciu metody delty można znaleźć w książce van der Vaart [44].

Fakt 1.4 *Jeśli funkcja* $f : \mathbb{R}^d \to \mathbb{R}$ *jest ciągle różniczkowalna w otoczeniu punktu* $\boldsymbol{I}$ *oraz*

$$\sqrt{n}(\hat{\boldsymbol{Y}}_n - \boldsymbol{I}) \xrightarrow{\mathcal{D}} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

dla pewnej skończonej macierzy kowariancji $\boldsymbol{\Sigma}$ *to*

$$\sqrt{n}\left(f(\hat{\boldsymbol{Y}}_n) - f(\boldsymbol{I})\right) \xrightarrow{\mathcal{D}} \mathrm{N}(\mathbf{0}, (\nabla f)(\boldsymbol{I})\boldsymbol{\Sigma}(\nabla f)(\boldsymbol{I})^{\mathrm{T}}).$$

W szczególności więc jeśli $\boldsymbol{Y}$ ma skończoną macierz kowariancji $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1,\ldots,d}$ oraz $\hat{\boldsymbol{Y}}_n = (1/n)\sum_{j=1}^n \boldsymbol{Y}_j$, gdzie $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ są niezależnymi replikacjami $\boldsymbol{Y}$ to dla estymatora $\hat{Z}_n = f(\hat{\boldsymbol{Y}}_n)$ przedział istotności na poziomie $\alpha = 0.05$ jest $f(\hat{\boldsymbol{Y}}_n) \pm 1.96\sigma/\sqrt{n}$, gdzie

$$\sigma^2 = (\nabla f)(I)\boldsymbol{\Sigma}((\nabla f)(I))^{\mathrm{T}} = \sum_{i,j=1}^{d} f^{(i)}(\boldsymbol{I}) f^{(j)}(\boldsymbol{I})\sigma_{ij} \ . \tag{1.6}$$

Niestety wzór na wariancję zawiera $\boldsymbol{I}$, którą to wielkość chcemy symulować. Jeśli więc $\boldsymbol{\Sigma}$ jest znana, to proponuje się przedziały ufności w postaci $f(\hat{Y}_n) \pm 1.96\hat{\sigma}/\sqrt{n}$, gdzie

$$\hat{\sigma}^2 = \nabla f)(I)\boldsymbol{\Sigma}((\nabla f)(\hat{\boldsymbol{Y}}))^{\mathrm{T}} = \sum_{i,j}^{d} f^{(i)}(\hat{\boldsymbol{Y}}_n) f^{(j)}(\hat{\boldsymbol{Y}}_n)\sigma_{ij} \ . \tag{1.7}$$

Estymator macierzy kowariancji $\boldsymbol{\Sigma}$ Naturalnym estymatorem wektora średniej $\boldsymbol{I}$ i macierzy kowariancji $\boldsymbol{\Sigma}$ wektora losowego $\boldsymbol{Y}$, gdy $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ są niezależnymi replikacjami $\boldsymbol{Y}$ jest

$$\begin{aligned} \hat{\boldsymbol{Y}}_n &= \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{Y}_j, \\ \hat{\sigma}_{i,j} &= \frac{1}{n-1}\sum_{m=1}^{n}(Y^i - \hat{Y}_n^i)(Y^j - \hat{Y}_n^j)\cdot \end{aligned}$$

Natomiast estymatorem σ^2 z (1.6) jest

$$\hat{\sigma^2} = \frac{1}{n-1}\sum_{m=1}^{n} \nabla f(\hat{\boldsymbol{Y}}_n)(\boldsymbol{Y}_m - \hat{\boldsymbol{Y}}_n)(\boldsymbol{Y}_m - \hat{\boldsymbol{Y}}_n)(\nabla f(\hat{\boldsymbol{Y}}_n))^{\mathrm{T}}\cdot$$

Zadania teoretyczne

[2]

1.1 Zaproponować algorytm obliczenia całki

$$\int_{-\infty}^{\infty} \psi(x) f(x)\, dx$$

(zakładamy, że jest ona absolutnie zbieżna), gdzie

$$f(x) = \begin{cases} e^{2x}, & -\infty < x < 0 \\ e^{-2x}, & 0 < x < \infty. \end{cases}$$

1.2 Przypuśćmy, że chcemy obliczyć I metodą Monte Carlo, gdzie $I = I^{'} I^{''}$ oraz wiemy, że $I^{'} = \mathbb{E}\, Y^{'}$, $I^{''} = \mathbb{E}\, Y^{''}$. Niech $Y_1^{'}, \ldots, Y_R^{'}$ będzie R niezależnych replikacji $Y^{'}$ oraz niezależnie od nich niech $Y_1^{''}, \ldots, Y_R^{''}$ będzie R niezależnych replikacji $Y^{''}$. Pokazać, że następujące estymatory są nieobciążone dla I:

$$\hat{Z}_1 = \left(\frac{1}{R} \sum_{i=1}^{R} Y_i^{'} \right) \left(\frac{1}{R} \sum_{i=1}^{R} Y_i^{''} \right)$$

$$\hat{Z}_2 = \frac{1}{R} \sum_{i=1}^{R} Y_i^{'} Y_i^{''} \ .$$

Pokazać, ze $\hat{Z}_1$ ma mniejszą wariancję. Wsk. Pokazać, że dla niezależnych zmiennych losowych X, Y mających drugie momenty mamy

$$\mathrm{Var}\,(XY) = \mathrm{Var}\, X \mathrm{Var}\, Y + (\mathbb{E}\, X)^2 \mathrm{Var}\, Y + (\mathbb{E}\, Y)^2 \mathrm{Var}\, X.$$

1.3 Zanalizować przykład II.3.6 pod kątem odpowiedniości liczby replikacji. Jak dobrać δ aby poza $k = 27$ na wykresie II.3.2 trudno byloby rozróżnić krzywą wysumulowaną od teoretycznej. Czy lepiej byłoby przyjąć $\delta = 0.05$.

[2] 1.2.2016

1.4 [Igła Buffona a liczba π] W zadaniu Buffona o liczbie π oblicza się prawdopodobieństwo

$$p = \frac{2L}{\pi}$$

przecięcia jednej z równoległych prostych oddalonych od siebie o 1 przez igłę o długości L. Wiadomo, że przeprowadzając niezależne eksperymenty mamy nieobciążony estymator $\hat{p}$ prawdopodobieństwa p. Wtedy $\pi = (2L)/p$. Zaplanować obliczenie $f(I)$ z dokładnością $b = 0.01$ na poziomie $\alpha = 0.05$, gdy $f(x) = (2L)/x$ oraz $I = \mathbb{E}\, Y$, gdzie $Y = 1$, gdy igła przecina linię, natomiast $Y = 0$ w przeciwnym razie.

1.5 Ile należy zrobić replikacji na poziomie ufności $\alpha = 0.05$ aby błąd $b = \sigma/50$.

1.6 Wiadomo, że jeśli $Y_1, \ldots, Y_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie N(0,1) to

$$\frac{(n-1)\hat{S}^2/\sigma^2 - n + 1}{\sqrt{2n-2}} \xrightarrow{\mathcal{D}} \mathrm{N}(0,1)\ , \tag{1.8}$$

gdy $n \to \infty$; patrz [44], str. 27. Natomiast jeśli zmienne są o rozkładzie dowolnym , to

$$\sqrt{n}\left(\hat{S}_n^2 - \sigma^2\right) \xrightarrow{\mathcal{D}} \mathrm{N}(0, \mu_4 - \sigma^4), \tag{1.9}$$

gdzie $\mu_4 = \mathbb{E}\,(Y - \mathbb{E}\, Y)^4$ jest momentem centralnym rzędu 4. Ile należy zrobić prób aby wyestymować σ^2 z dokładnością do 5-ciu % na poziomie $\alpha = 0.05$. Rozważyć oba warianty przy użyciu (1.8), (1.9).

Zadania laboratoryjne

1.7 a. Napisać algorytm na wygenerowanie punktu losowego o rozkładzie jednostajnym $\mathrm{U}(\Delta)$ w trójkącie Δ o wierzchołkach $(1,1), (-1,-1), (1,-1)$. Policzyć średnią liczbę liczb losowych $U_1, U_2, \ldots$ potrzebnych do wygenerowania jednej liczby V o rozkładzie $\mathrm{U}(\Delta)$.
b. Napisać algorytm na obliczenia całki

$$\int_\Delta e^{-(x^2+y^2)}\, dx\, dy$$

zgrubną metodą MC.

1.8 Wyprodukować 10 000 przedziałów ufności postaci $(\hat{Y}_n \pm \frac{z_{1-\alpha/2}\hat{S}_n}{\sqrt{n}})$ dla $Y = \mathbf{1}(U \leq u)$, $n = 1000$ oraz $I = \mathbb{E}\,Y$ następnie sprawdzić jaki procent tych przedziałów pokryje $I = 0.5, 0.9, 0.95$ gdy odpowiednio $u = 0.5, 0.9, 0.95$. Wyjaśnić otrzymane wyniki.

Projekt

1.9 Obliczyć składkę netto z dokładnością do czterech miejsc po przecinku, dla 40-latka w ubezpieczeniu na całe życie, gdy techniczna stopa procentowa wynosi $r = 0.05$, przyszły czas życia $T = T_{40}$ ma rozkład Makehama (można użyć funkcji Makeham z zadania III.9.21). Wzór na składkę netto gdy suma ubezpieczenia wynosi 1 zł jest

$$\pi = \frac{\mathbb{E}\, e^{-rT}}{\mathbb{E} \int_0^T e^{-rt}\, dt} \,.$$

Rozdział V

Techniki redukcji wariancji

[1] W tym rozdziale zajmiemy się konstrukcją i analizą metod prowadzących do redukcji estymatora wariancji. Jak zauważyliśmy w poprzednim rozdziale, metoda MC polega na tym, że aby obliczyć powiedzmy I, znajdujemy zmienną losową Y taką, że $I = \mathbb{E}\, Y$. Wtedy $\hat{Y} = \sum_{j=1}^{n} Y_j/n$ jest estymatorem MC dla I, gdzie $Y_1, \ldots$ są niezależnymi replikacjami Y. Wariancja $\operatorname{Var} \hat{Y} = \operatorname{Var} Y/n$, a więc widzimy, że musimy szukać takiej zmiennej losowej Y dla której $\mathbb{E}\, Y = I$, i mającej możliwie mała wariancję. Ponadto musimy wziąć pod uwagę możliwość łatwego generowania replikacji Y. Zauważmy też, że czasami pozbycie się założenia, że $Y_1, \ldots$ są niezależnymi replikacjami Y może też być korzystne.

W tym rozdziale zajmiemy się następującymi technikami:

1. metoda warstw,
2. metoda zmiennych antytetycznych,
3. metoda wspólnych liczb losowych,
4. metoda zmiennych kontrolnych,
5. warunkowa metoda MC,
6. metoda losowania istotnościowego.

[1] Poprawiane: 1.2.2016

1 Metoda warstw

Jak zwykle chcemy obliczyć $I = \mathbb{E}\, Y$. Przyjmijmy, że Y jest o wartościach w E (może być $\mathbb{R}$, $\mathbb{R}^d$ lub podzbiór) i niech $A^1, \ldots, A^m$ będą rozłącznymi zbiorami takimi, że $\mathbb{P}(Y \in \bigcup_{j=1}^m A^j)$. Rozbicie definiuje nam *warstwy* $W^j = \{Y \in A^j\}$ Możemy się spodziewać, że przy odpowiednim doborze warstw i zaprojektowaniu liczby replikacji w warstwach będzie możliwa redukcja wariancji.

Oznaczmy

- prawdopodobieństwo wylosowania warstwy $p_j = \mathbb{P}(Y \in A^j)$ (to przyjmujemy za znane i dodatnie)
- oraz $I^j = \mathbb{E}\,[Y|Y \in A^j]$ $(j = 1, \ldots, m)$ (oczywiście to jest nieznane bo inaczej nie trzeba by nic obliczać).

Ze wzoru na prawdopodobieństwo całkowite

$$\mathbb{E}\, Y = p_1 I^1 + \ldots + p_m I^m \ .$$

Metoda polega na losowaniu na poszczególnych warstwach, korzystając z tego, że potrafimy losowac (o tym poźniej)

$$Y^j =_{\mathrm{d}} (Y|Y \in A_j).$$

Niech teraz n będzie ogólną liczbą replikacji, którą podzielimy na odpowiednio n_j replikacji Y^j w warstwie j. Niech $Y_1^j, \ldots, Y_{n_j}^j$ będą więc niezależnymi replikacjami Y^j w warstwie j. Zakładamy również niezależność replikacji pomiędzy poszczegolnym warstwami. Oczywiście $n = n_1 + \ldots + n_m$. Definiujemy

$$\hat{Y}_{n_j}^j = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_i^j$$

jako estymator I^j. Niech $\sigma_j^2 = \operatorname{Var} Y^j$ będzie wariancją w warstwie j. Zauważmy, że $\hat{Y}^j$ jest estymatorem nieobciążonym I^j oraz

$$\operatorname{Var} \hat{Y}_{n_j}^j = \frac{\sigma_j^2}{n_j}.$$

Następnie definiujemy *estymator warstwowy*

$$\hat{Y}_n^{\mathrm{str}} = p_1 \hat{Y}^1 + \ldots + p_m \hat{Y}^m \ .$$

Zakładamy, że wiadomo jak generować poszczególne Y^j. Zauważmy, że estymator warstwowy jest nieobciążony

$$\begin{aligned}\mathbb{E}\,\hat{Y}^{\rm str} &= p_1\mathbb{E}\,\hat{Y}^1 + \ldots + p_m\mathbb{E}\,\hat{Y}^m \\ &= \sum_{i=1}^m p_i\frac{\mathbb{E}\,Y\mathbf{1}_{\{Y\in A_i\}}}{p_i} = \mathbb{E}\,Y = I.\end{aligned}$$

Niech dla $\boldsymbol{x} = (x_1, \ldots, x_m)$

$$\sigma^2_{\rm str}(\boldsymbol{x}) = \sum_{j=1}^m \frac{p_j}{x_j}\sigma_j^2.$$

Pamiętając o niezależności poszczególnych replikacji mamy

$$\begin{aligned}&\sigma^2_{\rm str}((n_1/n, \ldots, n_m/n)) \qquad (1.1)\\ &= n{\rm Var}\,\hat{Y}^{\rm str}_n \\ &= n\sum_{j=1}^m p_j^2{\rm Var}\,\hat{Y}^j = p_1^2\frac{n}{n_1}\sigma_1^2 + \ldots + p_m^2\frac{n}{n_m}\sigma_m^2 \\ &= \frac{p_1^2}{n_1/n}\sigma_1^2 + \ldots + \frac{p_m^2}{n_m/m}\sigma_m^2\ . \qquad (1.2)\end{aligned}$$

Zastanowimy się co się stanie gdy przechodzimy z $n \to \infty$ tak aby

(*) $\quad q_j(n) = n_j/n$ były zbieżne, powiedzmy do q_j $(j = 1, \ldots, m)$.

Oczywiście $\boldsymbol{q} = (q_1, \ldots, q_m)$ jest funkcją prawdopdobieństwa. Wtedy

$$\sigma^2_{\rm str}((n_1/n, \ldots, n_m/n)) \to \sigma^2(\boldsymbol{q})\ .$$

Kluczowym dla analizy wyników jest następujący fakt.

Fakt 1.1 *Jeśli z n i $n_1, \ldots, n_m$ przechodzimy do nieskończoności w sposób (*), to*

$$\frac{\sqrt{n}}{\sigma_{\rm str}(\boldsymbol{q})}(\hat{Y}^{\rm str}_n - I) \xrightarrow{\mathcal{D}} {\rm N}(0,1)\ .$$

Dowód Mamy

$$\begin{aligned}\sqrt{n}(\sum_{j=1}^{m} p_j \hat{Y}^j - I) &= \sqrt{n}(\sum_{j=1}^{m} p_j \hat{Y}^j - \sum_{j=1}^{m} p_j I^j) \\ &= \sum_{j=1}^{m} p_j \sigma_j \sqrt{\frac{n}{n_j}} \left[\frac{\sqrt{n_j}}{\sigma_j} \left(\hat{Y}^j - y_j \right) \right] \\ &= \sum_{j=1}^{m} \frac{p_j \sigma_j}{\sqrt{q_j(n)}} \left[\frac{\sqrt{n_j}}{\sigma_j} \left(\hat{Y}^j - y_j \right) \right] \\ &\xrightarrow{\mathcal{D}} \mathrm{N}(0, \sum_{j=1}^{m} p_j^2 \frac{\sigma_j^2}{q_j}),\end{aligned}$$

ponieważ

$$\frac{\sqrt{n_j}}{\sigma_j} \left(\hat{Y}^j - y_j \right) \xrightarrow{\mathcal{D}} N_j$$

gdzie $N_1, \ldots, N_m$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie N(0,1). □

Dobór liczby replikacji w warstwach Możemy teraz postawić dwa pytania.

- Mając zadane warstwy (a więc znając (p_j) oraz (σ_j^2)) jak dobrać dla zadanej całkowitej liczby replikacji n, liczby replikacji (n_j) na poszczególnych warstwach aby wariancja $n\mathrm{Var}\,\hat{Y}_n^{\mathrm{str}}$ była minimalna.

- Jak dobrać warstwy aby zmniejszyć (zminimizować) wariancję.

Zajmiemy się odpowiedzią na pierwsze pytanie. Okazuje się, że optymalny dobór liczby replikacji n_j można wyliczyć z następującego twierdzenia.

Twierdzenie 1.2 *Niech n będzie ustalone i $n_1 + \ldots + n_m = n$. Wariancja $p_1^2 \frac{n}{n_1}\sigma_1^2 + \ldots + p_m^2 \frac{n}{n_m}\sigma_m^2$ estymator warstwowego $\hat{Y}^{\mathrm{str}}$ jest minimalna gdy*

$$n_j = \frac{p_j \sigma_j}{\sum_{j=1}^{m} p_j \sigma_j} n \, .$$

Dowód Musimy zminimizować $\sigma^2_{\rm str}(\boldsymbol{x})$ przy warunku $x_1 + \ldots + x_m = 1$, $x_j \geq 0$. Jeśli podstawimy do wzoru (1.2)

$$q^*_j = \frac{p_j \sigma_j}{D}$$

gdzie $D = \sum_{j=1}^m p_j \sigma_j$ to $\sigma^2_{\rm spr}(\boldsymbol{q}^*) = D^2$. Ale z nierówności Cauchy'ego-Schwarza

$$(\sum_{j=1}^n x_j z_j)^2 \leq \sum_{j=1}^n x_j^2 \sum_{j=1}^n z_j^2,$$

mamy

$$\begin{aligned} D^2 = (\sum_{j=1}^m p_j \sigma_j)^2 &= \left(\sum_{j=1}^m \sqrt{x_j} \frac{p_j \sigma_j}{\sqrt{x_j}} \right)^2 \\ &\leq \sum_{j=1}^m x_j \sum_{j=1}^m \frac{p_j^2 \sigma_j^2}{x_j} \\ &= \sigma^(\boldsymbol{x}) \ . \end{aligned}$$

□

Niestety ten wzór jest mało użyteczny ponieważ nie znamy zazwyczaj wariancji σ_j^2, $j = 1, \ldots, m$. Chyba, że jak poprzednio przyjmiemy empiryczną wariancję $\hat{S}_j^2$ parametru σ_j^2 w warstwie j. Mamy za to następujące fakt, że tzw. *alokacja proporcjonalna* zmniejsza wariancję. Mówimy o takiej alokacji, jeśli $n_i = np_i$ $i = 1, \ldots, m$. Wtedy z wzoru (**[??e.war:str??]**) mamy

$$\sigma^2_{\rm str}(n_1/n, \ldots, n_m/n) = \sum_{j=1}^m p_j \sigma_j^2. \tag{1.3}$$

W przypadku alokacji proporcjonalnej mamy $\boldsymbol{p} = (n_1/n, \ldots, n_m/n)$.

Fakt 1.3 *Jesli alokacja jest proporcjonalna, to*

$$\sigma^2_{\rm str}(\boldsymbol{p}) \leq \sigma^2_Y \ .$$

Dowód Z wzoru na prawdopodobieństwo całkowite

$$\mathbb{E}\, Y^2 = \sum_{j=1}^m p_j \mathbb{E}\,[Y^2 | Y \in A^j] = \sum_{j=1}^m p_j (\sigma_j^2 + (I^j)^2).$$

Ponieważ $I = \sum_{j=1}^m p_j I^j$ więc

$$\begin{aligned} \operatorname{Var} Y &= \sum_{j=1}^m p_j(\sigma_j^2 + (I^j)^2) - (\sum_{j=1}^m p_j I^j)^2 \\ &= \sum_{j=1}^m p_j \sigma_j^2 + \sum_{j=1}^m p_j (I^j)^2 - (\sum_{j=1}^m p_j I^j)^2. \end{aligned}$$

Z nierówności Jensena

$$\left(\sum_{j=1}^m p_j I^j \right)^2 \leq \sum_{j=1}^m p_j (I^j)^2,$$

skąd

$$\operatorname{Var} Y \geq \sum_{j=1}^m p_j \sigma_j = \sigma^2(\boldsymbol{p}).$$

□

A więc przy alokacji proporcjonalnej bierzemy $n_j = \lfloor n p_j \rfloor$, i wtedy wariancja się zmniejszy.

Modyfikacje Zamiast jednej zmiennej Y możemy rozważyć parę Y, X i warstwy zdefiniować następująco. Niech $A^1, \ldots, A^m$ będą rozłącznymi zbiorami takimi, że $\mathbb{P}(X \in \bigcup_{j=1}^m A^j) = 1$. Wtedy j-ta warstwa jest zdefiniowana przez $W^j = \{X \in A^j\}$ i

$$I = \mathbb{E}\, Y = \sum_{j=1}^m p_j \mathbb{E}\,[Y | X \in A^j],$$

gdzie $p_j = \mathbb{P}(X \in A_j)$. Aby napisać wzór na estymator warstwowy musimy przedefiniowac pojęcie

$$Y^j =_{\mathrm{d}} (Y | X \in A^j).$$

Zostawiamy czytelnikowi napisanie wzoru na estymator warstwowy w tym przypadku. Zauważmy, że w przypadku ogólnym warstwy W^j $(j = 1, \ldots, m)$ tworzą partycję przestrzeni zdarzeń elementarnych Ω. Najogolniej mówiąc przez warstwy rozumiemy rozbicie Ω na zdarzenia $W^1, \ldots, W^m$.

Przykłady

Przykład 1.4 [Asmussen & Glynn [4]] Rozważmy problem obliczenia całki

$$I = \int_0^1 g(x)\,dx. \tag{1.4}$$

Niech $Y = g(U)$ oraz $X = U$. Wtedy możemy wprowadzić warstwy następująco: $W^j = \{U \in (\frac{j-1}{m}, \frac{j}{m}]\}$, $j = 1, \ldots, n$. W tym przypadku bardzo prosto można przeprowadzić losowanie $Y^j =_{\mathrm{d}} (Y|X \in (j-1/m, j/m])$. Mianowicie zmienna

$$V^j = \frac{j-1}{m} + \frac{U}{m}$$

ma rozkład jednostajny na $(\frac{j-1}{m}, \frac{j}{m}]$. Ponadto mamy $p_j = 1/m$. Gdybyśmy zastosowali alokację proporcjonalną z $m = n$ (wtedy mamy $n_j = 1$), to estymatorem warstwowym I jest

$$\hat{Y}_n^{\mathrm{str}} = \frac{1}{n}\sum_{j=1}^{n} g(V^j).$$

Porównajmy z całkowaniem numerycznym według wzoru:

$$\hat{g}_n = \frac{1}{n}\sum_{j=1}^{n} g(\frac{2j-1}{2n}).$$

Zauważmy, że $\mathbb{E}\,V^j = \frac{2j-1}{2n}$ jest środkiem przedziału $(\frac{j-1}{n}, \frac{j}{n}$.

Zobaczymy teraz jak estymator warstwowy zmniejsza wariancję przy obliczeniu całki (1.4), gdzie g jest funkcją ciągle różniczkowalną z

$$M = \sup_{x\in[0,1]} |g'(x)| < \infty.$$

Zauważmy, że

$$\begin{aligned} Y^j &=_d \left(g(U)|\frac{j-1}{n} \le U < \frac{j}{n}\right) \\ &=_{\mathrm{d}} g(\frac{j-1}{n} + \frac{U}{n}). \end{aligned}$$

Wariancja naszego estymatora warstwowego wynosi

$$\sigma_{\mathrm{str}}(1/n, \ldots, 1/n) = \frac{1}{n}\sum_{j=1}^{n} \sigma_j^2,$$

gdzie

$$\sigma_j^2 = \mathrm{Var}\,(g(\frac{j-1}{n} + \frac{1}{n}U)).$$

Korzystając z twierdzenia o wartości średniej

$$g(\frac{j-1}{m} + \frac{U}{n}) = g(\frac{j-1}{n}) + g^{'}(\theta_j)\frac{U}{n}$$

gdzie θ_j jest punktem z odcinka $((j-1)/n, (j-1+U)/n)$. mamy

$$\sigma_j^2 = \mathrm{Var}\,(g(\frac{j-1}{n} + \frac{U}{n}) = \frac{\mathrm{Var}\,(g^{'}(\theta_j)U)}{n^2}.$$

Teraz

$$\sigma_j^2 \leq \frac{\mathbb{E}\,(g^{'}(\theta_j)^2U^2)}{m^2} \leq \frac{M^2}{n^2}.$$

A więc wariancja estymatora warstwowego z alokacją proporcjonalną jest rzędu

$$\sigma_{\mathrm{str}}^2 = \sum_{j=1}^{n} p_j\sigma_j^2 \leq \sum_{j=1}^{n} \frac{1}{n}\frac{M^2}{n^2} = \frac{M^2}{n^2} = O(n^{-2}).$$

Korzystając z rozważań z Uwagi **[??re:szynbkosc??]** możemy je uogolnić na nasz przyapadek. Mamy bowiem

$$\mathbb{P}(|\hat{Y}_n^{\mathrm{str}} - I| \geq b_n) \leq \frac{\mathrm{Var}\,\hat{Y}_n^{\mathrm{str}}}{b_n^2} \leq \frac{O(n^{-2})}{b_n^2}.$$

Z równania $\frac{O(n^{-2})}{b_n^2} = \alpha$ mamy, że $b_n = O(n^{-1})$, co jest lepszą szybkościa zbieżności niż kanoniczna $O(n^{-1/2})$.

Przykład 1.5 [Asmussen & Glynn [4]] Znajdziemy teraz estymator dla $I = \int_0^1 g(x)\,dx$, który ma zbieżność szybszą niż kanoniczna $n^{-1/2}$, gdzie n jest liczba replikacji. To nie będzie estymator wartwowy ale wykorzystamy pomysły z poprzedniego przykładu 1.4. Zakładamy, że g jest funkcją ciągle różniczkowalną z

$$M = \sup_{x\in[01,]} g^{'}(x) < \infty.$$

Rozważmy więc estymator postaci

$$\hat{Y}_n = \frac{1}{n}\sum_{j=1}^{n} g\left(\frac{j-1+U_j}{n}\right).$$

Jest on nieobciążony bo

$$\begin{aligned}\mathbb{E}\,\hat{Y}_n &= \frac{1}{n}\sum_{j=1}^{n}\int_0^1 g\left(\frac{j-1+u}{n}\right)du \\ &= \sum_{j=1}^{n}\int_{(j-1)/n}^{j/n} g(u)\,du = \int_0^1 g(u)\,du.\end{aligned}$$

Teraz zanalizujemy wariancję. Z twierdzenia o wartości średniej

$$g\left(\frac{j-1}{n}+\frac{U_j}{n}\right) = g\left(\frac{j-1}{n}\right) + g^{'}(\theta_j)\frac{U_j}{n}.$$

A więc

$$\begin{aligned}\operatorname{Var} Y &= \frac{1}{n}\sum_{j=1}^{n}\operatorname{Var} g\left(\frac{j-1+U_j}{n}\right) \\ &= \frac{1}{n^2}\sum_{j=1}^{n}(\frac{g^{'}(\theta_j)U_j}{n})^2 \le \frac{M^2}{n^3}.\end{aligned}$$

W tym przypadku błąd $b_n = O(1/n^{3/2})$.

Przykład 1.6 Losowanie warstwowe dla dowolnych rozkładów Teraz omówimy metody symulacji zmiennej o zadanym rozkładzie, używające koncepcji wartw. Przypomnijmy, że

$$V^i = \frac{i-1}{m}+\frac{U_i}{m}, \qquad i = 1,\ldots,m$$

ma rozkład jednostajny w $(i-1)/m, i/m]$. Je sli więc położymy $n = m$ to wtedy $n_j = 1$ i $V^1,\ldots,V^n$ tworzy warstwową próbę z rozkładu jednostajnego.

Rozważymy symulację zmiennej losowej Y z dystrybuantą $F(x)$ metodą warstw. Dla ułatwienia założmy, ze F jest ścisle rosnąca. Rozważania przeprowadzamy na prostej $\mathbb{R}$ ale prosta modyfikacja pozwala rozważać rozkłady które są skoncentrowane na podzbiorze $\mathbb{R}$. Niech $p_1,\ldots,p_m$ będą zadanymi rozmiarami warstw, gdzie $p_j > 0$ oraz $\sum_{j=1}^m p_j = 1$. Prostą $\mathbb{R}$ rozbijamy na m odcinków $A_1 = (a_0, a_1], (a_1, a_2], \ldots, A_m = (a_{m-1}, a_m)$, gdzie $a_0 = -\infty$

oraz $a_m = \infty$, tak aby $\mathbb{P}(Y \in (a_{j-1}, a_j]) = p_j$ (przyjmujemy, ze rozkład jest skoncentrowany na całej prostej). Mianowicie

$$\begin{aligned} a_0 &= -\infty \\ a_1 &= F^{-1}(p_1) \\ \vdots &= \vdots \\ a_m &= F^{-1}(p_1 + \ldots + p_m) = F^{-1}(1). \end{aligned}$$

W przypadku gdy nośnik E rozkładu F jest właściwym podzbiorem $\mathbb{R}$ to jak już zauważyliśmy trzeba rozbicie zmodyfikować.

Teraz $Y^j =_{\mathrm{d}} (Y|Y \in A_j)$ możemy symulować następująco, mając procedurę obliczania funkcji odwrotnej dystrybuanty

$$F^{-1}(u) = \inf\{x : F(x) \leq u\}.$$

Mamy $p_j = \mathbb{P}(A^j)$

Fakt 1.7 *Niech* $V^j = a_{j-1} + (a_j - a_{j-1})U)$. *Wtedy*

$$F^{-1}(V^j) =_{\mathrm{d}} (Y|Y \in A^j).$$

W pierwszym eksperymencie przprowadzić 100 replikacji i narysować histogram z klasami $A_1, \ldots, A_{10}$. Natomiast drugi eksperyment przeprowadzić przy losowaniu w warstwach losując w każdej 10 replikacji. Porównąc otrzymane rysunki.

Przykład 1.8 Obliczenie prawdopodobieństwa awarii sieci. Rozważmy schemat połączeń z podrozdziału VI.4. Ponieważ typowo prawdopodobieństwo I jest bardzo małe, więc aby je obliczyć używając estymatora CMC musimy zrobić wiele replikacji. Dlatego jest bardzo ważne optymalne skonstruowanie dobrego estymatora dla I. Można na przykład zaproponować estymator warstwowy, gdzie j-tą warstwą $\mathcal{S}(j)$ jest podzbiór składających się z tych B takich, że $|B| = j$. Zauważmy, że

$$p_j = \mathbb{P}(X \in \mathcal{S}(j)) = \mathbb{P}(\text{dokładnie } j \text{ krawędzi ma przerwę}) = \binom{n}{j} q^j (1-q)^{n-j} .$$

Każdy $A \subset \mathcal{S}$ ma jednakowe prawdopodobieństwo $q^j(1-q)^{|\mathcal{S}|-j}$ oraz

$$\mathbb{P}(X = A | X \in \mathcal{S}(j)) = 1/\binom{n}{j} \qquad A \subset \mathcal{S} .$$

Aby więc otrzymać estymator warstwowy generujemy np_j replikacji na każdej warstwie $\mathcal{S}(j)$ z rozkładem jednostajnym.

2 Zależności zmniejszają wariancję

2.1 Zmienne antytetyczne

Rozpatrzymy teraz metody symulacji $I = \mathbb{E}\,Y$, w których replikacje nie muszą być niezależne. Będzie nam potrzebna następująca nierówność, która jest natychmiastowym wnioskiem z twierdzenia XI.1.2.

Wniosek 2.1 *Jesli funkcja $\psi(x_1, \ldots, x_k)$, $0 \le x_1, \ldots, x_k \le 1$ jest monotoniczna (niemalejąca lub nierosnąca), to dla niezależnych zmiennych $U_1, \ldots, U_k$ o rozkładzie jednostajnym*

$$\mathrm{Cov}\,(\psi(U_1, \ldots, U_k), \psi(1 - U_1, \ldots, 1 - U_k)) \le 0\,.$$

Mówimy wtedy, że zmienne $\psi(U_1, \ldots, U_k), \psi(1 - U_1, \ldots, 1 - U_k)$ są ujemnie skorelowane. Zauważmy ponadto, że U i $1-U$ mają ten sam rozkład jednostajny $\mathrm{U}(0, 1)$. Rozpatrzmy teraz n zmiennych losowych $Y_1, Y_2, \ldots, Y_n$, przy czym n jest parzyste. Ponadto zakładamy, że

- pary $(Y_{2i-1}, Y_{2i})_i$, $i = 1, \ldots, n/2$ są niezależne o jednakowym rozkładzie,
- zmienne losowe $(Y_j)_{j=1}^n$ mają ten sam rozkład brzegowy co Y.

Jeśli rozważamy zgrubny estymator

$$\hat{Y}^{\mathrm{CMC}} = \frac{1}{n}\sum_{j=1}^{n} Y_j^{'}$$

gdzie $Y_1^{'}, Y_2^{'}, \ldots$ są niezależnymi replikacjami Y, to wariancja wynosi $\mathrm{Var}\,Y/n$. Zauważmy, że jeśli będziemy rozpatrywać estymator

$$\hat{Y}_n^{*} = \frac{1}{n}\sum_{j=1}^{n} Y_j$$

to jego wariancja

$$\begin{aligned}
\mathrm{Var}\,(\hat{Y}_n^{*}) &= \frac{\frac{n}{2}\mathrm{Var}\,(Y_1 + Y_2)}{n^2} \\
&= \frac{1}{2n}(2\mathrm{Var}\,(Y) + 2\mathrm{cov}(Y_1, Y_2)) \\
&= \frac{1}{2n}(2\mathrm{Var}\,(Y) + 2\mathrm{Var}\,(Y_1)\mathrm{Corr}(Y_1, Y_2)) \\
&= \frac{1}{n}\mathrm{Var}\,(Y)(1 + \mathrm{Corr}(Y_1, Y_2))\,.
\end{aligned}$$

A więc jeśli korelacja $\mathrm{Corr}(Y_1, Y_2)$ jest ujemna to redukujemy wariancję.

Przypuśćmy teraz, że dla pewnej funkcji $\psi : [0,1]^k \to \mathbb{R}$ mamy $Y_1 =_{\mathrm{d}} \psi(U_1, \ldots, U_k)$, i że ψ spełnia założenie z wniosku 2.1. Wtedy $Y_2 = \psi(1 - U_1, \ldots, 1 - U_k)$ i zmienne Y_1, Y_2 są ujemnie skorelowane. Dalej definiujemy $Y_3 = \psi(U_{k+1}, \ldots, U_{2k})$, $Y_4 = \psi(1 - U_{k+1}, \ldots, 1 - U_{2k})$ itd. Zauważmy, że tak zdefiniowany ciąg $Y_1, \ldots, Y_n$ spełnia nasze postulaty. Wtedy

$$\hat{Y}^{\mathrm{anth}} = \frac{\sum_{j=1}^{n} Y_j}{n}$$

nazywamy estymatorem anytetycznym i jeśli $\mathrm{corr}\,(Y_1, Y_2) < 0$, to z przedstawionych rozważań ma mniejszą wariancję od estymatora CMC, który ma wariancję $\mathrm{Var}\, Y/(n)$.

Przykład 2.2 Rozpatrzmy teraz następujący problem. Przypuśćmy, że mamy m zadań do wykonania. Czasy zadań są niezależnymi zmiennymi losowymi o jednakowym rozkładzie z dystrybuantą F. Mamy do dyspozycji c serwerów. Możemy te zadania wykonać na wiele sposobów.

- według najdłuższego zadania (Longest Processing Time First - LPTF),
- według najkrótszego zadania (Shortest Processing Time First - SPTF),
- w kolejności numerów (First Come First Served – FCFS),
- lub alternując – (ALT).

Następne zadanie zaczyna być opracowywane jedynie w momentach zakończenia poprzednego zadania zadań. Oznacza to, ze niedopuszczamy przerywań w trakcie pracy. Celem jest policzenie $\mathbb{E}\, C^{\mathrm{SPTF}}$, $\mathbb{E}\, C^{\mathrm{LPTF}}$, gdzie C jest czasem do zakończenia ostatniego zadania według procedury w super-indeksie. Powiedzmy, że $N = 5$ i $c = 2$ oraz zadania są wielkości $3, 1, 2, 4, 5$. Wtedy postępując wg SPTF zaczynamy z zadaniami wielkości 1 i 2 i po jednej jednostce czasu resztowe wielkości tych zadań są $3, 0, 1, 4, 5$ a więc mamy teraz zadanie z $N = 4$ i $c = 2$. Postępując tak dalej widzimy, że $C^{\mathrm{SPTF}} = 9$. Postępując wg. LPTF zaczynamy z zadaniami wielkości 4 i 5 i po 4-ch jednostkach czasu resztowe wielkości tych zadań są $3, 1, 2, 0, 1$ a więc mamy teraz zadanie z $N = 4$ i $c = 2$. Postępując tak dalej widzimy, że $C^{\mathrm{LPTF}} = 8$.

Możemy zastanowawiać się jaka dyscyplina jest optymalna. Przypuśćmy, że rozmiary zadań są niezależne o jednakowym rozkładzie wykładniczym

Exp(1). Procedura ALT jest łatwa do zanalizowania. Załóżmy, że N jest parzyste. Wtedy $\mathbb{E}\, C^{\mathrm{ALT}} = \mathbb{E}\, \max(X_1, X_2)$, gdzie

$$X_1 = \sum_{i=1}^{N/2} \xi_{1i}, \qquad X_2 = \sum_{i=1}^{N/2} \xi_{2i},$$

gdzie ξ_{ij} są niezależnymi zmiennymi losowym o jednakowym rozkładzie wykładniczym Exp(1). Niestety dla procedur LPTF i SPTF i nawet FCFS nie mamy wzorów teoretycznych.

W tablicy 2.1 podajemy wyniki symulacji C dla różnych dyscyplin, przy użyciu metody CMC oraz antytecznej (anthy), dla $N = 10$ zadań oraz $m = 1000$ replikacjami; I jest wysymulowaną wartością $\mathbb{E}\, C = I$, s jest odchyleniem standardowym $\sqrt{\mathrm{Var}\,(C)}$, oraz b jest połową długości przedziału ufności na poziomie $\alpha = 0.05$ (czyli można rzec błąd). Dla porównania

dyscyplina	I	s	b
FCFS	5.4820	1.8722	0.1160
SPTF	5.8993	2.0259	0.1256
LPTF	5.0480	1.6908	0.1048
FCFSanthy	5.5536	0.8999	0.0558
SPTFanthy	5.9673	1.0127	0.0628
LPTFanthy	5.0913	0.6998	0.0434.

Tablica 2.1: Szacowanie C przy użyciu CMC i anthy; $N = 10$, liczba replikacji $m = 1000$.

w tablicy 2.2 podane są wyniki symulacji z $m = 100000$ replikacji. Wszystkie symulacje są robione na tych samych liczbach pseudolosowych, ponieważ algorytmy zaczynają się od rand('state',0).[2]

2.2 Wspólne liczby losowe

Przypuśćmy, że chcemy sprawdzić różnicę $d = \mathbb{E}\, k_1(X) - \mathbb{E}\, k_2(X)$, gdzie X jest wektorem losowym. Można to zrobic na dwa sposoby. Najpierw estymujemy $\mathbb{E}\, k_1(X)$ a następnie, niezależnie $\mathbb{E}\, k_2(X)$ i wtedy liczymy d. To nie zawsze jest optymalne, szczególnie gdy różnica d jest bardzo mała.

[2]algorytmy: FCFS.m, FCFSanty.m, SPTF.m, SPTFanty.m, LPTF.m LPTFanty.m.

dyscyplina	I	s	b
FCFS	5.4909	1.7909	0.0111
SPTF	5.8839	1.9116	0.0118
LPTF	5.0472	1.5945	0.0099
FCFSanthy	5.5015	0.8773	0.0054
SPTFanthy	5.8948	0.9748	0.0060
LPTFanthy	5.0556	0.6858	0.0043.

Tablica 2.2: Szacowanie C przy użyciu CMC i anthy; $N = 10$, liczba replikacji $m = 100000$

Inny sposobem jest zapisanie $d = \mathbb{E}\left(k_1(X) - k_2(X)\right)$ i następnie estymowania d. W pierwszym przypadku, $Y_1 = k_1(X_1)$ jest niezależne od $Y_2 = k_2(X_2)$, gdzie X_1, X_2 są niezależnymi replikacjami X. A więc wariancja $\mathrm{Var}\left(k_1(X_1) - k_2(X_2)\right) = \mathrm{Var}\left(k_1(X)\right) + \mathrm{Var}\left(k_2(X)\right)$. W drugim przypadku $\mathrm{Var}\left(k_1(X) - k_2(X)\right) = \mathrm{Var}\left(k_1(X)\right) + \mathrm{Var}\left(k_2(X)\right) - 2\mathrm{Cov}\left(k_1(X), k_2(X)\right)$. Jesli więc $\mathrm{Cov}\left(k_1(X), k_2(X)\right) > 0$ to istotnie zmiejszymy wariancję. Tak będzie w przypadku gdy obie funkcje są jednakowo monotoniczne.

Rozpatrzmy następujący przykład.[3] Niech $k_1(x) = e^{\sqrt{x}}$ i $k_2(x) = (1 + \frac{\sqrt{x}}{50})^{50}$ oraz $X = U$. Można oszacować, że $|k_1(x) - k_2(x)| \leq 0.03$, a więc $d \leq 0.03$. Przy piewszej metodzie wariancja Y_1 wynosi $0.195 + 0.188 = 0.383$ natomiast przy drugiej że jest mniejsza od 0.0009. Polecamy czytelnikowi przeprowadzenie symulacji tego przykładu i zweryfikowanie go empirycznie.

Przykład 2.3 Kontynuujemy przykład 2.2. Tym razem pytamy się o oczekiwaną różnicę $C^{\mathrm{SPTF}} - C^{\mathrm{LPTF}}$, tj. $I = \mathbb{E}\left(C^{\mathrm{SPTF}} - C^{\mathrm{LPTF}}\right)$. Można po prostu odjąć obliczone odpowiednie wielkości w poprzednim przykładzie, ale wtedy za wariancję trzeba by wziąć sumę odpowiednich wariancji, i przy tej samej długości symulacji błąd byłby większy. Można też wykorzystać metodę wspólnych liczb losowych. W tablicy 2.3 podajemy wyniki symulacji $I = \mathbb{E}\left(C^{\mathrm{SPTF}}\right) - C^{\mathrm{LPTF}}$) przy użyciu metody wspólnej liczby losowej (JCN – *Joint Common Number*). Podobnie jak poprzednio, robimy też symulacje przy użyciu zgrubnej metody Monte Carlo (CMC), dla $N = 10$ zadań, z $m = 1000$ replikacjami, s jest odchyleniem standardowym $\sqrt{\mathrm{Var}\left((C^{\mathrm{SPTF}} - C^{\mathrm{LPTF}}\right)}$, oraz b jest połową długości przedziału ufności na poziomie $\alpha = 0.05$. [4] Wszystkie

[3] Brak zalozenia o rozkładzie.

[4] skrypty SmLPTF.m, SmLPTFjoint.m.

dyscyplina	I	s	b
SmLPTFjoint	0.8513	0.5431	0.0352
SmLPTF	1.0125	2.5793	0.1599

Tablica 2.3: Szacowanie I przy użyciu JCN i CMC; $N = 10$, liczba replikacji $m = 1000$

symulacje są robione na tych samych liczbach pseudolosowych, ponieważ algorytmy zaczynają się od rand('state',0). Dla porównania podajemy w tablicy 2.4 wyniki dla 100000 replikacji.

dyscyplina	I	s	b
SmLPTFjoint	0.8366	0.5166	0.0032
SmLPTF	0.8455	2.4953	0.0155

Tablica 2.4: Szacowanie I przy użyciu JCN i CMC ; $N = 10$, liczba replikacji $m = 100000$

3 Metoda zmiennych kontrolnych

Zacznijmy, że jak zwykle aby obliczyć I znajdujemy taką zmienną losową Y, że $I = \mathbb{E}\, Y$. Wtedy $\hat{Y}^{\mathrm{CMC}} = \sum_{j=1}^{n} Y_j/n$ jest estymatorem MC dla I, gdzie $Y_1, \ldots, Y_n$ są niezależnymi replikacjami Y. Wariancja tego estymatora wynosi $\operatorname{Var} \hat{Y}_n = \operatorname{Var} Y/n$. Teraz przypuśćmy, że symulujemy jednocześnie drugą zmienną losową X (a więc mamy niezależne replikacje par $(Y_1, X_2), \ldots, (Y_n, X_n)$) dla której znamy $\mathbb{E}\, X$. Niech $\hat{X}_n = \sum_{j=1}^{n} X_j/n$ i definiujemy

$$\hat{Y}_n^{CV} = \hat{Y}_n^{\mathrm{CMC}} + c(\hat{X}_n - \mathbb{E}\, X).$$

Obliczmy wariancję nowego estymatora:

$$\operatorname{Var}(\hat{Y}_n^{CV}) = \operatorname{Var}(Y + cX)/n\ .$$

Wtedy $\operatorname{Var}(\hat{Y}^{CV})$ jest minimalna gdy c,

$$c = -\frac{\operatorname{Cov}(Y, X)}{\operatorname{Var}(X)}$$

i minimalna wariancja wynosi $\sigma_Y^2(1 - \rho^2)/n$, gdzie $\rho = \operatorname{corr}(Y, X)$. A więc wariancja jest zredukowana przez czynnik $1 - \rho^2$. Widzimy więc, ze należy

starać się mieć bardzo skorelowane Y z X. Ale musimy też umieć symulować zależną parę (Y, X).

W praktyce nie znamy $\mathrm{Cov}\,(Y, X)$ ani $\mathrm{Cov}\,(X)$ i musimy te wielkości symulować stosując wzory

$$\hat{c} = -\frac{\hat{S}^2_{YX}}{\hat{S}^2_X}$$

gdzie

$$\begin{aligned} \hat{S}^2_X &= \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \hat{Y})^2 \\ \hat{S}^2_{YX} &= \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \hat{Y})(X_j - \hat{X}). \end{aligned}$$

Przykład 3.1 Kontynuujemy przykład IV.1.1. Liczyliśmy tam π za pomocą estymatora $Y = 4\ \mathbf{1}(U_1^2 + U_2^2 \leq 1)$. Jako zmienną kontrolną weźmiemy $X = \mathbf{1}(U_1 + U_2 \leq 1)$, które jest mocno skorelowane z Y oraz $\mathbb{E}\,X = 1/2$. Możemy policzyć $1 - \rho^2 = 0.727$. Teraz weźmiemy inną zmienną kontrolną $X = \mathbf{1}(U_1 + U_2 > \sqrt{2})$. Wtedy $\mathbb{E}\,X = (2 - \sqrt{2})^2/2$ i $1 - \rho^2 = 0.242$. Zauważmy (to jest zadanie) że jeśli weźmiemy za zmienną kontrolną X lub $a + bX$ to otrzymamy taką samą redukcję wariancji.

Przykład 3.2 Chcemy obliczyć całkę $\int_0^1 g(x)\,dx$ metodą MC. Możemy położyć $Y = g(U)$ i jako kontrolną zmienną przyjąć $X = f(U)$ jeśli znamy $\int_0^1 f(x)\,dx$.

4 Warunkowa metoda MC

Punktem wyjścia jest jak zwykle obliczanie $I = \mathbb{E}\,Y$. Estymatorem CMC jest wtedy

$$\hat{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j,$$

gdzie $Y_1, \ldots, Y_n$ są niezależnymi kopiami Y. Oprócz zmiennej Y mamy też X. Dla ułatwienia rozważań załóżmy, że wektor (Y, X) ma łączną gęstość $f_{Y,X}(y, x)$. Wtedy brzegowa gęstość X jest

$$f_X(x) = \int f_{Y,X}(y, x)\,dy.$$

Niech

$$\begin{aligned}\phi(x) &= \mathbb{E}\left[Y|X=x\right] \\ &= \frac{\int y f_{Y,X}(y,x)\,dy}{f_X(x)}.\end{aligned}$$

Z twierdzenia o prawdopodobieństwie całkowitym

$$I = \mathbb{E}\, Y = \mathbb{E}\,\phi(X).$$

Wtedy definiujemy estymator warunkowy

$$\hat{Y}_n^{\mathrm{Cond}} = \frac{1}{n}\sum_{j=1}^{n} Y_j^{\mathrm{cond}}$$

gdzie $Y_1^{\mathrm{cond}}, \ldots, Y_n^{\mathrm{cond}}$ są replikacjami $Y^{\mathrm{cond}} = \phi(X)$. Estymator $\hat{Y}_n^{\mathrm{Cond}}$ jest nieobciążony. Pokażemy, że ma mniejszą wariancję. W tym celu skorzystamy z nierówności Jensena

$$\mathbb{E}\left[Y^2|V=x\right] \geq \left(\mathbb{E}\left[Y|V=x\right]\right)^2 .$$

Jeśli X ma gęstość f_X, to

$$\begin{aligned}\mathrm{Var}\left(Y\right) &= \int \mathbb{E}\left[Y^2|X=x\right] f_X(x)\,dx - I^2 \\ &\geq \int \left(\mathbb{E}\left[Y|X=x\right]\right)^2 f_X(x)\,dx - I^2 \\ &= \mathrm{Var}\left(\phi(V)\right)\end{aligned}$$

Dla czytelników znających ogólną teorię warunkowych wartości oczekiwanych wystarczy przytoczyć ogólny wzór

$$\mathrm{Var}\, Y = \mathbb{E}\,\mathrm{Var}\left(Y|X\right) + \mathrm{Var}\,\mathbb{E}\left(Y|X\right),$$

który można uzasadnić następująco. Wychodząc prawej strony

$$\begin{aligned}&\mathbb{E}\left[\mathrm{Var}\left(Y|X\right)\right] + \mathrm{Var}\left(\mathbb{E}\left[Y|X\right]\right) \\ &\quad= \mathbb{E}\left(\mathbb{E}\left[Y^2|X\right] - \left(\mathbb{E}\left[Y|X\right]\right)^2 + \mathbb{E}\left(\mathbb{E}\left[Y|X\right]\right)^2 - \left(\mathbb{E}\left(\mathbb{E}\left[Y|X\right]\right)^2\right. \\ &\quad= \mathbb{E}\, Y^2 - \left(\mathbb{E}\, Y\right)^2.\end{aligned}$$

Stąd mamy od razu

$$\mathrm{Var}\, Y \geq \mathrm{Var}\,\mathbb{E}\left(Y|X\right).$$

Przykład 4.1 Kontynuujemy przykład IV.1.1. Liczyliśmy tam π za pomocą estymatora $Y = 4\mathbf{1}(U_1^2 + U_2^2 \leq 1)$. Przyjmując za $X = U_1$ mamy

$$\phi(x) = \mathbb{E}\,[Y|X = x] = 4\sqrt{1 - x^2}.$$

Estymator zdefiniowany przez $Y^{\mathrm{Cond}} = 4\sqrt{1 - U_1^2}$ redukuje wariancję 3.38 razy, ponieważ jak zostało podane w przykładzie 5.1, mamy $\mathrm{Var}\,(Y^{\mathrm{Cond}}) = 0.797$ a w przykładzie IV.1.1 policzyliśmy, że $\mathrm{Var}\,(Y) = 2.6968$.

Przykład 4.2 Niech $I = \mathbb{P}(\xi_1 + \xi_2 > y)$, gdzie ξ_1, ξ_2 są niezależnymi zmiennymi losowymi z jednakową dystrybuantą F, która jest znana, a dla której nie mamy analitycznego wzoru na F^{*2}. Oczywiście za Y możemy wziąć

$$Y = \mathbf{1}(\xi_1 + \xi_2 > y)$$

a za $X = \xi_1$. Wtedy

$$Y^{\mathrm{Cond}} = \overline{F}(y - \xi_1).$$

Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o takim samym rozkładzie jak ξ_1. Wtedy $\hat{Y}_n^{\mathrm{cond}} = (1/n)\sum_{j=1}^n \overline{F}(y - X_j)$ będzie niobciążonym estymatorem I z wariancją mniejszą niż w zgrubnym estymatorze.

Przykład 4.3 Będziemy w pewnym sensie kontynuować rozważania z przykładu 4.2. Niech $S_n = \xi_1 + \ldots + \xi_n$, gdzie $\xi_1, \xi_2, \ldots, \xi_n$ są niezależnymi zmiennymi losowymi z jednakową gęstością f. Estymator warunkowy można też używać do estymowania gęstości $f^{*n}(x)$ dystrybuanty $F^{*n}(x)$, gdzie jest dana jawna postać f. Wtedy za estymator warunkowy dla $f^{*n}(y)$ zbudujemy na podstawie

$$Y^{\mathrm{cond}} = f(y - S_{n-1})\ .$$

5 Losowanie istotnościowe

Przypuśćmy, że chcemy obliczyć I, które możemy przedstawić w postaci

$$I = \int_E k(x) f(x)\, dx = \mathbb{E}\, k(Y),$$

gdzie E jest podzbiorem $\mathbb{R}$ lub $\mathbb{R}^d$, f jest funkcją gęstości na E, oraz funkcja $k(x)$ spełnia $\int_E (k(x))^2 f(x)\, dx < \infty$. Na przykład E jest odcinkiem $[0, 1]$ lub

półprostą $[0,\infty)$. Zakładamy, że Y określona na przestrzeni probabilistycznej $(\Omega,\mathcal{F},\mathbb{P})$ jest o wartościach w E. Estymatorem CMC dla I jest

$$\hat{Y}_n^{\mathrm{CMC}} = \frac{1}{n}\sum_{j=1}^{n} Y_j,$$

gdzie $Y_1,\ldots,Y_n$ są niezależnymi kopiami Y. Niech $\tilde{f}(x)$ będzie gęstością na E, o której będziemy zakładać, że jest ściśle dodatnia na E i niech X będzie zmienną losową z gęstością $\tilde{f}(x)$. Wtedy

$$I = \int_E k(x)\frac{f(x)}{\tilde{f}(x)}\tilde{f}(x)\,dx = \mathbb{E}\,k(X)\frac{f(X)}{\tilde{f}(X)} = \mathbb{E}\,k_1(X),$$

gdzie $k_1(x) = k(x)f(x)/\tilde{f}(x)$.

Teraz definiujemy *estymator istotnościowy* (*importance sampling*)

$$\hat{Y}_n^{\mathrm{IS}} = \frac{1}{n}\sum_{j=1}^{n} k(X_j)\frac{f(X_j)}{\tilde{f}(X_j)},$$

gdzie $X_1,\ldots,X_n$ są niezależnymi kopiami X. Oczywiście $\hat{Y}_n^{\mathrm{IS}}$ jest nieobciążony bo

$$\mathbb{E}\,k_1(X) = \int_E k(x)\frac{\tilde{f}(x)}{f(x)}\tilde{f}(x)\,dx = I,$$

z wariancją

$$\mathrm{Var}\,(\hat{Y}_n^{\mathrm{IS}}) = \frac{1}{n}\mathrm{Var}\,\left[k(X)\frac{f(X)}{\tilde{f}(X)}\right] = \frac{1}{n}(\int_E \left(k(x)\frac{f(x)}{\tilde{f}(x)}\right)^2 \tilde{f}(x)\,dx - I^2).$$

Okazuje się, że odpowiednio wybierając $\tilde{f}(x)$ można istotnie zmniejszyć wariancję. Niestety, jak pokazuje następujący przykład, można też istotnie sytuację pogorszyć.

Przykład 5.1 Funkcja ćwiartki koła jest zadana wzorem

$$k(x) = 4\sqrt{1-x^2}, \qquad x\in[0,1],$$

i oczywiście

$$\int_0^1 4\sqrt{1-x^2}\,dx = \pi.$$

W tym przykładzie $Y = U$ jest o rozkładzie jednostajnym na $E = [0, 1]$. Zauważmy, że $f(x) = \mathbf{1}_{[0,1]}(x)$. Wtedy zgrubny estymator jest dany przez

$$\hat{Y}_n^{\mathrm{CMC}} = \frac{1}{n}\sum_{j=1}^{n} 4\sqrt{1-U_j^2}$$

i ma wariancję

$$\operatorname{Var}\hat{Y}_n^{\mathrm{CMC}} = \frac{1}{n}(\int_0^1 16(1-x^2)\,dx - \pi^2) = \frac{0.797}{n}.$$

Niech teraz $\tilde{f}(x) = 2x$. Wtedy estymatorem istotnościowym jest

$$\hat{Y}_n^{\mathrm{IS}} = \frac{1}{n}\sum_{i=1}^{n} \frac{4\sqrt{1-X_i^2}}{2X_i},$$

gdzie $X_1, \ldots, X_n$ są niezależnymi replikacjami liczby losowej X z gęstością $2x$, ma wariancję

$$\operatorname{Var}(\hat{Y}_n^{\mathrm{IS}}) = \frac{1}{n}\operatorname{Var}\frac{k(X)}{\tilde{f}(X)}.$$

Na koniec policzymy, że wariancja

$$\operatorname{Var}\frac{k(X)}{\tilde{f}(X)} = \mathbb{E}\left(\frac{4\sqrt{1-X^2}}{2X}\right)^2 - I^2 = 8\int_0^1 \frac{1-x^2}{x}\,dx - I^2 = \infty.$$

Teraz pokażemy inną gęstość $\tilde{f}$ dla której estymator istotnościowy, jest z mniejszą wariancją niż dla estymatora CMC. Niech

$$\tilde{f}(x) = (4-2x)/3.$$

Wtedy

$$\hat{Y}_n^{\mathrm{IS}} = \frac{1}{n}\sum_{i=1}^{n} \frac{4\sqrt{1-X_i^2}}{(4-2X_i)/3},$$

i

$$\operatorname{Var}\frac{k(X)}{\tilde{f}(X)} = \int_0^1 \frac{16(1-x^2)}{(4-2x)/3} - \pi^2 = 0.224$$

a więc ponad 3 razy lepszy od estymatora CMC.

Załóżmy na chwilę, że funkcja k jest nieujemna i rozważmy nową gęstośc

$$\tilde{f}(x) = k(x)f(x)/\int_E k(x)f(x)\,dx = k(x)f(x)/I.$$

Wtedy wariancja estymatora IS jest zero, ale taki estymator nie ma praktycznego sensu, bo do jego definicji potrzebujemy I, które właśnie chcemy obliczyć. Ale ta uwaga daje nam wskazówkę jak dobrać $\tilde{f}$. Mianowicie ma to być gęstość przypominająca kształtem $k(x)f(x)$. Proponujemy porównać $k(x)f(x) = k(x)$ z odpowienio z gęstościami $\tilde{f}$ z przykładu 5.1.

Podany teraz materiał jest opcjonalny. Zauważmy, że te same rozważania można przeprowadzić formalnie inaczej. Niech $(\Omega, \mathcal{F})$ będzie przestrzenią mierzalną i $X_1, X_2, \ldots$ będą zmiennymi losowymi na tej przestrzeni. Przyjmijmy, że $\mathcal{F}_n$ jest generowana przez $\{X_1 \in B_1, \ldots, X_n \in B_n\}$, gdzie $B_j \in \mathcal{B}(\mathbb{R})$ $(j = 1, \ldots, n)$ oraz, że $\mathcal{F} = \sigma(\bigcup_{n\geq 1} \mathcal{F}_n)$. Teraz $\mathbb{P}$ i $\tilde{\mathbb{P}}$ są miarami probabilistycznymi na Ω takimi, że odpowiednio $X_1, X_2, \ldots$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie odpowiednio z gęstością f lub $\tilde{f}$.

Fakt 5.2 *Niech $\tilde{f} > 0$ p.w.. Wtedy na $\mathcal{F}_n$*

$$\mathbb{P}(d\omega) = L_n(\omega)\tilde{\mathbb{P}}(d\omega),$$

gdzie

$$L_n = \prod_{j=1}^{n} \frac{f(X_j)}{\tilde{f}(X_j)}. \tag{5.5}$$

Jeśli Y jest zmienną losową mierzalną $\mathcal{F}_n$, to

$$I = \mathbb{E}\,Y = \tilde{\mathbb{E}}\,[YL_n]. \tag{5.6}$$

Dowód Niech $A = \{X_1 \in B_1, \ldots, X_n \in B_n\} \in \mathcal{F}_n$. Wtedy

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \int_{B_1} \ldots \int_{B_n} f(x_1)\cdots f(x_n)\,dx_1 \ldots dx_n \\ &= \int_{B_1} \ldots \int_{B_n} \frac{f(x_1)\cdots f(x_n)}{\tilde{f}(x_1)\cdots \tilde{f}(x_n} \tilde{f}(x_1)\cdots\tilde{f}(x_n)\,dx_1 \ldots dx_n.\end{aligned}$$

To znaczy

$$\mathbb{E}\,\mathbf{1}(A) = \tilde{\mathbb{E}}\,L_n\mathbf{1}(A).$$

Stąd aproksymując Y funkcjami prostym otrzymujemy (5.6). □

Zauważmy, że L zdefiniowany przez (5.5) nazywa sie ilorazem wiarygodności.

Przykład 5.3 Policzymy teraz iloraz wiarygodności dla przypadku, gdy f jest gęstością standardowego rozkładu normalnego N(0,1), natomiast $\tilde{f}$ jest gęstością rozkładu normalnego N$(\mu, 1)$. Wtedy

$$L_n = \exp(-\mu \sum_{j=1}^{n} X_j + \frac{\mu^2}{2} n). \tag{5.7}$$

Postępując odwrotnie zdefiniujmy $\tilde{\mathbb{P}}$ na $\mathcal{F}_n$ przez

$$d\tilde{\mathbb{P}} = L_n^{-1} d\mathbb{P}.$$

Zauważmy, ze teraz musimy założyć, że $f(x) > 0$ p.w.. Wtedy przy nowej mierze probabilistycznej $\tilde{\mathbb{P}}$ zmienne $X_1, X_2, \ldots$ są niezależne o jednakowym rozkładzie N$(\mu$,1).

Zadania teoretyczne

[5]

5.1 Niech $F \sim \mathrm{Exp}(\lambda)$ i $Z_1 = F^{-1}(U)$, $Z_2 = F^{-1}(1-U)$. Pokazać, że

$$\mathbb{E}\, Z_i = 1/\lambda, \qquad \mathbb{E}\, Z_i^2 = 2/\lambda^2,$$

oraz

$$\mathrm{corr}\,(Z_1, Z_2) = -0.6449.$$

Wsk. $\int_0^1 \log x \log(1-x)\, dx = 0.3551$.

5.2 Obliczyć wartość oczekiwaną i wariancję rozkładu Pareto $\mathrm{Par}(\alpha)$.

5.3 Niech $Z_1 = F^{-1}(U)$, $Z_2 = F^{-1}(1-U)$. Oszacować korelację $\mathrm{corr}\,(Z_1, Z_2)$ gdy F jest dystrybuantą:
(a) rozkładu Poissona; $\lambda = 1, 2, \ldots, 10$,
(b) rozkładu geometrycznego $\mathrm{Geo}(p)$.

5.4 a) Napisać algorytm na generowanie zmiennej losowej Y z gęstością

$$f(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{w pozostałych przypadkach.} \end{cases}$$

Uzasadnić!

b. Przypuśćmy, że chcemy symulować $I = \mathbb{E}\, Y$, gdzie Y ma rozkład jak w częsci a. Ile należy zaplanować replikacji aby otrzymać I z dokładnością do jednego miejsca po przecinku, gdy

1. stosujemy zgrubny estymator Monte Carlo – CMC,
2. stosujemy metodę zmiennych antytetycznych.

Wsk: $\mathrm{corr}(\mathrm{rand}^{1/2}, (1-\mathrm{rand})^{1/2}) = -0.9245$.

5.5 Napisać wzór na estymator warstwowy, jeśli warstwy definoiowane są przez zmienną X.

[5] 1.2.2016

5.6 [Asmussen & Glynn [4]] Przedyskutować metodę redukcji wariancji dla obliczenia

$$\int_0^\infty (x + 0.02x^2) \exp(0.1\sqrt{1 + \cos x} - x)\, dx$$

metodą MC.

5.7 Dla przykładu 5.1 obliczyć wariancję $\hat{Y}_n^{\mathrm{IS}}$ gdy $\tilde{f}(x) = (4 - 2x)/3$, $0 \leq x \leq 1$.

5.8 [Madras [?]] (i) Chcemy symulować $I = \int_0^\infty k(x)\, dx$, gdzie

$$k(x) = \sin(x)(1 + x)^{-2} e^{-x/5}$$

za pomocą estymatora istotnościowego przy losowaniu zmiennych o rozkładzie wykładniczym z parametrem λ. Pokazać, że dla pewnych wielkośći λ wariancja estymatora istotnościowego jest nieskończona.
(ii) Pokazać, że jesli $k(x) = \sin(x)(1 + x)^{-2}$, to wariancja estymatora istotnościowego przy użyciu losowań o rozkładzie wykładniczym jest zawsze nieskończona. Zaprojektować dobrą metodą symulacji I.
(iii) Przypuśćmy, że całka I jest zbieżna warunkowo (tzn. nie jest zbieżna absolutnie). Pokazać, że wtedy każdy estymator istotnościowy ma nieskończoną średnią.

5.9 Niech $I = \mathbb{P}(L > x)$, gdzie

$$L = \sum_{j=1}^{N} D_n X_n,$$

$X_1, \ldots, X_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie N(0,1), niezależnymi od $D_1, \ldots, D_n$. Niech P ma rozkład Beta(1,19) z gęstością $(1 - p)^{18}/19$, $0 < p < 1$. Pod warunkiem $P = p$ zmienne $D_1, \ldots, D_n$ są niezależnymi próbami Bernoulliego z prawdopodobieństwem sukcesu p. Podać algorytm symulacji I warunkową metodą MC, z warunkowaniem względem $\sum_{j=1}^{N} D_j$. Wziąć x

$$x = 3\mathbb{E}\, L = 3N \exp P\mathbb{E}\, X = 3 \cdot 100 \cdot 0.05 \cdot 3 = 45.$$

(Asmussen&Glynn,p. 147).

5.10 W przypadku symulacji $I = \int_{[0,1]^d} g(x)\,dx$, gdy d jest duże, metoda warstw po każdej zmiennej będzie nieużyteczną ze względu na duża liczbę wastw. Remedium jest wykorzystanie tzw. kwadratów łacińskich. Niech $U_i^{(j)}$ będą niezależnymi zmiennym losowymi o rozkładzie U(0,1) oraz $\pi_i = (\pi_i(1), \ldots, \pi_i(n))$ będzie ciągiem niezależych losowych permutacji $(1, \ldots, n)$, $i = 1, \ldots, d$. Dla $j = 1, \ldots, n$

$$V^{(j)} = (V_1^{(j)}, \ldots, V_d^{(j)}) \in [0,1]^d,$$

gdzie

$$V_i^{(j)} = \frac{\pi_i(j) - 1 + U_i^{(j)}}{n}$$

Pokazać, że estymator

$$\hat{Y}_n^{LH} = \frac{1}{n} \sum_{j=1}^{n} g(V^{(j)})$$

jest estymatorem nieobciążonym I. Pokazać, że

$$\operatorname{Var} \hat{Y}_n^{LH} = \frac{\sigma_Y^2}{n} + \frac{n-1}{n} \operatorname{Cov}\left(g(V^{(1)}), g(V^2)\right)$$

gdzie $\sigma_Y^2 = \operatorname{Var} g(U_1, \ldots, U_d)$.

Zadania laboratoryjne

5.11 Niech $k_1(x) = e^{\sqrt{x}}$ i $k_2(x) = (1 + \frac{\sqrt{x}}{50})^{50}$ oraz $X = U$. Można oszacować, że $|k_1(x) - k_2(x)| \leq 0.03$, a więc $d \leq 0.03$. Przprowadzić eksperyment oblicznenia I oraz b obiema metodami rozpatrywanymi w podrozdziale 2.2.

5.12 Niech F będzie dystrybuantą rozkładu standardowego normalnego. Niech $p_i = 1/10$. Korzystając z tablic rozkładu normalnego wyznaczyć a_j,

takie, że

$$\begin{aligned} a_0 &= -\infty \\ a_1 &= F^{-1}(1/10) \\ a_2 &= F^{-1}(2/10) \\ \vdots &= \vdots \\ a_9 &= = F^{-1}(9/10) \\ a_{10} &= \infty. \end{aligned}$$

W pierwszym eksperymencie przeprowadzić 100 replikacji i narysować histogram z klasami $A_1, \ldots, A_{10}$. Natomiast drugi eksperyment przeprowadzić przy losowaniu w warstwach losując w każdej 10 replikacji. Porównąc otrzymane rysunki.

5.13 Pokazać, że estymator istotnościowy do obliczenia $I = \int_E k(x)f(x)$ zdefiniowany przez

$$k(x)f(x)/I$$

ma wariancję równą zero.

Rozdział VI

Symulacje stochastyczne w badaniach operacyjnych

[1] W tym rozdziale pokażemy wybrane przykłady zastosowań symulacji do konkretnych zadań w szeroko pojętej teorii badań operacyjnych, modelach telekomunikacyjnych, itd. Oczywiście przykładowe zadania podane będa w postaci uproszczonej. Nowym elementem pojawiającym się w tym rozdziale, jest to, że niektóre rozpatrywane procesy ewuulują losową w ciągłym czasie. Będziemy rozpatrywać pewne procesy stochastyczne, ale do symulacji będziemy odwoływać się do modelu a nie do konkretnej teorii procesów stochastycznych.

Aby dobrze zanalizować postawione zadanie należy

- zidentyfikować pytania które chcemy postawić,
- poznać system,
- zbudować model,
- zweryfikować poprawność założeń modelu,
- przeprowadzić eksperymenty z modelem,
- wybrać rodzaj prezentacji.

Ewolucja interesującej nas wielkości zmieniającej się wraz z czasem (np. ilość oczekujących zadań w chwili t, gdzie $0 \leq t \leq T_{\max}$, stan zapasów w

[1] 1.2.2016

chwili t, itd.) jest losowa. Zmienna, która nas interesuje może być skalarem lub wektorem. Zmienna ta może się zmieniać ciągle (patrz przykład realizacji ruchu Browna na rys. **[????]**) albo jedynie w chwilach dysktretnych. Wszystkie procesy możemy symulować w chwilach $0, \Delta, 2\Delta, \ldots$ i wtedy mamy do czynienia z *symulacją synchroniczną* . Jednakże dla specyficznej klasy zadań, gdy ewolucja zmienia się dyskretnie możemy przeprowadzać *symulację asychroniczną*, i to będzie głównym tematem rozważanym w tym rozdziale.

W wielu problemach mamy do czynienia ze zgłaszającymi się żądaniami/jednostkami itp. w sposób niezdeterminowany. Do opisu takiego strumienia zgłoszeń potrzebujemy modeli z teorii procesów stochastycznych. W następnym podorozdziale opiszemy tzw. proces Poissona (jednorodny i niejednorodny w czasie), który w wielu przypadkach jest adekwatym modelem.

1 Procesy zgłoszeń

1.1 Jednorodny proces Poissona

Zajmiemy się teraz opisem i następnie sposobami symulacji losowego rozmieszczenia punktów. Tutaj będziemy zajmowali się punktami na $\mathbb{R}_+$, które modelują *momenty zgłoszeń*. Proces zgłoszeń jest na przykład potrzebny do zdefiniowania procesu kolejkowego lub procesu ryzyka. Najprostszy sposób jest przez zadanie tak zwanego procesu liczącego $N(t)$. Jest to proces $N(t)$ $(t \geq 0)$, który zlicza punkty do momentu t. Zauważmy, że ponieważ momenty zgłoszeń na $\mathbb{R}_+$ są losowe, więc $N(t)$ jest procesem losowym, który ma niemalejące, kawałkami stałe z jenostkowymi skokami realizacje i $N(0) = 0$. Ponadto przyjmujemy, że jako funkcja czasu t jest prawostronnie ciągły. Interpretacją zmiennej losowej $N(t)$ jest liczba punktów do chwili t lub w w odcinku $(0, t]$. Zauważmy, że $N(t+h) - N(t)$ jest liczbą punktów w odcinku $(t, t+h]$. W dalszej części będziemy oznaczać przez $0 < A_1 < A_2 < \ldots$ momenty zgłoszeń na osi dodatniej $(0, \infty)$. Przez

$$\tau_1 = A_1, \quad \tau_i = A_i - A_{i-1}, \qquad (i = 2, \ldots) \tag{1.1}$$

oznaczamy *ciąg odstępów między zgłoszeniami*. Wtedy

$$N(t) = \#\{i : A_i \leq t\}$$

nazywamy procesem liczącym.

Przez proces Poissona (lub *jednorodny proces Poissona*) o intensywności $\lambda > 0$ nazywamy proces liczący dla którego (τ_j) są niezależnymi zmiennymi losowymi o jednakowym rozkładzie wykładniczym Exp(λ). Zauważmy, że $(N(t))_{t\geq 0}$ jest procesem Poissona o intensywności λ wtedy i tylko wtedy gdy

1. liczby punktów w rozłącznych odcinkach są niezależne,

2. liczba punktów w odcinku $(t, t+h]$ ma rozkład Poissona z średnią λh.

Zauważmy, że

$$\mathbb{P}(N(t, t+h] = 1) = \lambda h e^{-\lambda h} = \lambda h + o(h), \tag{1.2}$$

i dlatego parametr λ nazywa się intensywnością procesu Poissona. Inna interpretacja to $\mathbb{E}\, N(t, t+h] = \lambda h$, czyli λ jest średnią liczbą punktów na jednostkę czasu. Zauważmy, że tutaj ta średnia, którą często zwie się intensywnością λ, jest niezależna od czasu t.

Przypomnijmy, że w podrozdziale III.2 zauważyliśmy natępujący fakt. Niech $\tau_1, \tau_2, \ldots$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie Exp(1) oraz

$$N = \#\{i = 1, 2, \ldots : \tau_1 + \ldots + \tau_i \leq \lambda\}.$$

Wtedy zmienna losowa N ma rozkład Poissona z średnią λ. Okazuje się, że tak naprawdę kolejne sumy niezależnych zmiennych losowych o jednakowym rozkładzie wykładniczym Exp(λ) definiują coś więcej, czyli proces Poissona. A więc i-ty punkt ma współrzędną $A_i = \tau_1 + \ldots + \tau_i$, natomiast proces liczący można zapisać

$$N(t) = \#\{i = 1, 2, \ldots : \tau_1 + \ldots + \tau_i \leq t\}, \qquad t \geq 0.$$

Podamy teraz procedurę matlabowską generowania procesu Poissona w odcinku (0,tmax].

Procedura poipro

```
% poipro simulate  Poisson process
% of intensity lambda.
%
% Inputs: tmax - simulation interval
%         lambda - arrival intensity
```

```
%
% Outputs: n - number of arrivals in (0,tmax],
%          arrtime - time points of arrivals,
%                    if n=0 then arrtime='empty'

arrtime=-log(rand)/lambda;
if (arrtime>tmax)
  n=0; arrtime='empty'
  else
  i=1;
  while (arrtime(i)<=tmax)
    arrtime = [arrtime, arrtime(i)-log(rand)/lambda];
    i=i+1;
  end
  n=length(arrtime);      % arrival times t1,...,tn
end
```

1.2 Niejednorodny proces Poissona

Jednorodny w czasie proces Poissona nie zawsze oddaje dobrze rzeczywisty proces zgłoszeń, ze względu na to że intensywność jest stała. Istnieją potrzeby modelowania procesu zgłoszeń w którym intensywność zależy od momentu t. Na przykład w cyklu całodobowym jest naturalnym rozróżnienie intensywności przybyć ze względu na porę dnia. Zdefiniujemy teraz *niejednorodny proces Poissona* z *funkcją intensywności* $\lambda(t)$. Jeśli intensywność jednorodnego procesu Poissona λ jest $\mathbb{P}(N(dt) = 1) = \lambda\, dt$ (to jest inny zapis własności (1.2)), to w przypadku procesu niejednorodnego chcemy $\mathbb{P}(N(dt) = 1) = \lambda(t)\, dt$. Formalnie definiujemy niejednorodny proces Poissona z funkcją intensywności $\lambda(t)$ jako proces liczący, taki że

1. liczby punktów w rozłącznych odcinkach są niezależne,
2. liczba punktów w odcinku $(t, t + h]$ ma rozkład Poissona z średnią $\int_t^{t+h} \lambda(s)\, ds$.

Okazuje się, że jesli w interesującym nas odcinku $(0, t_{\max}]$, $\lambda(t) \leq \lambda_{\max}$ to proces niejednorodny można otrzymać z procesu Poissona o intensywności $\lambda_{\max}$ przez niezależne *przerzedzanie* punktów jednorodnego procesu Poissona; jeśli jest punkt w punkcie t to zostawiamy go z prawdopodobieństwem

$p(t) = \lambda(t)/\lambda_{\max}$. Mamy więc schemat dwukrokowy. Najpierw generujemy proces Poissona o stałej intensywności $\lambda_{\max}$ a następnie przerzedzamy każdy punkt tego procesu, niezależnie jeden od drugiego, z prawdopdobieństwami zależnymi od położenia danego punktu.

Podamy teraz procedurę matlabowską generowania procesu Poissona w odcinku $(0, t_{\max}]$.

Procedura nhpoipro

```
% nhpoipro simulate  nonhomogeneous Poisson process
% of intensity lambda(t).
%
% Inputs: tmax - simulation interval
%         lambda - arrival intensity function lambda(t)
%
% Outputs: n - number of arrivals in (0,tmax],
%          arrtime - time points of arrivals,
%                    if n=0 then arrtime='empty'

arrt=-log(rand)/lambdamax;
if (arrt>tmax)
  n=0; arrtime='empty';
  else
  i=1;
  while (arrt(i)<=tmax)
    arrtime=arrt;
    arrt = [arrt, arrt(i)-log(rand)/lambdamax];
    i=i+1;
  end
  m=length(arrtime);
  b= lambda(arrtime)/lambdamax;
  c= rand(1,m);
  arrtime=nonzeros(arrtime.*(c<b))
  n=length(arrtime);
end
```

1.3 Proces odnowy

Jeśli odstępy między zgłoszeniami $\tau_1, \tau_2, \ldots$ tworzą ciąg niezależnych nieujemnych zmiennych losowych o jednakowym rozkładzie, to mówimy że ciąg $A_1, A_2, \ldots$ zdefiniwany przez (1.1) jest *procesem odnowy*. W szczególności, jeśli o rozkładzie wykładniczym, to mamy proces Poissona. Niestety taki proces nie ma własności stacjonarności przyrostów, tj. gdy wektory

$$(N(s+t_2) - N(s+t_1), \ldots, N(s+t_n) - N(s+t_{n-1}))$$

mają rozkład niezależny od $s > 0$ ($n = 2, 3, \ldots, 0 \leq t_1 < t_2 < \ldots$).

2 Metoda dyskretnej symulacji po zdarzeniach

Wiele procesów stochastycznych modelujących interesujące nas ewolucje ma prostą strukturę. Przestrzeń stanów takiego procesu E jest przeliczalna a proces zmienia stan od czasu do czasu. Moment zmiany stanu nazywamy tutaj 'zdarzeniem'. Tak więc aby symulować system musimy przechodzić *od zdarzenia do zdarzenia*. Taka metoda nazywa sie w języku angielskim *discrete event simulation*. Zadania które możemy symulować w ten sposób spotykamy w wielu teoriach, między innymi

- zarządzania produkcją,
- zapasów,
- systemów telekomunikacyjnych,
- niezawodności.

Przykład 2.1 Najprostszym przykładem jest symulacja procesu *on-off*, który przyjmuje wartości $E = \{0, 1\}$. Jest to proces $L(t)$ alternujący pomiędzy 1 (stan *on* – włączenia) i 0 (stan *off* – wyłączenia). Kolejne czasy włączeń i wyłączeń są niezależne, czasy włączeń są o tym samym rozkładzie F_{on} i czasy włączeń są o tym samym rozkładzie F_{off}. Zdarzeniami są kolejne momenty włączeń i wyłączeń. Jednym z wskaźników jest tak zwany współczynnik gotowości – stacjonarne prawdopodobieństwo tego, że system jest sprawny. Jeśli $\mathbb{E}\, T^{\text{on}} < \infty$ jest średnim czasem bycia w stanie *on* a $\mathbb{E}\, T^{\text{off}} < \infty$ jest średnim czasem bycia w stanie *off* to wiadomo jest, że stacjonarny współczynnik gotowości wynosi $\rho = \mathbb{E}\, T^{\text{on}} / (\mathbb{E}\, T^{\text{on}} + \mathbb{E}\, T^{\text{on}})$. Symulacja $L(t)$ jest bardzo

prosta. Na zmianę musimy generować T^{on} i T^{off} tak długo aż przkroczymy horyzont $t_{\max}$. Teraz musimy obliczyć estymator

$$\hat{\rho} = \frac{1}{t_{\max}} \int_0^{t_{\max}} L(s)\, ds.$$

Jeśli założyć na przykład, że F_{on} i F_{off} mają gęstość, to estymator $\hat{\rho}$ współczynnika ρ jest zgodny, ale nie jest on nieobciążony.

Przykład 2.2 [Zagadnienie konserwatora] W urządzeniu jest N elementów, które mają bezawaryjny czas pracy o rozkładzie F; jesli któryś z elementów popsuje się, to czas naprawy przez konserwatora ma rozkład G. Załózmy, że jest jeden konserwator. Jeśli więc jest on zajęty, to pozostałe popsute muszą czekać. Przyjmujemy, że wszystkie zmienne są niezależne. Możemy się pytać, na przykład o liczbę niesprawnych elementów $L(t)$ w momencie $t \leq t_{\max}$. Jest to oczywiście zmienna losowa dla której możemy chcielibyśmy policzyć funkcję prawdopodobieństwa lub w szegółności średnią $\mathbb{E}\, L(t)$. Przy odpowiednich założeniach regularności ta wielkość dąży do pewnej stałe l, która jest oczekiwaną stacjonarną liczbą niesprawnych elementów. Możemy się równiez pytać o czas B do pierwszego momentu gdy wszystkie maszyny są zepsute. Ile wynosi $\mathbb{E}\, B$? Jeśli rozkłady F i G są dowolne, to nie znamy wzorów na postawione pytania. A więc pozostaje symulacja. Zdarzeniami są w momenty awarii i naprawy elementu. W przypadku awarii liczba niesprawnych zwiększa się o jeden natomiast naprawy zmniejsza się o jeden. Estymatorem l jest

$$\hat{l} = \frac{1}{t_{\max}} \int_0^{t_{\max}} L(s)\, ds.$$

Ogólny opis metody jest następujący. Przypuścmy, że chcemy symulować $\boldsymbol{n}$ – *stan systemu* do momentu $t_{\max}$ - horyzont symulacji. Jedyne stochastyczne zmiany systemu są możliwe w pewnych momentach $A_1, A_2, \ldots$. Pomiędzy tymi momentami system zachowuje się deterministycznie (przyjmujemy, że jest stały chociaż istnieją modele pozwalajające opuścić to założenie - tzw. procesy kawałkami deterministyczne Markowa). Przypuścmy, że mamy N zdarzeń $\mathcal{A}_i$ $(i = 1, \ldots, N)$. Pomysł polega na utworzeniu listy zdarzeniowej LZ w której zapisujemy momenty następnych zdarzeń po momencie obecnego; zdarzenie $\mathcal{A}_i$ będzie w momencie $t_{\mathcal{A}_i}$. Jeśli któreś zdarzenie nie może w następnym kroku nastąpić, to w liście zadaniowej nie będzie

zarejestrowany jego moment pojawienia się. Tak może być na przykład z momentami zgłoszeń do systemu po zadanym horyzoncie $t_{\max}$.

Algorytm znajduje najbliższe zdarzenie, podstawia nową zmienną czasu, wykonuje odpowiednie aktywności, itd. Oczywiście aby symulacja bya możliwa muszą być poczybione odpowiednie założenia niezależności.

Przykład 2.3 Narysujemy teraz realizacje systemu kolejkowego M/M/1 z intensywnością zgłoszeń λ i intensywnością obsługi μ. Interesuje nas proces $L(t)$ liczby zadań w systemie w chwili t. Odsyłamy do dodatku (**[????]**) po informacje z teorii kolejek, w szczeglólności systemu M/M/1. Realizacje takiego procesu można otrzymać korzystając z symulacji po zdarzeniach, jednakże my tutaj wykorzystamy ze specjalnej struktury procesów narodzin i śmierci. Proces $L(t)$ jest bowiem procesem narodzin i śmierci (B&D) z intesywnościa narodzin λ i intensywnością śmierci μ. Korzystając więc z teorii procesów B&D (patrz dodatek **[????]**) wiemy, że

- jeśli system jest niepusty to czasy do kolejnych skoków są wykładnicze z parametrem $\lambda+\mu$, skok górę jest z prawdopodobieństwem $\lambda/(\lambda+\mu)$, natomiast w dół z prawdopodobieństwem $\mu/(\lambda+\mu)$,
- jeśli system jest pusty to najbliższy skok (do jedynki) jest po czasie wykładniczym z parametrem λ.

Podamy teraz prosty algorytm na generowania realizacji długości kolejki gdy $\lambda = 8/17$ i $\mu = 9/17$ aż do 100-nego skoku. Zauważmy, że tutaj $\lambda + \mu = 1$, i że $L(0) = 5$. [2] Do wykreślenia realizacji będziemy używać matlabowskiego polecenia

```
stairs (x,y)
```

gdzie $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ które przy założeniu, że $x_1 < \ldots < x_n$, wykreśla łamaną łączącą punkty

$$(x_1, y_1), (x_2, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

```
clc;clear;
t=0;
s=5;
```

[2] mm1v1.m

```
jumptimes=[t];
systemsizes=[s];
    for i=1:1000
        if s>0
            t=t-log(rand);
            s=s+(2*floor((17/9)*rand)-1);
        else
            t=t-17*log(rand)/8;
            s=1;
        end
        jumptimes=[jumptimes,t];
        systemsizes=[systemsizes,s];
     end
stairs(jumptimes,systemsizes)
```

Na rysunkach 2.1 i 2.2 są przykładowe symulacje jednej realizacji dla wybranych parametrów λ i μ.

Zmodyfikujemy teraz algorytm tak aby otrzymać realizację na odcinku $[0, t_{\max}]$. Realizując algorytm otrzymamy dwa ciągi $(0, t_1, \ldots, t_{\max}), (s_0, s_1, \ldots, s_{\max})$, które następnie służą do wykreślenia realizacji za pomocą polecenia "stairs".

2.1 Modyfikacja programu na odcinek $[0, tmax]$.

```
clc;clear;
t=0;
s=5;
tmax=1000;
jumptimes=[t];
systemsizes=[s];
    until t<tmax;
        if s>0
            t=t-log(rand);
            s=0,s+(2*floor((17/9)*rand)-1);
        else
            t=t-7*log(rand)/8;
            s=1;
        end
        jumptimes=[jumptimes,t];
```

```
        systemsizes=[systemsizes,s];
    end
jumptimes =[jumptimes, tmax];
systemsizes =[systemsizes,s];
stairs(jumptimes,systemsizes)
```

Teraz zastanowimy się jak obliczyć $\int_0^{t_{\max}} L(s)\,ds$ w celu obliczenia średniej długości kolejki

$$\text{srednia} = \frac{\int_0^{t_{\max}} L(s)\,ds}{t_{\max}}.$$

[3]

```
clc;clear;
t=0;
s=5; Lint=0;tdiff=0;
tmax=1000000;
    while t<tmax
        told=t;
        if s>0
        t=t-log(rand);
        tnew=min(t,tmax);
        tdiff=tnew-told;
        Lint=Lint+tdiff*s;
        s=s+(2*floor((17/9)*rand)-1);
        else
        t=t-17*log(rand)/8;
        s=1;
        end
    end
    srednia=Lint/tmax
```

Przeprowadzono eksperymetr otrzymując następujące wyniki
tmax=1000;srednia =8.2627
tmax=100000;srednia = 8.1331
tmax=10000000;srednia = 7.9499

[3] mm1sredniav2

Z teorii (patrz dodatek **[????]**) wiadomo, że z prawdopdobieństwem 1

$$\lim_{\to\infty} \frac{\int_0^t L(s)\,ds}{t} = 8.$$

W rozdziale VII (patrz przykład **[??ex.mm1.analiza??]**)zastanowimy się jak odpowiednio dobrać $t_{\max}$.

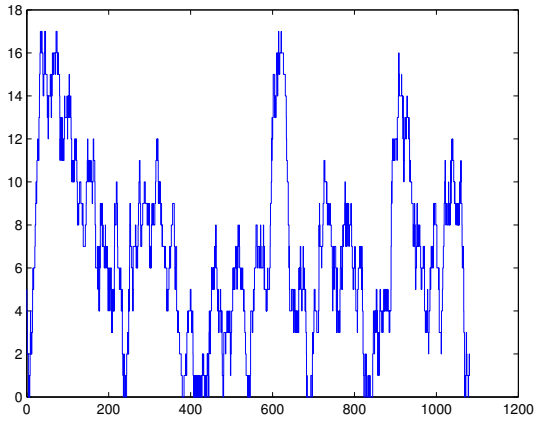
Ogólny algorytm dyskretnej symulacji po zdarzeniach

```
INICJALIZACJA

podstaw zmienną czasu równą 0;
ustaw parametry początkowe 'stanu systemu' i liczników;
ustaw początkową listę zdarzeniową LZ;

PROGRAM

podczas gdy zmienna czasu < tmax;
wykonaj
   ustal rodzaj następnego zdarzenia;
   przesuń zmienną czasu;
   uaktualnij 'stan systemu' & liczniki;
   generuj następne zdarzenia & uaktualnij listę zdarzeniową LZ
koniec

OUTPUT
Obliczyć estymatory & przygotować dane wyjściowe;
```

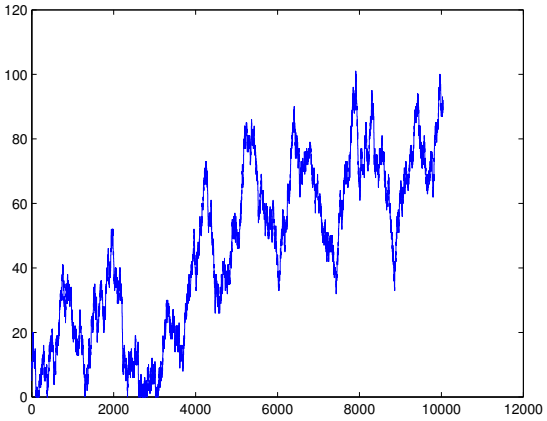
2.2 Symulacja kolejki z jednym serwerem

Metodę zilustrujemy na przykładzie algorytmu symulacji liczby zadań w kolejce GI/G/1. W dodatku XI.6 znajdują się fakty z teorii kolejek, a w szczególności opis systemu z jednym serwerem typu GI/G/1. Rozpatrujemy system w przedziale czasowym $[0, tmax]$. Dalej będziemy pisać $t_{\max}$ =tmax. Interesują nas dwie ewolucje:

- proces liczby zadań w systemie w chwili t; $0 \leq t \leq tmax$,
- ciąg czasów odpowiedzi $D(i)$.



Rysunek 2.1: Realizacja liczby zadań; $\lambda = \mu = .5$.



Rysunek 2.2: Realizacja liczby zadań; $\lambda = 8/17, \mu = 9/17$.

Metoda opisana poniżej jest odpowiednia dla symulacji długości kolejki, a przy okazji zbadania czasu odpowiedzi. W przypadku symulacji jedynie kolejnych czasów odpowiedzi, dla systemu GI/G/1 z dyscypliną FIFO możemy wykorzystać prostą rekurencję. Zajmiemy się więc symulacją procesu liczby zadań $L(t)$ w systemie w chwili t. Zakładamy, że dyscyplina obsługi jest zachowująca pracę, tj. w momencie skończenia obsługi zadania przez serwer, nie jest ważne które zadanie jest wybierane do obsługi (o ile w tym momencie czekają przynajmniej dwa zadania) ale tylko to, że serwer natychmiast podejmuje pracę.

Dla systemu GI/G/1 mamy dwa zdarzenia: *zgłoszenie* – $\mathcal{A}_1 = A$, *odejście* – $\mathcal{A}_2$=O. Poszczególne elementy algorytmu zależą od stanu *listy zdarzeniowej* – LZ, tj. od wielkości tA,tO - czasów następnego odpowiednio zgłoszenia i odejścia. Możliwe stany listy zdarzeniowej LZ to {tA},{tO}, {tA,tO}, $\emptyset$.

Potrzebujemy

- t – zmienna czasu,
- n – zmienna systemowa; liczba zadań w chwili t,

Ponadto zbieramy też następujące dane (output variables):

- t(i),n(i) - zmienne otrzymane (output variables) – moment kolejnego i-tego zdarzenia oraz stan systemu tuż przed tym momentem; zbierane w macierzy JS (jumptimes,systemsizes). Początkowe wartości JS to $t = 0$ oraz stan początkowy s.

Niech F będzie dystrybuantą odstępu τ_i między zgłoszeniami oraz G dystrybuanta rozmiaru zadania. Na przykład w systemie M/M/1, F jest rozkładem wykładniczym z parametrem lambda natomiast G jest rozkładem wykladniczym z parametrem mi. Zauważmy, że:

- Na podstawie znajomości JS={(t(j),n(j))} możemy wykreślić realizację $L(s),\ 0 \leq s \leq tmax$.

Podamy teraz algorytm na symulację stanów kolejki z jednym serwerem. Będziemy zakładać, że rozkłady F i G są ciągłe. Na początku przyjmujemy, że w chwili $t = 0$ system jest pusty (n=0). Jeśli $n = 0$, to odpowiada mu przypadek LZ={tA} jesli tA<tmax, lub LZ=
empty} w przeciwnym razie.

Algorytm

```
POCZĄTEK (gdy w chwili 0 system jest pusty).

utwórz wektory/macierze:
   JS -- macierz dwu-wierszowa; JS=[0;0]

 Losujemy Tau o rozkładzie F.
  Jeśli Tau>tmax, to  LZ=emptyset.
  Jeśli Tau=< tmax,
      to dołączamy do JS  t=n=0,
      oraz podstawiamy tA=Tau i LZ={tA}.
```

Następna część zależy już od stanu listy LZ.

```
PRZYPADEK 1. LZ={tA}.

dołącz do JS: t:=tA, n:=n+1
generuj: zm. los. Tau o rozkładzie F do najbliższego zgłoszenia
generuj: zm. los. S o rozkładzie G (rozmiar następnego zadania)
resetuj: tO=t+S
resetuj: jeśli t+Tau>tmax to:  nowy LZ={tO}
resetuj: jesli t+Tau=< tmax, to tA=t+Tau; nowy LZ={tA,tO}

PRZYPADEK 2. LZ={tA,tO}, gdzie tA =< tO.

dołącz do JS: t=tA, n=n+1
generuj: Tau o rozkladzie F
resetuj: jeśli t+Tau>tmax to podstaw tO; nowy LZ={tO}
resetuj: jesli t+Tau < tmax
         to  podstaw  tA=t+Tau; nowy LZ={tA,tO}
```

```
PRZYPADEK 3. LZ={tA,tO}, gdzie tA > tO.

dołącz do JS: t=tO, n=n-1
jeśli n>0: jeśli to generuj: Y o rozkladzie  G i resetuj  tO=t+Y,
nowa lista zdarzeniowa LZ={tA,tO}
jeli n=0 to nowa  LZ={tA}

PRZYPADEK 4:  LZ={tO}

dołącz do JS: t=t0, n=max(0,n-1),  nowa lista zdarzeniowa
LZ={emptyset}.

PRZYPADEK 5:  LZ={emptyset}

dołącz do JS: t=tmax, n=n.

KONIEC -- skończ symulacje.
```

Teraz zrobimy kilka uwag o rozszerzeniu stosowalności algorytmu.

Dowolne warunki początkowe W przypadku, gdy w chwili 0 system nie jest pusty, musimy zadać/znać liczbę zadań w systemie w chwili 0, czas do zgłoszenia tA oraz czas do zakończenia obslugi tO. To znaczy musimy zadać listę zdarzeniową {tA,tO} oraz zmienną 'n'. Zauważmy, że w najogólniejszym przypadku gdy system w chwili 0 jest juz od pewnego czasu pracujący, to (n,tA,tO) jest wektorem losowym o zadanym rozkładzie. Oczywiście możemy przyjąć, że (n,tA,tO) jest deterministyczne. W ogólnym przypadku podanie tego rozkładu jest sprawą bardzo trudną. Ale na przykład w przypadku systemu M/M/1, tA ma rozkład wykładniczy z parametrem lambda (jeśli $n > 0$) oraz tO ma rozkład wykladniczy z parametrem mi. Jest to konsekwencja braku pamięci rozkładu wykładniczego; patrz XI.5.6.

2.3 Czasy odpowiedzi, system z c serwerami

Wyobraźmy sobie, że do systemu wpływają zadania – k-te w chwili A_k rozmiaru S_k. Jest c serwerów. Serwer i-ty ($i = 1, \ldots, c$) pracuje z szybkością $h(i)$. Bez straty ogólności możemy założyć, że $h(1) \geq h(2) \geq \ldots \geq h(c)$. A więc

zadanie rozmiaru S_k na i-tym serwerze jest jest obsługiwane przez $S_k/h(i)$ jednostek czasu. Teraz będzie ważna dyscyplina obsługi – *zgodna z kolejnością zgłoszeń*. Zauważmy, że w przypadku większej liczby kanałów nie ma równoważności z dyscypliną FIFO. W przypadku gdy kilka serwerów jest pustych, zostaje wybrany najszybszy, czyli z najmniejszym numerem. Interesującą nas charakterystyką jest czas odpowiedzi (response time) – czas od momentu zgłoszenia zadania do opuszczenia przez niego systemu.

W tym wypadku mamy $c+1$ zdarzeń: *zgłoszenie*, *odejście z pierwszego serwera*,...,*odejście z c-tego serwera*. Poszczególne elementy algorytmu zależą od stanu *listy zdarzeniowej* – LZ, tj. od wielkości tA,tO(1),...tO(c) - czasów następnego odpowiednio zgłoszenia i odejścia z odpowiedniego serwera. W liście zdarzeniowej musimy też przechowywać razem z $tO(i)$ numer $K(i)$ zadania, które w tym momencie opuszcza system. W przypadku $c = 1$ wystarczyłoby rejestrowanie kolejnych momentów głoszeń A_n i kolejnych momentów odejść z systemów, ponieważ w tym przypadku kolejność zgłoszeń odpowiada kolejności odejść. Przez krok będziemy rozumieli przejscie od jednego zdarzenia do następnego.

Możliwe stany listy zdarzeniowej LZ są podzbiorami {tA,tO(1),...,tO(c)} (włącznie ze zbiorem pustym $\emptyset$).

Będziemy potrzebować

- nmax – liczba zadań symulowanych,

oraz

- n – zmienna systemowa; liczba zadań w chwili t,
- NK – numer aktualnie czekającego na obsługę zadania,
- NA – liczba zgłoszeń do aktualnego kroku,
- NO – liczba odejść do aktualnego kroku.

Będziemy też zbierać następujące dane (output variables):

- (A(i),O(i)) - zmienne otrzymane (output variables) – moment zgłoszenia oraz moment wykonania i-tego zadania; zbierane w wektorach A, O.

Niech F będzie dystrybuantą odstępu Tau między zgłoszeniami oraz G dystrybuanta rozmiaru zadania S. Zakładamy, że dystrybuanty F, G są ciągłe. Zauważmy:

- Na podstawie A(i),O(i) możemy znaleźć ciąg czasów odpowiedzi $D(i) = O(i) - A(i)$.

3 ALOHA

W chwilach n ($n = 0, 1, 2, \ldots$) 'użytkownicy' komunikują się przez wspólnie użytkowany kanał. Kiedy użytkownik ma wiadomość do nadania może ją przesyłać przez kanał. Wiadomość jest podzielona na pakiety jednakowej długości, które wysyłane są w 'szczelinach' (slots) czasowych (tej samej długości co pakiety). Jeśli w tej samej szczelinie chce transmitować więcej niż jeden użytkownik, wtedy następuje 'kolizja' o której dowiadują się użytkownicy i transmisja jest nieudana. Udana transmisja jest wtedy i tylko wtedy gdy dokładnie jeden pakiet jest transmitowany w szczelnie. Oczywiście, gdyby w następnej szczelinie założyć, że dotychczasowi użytkownicy będą znowu próbować nadawać, to nawet gdyby nie pojawiły się nowe wiadomości, na pewno nastąpiłaby kolizja. Aby rozwiązać ten problem i moć nadawać w takim kanale opracowano rozmaite protokoły, na przykład tak zwany *szczelinowa ALOHA* slotted ALOHA. W tych protokołach zakłada się, że użytkownicy mogą jedynie słyszeć kolizje, ale nie znają liczby pakietów oczekujących u poszczególnych użytkowników. Jeśli więcej niż jedna stacja zdecyduje się transmitować w tej samej szczeline, wtedy mamy kolizję i żaden pakiet nie został przesłany, i wszyscy muszą czekać do momentu $n+1$. Jeśli natomiast tylko jedna stacja zdecyduje się na transmisję, to trwa ona 1 jednostkę czasu i kończy się sukcesem.

Pomysł jest taki, żeby nadawanie pakietu po kolizji było opóźnione. Jak to zrobić zależy od protokółu.

Aby moć zbadać protokóły zbudujemy model matematyczny. Oś czasowa jest podzielona na szczelny; i-ta jest $[i, i+1)$, $i = 0, 1, \ldots$. Aby opisać model potrzebujemy następujące składniki:

- X_n – liczba pakietów czekających na transmisję na początku szczeliny n-tej
- A_n – liczba nowych pakietów gotowych do nadawania na początku szczeliny n-tej.
- $\{h_i, 1 \leq i \leq K\}$ – strategia nadawania, gdzie $0 < h_i < 1$, $K \leq \infty$

Powyżej $0 < h_i < 1$ oznacza, że po i nieudanych transmisjach pakiet jest nadawany w następnej szczelinie z prawdopodobieństwem h_i. Zakłada się, że $\{A_n\}$ są niezależne o jednakowym rozkładzie z wspólną funkcją prawdopodobieństwa $\mathbb{P}(A_n = i) = a_i$, $i = 0, 1, \ldots$ z $\mathbb{E}\, A_n = \lambda < \infty$. Możemy rozpatrywać następujące strategie retransmisji.

- $h_i = h$ – szczelinowa ALOHA; $h_i = h$
- $h_i = 2^{-i}$ – ETHERNET (exponential back-off).

Niech

- Z_n – liczba pakietów podchodzących do transmisji w i-tej szczelinie.

Zauważmy, że

$$X_{i+1} = X_i + A_i - \mathbf{1}(Z_i = 1), \qquad i = 0, 1 \ldots \tag{3.3}$$

Niech

$$N_n^{\mathrm{s}} = \sum_{i=1}^{n} \mathbf{1}(Z_i = 1), \qquad n = 1, \ldots$$

będzie liczbą pakietów przetransmitowanych w pierwszych n szczelinach; $n = 1, 2, \ldots$. Definiujemy *przepustowość (throughput)* jako

$$\lim_{n\to\infty} \frac{N_n^{\mathrm{s}}}{n} = \gamma.$$

Zauważmy, że w protokóle szczelinowa ALOHA, rozkład Z_n pod warunkiem $X_n = k$ jest taki sam jak $A + B(k)$, gdzie $\{B(k)\}$ jest zmienną losową niezależną od A z funkcją prawdopodobieństwa

$$\mathbb{P}(B_n(k) = i) = b_i(k) = \binom{k}{i} h^i (1-h)^{k-i}, \quad i = 0, 1, \ldots, k$$

Warunkowy dryf definiuje się jako

$$\mathbb{E}\left[X_{n+1} - X_n | X_n = k\right] = \lambda - b_1(k)a_0 - b_0(k)a_1. \tag{3.4}$$

Gdy $k \to \infty$ to

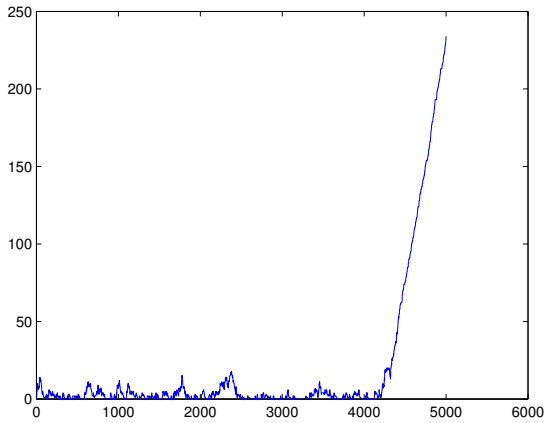
$$\mathbb{E}\left[X_{n+1} - X_n | X_n = k\right] \to \lambda > 0$$

co sugeruje, że jak ciąg osiągnie wysoki pułap to ma dryf do nieskończoności. W teorii łańcuchów Markowa formalnie dowodzi się chwilowości ciągu (X_n).

Do symulacji będziemy zakładać, że A_n mają rozkład Poi(λ). Przykładowa symulacja (X_n) jest pokazana na rys. 3.3.

Natomiast na rys. 3.4 pokazane są lokalne przepustowości

$$B_n = \sum_{i=1}^{n} \mathbf{1}(Z_i = 1)/n; \qquad n \le N.$$



Rysunek 3.3: Liczba oczekujących pakietów X_n ($n \leq 5000$) w ALOHA; $\lambda = 0.31$, $h = 0.1$.

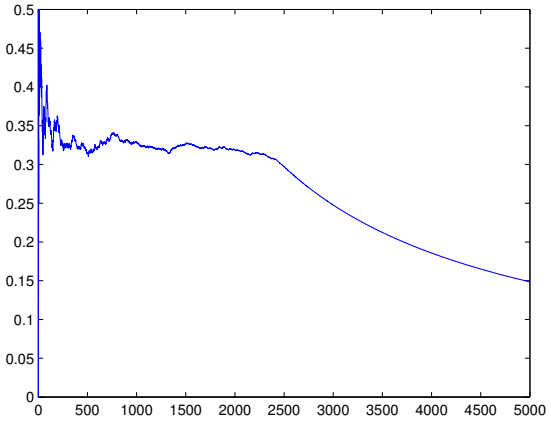
4 Prawdopodobieństwo awarii sieci

Rozważmy sieć połaczeń, która jest spójnym grafem składającym się z węzłów, z których dwa s i t są wyróżnione. Węzły są połaczone krawędziami ponumerowanymi $1, \ldots, n$. Każda krawędź może być albo sprawna – 0 lub nie – 1. W przypadku niesprawnej krawędzi będziemy mówili o przerwaniu. możemy też mówić o usunięciu krawędzi, i wtedy nowy graf nie musi byc dalej spójny.

Stany krawędzi opisuje wektor $e_1, \ldots, e_n$, gdzie $e_i = 0$ lub 1. Sieć w całości może być sprawna – 0 jeśli jest połączenie od s do t lub nie (t.j. z przerwą) – 1, tzn graf powstały po usunięciu niesprawnych krawędzi nie jest spójny. Niech $\mathcal{S} = \{0,1\}^n$ i $\boldsymbol{e} = (e_1, \ldots, e_n)$. Formalnie możemy opisać stan systemu za pomocą funkcji $k : \mathcal{S} \to \{0,1\}$. Mówimy, że sieć jest niesprawna i oznaczamy to przez 1 jeśli nie ma połaczenia pomiędzy węzłami s i t, a 0 w przeciwnym razie. To czy sieć jest niesprawna (lub sprawna) decyduje podzbiór B niesprawnych krawędzi, tj $e_j = 1, j \in B, e_j = 0, \notin B$. Podamy teraz przykłady. Oznaczmy $|\boldsymbol{e}| = \sum_{j=1}^{n} e_j$.

- *Układ szeregowy* jest niesprawny gdy jakakolwiek krawędź ma przerwanie. A więc $k(\boldsymbol{e}) = 1$ wtedy i tylko wtedy gdy $|\boldsymbol{e}| \geq 1$.

- *Układ równoległy* jest niesprawny jeśli wszystkie elementy są niesprawne. A więc $k(\boldsymbol{e}) = 1$ wtedy i tylko wtedy gdy $|\boldsymbol{e}| = n$.



Rysunek 3.4: Szczelinowa ALOHA; Przepustowość B_n/n ($n \leq 5000$); $\lambda = 0.31$, $h = 0.1$.

Teraz przypuśćmy, że każda krawędź może mieć przerwanie z prawdopodobieństwem q niezależnie od stanu innych krawędzi. Opiszemy to za pomocą wektora losowego X przyjmującego wartości z $\mathcal{S}$. Zauważmy, że X jest dyskretną zmienną losowa z funkcją prawdopodobieństwa

$$\mathbb{P}(X = \boldsymbol{e}) = p(\boldsymbol{e}) = q^{\sum_{j=1}^{n} e_j}(1-q)^{n-\sum_{j=1}^{n} e_j},$$

dla $\boldsymbol{e} \in \mathcal{S}$. Naszym celem jest obliczenie prawdopodobieństwa, że sieć jest niesprawna. A więc

$$I = \mathbb{E}\, k(X) = \sum_{\boldsymbol{e}\in\mathcal{S}} k(\boldsymbol{e})p(\boldsymbol{e})$$

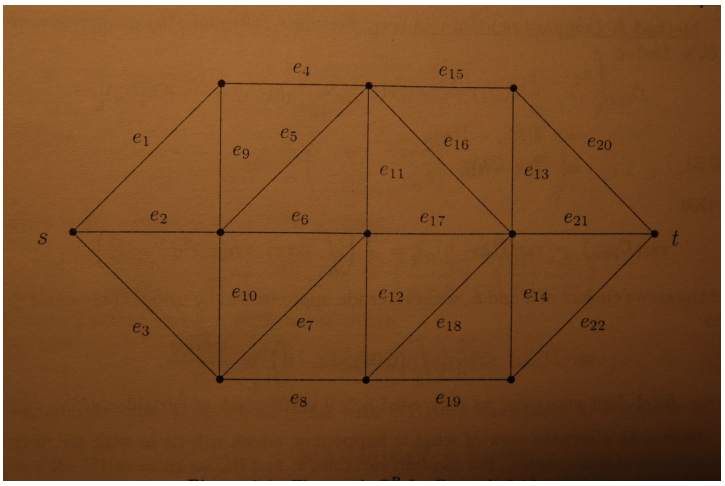
jest szukanym prawdopodobieństwem, że sieć jest niesprawna. Z powyższego widzimy, że możemy używać estymatora dla I postaci

$$\hat{X}_n^{\mathrm{CMC}} = \frac{1}{n}\sum_{i=1}^{n} k(X_i),$$

gdzie $X_1, \ldots, X_n$ są niezależnymi replikacjami X. Ponieważ $k(X) = 0$ lub 1,

$$I = \mathbb{P}(k(X) = 1) = \mathbb{E}\, k(X) \quad \text{oraz} \quad \operatorname{Var} k(X) = I(1-I).$$

Proponujemy zrobić symulację dla konkretnego przykładu pochodzącego z książki Madrasa [**?**]. Sieć ma 11 węzłów połączonych 22 krawędziami i jest przedstawiona na rys. 4.5.



Rysunek 4.5: Schemat połączeń w sieci.

Ponieważ typowo prawdopodobieństwo I jest bardzo małe, więc aby wyliczyć szukane prawdopodobieństwo awarii musimy zrobić wiele replikacji. Dlatego jest bardzo ważne optymalne skonstruowanie dobrego estymatora dla I.

Zanalizujemy teraz estymator istotnościowy z $\tilde{p}(\boldsymbol{e}) = \beta^{|\boldsymbol{e}|}(1-\beta)^{n-|\boldsymbol{e}|}$. Mamy

$$I = \sum_{\boldsymbol{e}\in\mathcal{S}} k(\boldsymbol{e})p(\boldsymbol{e}) = \sum_{\boldsymbol{e}\in\mathcal{S}} \frac{k(\boldsymbol{e})p(\boldsymbol{e})}{\tilde{p}(\boldsymbol{e})}\tilde{p}(\boldsymbol{e}).$$

Teraz zauważmy, że $k(\boldsymbol{e}) = 1$ wtedy i tylko wtedy gdy sieć jest niesprawna a w pozostałych przypadkach równa się zero. A więc

$$I = \sum_{\boldsymbol{e}:k(\boldsymbol{e})=1} \frac{p(\boldsymbol{e})}{\tilde{p}(\boldsymbol{e})}\tilde{p}(\boldsymbol{e}),$$

gdzie

$$\frac{p(\boldsymbol{e})}{\tilde{p}(\boldsymbol{e})} = \frac{q^{|\boldsymbol{e}|}(1-q)^{n-|\boldsymbol{e}|}}{\beta^{|\boldsymbol{e}|}(1-\beta)^{n-|\boldsymbol{e}|}} = \left(\frac{1-q}{1-\beta}\right)^n \left(\frac{q(1-\beta)}{\beta(1-q)}\right)^{|\boldsymbol{e}|},$$

gdzie $|\boldsymbol{e}| = \sum_{j=1}^{n} e_j$. Będziemy przyjmowali, ze $\beta \geq q$ co pociąga

$$\frac{q(1-\beta)}{\beta(1-q)} \leq 1.$$

Niech teraz d będzie najmniejszą liczbą krawędzi z przerwaniem, powodująca niesprawność całej sieci. Na przykład dla sieci szeregowej d wynosi 1, a dla

równoległej n. Dla sieci z Rys. 4.5 mamy $d = 3$. A więc, jesli $k(\boldsymbol{e}) = 1$, to $|\boldsymbol{e}| \geq d$.

Przy założeniu $|\boldsymbol{e}| \geq d$

$$\begin{aligned}\left(\frac{1-q}{1-\beta}\right)^n \left(\frac{q(1-\beta)}{\beta(1-q)}\right)^{|\boldsymbol{e}|} &\leq \left(\frac{1-q}{1-\beta}\right)^n \left(\frac{q(1-\beta)}{\beta(1-q)}\right)^d, \qquad |\boldsymbol{e}| \geq d \\ &= \frac{(1-q)^{n-d}q^d}{\beta^d(1-\beta)^{n-d}}.\end{aligned}$$

Teraz pytamy dla jakiego $\beta \geq q$ prawa strona jest najmniejsza. Odpowiedź na to pytanie jest równoważna szukaniu maksimum funkcji $\beta^d(1-\beta)^{n-d}$. Przyrównując pochodną do zera mamy

$$d\beta^{d-1}(1-\beta)^{n-d} - (n-d)\beta^d(1-\beta)^{n-d-1} = 0.$$

Stąd $\beta = d/n$. A więc dla $|\boldsymbol{e}| \geq d$ mamy

$$\left(\frac{1-q}{1-\beta}\right)^n \left(\frac{q(1-\beta)}{\beta(1-q)}\right)^{|\boldsymbol{e}|} \leq \left(\frac{1-q}{1-\frac{d}{n}}\right)^n \left(\frac{q(1-\frac{d}{n})}{\frac{d}{n}(1-q)}\right)^d = \delta.$$

Jeśli $d/n \geq q$, to oczywiście $\delta \leq 1$. Teraz

$$\begin{aligned}\tilde{\mathbb{E}}\, k^2(X)\frac{p^2(X)}{\tilde{p}^2(X)} &= \sum_{\{\boldsymbol{e}:k(\boldsymbol{e})=1\}} \frac{p^2(\boldsymbol{e})}{\tilde{p}^2(\boldsymbol{e})}\tilde{p}(\boldsymbol{e}) \\ &= \sum_{\{\boldsymbol{e}\kappa(\boldsymbol{e})=1\}} \frac{p(\boldsymbol{e})}{\tilde{p}(\boldsymbol{e})}p(\boldsymbol{e}) \\ &= \sum_{\{\boldsymbol{e}:\ k(\boldsymbol{e})=1\}} \left(\frac{1-q}{1-\beta}\right)^n \left(\frac{q(1-\beta)}{\beta(1-q)}\right)^{|\boldsymbol{e}|} p(\boldsymbol{e}) \\ &\leq \delta \sum_{\{\boldsymbol{e}:\ |\boldsymbol{e}|\geq d\}} p(\boldsymbol{e}) = \delta I.\end{aligned}$$

Ponieważ I zazwyczaj jest małe, więc

$$\operatorname{Var}\hat{Y}_n = \frac{I(1-I)}{n} \approx \frac{I}{n}.$$

Z drugiej strony

$$\operatorname{Var}\hat{Y}_n^{\mathrm{IS}} \sim \frac{\delta}{n}I.$$

A więc estymator IS ma $1/\delta$ razy mniejszą wariancję aniżeli by to robić za pomocą zgrubnego estymatora.

W szczególności przyjmijmy dla sieci z Rys. 4.5 prawdopodobieństwo przerwania $q = 0.01$. Dla tej sieci mamy $\beta = 3/22$ i stąd $\delta = 0.0053$. Czyli używając estymatora IS potrzebne będzie $\sqrt{1/\delta} \approx 14$ razy mniej replikacji.

Przykład 4.1 Obliczenie prawdopodobieństwa awarii sieci metodą warstw. Rozważmy schemat połączeń z rys. 4.5. Ponieważ typowo prawdopodobieństwo I jest bardzo małe, więc aby je obliczyć używając estymatora CMC musimy zrobić wiele replikacji. Dlatego jest bardzo ważne optymalne skonstruowanie dobrego estymatora dla I. Można na przykład zaproponować estymator warstwowy, gdzie j-tą warstwą $\mathcal{S}(j)$ jest podzbiór $\mathcal{S}(j)$ składających się z tych B takich, że $|B| = j$. Zauważmy, że

$$p_j = \mathbb{P}(X \in \mathcal{S}(j)) = \mathbb{P}(\text{dokładnie } j \text{ krawędzi ma przerwę}) = \binom{n}{j} q^j (1-q)^{n-j} .$$

Każdy $A \subset \mathcal{S}$ ma jednakowe prawdopodobieństwo $q^j(1-q)^{|\mathcal{S}|-j}$ oraz

$$\mathbb{P}(X = A | X \in \mathcal{S}(j)) = 1/\binom{n}{j} \qquad A \subset \mathcal{S} .$$

Aby więc otrzymać estymator warstwowy generujemy np_j replikacji na każdej warstwie $\mathcal{S}(j)$ z rozkładem jednostajnym.

5 Uwagi bibliograficzne

O procesach Poissona można przeczytać w wielu książkach z procesów stochastycznych. Jednakże warto polecić książki specjalistycznej, jak na przykład Kingmana [20, 21]. Metoda symulacji niejednorodnego procesu Poisson przez przezedzanie pochodzi z pracy [32]. Metoda symulacji po zdarzeniach (discrete event simulation) ma bogatą literaturę. W szczególności można polecić podręcznika Fishmana [15], Bratleya i współautorów [7]. Jesli procesy pomiędzy zdarzeniami nie są stałe ale zmieniają się w sposób deterministyczny, to takie procesy można modelować za pomocą procesów kawałkami deterministycznie Markowa (piecewise deterministic Markov process); patrz monografia Davisa [10].

O zagadnieniu konserwatora i innych problemach teorii niezawodności można poczytać w książce Kopocińskiego [25]

O protokołach w transmisji rozproszonej można poczytać w książkach: Ross [1], str. 148, Bremaud [8], str. 108–110, 173–178, lub w pracach Kelly (1985) i Kelly & MacPhee [18, 19], Aldous (1987) [3].

Analiza przepływu w sieci z rys. 4.5 jest wzięta z książki Madrasa [**?**].

Zadania teoretyczne

5.1 W szczelinowym modelu ALOHA z $\lambda = 0.31$ oraz $h = 0.1$ policzyć funkcję

$$\phi(k) = \mathbb{E}\left[X_{n+1} - X_n | X_n = k\right] = \lambda - b_1(k)a_0 - b_0(k)a_1, \qquad k = 0, 1, \ldots,$$

gdy A_i mają rozkład Poissona. Zrobić wykres. Czy można na podstawie tego wykresu skomentować rys. 3.3 i 3.4.

5.2 Dla sieci z rys. 4.5 oszacować dwustronnie I i następnie przeprowadzić analizę symulacji dla zgrubnego estymatora.

Zadania laboratoryjne

5.3 Wygenerować $A_1, A_2, \ldots$, w odcinku $[0, 1000]$, gdzie $A_0 = 0$ oraz $A_1 < A_2 < \ldots$ są kolejnymi punktami w niejednorodnym procesie Poissona z funkcją intensywności $\lambda(t) = a(2 - \sin(\frac{2\pi}{24}t))$. Przyjąć $a = 10$. Niech $A(t)$ będzie liczbą punktów w odcinku $[0, t]$. Zastanowić się jaki ma rozkład $A(t)$ i znaleźć jego średnią. Policzyć z symulacji $A(1000)/1000$ – średnią liczbą punktów na jednostkę czasu, zwaną asymptotyczną intensywnościa $\bar{\lambda}$. Porównać z

$$\int_0^{24} \lambda(t)\, dt/24 \; .$$

5.4 Kontynuacja zad. refzad.proc.Poissona. Wygenerować $\tau_1, \tau_2, \ldots, \tau_{1000}$, gdzie $\tau_i = A_i - A_{i-1}$, $A_0 = 0$ oraz $A_1 < A_2 < \ldots$ są kolejnymi punktami w niejednorodnym procesie Poissona z funckcją intensywności $\lambda(t) = a(2 - \sin(\frac{2\pi}{24}t))$. Przyjąć $a = 10$. Obliczyć średni odstęp między punktami $\hat{\tau} = \sum_{j=1}^{1000} \tau_i/1000$.

5.5 Zbadać jak szybko $P(L(t) = 1)$ w alternującym procesie *on – off* zbiega do współczynnika gotowości. Rozpatrzyć kilka przypadków:

- F_{on} F_{off} są wykładnicze,
- F_{on} F_{off} są Erlanga odpowiednio Erl(3,λ) i Erl(3,μ),
- F_{on} F_{off} są Pareto.

Przyjąć $\mathbb{E}\, T^{\text{on}} = 1$ i $\mathbb{E}\, T^{\text{off}} = 2$.

5.6 Obliczyć średnią $\mathbb{E}\, L(i)$ $(i = 1, \ldots, 10)$ w systemie M/M/1 z
a. $\lambda = 1/2$ i $\mu = 1$,
b. $\lambda = 1$ i $\mu = 1$, jeśli $L(0) = 0$.

5.7 Przeprowadzić symulacje dobroci *czystego protokółu ALOHA*. Bardziej dokładnie, przypuśćmy, że pakiety przybywają zgodnie z procesem Poissona z intensywnością λ i wszystkie są długości 1. W przypadku kolizji, każdy z pakietów jest retransmitowany po czasie wykładniczym z parametrem μ. Zakładamy, że wszystkie zmienne są niezależne. W modelu musimy obliczyć następujące zmienne. Na podstawie symulacji zastanowić się nad przepustowością γ przy tym protokóle.

5.8 Napisać algorytm na symulację protokółu ETHERNET.

5.9 Obliczyć średni czas do awarii systemu składającego się z N elementów połączonych równolegle z jednym konserwatorem, jeśli czas życia elementu ma rozkład wykładniczy Exp(1/2), natomiast czas naprawy przez konserwatora Exp(1)). Policzyć dla $N = 1, \ldots, 10$. Porównać z średnim czasem do awarii systemu składającego się z N elementów połączonych równolegle, ale bez konserwatora.

Projekt

Projekt 3 Zbadać dobroć protokółu szczelinowa ALOHA ze względu na parametr λ. Przez symulację znaleźć wartości krytyczne dla λ i określić typowe zachowanie się protokółu. Zbadać inną modyfikację protokółu szczelinowa ALOHA, w której dopuszcza się, że użytkownicy mają też dodatkową wiedzę o stanie X_n, i wtedy w n-tej szczelinie przyjmować $h = 1/X_n$. Czy ten protokół może być stabilny, tzn. mający przepustowość $\gamma > 0$; jeśli tak to kiedy.

Projekt 4 Produkt jest składany z komponentów na dwóch stanowiskach. Po zakończeniu prac na pierwszym stanowisku jest dalej opracowywany na drugim, z którego wychodzi gotowy produkt. Produkcja trwa każdego dnia przez 8 godzin. Na początku dnia w magazynie jest przygotowane n_1 komponentów, które sukcesywnie są dostarczane na stanowisko pierwsze. Przed stanowiskiem drugim może oczekiwać co najwyżej k_1, w przeciwnym razie praca na stanowisku pierwszym jest zablokowana. Ile należy przygotować komponentów n_1 i jak duży bufor k_1 aby z prawdopodobieństwem nie mniejszym 0.9 była zapewniona produkcja przez cały dzień, tj. 8 godzin. Przez aprzestanie produkcji rozmumiemy, że zarówno magazyn jak i bufor przestają pracować. Jak duży należy przygotować magazyn $n_1 + k_1$. Przyjąć, że

- czas pracy na pierwszym stanowisku ma rozkład Erlanga Erl(2,10),
- natomiast na drugim stanowisku jest mieszanką rozkladów wykładniczych z gęstością $0.8 \times 9e^{-9x} + 0.2 \times 3e^{-3x}$,
- koszty magazynowania przed rozpoczęciem pracy czy w oczekiwaniu na prace na drugim stanowisku są liniowe od $n_1 + k_1$.

Projekt 5 Mamy dwa następujące warianty pracy 2-ch procesorów. 1. Dwa procesory pracują osobno. Zadania napływają do i-tego procesu zgodnie z procesem Poissona z intensywnością λ_i i mają rozkład rozmiaru $G^{(i)}$. Niech $D^{(i)}(t)$ będzie liczbą zadań obsłużonych przez i-ty procesor to chwili t. Zakładamy, że $\lambda_i < \int_0^\infty x dG^{(i)}(x)$, $i = 1, 2$. Wtedy przepustowość

$$\lim_{t\to\infty} D^{(i)}(t)/t = \lambda_i,$$

a więc dwa procesory, jeśli pracują osobno mają przepustowość $\lambda_1 + \lambda_2$.
2. Możemy tak zmienić konstrukcję, że w przypadku gdy jeden z procesów nie ma pracy, to drugi pracuje z szybkością 2.

Zbadać o ile efektywniejszy jest zparowany procesor, w zależności od intesywności wejścia i rozkładów rozmiarów zadań. Przyjąc $\lambda_1 = 1$.

Projekt 6 W banku jest c stanowisk obsługi. Klienci zgłaszają się zgodnie z procesem Poissona z intensywnością λ i ich czasy obsługi mają rozkład G. Zbadać następujące protokóły kolejkowania.
1. Przed każdym okienkiem jest osobna kolejka. Klienci wybierają okienko w sposób losowy.
2. Przed każdym okienkiem jest osobna kolejka. Klienci wybierają kolejne okienko, tj. $1, 2, \ldots, c, 1, 2, \ldots$.
3. Jest jedna kolejka i gdy nadchodzi czas klient idzie do akurat uwolnionego okienka.

Rozdział VII

Analiza symulacji stabilnych procesów stochastycznych.

[1] Przypuścmy, że chcemy obliczyć wartość I, gdy wiadomo, że z prawdopodobieństwem 1

$$I = \lim_{t\to\infty} \frac{\int_0^t X(s)\,ds}{t},$$

dla pewnego procesu stochastycznego $(X(s))_{s\geq 0}$. Wtedy naturalnym estymatorem jest

$$\hat{X}(t) = \frac{\int_0^t X(s)\,ds}{t}, \tag{0.1}$$

który jest (mocno) zgodnym estymatorem I. Jednakże zazwyczaj $\hat{X}(t)$ nie jest nieobciązonym estymatorem, tzn. $\mathbb{E}\,\hat{X}(t) \neq I$.

Modelem w pewien sposób idealnym jest sytuacja gdy proces stochastyczny jest stacjonarny. Wtedy estymator (0.1) jest nieobciążony. Zajmiemy się tym przypadkiem w następnym podrozdziale. Przypomnijmy, że przez proces stochastyczny $(X(t))_T$ z przestrzenią parametrów T nazywamy kolekcję zmiennych losowych. Za T możemy przyjąć np. $T = [0,\infty)$, $T = [a,b]$, itp. W tym rozdziale $T = [0,\infty)$.

[1] 1.2.2016

1 Procesy stacjonarne i analiza ich symulacji

Rozważmy proces stochastyczny $(X(t))_{t\geq 0}$ dla którego $\mathbb{E}\,X(t)^2 < \infty$, $(t \geq 0)$. Mówimy, że jest on *stacjonarny* jeśli spełnione są następujące warunki.

(1) Wszystkie rozkłady $X(t)$ są takie same. W konsekwencji wszystkie $\mathbb{E}\,X(t)$ są równe jeśli tylko pierwszy moment $m = \mathbb{E}\,X(0)$ istnieje.

(2) Dla każdego $h > 0$, łączne rozkłady $(X(t), X(t+h))$ nie zależą od t. W konsekwencji wszystkie kowariancje

$$R(h) = \mathrm{Cov}\,(X(t), X(t+h)),$$

(jeśli tylko drugie momenty istnieją) nie zależą od t.

(k) Dla każdego $0 < h_1 < \ldots < h_{k-1}$, łączne rozkłady $(X(t), X(t+h_1), \ldots, X(t+h_{k-1}))$ nie zależą od t.

Dla procesów stacjonarnych jest prawdziwe tzw. twierdzenie ergodyczne; dla procesu stacjonarnego $(X(t))_{t\geq 0}$ takiego, że $\mathbb{E}\,|X(t)| < \infty$, to z prawdopodobieństwem 1

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t X(s)\,ds$$

istnieje. Ponadto jeśli proces jest ergodyczny, granica jest stałą, a więc musi równać się $I = \mathbb{E}\,X(0)$.

Nie będziemy tytaj zajmowali się szczegółami teorii ergodycznej.[2] Łatwo jest podać przykład procesu stacjo narnego, który nie jest ergodyczny. Weźmy na przykład proces $X(t) = \xi$, gdzie ξ jest zmienną losową. Wtedy w trywialny sposób

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t X(s)\,ds = \xi\ .$$

Jesli ponadto są spełnione tzw. *warunki mieszania*, (to jest mocniejszy warunek niż egodyczność) to wtedy dla pewnej stałej $\bar{\sigma}^2 > 0$ (to nie jest $\mathrm{Var}\,X(t)$) proces

$$\frac{\int_0^t X(s)\,ds - tI}{\bar{\sigma}\sqrt{t}} \tag{1.2}$$

[2] referencje ?

dla $t \to \infty$, zbiega według rozkładu do $\mathrm{N}(0,1)$. Zastanówmy się, ile wynosi $\bar{\sigma}^2$. Zmienna

$$W = \frac{\int_0^t X(s)\,ds - tI}{\bar{\sigma}\sqrt{t}}$$

musi być asymptotycznie ustandaryzowana, to jest z średnią zero (trywialnie spełnione) oraz wariancją dążąca do 1. Mamy

$$\operatorname{Var} W = \frac{\operatorname{Var} \int_0^t X(s)\,ds}{\bar{\sigma}^2 t}. \tag{1.3}$$

Policzymy teraz mianownik.

Przepiszmy (1.2) w wygodnej dla nas formie

$$\frac{\int_0^t X(s)\,ds - tI}{\bar{\sigma}\sqrt{t}} = \frac{\sqrt{t}}{\bar{\sigma}}(\hat{X}(t) - I).$$

Chcemy aby $(\hat{X}(t) - I)$ było ustandaryzowane. Oczywiście $\mathbb{E}\,(\hat{X}(t) - I) = 0$, natomiast aby wariancja była równa 1 musimy unormować

$$\frac{\hat{X}(t) - I}{\sqrt{\operatorname{Var} \hat{X}(t)}}\,.$$

Musimy zbadamy zachowanie asymptotyczne mianownika. W tym celu udowodnimy następujący lemat.

Lemat 1.1 *Jeśli* $\int_0^\infty |R(v)|\,dv < \infty$ *i* $\int_0^\infty R(s)\,ds \neq 0$, *to dla* $t \to \infty$

$$\operatorname{Var}\,(\int_0^t X(s)\,ds) \sim 2t \int_0^\infty R(v)\,dv.$$

Dowód. Mamy

$$\mathbb{E} \int_0^t X(s)\,ds = \int_0^t \mathbb{E}\, X(s)\,ds = mt\,.$$

Dalej bez straty ogólności możemy założyć $m = 0$. Wtedy

$$\begin{aligned}
&\mathrm{Var}\,(\int_0^t X(s)\,ds) \\
&= \mathbb{E}\,(\int_0^t X(s)\,ds)^2 \\
&= \mathbb{E}\,[\int_0^t \int_0^t X(w)X(v)\,dw\,dv] \\
&= \int_0^t \int_0^t \mathbb{E}\,((X(w)X(v))\,dw\,dv \\
&= \int_{(w,v)\in[0,t]^2:w\le v} \mathbb{E}\,(X(w)X(v))\,dw\,dv + \int_{(w,v)\in[0,t]^2:w>v} \mathbb{E}\,(X(w)X(v))\,dw\,dv.
\end{aligned}$$

Teraz dla $v \le w$ mamy $\mathbb{E}\,X(w)X(v)) = R(w-v)$. Mamy również dla $h > 0$ ponieważ

$$R(-h) = \mathbb{E}\,X(x)X(x+h) == \mathbb{E}\,X(x-h)X(x) = R(h).$$

i stąd

$$\begin{aligned}
\int_{(w,v)\in[0,t]^2:w\le v} \mathbb{E}\,X(w)X(v)\,dw\,dv \;+\; &\int_{(w,v)\in[0,t]^2:w>v} \mathbb{E}\,X(w)X(v)\,dw\,dv \\
&= 2\int_0^t dv \int_0^v R(v-w)\,dv \\
&\sim t\,2\int_0^\infty R(w)\,dw\ .
\end{aligned}$$

□

Ponieważ

$$\mathrm{Var}\,W = \frac{\mathrm{Var}\,\int_0^t X(s)\,ds}{\bar{\sigma}^2 t} \to 1,$$

więc

$$\bar{\sigma}^2 = 2\int_0^\infty R(s)\,ds.$$

Stąd mamy następujący fakt.

Fakt 1.2 *Jeśli* $(X(t)_{t\geq 0}$ *jest procesem stacjonarnym i ergodycznym,* $\mathbb{E}\,X(t)^2 < \infty$ *oraz* $\int_0^\infty |R(v)|\,dv < \infty$ *i* $\int_0^\infty R(s)\,ds \neq 0$*, to estymator* $\hat{X}(t)$ *jest nieobciązony (ma wartość oczekiwaną* I*) oraz*

$$\lim_{t\to\infty} t\mathrm{Var}\,(\hat{X}(t)) = \bar{\sigma}^2 = 2\int_0^\infty R(s)\,ds\ . \tag{1.4}$$

Dowód. Mamy

$$\mathbb{E}\,\hat{X}(t) = \frac{1}{t}\int_0^t \mathbb{E}\,X(s)\,ds = \mathbb{E}\,X(0) = I.$$

Aby udowodnić (1.4) korzystawmy z (1.3) oraz Lematu 1.1. □

Zrobimy teraz kilka uwag.

- Przepiszmy (1.2) w wygodnej dla nas formie

$$\frac{\int_0^t X(s)\,ds - tI}{\bar{\sigma}\sqrt{t}} = \frac{\sqrt{t}}{\bar{\sigma}}(\hat{X}(t) - I).$$

  Oczywiście przy odpowiednich założeniach ten proces zbiega według rozładu do N(0,1).

- Wielkość $\bar{\sigma}^2 = 2\int_0^\infty R(s)\,ds$ zwie się TAVC (od *time average variance constant*) lub asymptotyczną wariancją.

Podsumujmy. Podobnie jak w rozdziale IV możemy napisać *fundamentalny związek*

$$b = z_{1-\epsilon/2}\frac{\bar{\sigma}}{\sqrt{t}}\ . \tag{1.5}$$

pomiędzy błędem b, poziomem ufności α oraz horyzontem symulacji t. Natomiast przedział ufności na poziome α wynosi

$$\left(\hat{X}(t) - \frac{z_{1-\alpha/2}\bar{\sigma}}{\sqrt{t}}, \hat{X}(t) + \frac{z_{1-\alpha/2}\bar{\sigma}}{\sqrt{t}}\right)\ .$$

co zapisujemy krótko

$$(\hat{X}(t) \pm \frac{z_{1-\alpha/2}\bar{\sigma}}{\sqrt{t}}).$$

Niestety w typowych sytuacjach nie znamy $\bar{\sigma}^2$ a więc musimy tą wielkość wyestymować z symulacji.[3]

[3]Czytaj dalej str. 540 Fishmana

2 Przypadek symulacji obciążonej

Będziemy teraz analizowali przypadek symulacji liczby I dla estymatora $\hat{X}$ spełniającego:

1. $I = \lim_{t\to\infty} \hat{X}(t)$ (z prawdopodobieństwem 1),
2. $\frac{\sqrt{t}}{\bar{\sigma}}(\hat{X}(t) - I)$ zbiega wg rozkładu do $\mathcal{N}(0,1)$, gdzie

$$t\mathrm{Var}\,(\int_0^t X(s)\,ds) \to \bar{\sigma}^2,$$

   gdy $t \to \infty$ dla pewnej skończonej liczby $\bar{\sigma}^2 > 0$.

Ponieważ nie zakładamy stacjonarności, więc estymator $\hat{X}(t)$ jest zgodny ale nie jest nieobciążony. Powstaje więc problem analizy jak szybko $b(t) = \mathbb{E}\,\hat{X}(t) - I$ zbiega do zera. Zanalizujemy przypadek gdy $\beta = \int_0^\infty (\mathbb{E}\,X(s) - I)\,ds$ zbiega absolutnie, tzn. $\int_0^\infty |\mathbb{E}\,X(s) - I|\,ds < \infty$. Wielkość β można też przedstawić jako

$$\begin{aligned}
\beta &= \lim_{t\to\infty} \int_0^t (\mathbb{E}\,X(s) - I)\,ds \\
&= \lim_{t\to\infty} (\mathbb{E} \int_0^t X(s)\,ds - tI) \\
&= \lim_{t\to\infty} t(\mathbb{E}\,\hat{X}(t) - I)
\end{aligned}$$

nazywamy *asymptotycznym obciążeniem*. Wtedy

$$\begin{aligned}
&\mathbb{E}\,\hat{X}(t) - I \\
&\quad = \frac{\int_0^t (\mathbb{E}\,(X(s)) - I)}{t} \\
&\quad = \frac{\int_0^\infty (\mathbb{E}\,(X(s)) - I)\,ds - \int_t^\infty (\mathbb{E}\,(X(s)) - I)}{t} \\
&\quad = \frac{\beta}{t} + o(1/t)\ .
\end{aligned}$$

Ponieważ $\sqrt{\mathrm{Var}\,\hat{X}(t)} \sim \bar{\sigma}/\sqrt{t}$ a więc zmienność naszego estymatora jest rzędu $\mathrm{O}(t^{-1/2})$ czyli większa niż jego obciążenie.

Przykład 2.1 Aby uzmysłowić czytelnikowi trudności jakie można napotkać podczas symulacji rozpatrzymy przypadek symulacji średniej liczby zadań w systemie M/G/∞ (po więcej informacji patrz dodatek XI.7.1). Będziemy rozpatrywać liczbę zadań $L(t)$ dla

1. $\lambda = 5$, i $S =_{\rm d} \mathrm{Exp}(1)$, $T_{\max} = 1000$,
2. $\lambda = 5$, i $S =_{\rm d} 0.2X$, gdzie $X =_{\rm d} \mathrm{Par}(1.2)$,$T_{\max} = 1000$,
3. $\lambda = 5$, i $S =_{\rm d} 0.2X$, gdzie $X =_{\rm d} \mathrm{Par}(1.2)$,$T_{\max} = 10000$,
4. $\lambda = 5$, i $S =_{\rm d} 0.2X$, gdzie $X =_{\rm d} \mathrm{Par}(1.2)$,$T_{\max} = 100000$,
5. $\lambda = 5$, i $S =_{\rm d} 1.2X$, gdzie $X =_{\rm d} \mathrm{Par}(2.2)$,$T_{\max} = 1000$,
6. $\lambda = 5$, i $S =_{\rm d} 2.2X$, gdzie $X =_{\rm d} \mathrm{Par}(3.2)$,$T_{\max} = 1000$,

We wszystkich tych przykładach $\rho = \lambda \mathbb{E}\, S = 5$ a więc powinniśmy otrzymać zawsze średnią oczekowaną liczbę zgłoszeń w systemie w warunkach stacjonarnych $\bar{l} = 5$ (patrz dodatek XI.7.1). Na rysunkach odpowiednio 2.1-2.6 pokazana jest wysymulowana realizacja

$$\hat{L}(t) = \frac{\int_0^t L(s)\, ds}{t}$$

w przedziale $1 \le t \le t_{\max}$.

Możemy zauważyć, że symulacja 1-sza szybko zbiega do granicznej wartości 5, chociaż horyzont $t_{\max}$ jest krótki. W tym przypadku $R(t) = 5\exp(-t)$, a więc $\bar{\sigma}^2 = 2\int_0^\infty R(s)\, ds = 10$, skąd $\bar{\sigma} = 3.1622$.

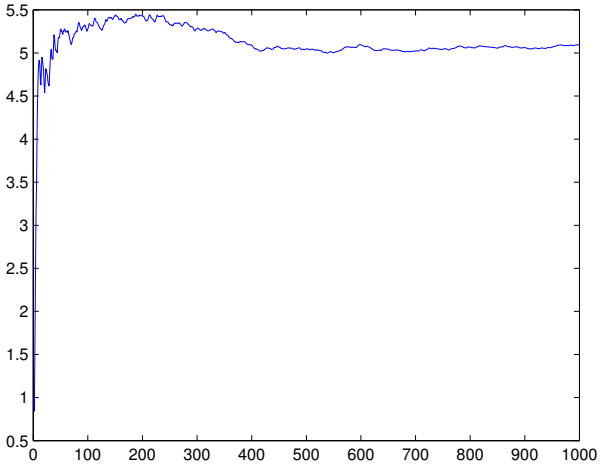
Natomiast z 2-gą jest coś dziwnego i dlatego warto ten przypadek bliżej przeanalizować. Policzymy najpierw $\bar{\sigma}^2$ ze wzoru (XI.7.8). Mamy

$$R(t) = 5\int_t^\infty (1 + \frac{s}{0.2})^{-1.2}\, ds = 5(1 + \frac{t}{0.2})^{-0.2}$$

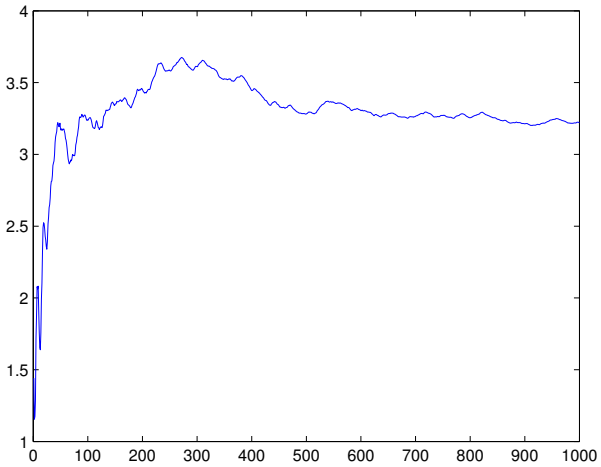
skąd mamy, że

$$\bar{\sigma}^2 = 2\int_0^\infty R(t)\, dt = \infty\ .$$

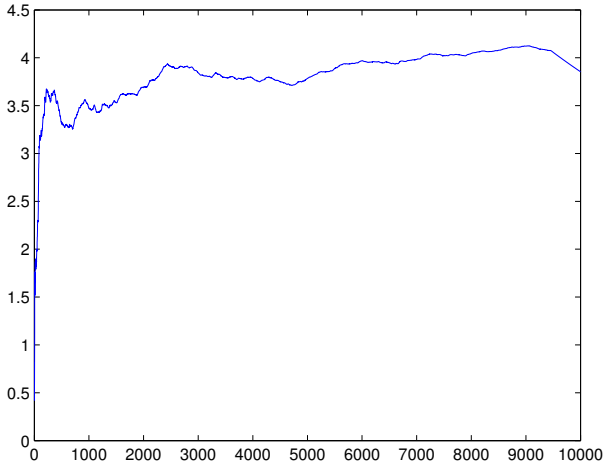
A więc nasze założenie o bezwzględnej całkowalności $R(t)$ nie jest spełnione. Nie można zatem stosować fundamentalnego związku, chociaż oczywiście ten estymator zbiega do 5. Zobaczymy to na dalszych przykładach na rys. 2.3, 2.4 (symulacja 3-cia i 4-ta).



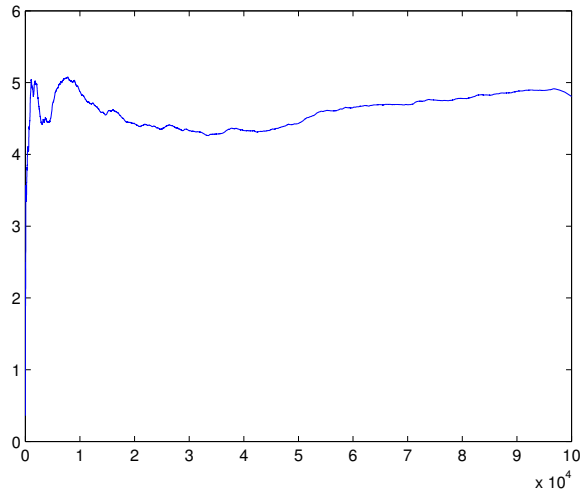
Rysunek 2.1: $\lambda = 5$, Exp(1), $T_{\max} = 1000$; $\bar{l} = 5.0859$



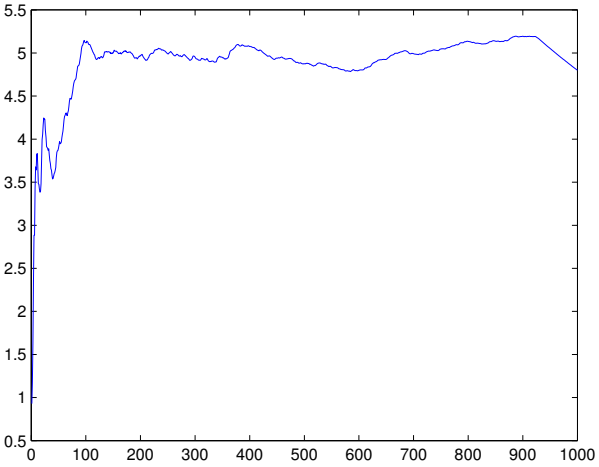
Rysunek 2.2: $\lambda = 5$, Par(1.2) $\times$ 0.2,$T_{\max} = 1000$; $\bar{l} = 3.2208$



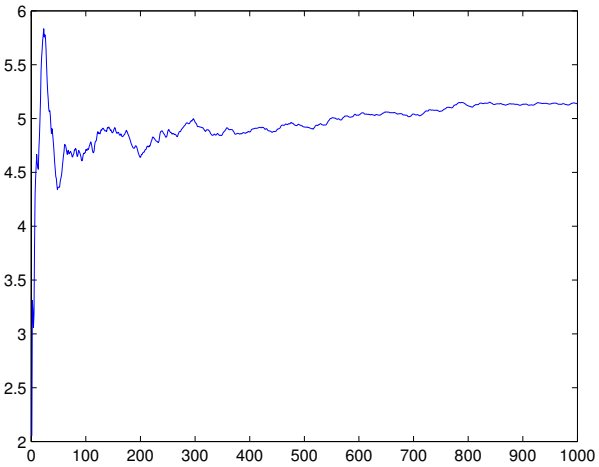
Rysunek 2.3: $\lambda = 5$, Par(1.2) $\times$ 0.2,$T_{\max} = 10000$; $\bar{l} = 3.9090$



Rysunek 2.4: $\lambda = 5$, Par(1.2) $\times$ 0.2, $T_{\max} = 100000$; $\bar{l} = 4.8395$



Rysunek 2.5: $\lambda = 5$, $\mathrm{Par}(2.2) \times 1.2$, $T_{\max} = 1000$; $\bar{l} = 4.8775$



Rysunek 2.6: $\lambda = 5$, $\mathrm{Par}(3.2) \times 2.2$, $T_{\max} = 1000$; $\bar{l} = 5.1314$

W przypadku sumulacji 5-tej mamy $\alpha = 2.2$ i wtedy $S = 1.2X$, gdzie $X =_{\rm d}{\rm Par}(2.2)$. Stąd

$$R(t) = 5\int_t^\infty (1+\frac{t}{1.2})^{-2.2}\,ds = 5(1+\frac{t}{1.2})^{-1.2}\,.$$

Widzimy więc, ze $\bar\sigma^2 = 2\int_0^\infty R(t)\,dt = 60$, skąd mamy że aby $b = 0.5$ (jedna cyfra wiodąca) wystarczy $T_{\max} = 922$. Należy ostrzec, że nie wzięliśmy pod uwagę rozpędówki, przyjmując do obliczeń, że proces $L(t)$ jest od początku stacjonarny.

Jesli $\alpha = 3.2$ i wtedy $S = 2.2X$, gdzie $X =_{\rm d}{\rm Par}(3.2)$. Stąd

$$R(t) = 5\int_t^\infty (1+\frac{s}{2.2})^{-3.2}\,ds = 5(1+\frac{t}{2.2})^{-2.2}\,.$$

Widzimy więc, ze $\bar\sigma^2 = 2\int_0^\infty R(t)\,dt = 18.3$, skąd mamy że aby $b = 0.5$ (jedna cyfra wiodąca) wystarczy $T_{\max} = 281$. Natomiast aby $b = 0.05$ to musimy mięc $T_{\max}$ większe od 28100.

3 Symulacja gdy X jest procesem regenerującym się.

Będziemy zakładać następującą strukturę procesu stochastycznego X. Przypuśćmy, że istnieją chwile losowe $0 \le T_1 < T_2 < \ldots$ takie, że pary zmiennych losowych $((Z_i, C_i))_{i\ge1}$ są niezależne i począwszy od numeru $i \ge 2$ niezależne, gdzie

$$Z_0 = \int_0^{T_1} X(s)\,ds, Z_1 = \int_{T_1}^{T_2} X(s)\,ds, \ldots$$

oraz $C_0 = T_1, C_1 = T_2 - T_1, \ldots$. Takie procesy nazywamy procesami *regenerującymi się*. Wtedy mamy następujący rezultat (patrz Asmussen & Glynn [4]):

Fakt 3.1 *Z prawdopodobieństwem 1, dla $t \to \infty$*

$$\hat X(t) \to I = \frac{\mathbb{E}\int_{T_1}^{T_2} X(s)\,ds}{\mathbb{E}\,C_1}, \tag{3.6}$$

oraz

$$t^{1/2}(\hat X(t) - I) \to_{\rm d} \mathcal{N}(0, \bar\sigma^2), \tag{3.7}$$

i TAVC wynosi

$$\bar{\sigma}^2 = \frac{\mathbb{E}\,\zeta_1^2}{\mathbb{E}\,Z_1}$$

gdzie

$$\zeta_k = Z_k - IC_k\ .$$

Patrząc na postać (3.6) widzimy, że można też użyć innego estymatora, który jest krewniakiem estymatora $\hat{X}$. Mianowicie możemy wziąć

$$\hat{I}_k = \frac{Z_1 + \ldots + Z_k}{C_1 + \ldots + C_k}.$$

Jest oczywiste, że z prawdopodobieństwem 1

$$\hat{I}_k \to I$$

ponieważ

$$\frac{Z_1 + \ldots + Z_k}{C_1 + \ldots + C_k} = \frac{(Z_1 + \ldots + Z_k)/k}{(C_1 + \ldots + C_k)/k} \to \frac{\mathbb{E}\,Z_1}{\mathbb{E}\,C_1}\ .$$

Wtedy mamy następujące twierdzenie (patrz Asmussen & Glynn [4], Prop. 4.1).

Twierdzenie 3.2 *Dla* $k \to \infty$

$$k^{1/2}(\hat{I}_k - I) \to_{\mathrm{d}} \mathcal{N}(0, \eta^2),$$

gdzie

$$\eta^2 = \frac{\mathbb{E}\,\zeta_1^2}{\mathbb{E}\,C_1}$$

i

$$\zeta_1 = Z_1 - IC_1\ .$$

Stąd asymptotyczny $100(1-\epsilon)\%$ *przedział ufności jest dany przez*

$$\hat{I}_k \pm z_{1-\epsilon/2}\hat{\eta}^2/\sqrt{k}$$

gdzie

$$\hat{\eta}^2 = \frac{1}{k-1}\sum_{l=1}^{k}(C_l - \hat{I}_k Z_l)^2/(\frac{1}{k}\sum_{l=1}^{k} C_l)\ .$$

Przykład 3.3 Proces $L(t)$ w M/M/1 jest regenerujący ponieważ

$$T_1 = \inf\{t > 0 : L(t-) > 0, L(t) = 0\}$$

oraz

$$T_{n+1} = \inf\{t > T_n : L(t-) > 0, L(t) = 0\}$$

definiuje momenty regeneracji procesu $L(t)$. Wynika to z własności braku pamięci rozkładu wykładniczego (patrz XI.5.6). Po osiągnięciu przez proces $L(t)$ zera, niezależnie jak długo czekaliśmy na nowe zgłoszenie, czas do nowego zgłoszenia ma rozkład wykładniczy Exp(λ).

4 Przepisy na szacowanie rozpędówki

W wielu przypadkach proces po *pewnym czasie* jest już w przybliżeniu w warunkach stacjonarności. Następujący przykład ilustruje ten pomysł w sposób transparentny.

Przykład 4.1 Z teorii kolejek wiadomo, że jeśli w systemie M/M/1 wystartujemy w chwili $t = 0$ z rozkładu geometrycznego Geo(ρ), to proces liczby zadań $L_1(t)$ jest stacjonarny. Dla uproszczenia wystartujmy proces niestacjonarny z $L_2(0) = 0$. Zakładamy przypadek $\rho < 1$. Wtedy proces $L_1(t)$ osiąga stan 0 w skończonym czasie. Ponieważ oba procesy mają tylko skoki ± 1, więc mamy że procesy $L_1(t)$ i $L_2(t)$ muszą się spotkać, tzn.

$$T = \inf\{t > 0 : L_1(t) = L_2(t)\}$$

jest skończone z prawdopodobieństwem 1. Od tego momentu procesy ewoluują jednakowo. To pokazuje, że od momentu T proces $L_2(t)$ jest stacjonarny. Jednakże jest to moment losowy.

Bardzo ważnym z praktycznych względów jest ustalenie początkowego okresu s którego nie powinniśmy rejestrować. Ten okres nazywamy *długością rozpędówki* lub krótko *rozpędówką* (warm-up period). W tym celu definiujemy nowy estymator

$$\hat{X}(s,t) = \frac{\int_s^t X(v)\,dv}{t-s}\,. \tag{4.8}$$

Jedna realizacja – metoda średnich pęczkowych Dla tych czytelników, którzy znają teorię procesów stochastycznych, kluczową własnością zakładaną przez proces $X(t)$ jest spełnienie funkcjonalnej wersji centralnego twierdzenia granicznego

$$(\epsilon^{1/2}(X(t/\epsilon) - It/\epsilon))_{t\geq 0} \xrightarrow{\mathcal{D}} (\bar{\sigma}B(t))_{t\geq 0}, \tag{4.9}$$

gdzie $(B(t))_{t\geq 0}$ jest standardowym ruchem Browna.

Dzielimy odcinek symulacji $[0,t]$ na l pęczków jednakowej długości t/l (dla ułatwienia załóżmy, że l dzieli t) i zdefiniujmy *średnie pęczkowe*

$$\hat{X}_k(t) = \hat{X}((k-1)t/l, kt/l) = \frac{1}{t/l}\int_{(k-1)t/l}^{kt/l} X(s)\,ds,$$

gdzie $k = 1, 2, \ldots$. Jeśli t ale również t/l są duże, to można przyjąć, że

(P1) $\hat{X}_k(t)$ $(k = 1, \ldots, l)$ ma w przybliżeniu rozkład normalny $\mathcal{N}(I, \bar{\sigma}^2 l/t)$; $(k = 1, \ldots, l)$,

(P2) $\hat{X}_k(t)$ $(k = 1, \ldots, l)$ są niezależne.

Formalne uzasadnienie można przeprowadzić korzystając z założenia, że spełniona jest funkcjonalna wersja centralnego twierdzenia granicznego (4.9). Ponieważ

$$\hat{X}(t) = \frac{\hat{X}_1(t) + \ldots + \hat{X}_l(t)}{l},$$

więc dla $t \to \infty$

$$l^{1/2}\frac{\hat{X}(t) - I}{\hat{s}_l(t)} \to_{\mathrm{d}} T_{l-1}$$

gdzie T_{l-1} zmienną losową o rozkładzie t-Studenta z $l-1$ stopniami swobody i

$$\hat{s}_l^2(t) = \frac{1}{l-1}\sum_{i=1}^{l}(\hat{X}_i(t) - \hat{X}(t))^2\ .$$

A więc jeśli $t_{1-\epsilon/2}$ jest $1-\epsilon/2$ kwantylem zmiennej losowej T_{l-1}, to asymptotyczny przedział ufności dla I jest

$$\hat{X}(t) \pm t_{1-\epsilon/2}\frac{\hat{s}_l(t)}{\sqrt{l}}\ .$$

Asmussen & Glynn [4] sugerują aby $5 \leq l \leq 30$.

4.1 Metoda niezależnych replikacji

Efekty symulacji podczas rozpędówki nie musimy rejestrować a dopiero po tym czasie zaczynamy rejestrować i następnie opracować wynik jak dla procesu stacjonarnego. Problemem jest, że nie wiemy co to znaczy po *pewnym czasie*. Fishman [14, 15] zaproponował następującą graficzną metodę szacowania rozpędówki. Zamiast symulowania jednej realizacji $X(t)$ $(0 \le t \le t_{\max})$ symulujemy teraz n realizacji $X^i(t)$, $(0 \le t \le t_{\max})$, gdzie $i = 1, \ldots, n$. Każdą zaczynamy od początku z tego samego punktu, ale z rozłącznymi liczbami losowymi. Niech

$$\hat{X}^i(s,t) = \frac{\int_s^t X^i(v)\,dv}{t-s}$$

będzie opóźnioną średnią dla i-tej replikacji realizacji procesu $(i = 1, \ldots, n)$. Definiujemy

$$\hat{X}^{\cdot}(t) = \frac{1}{n}\sum_{i=1}^{n} X^i(t), \qquad 0 \le t \le t_{\max},$$

oraz

$$\hat{X}^{\cdot}(t, t_{\max}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\int_t^{t_{\max}} X^i(v)\,dv}{t_{\max}-t}, \qquad 0 \le t \le t_{\max}\,.$$

Jeśli proces X jest stacjonarny to obie krzywe $\hat{X}^{\cdot}(t)$ i $\hat{X}^{\cdot}(t, t_{\max})$ powinny wyglądać podobnie. Pomijamy fluktuacje spowodowane niedużym n dla $\hat{X}^{\cdot}(t)$ oraz krótkością odcinka dla $\hat{X}^{\cdot}(t, t_{\max})$. Dlatego porównując na jednym wykresie

$$(\hat{X}^{\cdot}(s), \hat{X}^{\cdot}(s, t_{\max})),\ 0 \le s \le t_{\max}$$

możemy oszacować jakie s należy przyjąć.

4.2 Efekt stanu początkowego i załadowania systemu

Na długośc rozpędówki mają istotny wpływ przyjęty stan początkowy symulacji oraz załadowanie systemu.

Przykład 4.2 Tytułowy problem zilustrujemy za pracą Srikanta i Whitta [43] o symulacji prawdopodobieństwa straty

$$p_{\text{block}} = \frac{\rho^c/c!}{\sum_{j=0}^{c} \rho^j/j!}$$

(jest to tzw. wzór Erlanga) w systemie M/M/c ze stratami (patrz dodatek XI.7.3). Zadania wpływają z intensywnością λ natomiast ich średni rozmiar wynosi μ^{-1}. Jak zwykle oznaczamy $\rho = \lambda/\mu$. Dla tego systemu oczywiście znamy prawdopodobieństwo straty jak i wiele innych charakterystyk, co pozwoli nam zrozumieć problemy jakie mogą występować przy innych symulacjach. Zacznijmy od przeglądu kandydatów na estymatory. Niech $A(t)$ będzie liczbą wszystkich zadań (straconych i zaakceptowanych) które wpłynęły do chwili t a $B(t)$ jest liczba straconych zadań do chwili w odcinku $[0, t]$. Wtedy

$$B(t) = \sum_{j=1}^{A(t)} J_j$$

gdzie J_j jest zdefiniowane w (XI.7.13). A więc

$$B(t)/A(t) \to p_{\text{block}}$$

z prawdopdobieństwm 1. To uzasadnia propozycję estymatora p_{block}

$$\bar{B}_{\text{N}}(t) = \frac{B(t)}{A(t)}.$$

Estymator ten nazwiemy *estymatorem naturalnym*. Innym estymatorem jest tak zwany *prosty estymator*

$$\bar{B}_{\text{S}}(t) = \frac{B(t)}{\lambda t},$$

lub *niebezpośredni estymator* (indrirect estimator)

$$\bar{B}_{\text{I}}(t) = 1 - \frac{\hat{L}(t)}{\rho}.$$

Podamy teraz krótkie uzasadnienia dla tych estymatorów. Czy są one zgodne? Pierwszy jest oczywiście zgodny. Podobnie wygląda sytuacja z estymatorem prostym B_{S}. Mianowicie

$$\begin{aligned} \lim_{t\to\infty} \frac{B(t)}{A(t)} &= \lim_{t\to\infty} \frac{B(t)}{t} \frac{t}{A(t)} \\ &= \lim_{t\to\infty} \frac{B(t)}{\lambda t}\,. \end{aligned}$$

ponieważ dla procesu Poissona, mamy $t/A(t) \to \lambda$ z prawdopdobieństwem 1. Natomiast dla estymatora niebezpośredniego aby pokazać zgodność trzeba wykorzystać tzw. wzór Little'a [4]:

$$\hat{l} = \lambda(1 - p_{\text{block}})/\mu \ ,$$

gdzie $\hat{l}$ jest średnią Jeśli natomiast by założyć, że system jest w warunkach stacjonarności, to można pokazać, że estymatory prosty i niebezpośredni są nieobciążone. W pracy Srikanta i Whitta [43] opisano następujący eksperyment z $\lambda = 380$, $\mu = 1$ i $c = 400$. Dla tego systemu wysymulowano, startując z pustego systemu, przy użyciu estymatora niebezpośredniego, z $t_{\max} = 5400$, mamy $p_{\text{block}} = 0.00017$. Tutaj TAVC wynosi około 0.00066. W tym zadaniu można policzyć numerycznie asymptotyczne obciążenie co przedstawione jest w tabl. 4.1 z $\alpha = \rho/c$. Zauważmy, że w przypadku $N(0) = 0$, TAVC jest

α	$N(0) = 0$	$N(0) = c$
0.7	1.00	-0.42
0.8	1.00	-0.24
0.9	0.99	-0.086
1.0	0.85	-0.0123
1.1	0.66	-0.0019
1.2	0.52	-0.0005

Tablica 4.1: Asymptotyczne obciążenie β dla $\hat{B}_{\text{I}}$

tego samego rzędu co obciążenia ($\beta/5400$ i 0.00066).

W następnym przykładzie zajmiemy się wpływem załadowania systemu.

Przykład 4.3 Pytanie o to jak długi należy dobrać horyzont symulacji $t_{\max}$ aby osiągnąć zadaną precyzję zależy często od współczynnika załadowania systemu który symulujemy. W systemie M/M/1 mozna policzyć TAVC $\bar{\sigma}^2$, która to wielkość poprzez fundamentalny związek rzutuje na $t_{\max}$. Mianowicie (patrz Whitt [47]) mamy

$$\bar{\sigma}^2 = \frac{2\rho(1+\rho)}{(1-\rho)^4} \ .$$

Oczywiście musimy założyć $\rho < 1$.

[4] Sprawdzić tożsamość.

Stąd widać, że $\bar{\sigma}^2$ gwałtownie rośnie, gdy $\rho \to 1$, czyli w *warunkach wielkiego załadowania*. Wzór powyższy jest szczególnym przypadkiem wzoru otrzymanego dla procesów narodzin i śmierci bez obliczania funkcji $R(t)$; patrz Whitt [47]. Stąd, korzystając z fundamentalnego związku, mamy że tak samo szybko rośnie horyzont symulacji aby utrzymać zadaną dokładność. Przeprowadzimy teraz analizę jak dobrać $t_{\max}$ do symulacji średniej liczby zadań w systemie. Takie symulacje przeprowadzaliśmy w podrozdziale VI.2.1. Przypomnijmy, że do eksperymentu przyjęto $\lambda = 8/17$ oraz $\mu = 9/17$. Wtedy średnia liczba zadań w systemie wynosi $\bar{l} = 8$ (patrz wzór XI.7.10). Ponieważ $\rho = 8/9$ więc

$$\bar{\sigma}^2 = 22032.$$

Jeśli ponadto przyjmiemy $b = 0.5$ i $\alpha = 0.05$ to wtedy $t_{\max} = 338550$. Natomiast gdy w jednym naszym z eksperymentów przyjęliśmy $t_{\max} = 10000000$ przy którym otrzymaliśmy wynik $\bar{l} = 7.9499$, to korzystając z fundamentalnego wzoru **[??eq.stac.fund.zw??]** orzymamy, że połowa przedziału ufności wynosi 0.0920.

5 Uwagi bibliograficzne

Cały cykl prac na temat planowania symulacji w teorii kolejek napisał Whitt z współpracownikami; [45, 46, 47, 43]. Tym tematem zajmował się również Grassmann [49]

Zadania teoretyczne

5.1 Uzasadnić zgodność estymatorów z podrozdziału **[??whit.srinikant??]**. Pokazać, że estymatory B_I oraz B_S sa nieobciążone jeśli proces stanu systemu $L(t)$ jest stacjonarny.

Zadania laboratoryjne

5.2 Przeprowadzić następujące eksperymenty dla M/M/1. Napisać program kreślący realizację $L(t)$ dla $0 \leq t \leq 10$. Eksperyment przeprowadzić dla
(a) $\lambda = 1, \mu = 1.5$ oraz $L(0) = 0$.
(b) $\lambda = 1, \mu = 1.5$ i $L(0) = 2$.
(c) $\lambda = 1, \mu = 1.5$ i $L(0) = 4$.
(d) $L(0) \sim \text{Geo}(\rho)$, gdzie $\rho = \lambda/\mu = 2/3$.

5.3 Kontynuacja zad. 5.2. Obliczyć $\widehat{L}(t_{\max})$, gdzie $\widehat{L}(t) = (\int_0^t L(s)\,ds)/t$ gdy $L(0) = 0$.
(a) Przeprowadzić eksperyment przy $\lambda = 1, \mu = 1.5$ oraz $L(0) = 0$ dla $t_{\max} = 5, 10, 50, 100, 1000$. Policzyć dla każdej symulacji połowy długości przedziały ufności na poziomie 95%.
(b) Przeprowadzić eksperyment przy $\lambda = 1, \mu = 1.1$ oraz $L(0) = 0$ dla $t_{\max} = 5, 10, 50, 100, 1000$.

5.4 Kontynuacja zad. 5.2. Przez replikacje rozumiemy teraz pojedynczą realizację $L(t)$ dla $0 \leq t \leq 10$. Dla replikacji $L_i(t)$ zachowujemy wektor $\boldsymbol{l}_i = (L_i(1), L_i(2), \ldots, L_i(10))$ i niech 1000 będzie liczbą replikacji. Obliczyć

$$\widehat{\boldsymbol{l}} = (\frac{1}{1000}\sum_{j=1}^{1000} L_j(0.5), \frac{1}{1000}\sum_{j=1}^{1000} L_j(1), \ldots, \frac{1}{1000}\sum_{j=1}^{1000} L_j(10)) \ .$$

Zrobić wykres średniej długości kolejki w przedziale $0 \leq t \leq 10$.
(a) Przeprowadzić eksperyment dla $L(0) = 0$.
(b) Przeprowadzić eksperyment dla $L(0) \sim \text{Geo}(\rho)$.

5.5 Napisać algorytm na generowanie dla systemu M/G/∞ estymatora

$$\frac{1}{t_{\max} - t_{\min}} \int_{t_{\min}}^{t_{\max}} L(t)\,dt$$

(patrz wzór (4.8)). Wsk. Zobacz algorytm w dodatku XI.7.1 z $t_{\min} = 0$.
Przeprowadzić eksperymenty omówione w przykładzie 2.1 z rozmaitymi $t_{\min}$.

Projekt

5.6 Napisać program symulujący realizację procesu liczby zadań $L(t)$ w systemie M/G/∞. Przeprowadzić eksperyment z
(a) $\lambda = 5$ oraz $G = \mathrm{Exp}(\mu)$ z $\mu = 1$.
(b) $\lambda = 5$ oraz rozmiary zadań mają rozkład Pareto z $\alpha = 1.2$ przemnożonym przez 0.2 (tak aby mieć średnią 1).
Przyjąć $L(0) = 0$, $L(0) = 10$ oraz $L(0)$ ma rozkład Poissona Poi(5). Dla wszystich wypadków oszacować rozpędówkę.
Na jednym wykresie umieścić $\int_0^t L(s)\,ds/t$ dla $1 \le t \le 100$ z symulacji (a) i (b). Przyjąć $L(0) = 0$.

Rozdział VIII

MCMC

[1]

1 Łańchuchy Markowa

Przez łańcuch Markowa rozumiemy pewnien ciąg zmiennych (elementów) losowych $Y_0, Y_1, Y_2, \ldots$ o wartościach w przestrzeni S i związany z sobą. W tym podrozdziale będziemy zakładać, że $\mathcal{E}$ jest skończona z elementami $\mathcal{E} = \{s_1, s_2, \ldots, s_m\}$. Dla wygody będziemy dalej przyjmować $\mathcal{E} = \{1, \ldots, m\}$. Przez macierz przejścia $\boldsymbol{P} = (p_{ij})_{i,j\in\mathcal{E}}$ rozumiemy macierz, której elementy spełniają:

$$p_{ij} \geq 0, \qquad i, j \in \mathcal{E}$$

oraz

$$\sum_{j\in\mathcal{E}} p_{ij} = 1, \qquad i \in \mathcal{E}.$$

Niech $\boldsymbol{\alpha} = (\alpha_j)_{j\in\mathcal{E}}$ będzie rozkładem (funkcją prawdopdobieństwa) na $\mathcal{E}$. Mówimy, że $Y_0, Y_1, \ldots$ jest łańcuchem Markowa z rozkładem początkowym $\boldsymbol{\alpha}$ i macierzą przejścia $\boldsymbol{P}$ jeśli dla $i_0, \ldots, i_k \in \mathcal{E}$

$$\mathbb{P}(Y_0 = i_0, Y_1 = i_1, \ldots, Y_k = i_k) = \alpha_{i_0} p_{i_0 i_1} \cdots p_{i_{k-1} i_k}.$$

Ważnym pojęciem jest *rozkład stacjonarny*, tj. $\boldsymbol{\pi} = (\pi_j)_{j\in\mathcal{E}}$ (oczywiście jest to ciąg liczb nieujemnych sumujących się do 1) i spełniający

$$\pi_j = \sum_{i\in\mathcal{E}} \pi_i p_{ij}, \qquad j \in \mathcal{E}.$$

[1] 1.2.2016

Podamy teraz kilka potrzebnych dalej definicji. Elementy potęgi macierzy $\boldsymbol{P}^n$ oznaczamy $p_{ij}^{(n)}$. Dwa stany $i, j \in \mathcal{E}$ sa skomunikowane, jeśli istnieją n_0, n_1 takie, że $p_{ij}^{(n_0)} > 0$ oraz $p_{ji}^{(n_1)} > 0$. Macierz $\boldsymbol{P}$ jest nieredukowalna, jeśli wszystkie stany są skomunikowane.

Mówimy, że macierz $\boldsymbol{P}$ jest regularna (ergodyczna), jeśli istnieje n_0, takie że $\boldsymbol{P}^{n_0}$ ma wszystkie elementy ściśle dodatnie ($\boldsymbol{P}^{n_0} > \boldsymbol{0}$). Można pokazać, że $\boldsymbol{P}$ jest regularna wtedy i tylko wtedy gdy jest nieredukowalna i nieokresowa (odsyłamy po definicję okresowości do specjalistycznych książek z łańcuchów Markowa lub [**?**]).

Fakt 1.1 *Jeśli $\boldsymbol{P}$ jest regularna, to istnieje tylko jeden rozkład stacjonarny $\boldsymbol{\pi}$, którego elementy są ściśle dodatnie, oraz*

$$\boldsymbol{P}^n \to \boldsymbol{\Pi},$$

gdy $n \to \infty$, gdzie

$$\boldsymbol{\Pi} = \begin{pmatrix} \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{pmatrix}$$

Fakt 1.2 *Niech $\boldsymbol{P}$ i $\boldsymbol{\alpha}$ będą dane. Istnieje funkcja $\varphi : \mathcal{E} \times [0,1] \to \mathcal{E}$ taka, że jeśli Y_0 ma rozkład $\boldsymbol{\alpha}$, to $Y_0, Y_1, \ldots$ zdefiniowane rekurencyjnie*

$$Y_{n+1} = \varphi(Y_n, U_{n+1}), \qquad n = 0, 1, \ldots$$

gdzie $U_1, \ldots$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie jednostajnym na $[0,1]$, jest łańcuchem Markowa z rozkładem początkowym $\boldsymbol{\alpha}$ i macierzą przejścia $\boldsymbol{P}$.

Dowód Zdefiniujmy

$$\varphi(i,x) = \begin{cases} 1, & x \in [0, p_{i1}) \\ 2, & x \in [p_{i1}, p_{i1} + p_{i2}) \\ \vdots & \vdots \\ m, & x \in [p_{i1} + \cdots + p_{i,m-1}, 1]. \end{cases}$$

□

Reprezentacja powyższa nie jest jednoznaczna.

Oznaczmy

$$\hat{Y}_n = \frac{1}{n} \sum_{j=0}^{n-1} f(Y_j). \tag{1.1}$$

Fakt 1.3 *Jeśli macierz prawdopodobieństwa przejścia jest nieredukowalna, to*

$$\hat{Y}_n \to \bar{y} = \sum_{j=1}^{m} \pi_j f(j), \qquad p.n.p. \tag{1.2}$$

gdy $n \to \infty$. *Ponadto*

$$\sqrt{n}(\hat{Y} - \bar{y}) \to_{\mathrm{d}} \mathrm{N}(0, \bar{\sigma}^2),$$

gdzie

$$\bar{\sigma}^2 = \lim_{n\to\infty} n\mathrm{Var}\,\hat{Y}_n. \tag{1.3}$$

Należy zauważyć, że granice w (1.2) i (1.3) nie zależą od rozkładu początkowego $\boldsymbol{\alpha}$.

Uwaga Zamiast formalnego dowodu pokażemy jego pomysł. Niech $i \in \mathcal{E}$ oraz podstawmy $Y_j = f(Y_j)$. Ponieważ $\boldsymbol{P}$ jest nieredukowlana, więc łańcuch (Y_n) nieskończenie często odwiedza stan i. Niech $\gamma_j, j = 1, 2, \ldots$ będą momentami odwiedzin (Y_n) w i. Ponieważ (Y_n) jest łańcuchem Markowa, więc cały ciąg Y_n rozpada się na niezależne cykle $Y_0, \ldots, Y_{\gamma_1-1}$, $Y_{\gamma_1}, \ldots, Y_{\gamma_2}, \ldots$. Zauważmy teraz, że

$$W_i = \sum_{j=\gamma_i}^{\gamma_{i+1}-1} Y_j, \qquad i = 1, 2,$$

są niezaleznymi zmiennym losowymi o jednakowym rozkładzie. Ponadto z teorii łańcuchów Markowa wiadomo, że $\mathbb{E}\, W_i^2 < \infty$. Kluczem do dalszych rozważań jest przedstawienie

$$Y_1 + \cdots + Y_n = W_1 + \cdots + W_{\nu(n)} + R_n$$

gdzie $\nu(n) = \max\{j : \gamma(j) \le n\}$, oraz R_n nie zależy od W_j-ów. Teraz należy wykorzystać wariant mocnego prawa wielkich liczb oraz centralnego twierdzenia granicznego z losowym indeksem.

W rozdziale IV zdefiniowaliśmy pojęcie stacjonarności procesu stochastycznego $X(t)$. Jest to pojęcie ważne gdy parametr procesu jest $\mathbb{R}$ lub $\mathbb{R}_+$ ale też gdy $\mathbb{Z}$ lub $\mathbb{Z}_+$. Przypomnijmy więc to pojęcie dla ciagu zmiennych losowych $Z_n, n \in \mathbb{Z}_+$. Mówimy, że jest on *stacjonarny* jeśli spełnione są następujące warunki.

(1) Wszystkie rozkłady Z_j są takie same ($j \in \mathbb{Z}_+$). W konsekwencji wszystkie $\mathbb{E}\, Z_j$ są równe jeśli tylko pierwszy moment $m = \mathbb{E}\, Z_0$ istnieje.

(2) Dla każdego $h = 0, 1, \ldots$, łączne rozkłady (Z_j, Z_{j+h}) nie zależą od j. W konsekwencji wszystkie kowariancje $\rho_h = \mathrm{Cov}\,(Z_j, Z_{j+h})$, jeśli tylko drugi moment istnieje.

(k) Dla każdego $0 < h_1 < \ldots < h_{k-1}$, łączne rozkłady $(Z_i, Z_{j+h_1}, \ldots, Z_{j+h_{k-1}})$ nie zależą od j.

Niech $\boldsymbol{\pi}$ będzie rozkładem stacjonarnym i $(X_n)_{n=0,1,\ldots}$ będzie łańcuchem Markowa z rozkładem początkowym $\boldsymbol{\alpha} = \boldsymbol{\pi}$ i macierzą przejścia $\boldsymbol{P}$. Wtedy $(Y_n)_{n=0,1,\ldots}$, gdzie $Y_n = f(X_n)$ jest ciągiem stacjonarnym i funkcją kowariancji jest ciąg

$$\rho_k = \mathrm{Cov}\,(Y_n, Y_{n+k}).$$

Jesli przestrzeń stanów jest skończona i $\boldsymbol{P}$ jest regularna, to mamy $\sum_j |\rho_j| < \infty$.

Zauważmy, że podobnie jak rozdziale IV stałą $\bar{\sigma}^2$ możemy nazywać TAVC. Pokażemy teraz inny wzór na TAVC.

Lemat 1.4 *Jeśli $\boldsymbol{P}$ jest regularna to*

$$\bar{\sigma}^2 = \rho_0 + 2\sum_{k=1}^{\infty} \rho_k.$$

Dowód Ponieważ $\bar{\sigma}^2$ zdefiniowane w (1.3) nie zależy od warunku początkowego $\boldsymbol{\alpha}$ więc rozważmy stacjonarny ciąg (Y_n). Bez straty ogólności możemy założyć $\mathbb{E}\, Y_j = 0$. Wtedy

$$\begin{aligned}
n\mathrm{Var}\,\hat{Y}_n &= \frac{1}{n}\mathrm{Var}\,(Y_1 + \cdots + Y_n) \\
&= \frac{1}{n}(n\mathrm{Var}\,Y_0 + 2\sum_{0\le i<j\le n} \mathrm{Var}\,(Y_i, Y_j)) \\
&= \rho_0 + \frac{2}{n}\sum_{0\le i<j\le n} \rho_{j-i} \\
&= \rho_0 + \frac{2}{n}\sum_{i=0}^{n}\sum_{h=1}^{i} \rho_h \\
&\sim \rho_0 + 2\sum_{h=1}^{\infty} \rho_h.
\end{aligned}$$

□

Jak estymować ρ_k i $\bar{\sigma}^2$? Możemy to zrobić przy użyciu estymatorów

$$\hat{\rho}_k = \frac{1}{n-k}\sum_{j=1}^{n-k}(Y_j - \hat{Y}_n)(Y_{j+k} - \hat{Y}_n)$$

oraz

$$\hat{\sigma}^2 \approx \sum_{k=-N}^{N} \hat{\rho}_k$$

Estymator $\hat{Y}_n$ jest oczywiście zgodny (dąży z prawdopodobieństwem 1 do $\bar{y} = \sum_{j=1}^{m} \pi_j f(j)$), ale jeśli rozkład początkowy $\boldsymbol{\alpha}$ zmiennej X_0 nie jest rozkładem stacjonarnym, to nie jest nieobciążonym. Jak szybko $\mathbb{E}\,\hat{Y}_n$ zbiega do $\bar{y}$ mówi następujacy fakt.

Fakt 1.5 *Dla pewnej skończonej stałej A*

$$\mathbb{E}\,Y_n = \bar{y} + \frac{A}{n} + o(\frac{1}{n}).$$

Dowód Podstawmy $a_j = \mathbb{E}\,Y_j - \bar{y}$. Ponieważ $\mathcal{E}$ jest skończona, więc $\sum_j |a_j| < \infty$ (tego nie będziemy dowodzić). Teraz

$$\mathbb{E}\,\hat{Y}_n = \frac{1}{n}\sum_{j=0}^{n}(\mathbb{E}\,Y_j - \bar{y} + \bar{y}) = \bar{y} + \frac{1}{n}\sum_{j=0}^{n-1} a_j = \bar{y} + \frac{1}{n}\sum_{j=0}^{\infty} a_j - \frac{1}{n}\sum_{j=n}^{\infty} a_j$$

a ponieważ $\frac{1}{n}\sum_{j=n}^{\infty} a_j = o(1/n)$ więc dowód jest skończony. □

Uwaga Fakt 1.5 jest prawdziwy tylko dla łanczuchów ze skończoną liczba stanów.

Fakt 1.6 *Niech (Y_n) będzie łancuchem Markowa z macierzą prawdopodobieństwa $\boldsymbol{P}$. Załóżmy, że istnieją ściśle dodatnie liczby $v_j, j \in \mathcal{E}$ takie, że*

$$v_i p_{ij} = v_j p_{ji}, \qquad i, j \in \mathcal{E} \tag{1.4}$$

Wtedy

$$\pi_j = \frac{v_j}{\sum_{i\in\mathcal{E}} v_i}, \qquad j \in \mathcal{E}$$

jest rozkładem stacjonarnym.

Dowód Bez straty ogólności możemy założyć, że v_j w (1.4) sumują się do jedności. Teraz

$$\sum_{i\in\mathcal{E}} v_i p_{ij} = \sum_{i\in\mathcal{E}} v_j p_{ji} = v_j \sum_{i\in\mathcal{E}} p_{ij} = 1.$$

□

Warunek (1.4) będziemy nazywać DBC (od *detailed balance condition*). Dla stacjonarnych łancuchów Markowa jest on równoważny odwracalności. Mówimy, że stacjonarny łancuch Markowa $Y_0, Y_1, \ldots$ jest odwracalny, jesli dla $n = 0, 1, 2, \ldots$

$$(Y_0, Y_1, \ldots, Y_n) =_{\mathrm{d}} (Y_n, Y_{n-1}, \ldots, , Y_0).$$

Wniosek 1.7 *Jeśli macierz przejścia* $\boldsymbol{P}$ *jest symetryczna, tj.* $p_{ij} = p_{ji}$*, to rozkład stacjonarny jest jednostajny na* $\mathcal{E}$ *czyli* $\pi = 1/|\mathcal{E}|$*,* $j \in \mathcal{E}$*.*

Przykład 1.8 Proste błądzenie przypadkowe na grafie G**.** Graf G składa się z wierzchołków V oraz krawędzi E. Zbiór krawędzi można utożsamić z jakimś podbiorem zbioru par (nieuporządkowanych) $(i, j), i \neq j, i, j \in V$. Mówimy, że (i, j) sa sąsiadami, jeśli jest krawędź łącząca i z j. Liczbę sąsiadów wierzchołka i oznaczamy przez $\deg(i)$. W prostym (symetrycznym) błądzeniu przypadkowym z stanu i przechodzimy do dowolnego sąsiada z tym samym prawdopodobieństwem. A więc

$$q_{ij} = \begin{cases} \frac{1}{\deg(i)}, & \text{jesli } j \text{ jest sąsiadem } i, \\ 0, & \text{w przeciwnym razie.} \end{cases} \tag{1.5}$$

Łatwo sprawdzić, że jesli $v_i = \deg(i)$ to równość (1.4) jest spełniona. A więc

$$\pi_i = \frac{\deg(i)}{\sum_{j\in V} \deg(j)}\,, \qquad j \in \mathcal{E}$$

jest rozkładem stacjonarnym. Polecamy się zastanowić kiedy macierz $\boldsymbol{Q}$ jest regularna a kiedy nieredukowalna. Podstawą jest pojęcie grafu spójnego, w którym jest możliwość przejścia po krawędziach z każdego wierzchołka do wszystkich innych, niekoniecznie w jednym kroku. Innym problemem jest zauważenie, że okresowość jest związana z pojęciem *bipartite*.[2]

[2]A graph is bipartite if and only if it does not have cycles with an odd number of edges.

Przykład 1.9 Błądzenie przypadkowe na hiperkostce. Rozważmy zbiór wierzchołków $\{0,1\}^m$. Elementami są ciągi m-elementowe zer i jedynek. Sąsiadujące elementy są oddalone o 1 tj. jesli $s_1 = (e_{11}, \ldots, e_{1m})$ oraz $s_1 = (e_{21}, \ldots, e_{2m})$, są one sąsiadami jeśli $\sum_{j=1}^{m} |e_{1j} - e_{2j}| = 1$. A więc z każdego wierzchołka wychodzi m krawędzi, czyli $\deg(i) = m$. Rozkład stacjonarny jest więc rozkładem jednostajnym na V.

2 MCMC

Omówimy teraz pewną metodę Monte Carlo łańcuchów Markowa (Markov chain Monte Carlo; stąd MCMC). Naszym zadaniem jest generowanie obiektu losowego o wartościach w $\mathcal{E} = \{s_1, \ldots, s_m\}$ o rozkładzie $\boldsymbol{\pi}$, tj. o masie π_j na elemencie s_j (na razie to nie jest rozkład stacjonarny ponieważ jeszcze nie mamy macierzy przejścia). Bardziej dokładnie, chcemy obliczyć $\sum_{j\in\mathcal{E}} f(j)\pi_j$. Do tego celu skonstruujemy łańcuch Markowa z macierzą przejścia $\boldsymbol{P}$, dla którego $\boldsymbol{\pi}$ jest rozkładem stacjonarnym. Wtedy $\hat{Y}_n$ zdefiniowany w (1.1) będzie szukanym estymatorem $\sum_{j\in\mathcal{E}} f(j)\pi_j$.

2.1 Algorytmy Metropolisa

Niech $\boldsymbol{\pi}$ będzie naszym celowym rozkładem na $\mathcal{E}$. Niech $X_0, X_1, \ldots$ będzie jakimś łańcuchem Markowa na $\mathcal{E}$ z macierzą przejścia $\boldsymbol{Q} = (q_{ij})_{i,j\in\mathcal{E}}$. Zdefiniujmy też nową macierz prawdopodobieństw przejścia $\boldsymbol{P}$:

$$p_{ij} = \begin{cases} q_{ij} \min(1, \pi_j/\pi_i), & \text{jeśli } i \neq j \\ 1 - \sum_{k\neq i} q_{ik} \min(1, \pi_k/\pi_i), & \text{jeśli } i = j. \end{cases} \tag{2.6}$$

Zostawiamy czytelnikowi sprawdzenie, że $\boldsymbol{P} = (p_{ij})$ jest macierzą przejścia. Odpowiadający jej łańcuch Markowa oznaczymy przez (Y_n). Następujące twierdzenie jest podstawą algorytmu Metropolisa.

Fakt 2.1 *Załóżmy, że $\boldsymbol{Q}$ jest regularną symetryczną macierzą prawdopodobieństw przejścia oraz $\boldsymbol{\pi}$ jest rozkładem na $\mathcal{E}$ z ściśle dodatnimi elementami. Wtedy $\boldsymbol{P}$ jest macierzą regularną z stacjonarnym rozkładem $\boldsymbol{\pi}$.*

Dowód Z wzoru (2.6) widzimy, że $q_{ij} > 0$ wtedy i tylko wtedy gdy $p_{ij} > 0$. Ponieważ $\boldsymbol{Q}$ jest regularna, więc $\boldsymbol{P}$ też musi być. Sprawdzimy teraz (1.4) z

$v_j = \pi_j$. Mamy

$$\pi_i p_{ij} = \pi_i q_{ij} \min(1, \pi_j/\pi_i) = q_{ij} \min(\pi_i, \pi_j) = q_{ji} \min(\pi_j, \pi_i) = \pi_j p_{ji}$$

□

Na podstawie tego twierdzenia podamy teraz algorytm Metropolisa na generowanie łańcucha Markowa $Y_0, Y_1, \ldots$ z rozkładem stacjonarnym $\boldsymbol{\pi}$, na podstawie łańcucha Markowa (X_j), który umiemy prosto generować.

0. $Y_k = i$

1. generuj Y zgodnie z rozkładem $\boldsymbol{q}_i = (q_{i1}, \ldots, q_{im})$.

2. podstaw $\alpha = \min(1, \pi_Y/\pi_i)$

3. losuj rand; jeśli rand$\leq \alpha$ to podstaw $Y_{k+1} = Y$,

4. w przeciwnym razie idź do 0..

Chcemy się pozbyć założenia symetryczności. Zastanówmy się najpierw nad ogólnym pomysłem. Jest to w pewnym sensie metoda podobna do metody eliminacji z rozdziału III. Możemy przepisać (2.6) w postaci

$$p_{ij} = \begin{cases} q_{ij}\alpha_{ij}, & \text{jeśli } i \neq j, \\ 1 - \sum_{k \neq i} q_{ik}\alpha_{ik}, & \text{jeśli } i = j. \end{cases}$$

gdzie $\alpha_{ij} = \min(1, \pi_j/\pi_i)$. Teraz możemy się spytać czy istnieją inne podstawienia dla α_{ij} takie, że

(I) $\boldsymbol{P}$ jest macierzą przejścia regularną,

(II) π jest rozkładem stacjonarnym dla $\boldsymbol{P}$,

(III) macierz $\boldsymbol{P}$ jest macierzą łańcucha odwracalnego, tj. spełniającego warunek DBC (1.4).

Zauważmy, że jesli $q_{ij} = \pi_j$, $\alpha_i = 1$ $(i, j \in \mathcal{E})$, to mamy niezależne replikacje (Y_j) o rozkładzie $\boldsymbol{\pi}$. Uogólnieniem metody Metropolisa niewymagającej symetryczności $\boldsymbol{Q}$ jest podane w następującym fakcie

Fakt 2.2 *Niech* $\boldsymbol{Q}$ *będzie regularną macierzą przejścia taką, że* $q_{ij} > 0$ *wtedy i tylko wtedy gdy* $q_{ji} > 0$. *Jeśli*

$$\alpha_{ij} = \min\left(1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right), \qquad i \neq j$$

to $\boldsymbol{P}$ *spełnia postulaty (I)-(III) powyżej.*

Dowód To, że $\boldsymbol{P}$ jest macierzą stochastyczną jest oczywiste. Teraz sprawdzimy, ze dla $v_j = \pi_j$ jest spełniony warunek (1.4). Mamy

$$\pi_i q_{ij} \min\left(1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right) = \min\left(\pi_i q_{ij}, \pi_j q_{ji}\right) = \pi_j q_{ji} \min\left(1, \frac{\pi_i q_{ij}}{\pi_j q_{ji}}\right).$$

□

2.2 Przykład: Algorytm Metropolisa na grafie G.

Rozważmy graf G z przykładu 1.8. Niech $N = \max_{i \in V} \deg(i)$ oraz $\mathcal{N}(i)$ jest zbiorem wszystkich sąsiadów wierzchołka $i \in V$. Zdefiniujmy na przestrzeni stanów V łańcuch Markowa (X_n) błądzący po "sąsiednich" wierzcholkach z macierzą przejścia $\boldsymbol{Q}$:

$$q_{ij} = \begin{cases} \frac{1}{M}, & j \in \mathcal{N}(i) \\ 0, & i \neq j \& j \neq \mathcal{N}(i) \\ 1 - \frac{\deg(i)}{M}, & i = j \end{cases},$$

gdzie $M \geq N$. Zakładamy, że łańcuch Markowa z taką macierzą przejścia jest nieprzywiedlny i nieokresowy. To założenie wynika ze struktury sąsiedztwa. Można też nietrudno pokazać, że jest odwracalny. Stąd tez już łatwo widać, że jego rozkład stacjonarny $\boldsymbol{\pi}$ jest jednostajny $\mathcal{U}(V)$.

Chcemy teraz na przestrzeni stanów V skonstruować łańcuch Markowa (Y_n) z rozkładem stacjonarnym $\pi_j = b_j/B$, gdzie $b_j > 0$ oraz $B = \sum_{j \in V} b_j$, natomiast bez trudności policzymy $_j/b_k$. W zastosowaniach często przestrzeń stanów V jest duża i nie ma możliwości obliczyć B. Jeśli jednak chcemy tylko generować zmienne o wartościach w V z rozkładem $\boldsymbol{\pi}$, to możemy skorzystać z następującego lematu, na podstawie którego zbudujemy algorytm.

Lemat 2.3 *Rozważmy graf G z przykładu 1.8. Niech $N = \max_{i\in V} \deg(i)$ oraz $\mathcal{N}(i)$ jest zbiorem wszystkich sąsiadów wierzchołka $i \in V$. Zdefiniujmy na przestrzeni stanów V łancuch Markowa błądzący po sąsiednich wierzcholkach:*

$$p_{ij} = \begin{cases} \frac{1}{M}\min(1, \pi_j/\pi_i), & j \in \mathcal{N}(i) \\ 0, & i \neq j \& j \neq \mathcal{N}(i) \\ 1 - \sum_{j\neq i} p_{ij}, & i = j \end{cases},$$

gdzie $M \geq N$. Jeśli $\boldsymbol{Q}$ jest regularna, to $\boldsymbol{P}$ też z rozkładem stacjonarnym $\boldsymbol{\pi}$.

Dowód Pokażemy, że DBC jest spełnione i stąd z Faktu 1.6 wynika teza. Dla $i \neq j$, jesli $\pi_i \leq \pi_j$, to $p_{ij} = 1$ oraz $p_{ji} = \pi_i/\pi_j$ Stąd mamy DBC. Podobnie analizujemy przypadek $\pi_i \geq \pi_j$. □

Zauważmy teraz, że $M = |V| \geq \max_{i\in V} \deg(i)$. Wykorzystamy to w następującym algorytmie.

Algorytm ITM

```
1. Łańcuch jest w stanie i
2. Losuj j z sąsiedztwa i z V z prawdopodobieństwem 1/M
3. Akceptuj z prawdopodobieństwem min(1,pi_j/pi_i), gdzie j jest
   proponowanym nowym stanem
4. jeśli zaakceptowane, to nowy stan jest j,
    przeciwnym razie idź do 1.
```

Rozważmy teraz graf $G = (V, E)$. Na tym grafie rozpatrujemy błądzenie przypadkowe jak w przykładzie 1.8. Niech $\boldsymbol{Q}$ będzie macierzą przejścia tego błądzenia. Interesuje nas rozkład Bolzmana

$$\pi_i(T) = \frac{e^{-H(i)/T}}{\sum_{j\in V} e^{-H(j)/T}}, \qquad i \in V.$$

Z uogólnionionego algorytmu Metropolisa (patrz fakt 2.2) aby generować rozkład $\pi_i(T)$ na V musimy położyć $p_{ij}(T) = q_{ij}\alpha_{ij}$, gdzie

$$\alpha_{ij} = \alpha_{ij}(T) = \frac{1}{\deg(i)} \min\left(1, \frac{\deg(i)}{\deg(j)} e^{(H(i)-H(j))/T}\right), \qquad i, j \in V$$

gdzie $H : V \to \mathbb{R}_+ \cup \infty$. Rozkładem stacjonarnym dla macierzy $\boldsymbol{P}(T)$ zdefiniowanej przez

$$p_{ij}(T) = q_{ij}\alpha_{ij}(T) \tag{2.7}$$

jest

$$\pi_i(T) = \frac{e^{-H(i)/T}}{\sum_{j\in V} e^{-H(j)/T}}.$$

3 Simulated annealing; zagadnienie komiwojażera

W rozdziale I wprowadziliśmy zagadnienie komiwojażera. Jest to problem optymizacyjny, który ogólnie można sformułować następująco. Zadaniem jest znalezienie globalnego minimum funkcji $H : V \to \mathbb{R}_+ \cup \infty$. Niech $D = \{j \in V : H(j) = \min_{i\in V} H(i)\}$. Pomysł na algorytm podpowiada następujący fakt.

Fakt 3.1

$$\lim_{T\to 0} \pi_i(T) = \left\{ \begin{array}{ll} \frac{1}{|D|} & i \in D \\ 0 & w\ przeciwnym\ \ razie \end{array} \right\}$$

Dowód Niech $i \in D$ oraz $a^* = H(i)/T$. Mamy

$$\begin{aligned} \pi_i(T) &= \frac{e^{-H(i)/T}}{\sum_{j\in D} e^{-H(j)/T} + \sum_{j\in\mathcal{E}-D} e^{-H(j)/T}} \\ &= \frac{1}{\sum_{j\in D} 1 + \sum_{j\in\mathcal{E}-D} e^{-H(j)/T+a}} \to \frac{1}{|D|}. \end{aligned}$$

□

Powyższy fakt daje podstawy heurystyczne dla następującego algorytmu. Będziemy schładzać według schematu $(a^{(1)}, a^{(2)}, \ldots, n_1, n_2, \ldots)$, gdzie $a^{(1)} > a^{(2)} > \ldots$ oraz $n_1 < n_2 < \ldots$. Oznacza to, że konstruujemy niejednorodny w czasie łańcuch Markowa $X_0, X_1, \ldots$. Na V mamy łańcuch Markowa ewoluujący według nieredukowlanej i nieokresowej macierzy przejścia $\boldsymbol{Q}$. Macierz przejścia $\boldsymbol{P}(a)$ zadana jest wzorem (2.7). Przejście z stanu $X_n = i$ do $X_{n+1} = j$ jest z prawdopdobieństwem $p_{ij}(a_n)$, gdzie $a_j = a^{(i)}$, jesli $n_i \le n < n_{n+1}$. A więc przez n_1 kroków przechodzimy zgodnie z macierzą przejścia $\boldsymbol{P}(a^{(1)})$, potem przez następnych $n_2 - n_1$ kroków zgodnie z macierzą przejścia $\boldsymbol{P}(a^{(2)})$, itd. Przyjmujemy, że minimum jest osiągnięte jednoznacznie w j^*, tj. $|D| = 1$ i $a^* = H(j^*)$.

Można podejść do problemu ogólnie. Niech T_n będzie temperaturą w momencie po n krokach. Hajek (1988) [3] pokazał, że warunkim konicznym i dostatecznym na zbiżność do j^* według prawdopodobieństwa potrzeba i wystarcza aby dla pewnej stałej γ

$$\sum_{n=1}^{\infty} e^{-\gamma/T_k} = \infty.$$

Teraz wracając do schematu $(a_1, a_2, \ldots, n_1, n_2, \ldots)$, warunek powyższy będzie spełniony, jeśli $a_k = 1/k$ oraz $n_k = e^{\beta k}$. Mianowicie wtedy mamy $T_k = \beta/\log(k+1)$ i $\beta \geq \gamma$.

Przykład 3.2 W przykładzie II.5.1 wprowadziliśmy tzw. zagadnienie komiwojażera. Dane jest n miast $1, \ldots, n$ oraz miasto wyjściowe oznaczone numerem 0. Drogą jest $\xi = (0, \boldsymbol{e}, 0)$, gdzie $\boldsymbol{e}$ jest permutacją $1, \ldots, n$. Zbiór wszystkich dróg ξ będzie zbiorem wierzchołków V. Teraz wprowadzamy strukturę sąsiedztwa. Mówimy, że

$$\xi = (0, e_1, \ldots, e_i, \ldots, e_j, \ldots, e_n, 0)$$

jest sąsiadem ξ' jeśli jesli dla pewnych $i < j$

$$\xi' = (0, e_1, \ldots, e_{i-1}, e_j, e_{j-1}, \ldots, e_i, \ldots, e_{j+1}, e_{j+2}, \ldots, e_n, 0).$$

A więc każdy ξ ma $n(n-1)/2$ sąsiadów. Macierz przejścia $\boldsymbol{Q}$ jest zdefiniowana przez błądzenie przypadkowe po sąsiadach jak w przykładzie 1.8. A więc

$$p_{\xi\xi'}(T_n) = \begin{cases} \frac{2}{n(n-1)} \min(e^{(H(\xi)-H(\xi'))/T_n}, 1), & \xi' \in \mathcal{N}(\xi) \\ 0, & \text{w przeciwnym razie.} \end{cases}.$$

3.1 Przykład: Losowanie dozwolonej konfiguracji w modelu z twardą skorupą

Rozważmy teraz spójny graf (V, E). Chcemy pomalować niektóre wierzchołki grafu na czarno, pozostawiając resztę białymi. W wyniku otrzymamy tzw. konfigurację czyli elementu z $\{\text{c,b}\}^V$. Jednakże interesuje nas model z twardą

[3] Za Bremaud

skorupą (hard core), który nie dopuszcza dwóch sąsiednich wierzcholków pomalowanych na czarno. Taką konfigurację nazywamy *dozwoloną*. Niech z_G będzie liczbą dozwolonych konfiguracji.

Na zbiorze wszystkich konfiguracji definiujemy rozkład

$$\pi_G(\xi) = \begin{cases} \frac{1}{z_G} & \text{jesli } \xi \text{ jest konfiguracja dozwoloną,} \\ 0 & \text{w przeciwnym razie.} \end{cases} \tag{3.8}$$

Element losowy o takim rozkładzie przyjmuje wartości w zbiorze Λ konfiguracji dozwolonych. Pytaniem jest jak losować taki element losowy. Problemem jest, że w przypadku skomplikowanych grafów z_G jest duże i trudne do obliczenia. W tym celu posłużymy się więc metodą MCMC. Skonstruujemy nieredukowalny i nieokresowy łańcuch Markowa (X_n), którego rozkład stacjonarny jest π_G. Łańcuch ten startuje z dowolnej konfiguracji dozwolonej. Ponieważ rozkład X_n zbiega do $\boldsymbol{\pi}$ więc będziemy przyjmować, że X_n ma rozkład π. Jak zwykle pytaniem jest jakie duże należy wziąć n żeby można było rozsądnie przybliżać rozkład $\boldsymbol{\pi}$ przez rozkład X_n.

Łańcuch X_n będzie ewoluował na zbiorze konfiguracji dozwolonych $\boldsymbol{\Lambda}$. Będziemy oznaczać przez $X_n(v)$ kolor wierzchołka v. Niech $p_{\xi,\xi'}$ oznacza prawdopodobieństwo przejścia od konfiguracji dozwolonej ξ do $\xi^{'}$. Algorytm zaczynamy od wyboru dozwolonej konfiguracji X_0. Przejście z X_n do X_{n+1} otrzymamy następująco.

1. Wylosuj "losowo" wierzchołek $v \in V$ i rzuć monetą.

2. Jeśli pojawi się "orzeł" oraz wylosowany został wierzchołek v, którego wszyscy sąsiedzi są "biali", to pomaluj go na czarno; w przeciwnym razie w $n+1$-kroku wierzchołek v jest biały.

3. Kolory wszystkich pozostałych wierzchołkow są niezmienione.

Niech $\xi \in \boldsymbol{\Lambda}$ będzie konfiguracją. Dla wierzchołka v oznaczamy przez $\xi(v)$ kolor wierzchołka v oraz przez $m(\xi)$ oznaczamy liczbę czarnych wierzchołkow. Strukturę sąsiedztwa na zbiorze konfiguracji dozwolonych $\boldsymbol{\Lambda}$wprowadzamy następująco. Mówimy, że $\xi^{'} \in \mathcal{N}(\xi)$ jeśli $m(\xi)+1 = m(\xi^{'})$; dla pewnego v mamy $\xi(v) = 0, \xi^{'}(v) = 1$ oraz $\xi(w) = \xi^{'}(w) = 0$ dla $w \sim v$ lub jeśli $m(\xi^{'})+1 = m(\xi)$; dla pewnego v mamy $\xi(v) = 1, \xi^{'}(v) = 0$. Funkcja

prawdopodobieństwa przejścia $p_{\xi\xi'}$ jest następująca:

$$p_{\xi\xi'} = \begin{cases} \frac{1}{2|V|} & \text{jeśli} \xi^{'} \in \mathcal{N}(\xi) \\ \frac{|V|-m(\xi)}{2|V|} & \text{jeśli } \xi^{'} = \xi \\ 0 & \text{jeśli } |m(\xi) - m(\xi^{'})| \geq 2 \end{cases}$$

Fakt 3.3 *Łańcuch Markowa z macierzą prawdopodobieństwa* $\boldsymbol{P}$ *jest nieredukowalny i nieokresowy z rozkładem stacjonarnym* π_G *zadanym w (3.8).*

3.2 Rozkłady Gibbsa

Rozpatrzmy skończony graf (V, E). Na V definiujemy konfiguracje jako funkcje $\lambda : V \to D$, gdzie D jest skonczonym zbiorem. Wszystke dozwolone konfiguracje oznaczymy przez Λ. Na przykład $D = \{\text{c,b}\}$ określa kolory wierzchołków odpowiednio biały i czarny. Teraz definiujemy funkcję H od konfiguracji λ (nazywana energią tej konfiguracji). Na przykład

$$H(\lambda) = \sum_{v \in V} U(\lambda(v)),$$

lub

$$H(\lambda) = \sum_{v' \in \mathcal{N}(v)} W(v, v^{'}) + \sum_{v \in V} U(\lambda(v)),$$

gdzie U, W są nieujemnymi funkcjami. Rozkład Gibbsa na zbiorze konfiguracji Λ jest zdefiniowany prze funkcję prawdopdobieństwa

$$p(\lambda) = Z(T) e^{-H(\lambda)/T},$$

gdzie $Z(T)$ jest tak zwaną *funkcją partycji*

$$Z(T) = \sum_{\lambda \in \Lambda} p(\lambda).$$

W zastosowaniach problemem jest trudność obliczenia $Z(T)$. Ale nawet nieznając $Z(T)$ możemy generować (w przybliżeniu) element losowy na Λ o szukanym rozkładzie Gibbsa przy użyciu algorytmu Metropolisa w następującej formie.

4 Uwagi bibliograficzne

O metodzie MCMC mozna przeczytac w książce Asmussena i Glynna [4], Bremaud [**?**], Häggström [17]. Metoda simulated anneling jest omówiona w pracy Bersimas i Tsitsiklis [6].

5 Zadania

[4]

Zadania teoretyczne

5.1 Napisać algorytm na generowanie błądzenia przypadkowego na hiperkostce $\{0,1\}^m$.

5.2 Niech $B = \sum_{i=1}^{\infty} i^{-2}$. Ta wielkość jest znana ($B = \pi^2/6$) ale to jest nam niepotrzebne. Zaprojektować przy użyciu metody Metropolisa symulację rozkładu $\pi_i = 1/Bi^2$, $i = 1, 2, \ldots$. Wsk. Sąsiedzi dla dowolnego $i > 1$ to $i-1, i+1$ oraz dla $i = 1$ sąsiadem jest tylko 2.

[4] 1.2.2015

Rozdział IX

Symulacja procesów matematyki finansowej

W tym rozdziale pokażemy wybrane przykłady zastosowań symulacji w matematyce finansowej. Klasyczna matematyka finansowa opiera sie na ruchu Browna (RB) z którego łatwo otrzymać geometryczny ruch Browna (GRB). Bardziej skomplikowane procesy są opisywane przez stochastyczne równania różniczkowe (SDE). Niesposób tu formalnie wyłożyć dość zaawansowaną współcześnie analizę stochastyczną, ale postaramy się w sposób przystępny opisać symulowane procesy (tutaj i w dodatku XI.2).

Zauważmy na początku, że chociaż realizacje badanych tutaj procesów są ciągłe, to czegoś takiego nie można wylosować na konmputerze. Jedynie co możemy zrobić, to zamiast wylosowania realizacji $X(t), 0 \leq t \leq T$, losujemy realizację $(X(0), X(t_1), \ldots, X(t_n))$ w momentach $0 = t_0, t_1, \ldots, t_n \leq T$. Będziemy przyjmowali, ze $t_1 < \ldots < t_n$. Rozważana procesy są procesami dyfuzyjnymi, a więc z ciagłymi realizacjami procesy Markowa. Dla nas to oznacza, że rozkład $X(t_j)$ pod warunkiem $X(t_{j-1}), \ldots, X_0$ zależy jedynie od $X(t_{j-1})$ $(j = 1, \ldots, n)$. A więc będziemy losowali rekurencyjnie, wiedząc, że

$$X(t_{j-1}) = y$$

losujemy $X(t_j)$. Do tego nam jest potrzebna znajomość funckji przejścia $p_t(x, dy)$ i algorytm jak generować zmienną losową o rozkładzie $p_t(x, dy)$ dla dowolnych ustalonych t i x. Niestety znajomośc funkcji $p_t(x, dy)$ jest wyjątkowa, nie mówiąc już o algorytmie generowania tego rozkładu. Dlatego będziemy też omawiać metody numeryczne dla stochastycznych równań

różniczkowych, w szczególności tzw. metodę Eulera. Jednakże w odróżnieniu do rozważań poprzednich w wyniku zastosowania metod numerycznych otrzymamy realizację aproksymacji procesu, a więc $\tilde{X}(t_1), \ldots, \tilde{X}(t_n)$. Jeśli chcemy mieć proces stochastyczny z ciągłymi realizacjami, możemy połączyć punkty $X(0), \tilde{X}(t_1), \ldots, \tilde{X}(t_n)$ w sposób ciągły, na przykład prostymi, i taka losowa łamana będzie przybliżała proces $X(t), 0 \leq t \leq T$. Zwróćmy uwagę, że mamy do czynienia z dwoma błędami; błąd spowodowany metodą Eulera oraz błąd dyskretyzacji. Zacznikimy od przykładu.

Przykład 0.1 Niech $\{S(t), 0 \leq t \leq T\}$ będzie ewolucją ceny akcji. Przez opcję azjatycką nazywamy możliwośc zakupu akcji za cenę K jeśli cena w chwili T spełnia warunek $S(T) \geq K$, w przeciwnym razie rezygnację z zakupu. Jest to tzw. opcja europejska *call* . Inną opcją jest tzw. azjatycka, gdzie zamiast wartości $S(T)$ bierzemy po owagę uśrednioną cenę $\in_0^T S(s)\, ds/T$. W takimi razie korzyść kupującego jest odpowiednio

$$(S(T) - K)_+) \qquad \text{lub} \qquad \left(\frac{\int_0^T X(s)\, ds}{T} - K \right)_+ .$$

Jeśli więc r jest bezryzykownym natężeniem stopy procentowej, to cena takich opcji byłaby

$$e^{-rT} \mathbb{E}\, (S(T) - K)_+) \qquad \text{lub} \qquad e^{-rT} \mathbb{E} \left(\frac{\int_0^T X(s)\, ds}{T} - K \right)_+ .$$

Przypuśćmy, że znamy proces $S(t)$, który opisuje ewolucje ceny akcji. Matematycznie jest to zazwyczaj opisane przez stochasyczne równanie różniczkowe.Ponieważ nie mamy możliwości symulacji realizacji $\{S(t), 0 \leq t \leq T\}$ więc symulujemy $\tilde{S}(0), \tilde{S}(T/N), \ldots, \tilde{S}(T)$. Dla opcji azjatyckiej będziemy obliczali jej cenę za pomocą

$$Y = e^{-rT} \left(\frac{\sum_{j=1}^N \tilde{S}(jT/N)}{T} - K \right)_+ .$$

A więc estymator zdefiniowany przez Y nie będzie nieobciążony, mimo, że czasami rozkład $\tilde{S}(0), \tilde{S}(T/N), \ldots, \tilde{S}(T)$ pokrywa się z rozkładem $S(0), S(T/N), \ldots, S(T)$ (o takich przypadkach będziemy dalej mówili).

Jednakże nie zawsze jesteśmy w stanie symulować rozkład dokładny, ze względu na to, że nie znamy tego rozkładu ani procedury jego generowania. W przypadku gdy proces $S(t)$ jest zadany przez stochastyczne równanie

różniczkowe, to istnieje procedura numeryczna, generowania aproksymacji $\tilde{S}(0), \tilde{S}(T/N), \ldots, \tilde{S}(T)$, który to wektor ma jedynie rozkład przybliżony do $S(0), S(T/N), \ldots, S(T)$. Będzie to metoda Eulera, Milsteina, omawiane w dalszej części tego rozdziału. S Możemy interpolować $\tilde{S}(0), \tilde{S}(T/N), \ldots, \tilde{S}(T)$ do procesu $\{\tilde{S}(t), 0 \leq t \leq T\}$ w sposób ciągły i następnie badać jak szybko dąży do zera błąd postaci

$$\mathbb{E} \int_0^T |\tilde{S}(t) - S(t)|^p$$

lub

$$\mathbb{E} \sup_{0 \leq t \leq T} |\tilde{S}(t) - S(t)|.$$

Innym błędem zbadanym teoretycznie jest

$$e_s(N) = \mathbb{E}\,|S(1) - \tilde{S}(1)|.$$

1 Ruch Browna

Ze względu na to, że ruch Browna ma przyrosty niezależne (patrz dodatek XI.2.1) realizację $B(t)$ $(0 \leq t \leq T)$ możemy symulować w punktach $t_0 = 0, t_1 = \Delta, 2\Delta, \ldots, t_n = T$, gdzie $\Delta = T/n$. Jednakże dla algorytmu który podamy nie jest istotne, że $\Delta = t_i - t_{i-1}$. Możemy wysymulować w punktach $0 < t_1 < \ldots < t_{n-1} < t_n = T$. Korzystając z tego, że

$$B(t_i) - B(t_{i-1}) =_{\mathrm{d}} \sqrt{t_i - t_{i-1}} Z_i,$$

gdzie $Z_1, \ldots, Z_n$ są iid standardowe normalne. Bardziej ogólny model to ruch Browna z dryfem $\mathrm{BM}(\mu, \sigma)$, gdzie *dryf* $\mu \in \mathbb{R}$ i *współczynnikem dyfuzji* $\sigma > 0$. Taki proces definiujemy przez

$$X(t) = \sigma B(t) + \mu t \tag{1.1}$$

gdzie $B(t)$ jest standardowym ruchem Browna. W języku stochastycznych równań różniczkowych proces X zdefiniowany w (1.1) możemy zapisać

$$dX(t) = \mu\, dt + \sigma\, dB(t). \tag{1.2}$$

Proces $X(t)$ ma dalej niezależne przyrosty (pokazać samemu!) a więc w punktach $0 < t_1 < \ldots < t_{n-1} < t_n = T$ wartości tego procesu mozna otrzymać rekurencyjnie, zadając $X(0) = x$ oraz

$$X(t_{i+1}) = X(t_i) + \mu(t_i - t_{i-1}) + \sigma\sqrt{t_i - t_{i-1}} Z_{i+1}$$

dla $i = 1, \ldots, n$, i $Z_1, \ldots, Z_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie N(0,1). Wynika to z tego, że $X(t_{i+1}) - X(t_i) = \mu(t_{i+1} - t_i) + \sigma(W(t_{i+1} - W(t_i)))$ ma rozkład normalny $\mathrm{N}(\mu(t_{i+1} - t_i), \sigma^2(t_{i+1} - t_i))$.

Zostawiamy czytelnikowi zaproponowanie algorytmu symulacji procesu zadanego

$$dX(t) = a(t)\,dt + \sigma(t)\,dW(t)$$

z warunkiem początkowym $X(0) = x$, gdzie $\sigma(t) > 0$.

1.1 Konstrukcja ruchu Browna via most Browna

Przedstawimy teraz inną konstrukcję ruchu Browna, korzystając z mostu Browna. Rozpatrzmy $(B(s), B(t), B(u))$, gdzie $0 \leq u < s < t$.

Rozkład $B(s)$ pod warunkiem $B(u) = x$ i $B(t) = y$ jest normalny z średnią

$$\frac{(t-s)x + (s-u)y}{(t-u)}$$

oraz wariancją

$$\frac{(s-u)(t-s)}{(t-u)}.$$

W szczególności $B(t+h)$ pod warunkiem $B(t) = x$ i $B(t+2h) = y$ ma średnią $(x+y)/2$ i wariancję $h/2$.

Będziemy zakładać, że $T = 1$. Naszym celem jest wygenerowanie

$$b_0^k, \ldots, b_{2^k-1}^k, b_{2^k}^k$$

mający łączny rozkład taki jak

$$(B(0), \ldots, B((2^k - 1)/2^k), B(1)).$$

Algorytm:

1. Generuj b_0^0, b_1^0, gdzie $b_0^0 = 0$ oraz $b_0^1 \sim \mathcal{N}(0,1)$.
2. Mając wygenerowane $b_j^{k-1}, (j = 1, \ldots, 2^{k-1})$, zauważamy, że $b_{2j}^k = b_j^{k-1}, (j = 1, \ldots, 2^{k-1})$, natomiast dla $j = 2j+1$, generujemy $b_i^k \sim \mathcal{N}(y, 2^{-k-1})$, gdzie $y = \frac{1}{2}(b_j^{k-1} + b_{j+1}^{k-1})$.

Reprezentację standardowego ruchu Browna. Niech Z_0, $Z_{i,j}; l = 1, 2, \ldots, j = 1, \ldots, 2^{l-1}$ iid $\sim N(0,1)$. Definiujemy $b^k(t)$ jako

$$\Delta_0(t)Z_0 + \sum_{l=1}^{k} 2^{-(l+1)/2} \sum_{i=1}^{2^{l-1}} \Delta_{l,i}(t) Z_{l,i}, \qquad t \in [0,1].$$

Zauważmy, że

$$(b^k(0), b^k(1/2^k), \ldots, b^k(1))$$

ma taki sam rozkład, jak

$$(B(0), \ldots, B((2^k-1)/2^k), B(1)).$$

Można pokazać, że dla $k \to \infty$ mamy zbieżność prawie wszędzie do procesu

$$\Delta_0(t)Z_0 + \sum_{l=1}^{\infty} 2^{-(l+1)/2} \sum_{i=1}^{2^{l-1}} \Delta_{l,i}(t) Z_{l,i}, \qquad t \in [0,1].$$

Ten proces ma ciągłe trajektorie i spełnia warunki ruchu Browna.

2 Geometryczny ruch Browna

Przypuśmy, że

$$X(t) = X(0) \exp(\sigma B(t) + \eta t)$$

gdzie $B(t)$ jest standardowym ruchem Browna. Stosując wzór Ito do $X(t) = f(B(t), t)$, gdzie $f(x,t) = \xi \exp(\sigma x + \eta t)$ mamy

$$\begin{aligned} dX(t) &= \\ &= \frac{\partial}{\partial t} f(B(t),t)\, dt + \frac{\partial}{\partial x} f(B(t),t)\, dB(t) + \frac{1}{2}\frac{\partial^2}{\partial x^2} f(B(t),t) \\ &= \eta\xi \exp(\sigma B(t) + \eta t) + \sigma\xi \exp(\sigma B(t) + \eta t) + \frac{1}{2}\sigma^2 \exp(\sigma B(t) + \eta t) \\ &= (\eta + \sigma^2/2) X(t)\, dt + \sigma X(t)\, dB(t). \end{aligned}$$

Podstawiając $\mu = \eta + \sigma^2/2$ widzimy, że stochastyczne równanie różniczkowe

$$dX(t) = \mu X(t)\, dt + \sigma X(t)\, dB(t)$$

z warunkiem początkowym $X(0) = \xi$ ma rozwiązanie

$$X(t) = \xi \exp((\mu - \sigma^2/2)t + \sigma B(t)).$$

Ten proces nazywamy *geometrycznym ruchem Browna* $X(t)$ i oznaczamy go przez GBM(μ, σ). Zauważmy, że dla $u < t$

$$\frac{X(t)}{X(u)} = \exp((\mu - \sigma^2/2)(t-u) + \sigma(B(t) - B(u)))$$

skąd widzimy, że wzory w rozłącznych odcinkach

$$\frac{X(t_1) - X(t_0)}{X(t_0)}, \frac{X(t_2) - X(t_1)}{X(t_1)}, \ldots, \frac{X(t_n) - X(t_{n-1})}{X(t_{n-1})}$$

są niezaleznymi zmiennymi losowymi. Ponieważ przyrosty $B(t)$ są niezależne i normalnie rozłożone, oraz $B(t) - B(u) =_{\mathrm{d}} \sqrt{t-u}Z$, gdzie $Z \sim$N(0,1), więc wartości X w punktach $0 = t_0 < t_1 < \ldots < t_n$ spełniają rekurencję

$$X(t_{i+1}) = X(t_i) \exp((\mu - \frac{1}{2}\sigma^2)(t_{i+1} - t_i) + \sigma\sqrt{t_{i+1} - t_i} Z_{i+1})$$

z $Z_1, \ldots, Z_n$ niezależne o jednakowym rozkładzie N(0,1).

2.1 Model Vasicka

Rozważmy teraz proces zadany stochastycznym równaniem różniczkowym

$$dX(t) = \alpha(b - X(t))\, dt + \sigma dB(t)$$

gdzie $\alpha, b, \sigma > 0$. To równanie ma rozwiązanie

$$X(t) = X(0)e^{-\alpha t} + b(1 - e^{-\alpha t}) + \sigma \int_0^t e^{-\alpha(t-s)}\, dB(s).$$

o czym można się przekonać różniczkując (wg wzoru Ito) to rozwiązanie. Można też pokazać, że jesli $0 < u < t$ to

$$X(t) = X(u)e^{-\alpha(t-u)} + b(1 - e^{-\alpha(t-u)}) + \sigma \int_u^t e^{-\alpha(t-s)}\, dB(s). \qquad (2.3)$$

[1] A zatem rozkład $X(t)$ pod warunkiem ,że $X(u) = \xi$, jest normalny $\mathrm{N}(\mu(u,t), \sigma^2(u,t))$, gdzie

$$\mu(u,t) = e^{-\alpha u} + b(1 - e^{-\alpha u})$$

i

$$\sigma^2(u,t) = \frac{\sigma^2}{2}(1 - e^{-2\alpha(t-u)}).$$

Aby to uzasadnić trzeba wiedzieć, że $\int_0^T k(x)\, dB(x)$ ma rozkład normalny ze średnią 0 i wariancją $\int_0^T k^2(x)\, dx$. Liczymy teraz wariancję

$$\begin{aligned}
&\mathrm{Var} \int_u^t e^{-\alpha(t-s)}\, dB(s) = \\
&\int_0^t e^{-2\alpha(t-s)}\, ds \\
&= \frac{1}{2\alpha}(1 - e^{-2\alpha(t-u)}).
\end{aligned}$$

Powyższe rozważania uzasadniają następujący algorytm dający dokładny rozkład procesu X w punktach $0 = t_0 < t_1 < \ldots < t_n = T$:

$$X(t_{i+1}) = e^{-\alpha(t_{i+1}-t_i)} X(t_i) + \mu(t_i, t_{i+1}) + \sigma(t_i, t_{i+1}) Z_{i+1},$$

gdzie $Z_1, \ldots, Z_n$ są niezależnymi zmiennymi losowymi o jednakowym rzokladzie N(0,1). Po podstawieniu funkcji $\mu(t_i, t_{i+1})$ i $\sigma^2(t_i, t_{i+1})$ i dokonaniu obliczeń otrzymujemy

$$\begin{aligned}
&X(t_{i+1}) \\
&= e^{-\alpha(t_{i+1}-t_i)} X(t_i) + b(1 - e^{-\alpha(t_{i+1}-t_i)}) + \sigma \sqrt{\frac{1}{2\alpha}(1 - e^{-\alpha(t_{i+1}-t_i)})}\, Z_i
\end{aligned} \tag{4}$$

[1]

$$\begin{aligned}
X(t) &= X(u)e^{-\alpha(t-u)} + b(1 - e^{-\alpha(t-u)}) + \sigma \int_u^t e^{-\alpha s}\, dB(t) \\
&= (X(0)e^{-\alpha u} + b(1 - e^{-\alpha u}) + \sigma \int_0^u e^{-\alpha(u-s)}\, dB(s))e^{-\alpha(t-u)} \\
&\quad + b(1 - e^{-\alpha(t-u)}) + \sigma \int_u^t e^{-\alpha s}\, dB(t)
\end{aligned}$$

Ponieważ naszym celem jest symulacja realizacji procesu X a więc gdy $\max_i(t_{i+1}-t_i)$ jest małe, więc możemy skorzystać z aproksymacji

$$e^{-\alpha(t_{i+1}-t_i)} \approx 1 - \alpha(t_{i+1} - t_i).$$

Po dokonaniu uproszczeń otrzymamy

$$X(t_{i+1}) = \alpha(t_{i+1} - t_i)X(t_i) + \alpha(t_{i+1} - t_i))(b - X(t_i)) + \sigma\sqrt{t_{i+1} - t_i}Z_{i+1}. \quad (2.5)$$

Jak zobaczymy dalej powyższy wzór będzie szczególnym przypadkiem schematu Eulera.

2.2 Schemat Eulera

Formalnie możemy zapisać ogólne równanie stochastyczne jako

$$dX(t) = a(X(t))\,dt + b(X(t))\,dB(t).$$

Pozostawiamy na boku ważne pytania o istnienie rozwiązań i jednoznaczność. Będziemy milcząco zakładać pozytywną odpowiedź na te pytania. Wtedy aproksymacje $\hat{X(t_i)}$ wartości $X(t_i)$ otrzymamy przez $\hat{X}(0) = X(0)$ i dla $0 = t_0 < t_1 < \ldots < t_n = T$

$$\hat{X}(t_{i+1}) = \hat{X}(t_i) + a(\hat{X}(t_i))(t_{i+1} - t_i) + b(\hat{X}(t_i))\sqrt{t_{i+1} - t_i}Z_{i+1},$$

gdzie $Z_1, \ldots, Z_n$ są niezależnymi zmiennymi losowymi o jednakowym rzokladzie N(0,1).

Przykład 2.1 [Proces Coxa Ingersolla Rossa (CIR)] Proces w modelu Vasicka posiada podstawową wadę, że przyjmuje wartości ujemne. Tej wady nie posiada proces pierwiastkowej dyfuzji opisanej stochastycznym równaniem różniczkowym

$$dX9t) = \alpha(b - X(t))\,dt + \sigma\sqrt{X(t)}\,dB(t).$$

Dowodzi się, ze jesli $X(0) > 0$, to $X(t) \geq 0$ dla wszystkich $t > 0$ oraz jeśli ponadto $2\alpha b > \sigma^2$ to $X(t) > 0$ dla wszystkich $t > 0$. Zastosowanie schematu Eulera prowadzi to rekursji:

$$\hat{X}(t_{i+1}) = X(t_i) + \alpha(b - X(t_i))(t_{i+1} - t_i) + \sigma\sqrt{(X(t_i))_+}\sqrt{t_{i+1} - t_i}Z_{i+1},$$

gdzie $Z_1, \ldots, Z_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie N(0,1), i $(a)_+ = \max(0ma)$.

3 Uwagi bibliograficzne

Podstawowym źródłem informacji o symulacji procesów stochastycznych teorii matematyki finansowej jest książka Glassermana [16]. Inną podstawową monografią omawiającą symulacje procesów stochastycznych jest [4].

Zadania teoretyczne

3.1 Dla standardowego ruchu Browna $\{W(t), 0 \le t \le 1\}$ definiujemy proces

$$\{\tilde{W}^h(t), 0 \le t \le 1\}$$

który jest zgodny z procesem W w punktach $W(h), W(2h), \ldots, W(1)$ gdzie $h = 1/N$ i interpolowany liniowo do realizacji ciągłych. Pokazać, że

$$\mathbb{E} \int_0^1 |\tilde{W}^h(t) - W(t)| = c/N^{1/2},$$

gdzie $c = \sqrt{\pi/32}$.

3.2 Niech $X(t) = x + int_0^t a(t)\,dt + \int_0^t b(t)\,dW(t), \qquad 0 \le t \le T$. Pokazać, że jeśli $0 < t_1 < t_2 < \ldots < t_n$ można symulować rekurencyjnie, zadając $X(0) = x$ oraz

$$X(t_{i+1}) = X(t_i) + \int_{t_i}^{t_{i+1}} a(s)\,ds(t_i - t_{i-1}) + \sqrt{\int_{t_i}^{t_{i+1}} b^2(s)\,ds} Z_{i+1}$$

dla $i = 1, \ldots, n$, i $Z_1, \ldots, Z_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie N(0,1).

Projekt

3.3 Według Asmussen&Glynn (2007). Zaprojektować obliczenie p metodą Monte Carlo, gdzie p jest intensywnością spłaty kredytu w następującym modelu.

Udzielono kredytu w wysokości k na okres T lat, który będzie spłacany ze stałą intensywnością p. Krótkoterminowa stopa kredytowa jest $r(t)$, $0 \leq t \leq T$. Jeśli $s(t)$ oznacza kwotę spłaconą do chwili t; $s(0) = 0$ raz $q(t)$ oznacza ilość pieniędzy gdyby były trzymane na rachunku, to ignorując pozostałe koszty (administracja, podatki, zysk, itd) wielkość p wyznaczamy z równości $\mathbb{E}\, s(T) = \mathbb{E}\, q(T)$.
a) Uzasadnić, że

$$\begin{aligned} ds(t) &= (p + s(t)r(t))dt \\ dq(t) &= q(t)r(t) \end{aligned}$$

b) Jeśli przez $s_p(t)$ oznaczymy proces s obliczony przy intensywności spłaty p oraz $q_k(t)$ oznaczymy proces obliczony przy wartości początkowej $q(0) = k$, to pokazać, że

$$p = \frac{k\mathbb{E}\, q_1(T)}{\mathbb{E}\, s_1(T)}.$$

c) Przyjąć, że $r(t)$ jest procesem CIR. W pierwszym wariancie założyć, że jest on stacjonarny, ze średnią równą długoterminowej stopie procentowej c. Do obliczeń przyjąć, że długotermionowa stopa procentowa $c = 7\%$, $T = 5$, oraz dobrać α, σ^2 tak aby realizacje przypominały te z wykresu ...
d) Porównać z sytuacją, gdy $r(t) \equiv c$.

Rozdział X

Sumulacja zdarzeń rzadkich

1 Efektywne metody symulacji rzadkich zdarzeń

Niech będzie dana przestrzeń probabilistyczna $(\Omega, \mathcal{F}, \mathbb{P})$.

Rozważmy rodzinę zdarzeń $(A(x))$ poindeksowaną albo przez $x \in (0, \infty)$ lub $x \in \{1, 2, \ldots\}$ i niech

$$I(x) = \mathbb{P}(A(x)).$$

Mówmy, że $(A(x))$ jest *rodziną zdarzeń rzadkich* jeśli

$$\lim_{x \to \infty} I(x) = 0 \, .$$

Zauważmy, że dla bardzo małych $I(x)$ trzeba raczej kontrolować błąd względny niż bezwględny. Przez $Y(x)$ oznaczamy zmienną losową taką, że $I(x) = \mathbb{E}\, Y(x)$. Wtedy

$$\hat{Y}_n = \frac{1}{n} \sum_{j=1}^{n} Y_j(x),$$

gdzie $Y_1(x), \ldots, Y_n(x)$ są niezależnymi kopiami $Y(x)$, jest estymatorem nieobciążonym $I(x)$. A więc na poziomie istotności α (jak zwykle przyjmujemy $\alpha = 0.05$)

$$\begin{aligned} 1 - \alpha \sim \mathbb{P}(\hat{Y}_n(x) - 1.96 \frac{\sigma(x)}{\sqrt{n}} \leq I(x) \leq \hat{Y}_n(x) + 1.96 \frac{\sigma(x)}{\sqrt{n}}) \\ = \quad \mathbb{P}\left(-\frac{1.96}{\sqrt{n}} \frac{\sigma(x)}{I(x)} \leq \frac{I(x) - \hat{Y}_n(x)}{I(x)} \leq \frac{1.96}{\sqrt{n}} \frac{\sigma(x)}{I(x)} \right) \, . \end{aligned}$$

gdzie $\sigma^2(x) = \operatorname{Var} Y(x)$. Stąd widzimy, że aby kontrolować błąd względny musimy rozważać wielkość $\sigma^2(x)/I^2(x)$, gdzy $x \to \infty$. Mianowicie jeśli na przykład chcemy kontrolować błąd względny na poziomie b_r, to wtedy z równości

$$b_r = \frac{1.96}{\sqrt{n}} \frac{\sigma(I)}{I}$$

skąd mamy

$$n = \frac{1.96^2}{b_r} \frac{\sigma^2(I)}{I^2}.$$

Potrzebę bardziej wysublimowanej teorii przy symulacji zdarzeń rzadkich pokazuje następujący przykład.

Przykład 1.1 Niech $A(x) \in \mathcal{F}$ i $I(x) = \mathbb{P}(A(x))$ jest bardzo małe. Powtarzamy niezależnie zdarzenia $A_1(x), \ldots, A_n(x)$. Rozważmy estymator CMC dla $I(x)$ postaci

$$\hat{Y}_n = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{A_j(x)} \ .$$

Wtedy $\operatorname{Var}(\mathbf{1}_{A_j(x)})/I^2(x) = I(x)(1-I(x))/I^2(x) \sim I(x)^{-1}$. Jeśli więc chcemy mieć błąd względny nie większy niż b_r, to musi być spełniona równość

$$\frac{1.96}{\sqrt{n}} I(x)^{-1} = b_r$$

czyli

$$n = \frac{1.96^2}{b_r^2} I(x)^{-1} \ .$$

A więc dla bardzo małych $I(x)$ duża liczba replikacji jest wymagana. Problemem jest jak szybko zbiega $I(x)$ do zera.

Stąd mamy następującą naturalną definicję dla klasy estymatorów zdefiniowanyhch przez $Y(x)$.

Definicja 1.2 Mówimy, że estymator $\hat{Y}_n(x)$ ma *ograniczony błąd względny* jeśli

$$\limsup_{x \to \infty} \frac{\operatorname{Var}(Y(x))}{I^2(x)} < \infty.$$

Bardzo często estymatory z ograniczonym błędem względnym trudno znaleźć, lub trudno tą własność udowodnić. Dlatego wprowadza się trochę słabsze wymaganie, które jednakże często charakteryzuje zupełnie dobre estymatory.

Definicja 1.3 Mówimy, że estymator $\hat{Y}$ *logarytmicznie efektywny* jeśli

$$\liminf_{x\to\infty} \frac{\log \mathrm{Var}\,(Y(x))}{2\log I(x))} \geq 1\ .$$

Nastepny fakt pokazuje równoważne wysłowienie tej własności, z której natychmiast będzie widać, że własność logarytmicznej efektywności jest słabsza od własności ograniczonego błedu.

Fakt 1.4 *Własność logarytmicznej efektywności jest równoważna temu, że dla* $0 < \epsilon < 2$ *mamy*

$$\limsup_{x\to\infty} \frac{\mathrm{Var}\,(Y(x))}{I^{2-\epsilon}(x)} < \infty. \tag{1.1}$$

Do dowodu przyda się lemat.

Lemat 1.5 *Niech* $a(x) \to -\infty$. *Mamy* $\limsup a(x)b(x) < \infty$ *wtedy i tlko wtedy gdy* $\liminf b(x) \geq 0$.

Dowód Mamy (1.1) wtedy i tylko wtedy gdy

$$\limsup_{x\to\infty} \log\left(\frac{\mathrm{Var}\,(Y(x))}{I^{2-\epsilon}(x)}\right) < \infty.$$

To znaczy

$$\limsup_{x\to\infty} \log I(x)(\frac{\log \mathrm{Var}\,(Y(x))}{2\log I(x)} - (1-\frac{\epsilon}{2})) < \infty.$$

Teraz należy skorzystać z lematu 1.5 a $a(x) = \log I(x)$ i $b(x) = \log \mathrm{Var}\,(Y(x))/(2\log I(x)) - (1-\frac{\epsilon}{2})$.

2 Dwa przykłady efektywnych metod symulacji rzadkich zdarzeń

Niech $\xi_1, \xi_2, \ldots$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie z gęstością $f(x)$ i oznaczmy $S_n = \xi_1 + \ldots + \xi_n$. Celem naszym będzie obliczenie $I(x) = \mathbb{P}(S_n > (\mu+\epsilon)n)$, gdzie $\mu = \mathbb{E}\,\xi_1$ lub $I(x) = \mathbb{P}(S_n > x)$. Okaże się, że efektywne algorytmy zależą od ciężkości ogona rozkładu ξ_1. W przypadku lekko ogonowym przydatna będzie technika zamiany miary, którą trochę dokładniej omówimy poniżej.

2.1 Technika wykładniczej zamiany miary

Zacznijmy od pojedynczej kratowej zmiennej losowej ξ określonej na przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$ z funkcją prawdopodobieństwa $\{p_k\}$. Niech $\hat{m}(\theta) = \sum_k e^{\theta k} p_k$ będzie funkcją tworzącą momenty. Będziemy zakładać, że $\hat{m}(\theta)$ jest skończona w pewnym przedziale otwartym (θ_1, θ_2) takim, że $\theta_0 < 0 < \theta_1$. Dla tych wartości $\theta \in (\theta_0, \theta_1)$ definiujemy nową funkcję prawdopodobieństwa przez

$$\tilde{p}_k^\theta = \frac{e^{\theta k} p_k}{\hat{m}(\theta)}.$$

Można zawsze tak dobrać $(\Omega, \mathcal{F}, \mathbb{P})$ i zmienną losową ξ tak, że $\mathbb{P}(\xi = k) = p_k$ i dla pewnej innej miary probabilistycznej $\tilde{\mathbb{P}}^\theta$

$$\tilde{\mathbb{P}}^\theta(\xi = k) = p_k^\theta.$$

Teraz rozpatrzmy ciąg niezależnych zmiennych losowych $\xi_1, \ldots, \xi_n$ na $(\Omega, \mathcal{F}, \mathbb{P})$ o jednakowym rozkładzie takim jak ξ. Natomiast z nową miarą probabilistyczną $\tilde{\mathbb{P}}^\theta$ definiujemy następująco:

$$\begin{aligned} &\tilde{\mathbb{P}}^\theta(\xi_1 = k_1, \ldots, \xi_n = k_n) \\ &= \frac{e^{\theta(\theta(k_1+\ldots+k_n)}}{\hat{m}(\theta)} \mathbb{P}(\xi_1 = k_1, \ldots, \xi_n = k_n) \\ &= \tilde{p}_{k_1}^\theta \ldots \tilde{p}_{k_n}^\theta, \end{aligned}$$

czyli przy nowej mierze $\xi_1, \ldots, \xi_n$ są dalej niezależne. Oznaczając przez $\kappa(\theta) = \log \hat{m}(\theta)$ i podstawiając $S_n = \xi_1 + \ldots + \xi_n$ mamy podstawowy związek

$$\begin{aligned} &\mathbb{P}(\xi_1 = k_1, \ldots, \xi_n = k_n) = \\ &= \qquad e^{-\theta S_n + n\kappa(\theta)} \tilde{\mathbb{P}}^\theta(\xi_1 = k+1, \ldots, \xi_n = k_n). \end{aligned} \tag{2.2}$$

Niech teraz

$$A = \{(\xi_1, \ldots, \xi_n) \in B\},$$

gdzie $B \subset \mathbb{Z}^n$ i $\tilde{\mathbb{E}}^\theta$ oznacza wartość oczekiwaną zdefiniowaną przez $\tilde{\mathbb{P}}^\theta$. Wtedy mamy następujący lemat, który jest prawdziwy niezależnie od typu rozkładu.

Lemat 2.1

$$\mathbb{P}(A) = \hat{m}^n(\theta) \tilde{\mathbb{E}}^\theta[e^{-\theta(\xi_1+\ldots+\xi_n)}; A] = \tilde{\mathbb{E}}^\theta[e^{n\kappa(\theta)-\theta S_n)}; A]. \tag{2.3}$$

Dowód Udowodnimy lemat gdy ξ_1 ma rozkład kratowy. Korzystając z (2.2) mamy

$$\begin{aligned} \mathbb{P}(A) &= \hat{m}(\theta) \sum_{(k_1,\ldots,k_n)\in B} e^{-\theta(k_1+\ldots k_n)} \tilde{\mathbb{P}}^{\theta}(\xi_1 = k_1, \ldots, \xi_n = k_n) \\ &= \hat{m}^n(\theta) \tilde{\mathbb{E}}^{\theta}[e^{-\theta(\xi_1+\ldots+\xi_n)}; A]. \end{aligned} \tag{2.4}$$

□

Lemat 2.2 *Dla zmiennej losowej ξ_1 określonej na przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \tilde{\mathbb{P}}^{\theta})$ funkcja tworząca momenty*

$$\tilde{\mathbb{E}}^{\theta} e^{s\xi_1} = \hat{m}(s+\theta).$$

Stąd

$$\tilde{\mathbb{E}}^{\theta} \xi_1 = \frac{d\hat{m}(\theta+s)}{ds}_{|s=0} = \kappa'(\theta).$$

2.2 $\mathbb{P}(S_n > (\mu+\epsilon)n)$ - przypadek lekko-ogonowy

Będziemy obliczać

$$I(m) = \mathbb{P}(S_m > m(\mu+\epsilon)),$$

gdzie $\mathbb{E}\,\xi_1 = \mu$ i $\epsilon > 0$. Podstawiając $A = A(m) = \{S_m > m(\mu+\epsilon)\}$ w (2.4) mamy, że

$$Y(m) = e^{-\theta S_m + m\kappa(\theta)} \mathbf{1}_{A(m)}$$

przy mierze $\tilde{\mathbb{P}}^{\theta}$ definiuje nieobciążony estymator $\hat{Y}_n(m)$. Niech θ_0 spełnia

$$\mathbb{E}^{\theta_0} \xi_1 = \kappa'(\theta_0) = \mu + \epsilon.$$

Ponieważ κ jest funkcją wypukła (własność funkcji tworzących momenty) więc rozwiązanie θ_0 musi być ściśle dodatnie. Nietrywialnym rezultatem jest następujące twierdzenie.

Twierdzenie 2.3 *Estymator zdefiniowany przez*

$$Y(m) = e^{-\theta S_m + m\kappa(\theta)} \mathbf{1}_{A(m)}$$

przy nowe mierze $\tilde{\mathbb{P}}^{\theta_0}$ jest logarytmicznie efektywny.

Zauważmy ponadto, że optymalny estymator jest typu istotnościowego i można pokazać, że w tej klasie jest on jedyny.

2.3 $\mathbb{P}(S_n > x)$ - przypadek ciężko-ogonowy

Jak zwykle przez $\overline{F}(x) = 1 - F(x)$ oznacza ogon dystrybuanty ξ_1. Będziemy rozpatrywać dwa przypadki: Pareto z $\overline{F}(x) = 1/(1+x)^\alpha$ lub Weibull $\overline{F}(x) = e^{-x^r}$ z parametrem $r < 1$. Teoria idzie bez zmian jeśli zamiast Pareto będziemy mieli rozkład Pareto-podobny; patrz definicja w [39]. Niech $A(x) = \{S_n > x\}$. Tym razem n jest ustalone, natomiast parametrem jest $x \to \infty$. Niech

$$I(x) = \mathbb{P}(S_n > x).$$

W dalszej częsci przez $\xi_{(1)} < \xi_{(2)} < \ldots < \xi_{(n)}$ oznaczamy statystykę pozycyjną i $S_n = \xi_1 + \ldots \xi_n$, $S_{(n)} = \xi_{(1)} + \ldots \xi_{(n)}$ oraz $M_n = \xi_{(n)} = \max(\xi_1, \ldots, \xi_n)$. Zdefiniujmy cztery estymatory $\hat{Y}_n^i$ $(i = 1, \ldots, 4)$ oparte na

$$\begin{aligned} Y^1(x) &= \mathbf{1}_{A(x)} \\ Y^2(x) &= \bar{F}(x - S_{n-1}) \\ Y^3(x) &= \frac{\overline{F}((x - S_{(n-1)}) \vee \xi_{(n-1)})}{\overline{F}(\xi_{(n-1)})} \\ Y^4(x) &= n\overline{F}(M_{n-1} \vee (x - S_{n-1})). \end{aligned}$$

Estymatory oparte na $Y_i(x)$ $(i = 1, ..., 4)$ są nieobciążonymi estymatorami ponieważ

$$\begin{aligned} \bar{F}(x - S_{n-1}) &= \mathbb{P}(S_n > x | S_{n-1}) \\ \frac{\overline{F}((x - S_{(n-1)}) \vee \xi_{(n-1)})}{\overline{F}(\xi_{(n-1)})} &= \mathbb{P}(S_n > x | \xi_{(1)}, \ldots, \xi_{(n-1)}) \\ n\overline{F}(M_{(n-1)} \vee (x - S_{(n-1)}) &= n\mathbb{P}(S_n > x, M_n = \xi_n | \xi_1, \ldots, \xi_n). \end{aligned}$$

Twierdzenie 2.4 *W przypadku gdy rozkład ξ jest Pareto, to estymator $\hat{Y}_n^3(x)$ jest logarytmicznie efektywnym a estymator $\hat{Y}_n^4(x)$ ma ograniczony błąd względny, natomiast gdy ξ_1 jest Weibullowski z $r < \log(3/2)/\log 2$ to estymator $\hat{Y}_n^4(x)$ estymator jest logarytmicznie efektywny.*

2.4 Uwagi bibliograficzne

Kolekcja prac: Rare Event Simulation using Monte Carlo Methods. Gerardo Rubino i Bruno Tuffin, Editors. Wiley, 2009.

S.Juneja and P.Shahabuddin. Rare event simulation techniques: An introduction and recent advances. In S.G. Henderson and B.L. Nelson, eds, Simulation, Handbooks in Operations Research and Managemant Science, pp. 291–350. Elsevier, Amsterdam, 2006.

Podrozdziały 2.3 i 2.2 zostały opracowane na podstawie książki Asmussena i Glynna [4], do której odsyłamy po dowody twierdzeń i dalsze referencje.

Rozdział XI

Dodatek

1 Elementy rachunku prawdopdobieństwa

1.1 Rozkłady zmiennych losowych

W teorii prawdopodobieństwa wprowadza się zmienną losową X jako funkcję zdefiniowaną na przestrzeni zdarzeń elementarnych Ω. W wielu przypadka, ani Ω and X ni są znane. Ale co jest ważne, znany jest rozkład zmiennej losowej X, tj. przepis jak liczyć $F(B)$, że X przyjmie wartość z zbioru B, gdzie B jest podzbiorem (borelowskim) prostej $\mathbb{R}$.

Ze względu na matematyczną wygodę rozważa się rodzinę $\mathcal{F}$ podzbiorów Ω, zwaną rodziną zdarzeń, oraz prawdopodobieństwo $\mathbb{P}$ na $\mathcal{F}$ przypisujące zdarzeniu A z $\mathcal{F}$ liczbę $\mathbb{P}(A)$ z odcinka $[0,1]$.

Rozkład można opisać przez podanie *dystrybuanty*, tj. funkcji $F : \mathbb{R} \to [0,1]$ takiej, że $F(x) = \mathbb{P}(X \leq x)$. Prze $\overline{F}(x) = 1 - F(x)$ będziemy oznaczać *ogon* of F.

W praktyce mamy dwa typy zmiennych losowych: dyskretne i absolutnie ciągłe. Mówimy, że X jest *dyskretną* jeśli istnieje przeliczalny zbiór liczb $E = \{x_0, x_1, \ldots\}$ z $\mathbb{R}$ taki, że $\mathbb{P}(X \in E) = 1$. W tym przypadku definiujemy *funkcję prawdopodobieństwa* $p : E \to [0,1]$ przez $p(x_k) = \mathbb{P}(X = x_k)$; para (E, p) podaje pełny przepis jak liczyć prawdopopodobieństwa $\mathbb{P}(X \in B)$ a więc zadaje rozkład. Najważniejszą podklasą są zmienne losowe przyjmujące wartości w E który jest podzbiorem kraty $\{hk,\ k \in \mathbb{Z}_+\}$ dla pewnego $h > 0$. Jeśli nie będzie powiedziane inaczej to zakładamy, że $h = 1$, tzn $E \subset \mathbb{Z}_+$, i.e. $x_k = k$. Wtedy piszemy prosto $p(x_k) = p_k$ i mówimy, że rozkład jest *kratowy*.

Z drugiej strony mówimy, że X jest *absolutnie ciągłą* jeśli istnieje funkcja $f : \mathbb{R} \to \mathbb{R}_+$ taka, że $\int f(x)\,dx = 1$ i $\mathbb{P}(X \in B) = \int_B f(x)\,dx$ dla każdego $B \in \mathcal{B}(\mathbb{R})$. Mówimy wtedy, że $f(x)$ jest *gęstością* zmiennej X.

Rozkład dyskretnej zmiennej losowej nazywamy rozkładem dyskretnym. Jeśli X jest kratowy, to jego rozkład też nazywamy kratowy. Analogoicznie rozkład zmiennej losowej absolutnie ciągłej nazywamy rozkładem absolutnie ciągłym. Czasami dystrybuanta nie jest ani dyskretna ani absolutnie ciągła. Przykładem jest *mieszanka* $F = \theta F_1 + (1-\theta)F_2$, gdzie F_1 jest dyskretna z prawdopodobieństwem θ i F_2 jest absolutnie ciągła z prawdopodobieństwem $1-\theta$.

Aby zaznaczyć, że dystrybuanta, gęstość, funkcja prawdopodbieństwa, jest dla zmiennej losowej X, piszemy czasem $F_X, p_X, f_X, \ldots$.

Dla dwóch zmiennych losowych X i Y o tym samym rozkładzie piszemy $X \stackrel{\mathcal{D}}{=} Y$. To samo stosujemy do wektorów.

Mówoimy, że wektor losowy $(X_1, \ldots, X_n)$ jest *absolutnie ciągły* jeśli istnieje nieujemna i całkowalna funkcja $f : \mathbb{R}^n \to \mathbb{R}_+$ z

$$\int_{\mathbf{R}} \ldots \int_{\mathbf{R}} f(x_1, \ldots, x_n)\,dx_1 \ldots dx_n = 1$$

taka, że $\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B) = \int_{B_1} \ldots \int_{B_n} f(x_1, \ldots, x_n)\,dx_n \ldots dx_1$ dla wszystkich zbiorów borelowskich $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$. Funkcja $f(x)$ nazywa się *gęstością* $\boldsymbol{X} = (X_1, \ldots, X_n)^T$. Przytoczymy teraz pożyteczny rezultat dotyczący wzoru na gęstość przekształconego wektora losowego. Mianowicie niech będą dane funkcje $g_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, n$, które spełniają następujące warunki:

1. V jest otwartym pozdbiorem $\mathbb{R}^n$, takim że $\mathbb{P}(\boldsymbol{X} \in V) = 1$,
2. przekształcenie $\boldsymbol{g} = (g_1, \ldots, g_n)^T$ jest wzajemnie jednoznaczne na zbiorze A,
3. przekształcenie $\boldsymbol{g}$ ma na V ciągłe pierwsze pochodne cząstkowe ze względu na wszystkie zmienne,
4. jakobian $J_{\boldsymbol{g}}(\boldsymbol{x}) = \frac{\partial(g_1,\ldots,g_n)}{\partial(x_1,\ldots,x_n)}$ jest różny od zera na A.

Oznaczmy przez $\boldsymbol{h} = (h_1, \ldots, h_n)^T$ przekształcenie odwrotne do $\boldsymbol{g}$, tzn. $\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{y}) = (h_1(\boldsymbol{y}), \ldots, h_n(\boldsymbol{y}))$ dla każdego $\boldsymbol{y} \in \boldsymbol{g}(V)$. Rozważmy wektor $\boldsymbol{Y} = (\boldsymbol{X})$. Wektor ten ma gęstość $f_{\boldsymbol{Y}}$ zadaną wzorem:

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(h_1(y), \ldots, h_n(\boldsymbol{y}))|J_h(\boldsymbol{y})|\mathbf{1}(\boldsymbol{y} \in \boldsymbol{g}(V)). \tag{1.1}$$

Jeśli zmienne losowe $(X_1, \ldots, X_{n-1})$ $(X_1, \ldots, X_n)$ są absolutnie ciągłe z odpowiednio z gęstościami $f_{X_1,\ldots,X_{n-1}}$ i $f_{X_1,\ldots,X_n}$ i jeśli $f_{X_1,\ldots,X_{n-1}}(x_1, \ldots, x_{n-1}) > 0$, to

$$f_{X_n|X_1,\ldots,X_{n-1}}(x_n \mid x_1, \ldots, x_{n-1}) = \frac{f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)}{f_{X_1,\ldots,X_{n-1}}(x_1, \ldots, x_{n-1})}$$

nazywa się i *warunkową gęstością* X_n pod warunkiem $X_1 = x_1, \ldots, X_{n-1} = x_{n-1}$.

Dla ustalonego n i wszystkich $k = 1, 2, \ldots, n$ i $\omega \in \Omega$, niech

$X_{(k)}(\omega)$ oznacza k-tą najmniejszą wartość $X_1(\omega), \ldots, X_n(\omega)$. Składowe wektora losowego $(X_{(1)}, \ldots, X_{(n)})$ nazywają się *statystyka porządkową* $(X_1, \ldots, X_n)$.

1.2 Podstawowe charakerystyki

Niech X będzie zmienną losową. i $g : \mathbb{R} \to \mathbb{R}$ funkcją. Definiujemy $\mathbb{E}\, g(X)$ przez

$$\mathbb{E}\, g(X) = \begin{cases} \sum_k g(x_k)p(x_k) & \text{jeśli } X \text{ jest dyskretny} \\ \int_{-\infty}^{\infty} g(x)f(x)\,dx & \text{jeśli } X \text{ jest absolutnie ciągłe} \end{cases}$$

pod warunkiem odpowiednio, że $\sum_k |g(x_k)|p(x_k) < \infty$ i $\int_{-\infty}^{\infty} |g(x)|f(x)\,dx < \infty$. Jeśli g jest nieujemna, to używamy symbolu $\mathbb{E}\, g(X)$ również gdy równa się nieskończoność. Jeśli rozkład jest mieszanką dwóch typów to definiujemy $\mathbb{E}\, g(X)$ by

$$\mathbb{E}\, g(X) = \theta \sum_k g(x_k)p(x_k) + (1-\theta) \int_{-\infty}^{\infty} g(x)f(x)\,dx\,. \tag{1.2}$$

Czasami wygodne jest zapisanie $\mathbb{E}\, g(X)$ w terminach *całki Stieltjesa* t.j.

$$\mathbb{E}\, g(X) = \int_{-\infty}^{\infty} g(x)dF(x)\,,$$

która jest dobrze określona przy pewnych warunkach na g i F. W szczególności dla $g(x) = x$ wartość $\mu = \mathbb{E}\, X$ jest nazywana *średnią* , lub *wartością oczekiwaną* lub *pierwszy momentem* X. Dla $g(x) = x^n$, $\mu^{(n)} = \mathbb{E}\,(X^n)$ jest nazywana n-tym *momentem*. *Momentem centralnym rzędu* n nazywamy $\alpha_n = \mathbb{E}\,(X - \mu)^n$. *Wariancją* X jest $\sigma^2 = \mathbb{E}\,(X - \mu)^2$ i $\sigma = \sqrt{\sigma^2}$ jest

odchyleniem standardowym lub *dyspersją* . Czasami piszemy $\operatorname{Var} X$ zamiast σ^2. *Współczynnik zmienności* jest zdefiniowany przez $\mathrm{cv}_X = \sigma/\mu$, i *współczynnik asymetrii* przez $\alpha_3\sigma^{-3} = \mathbb{E}\,(X-\mu)^3\sigma^{-3}$. Dla dwóch zmiennych losowych X, Y definiujemy *kowariancję* $\operatorname{Cov}(X,Y)$ przez $\operatorname{Cov}(X,Y) = \mathbb{E}\,((X-\mathbb{E}\,X)(Y-\mathbb{E}\,Y))$ jeśli tylko $\mathbb{E}\,X^2, \mathbb{E}\,Y^2 < \infty$. zauważmy, że równoważnie mamy $\operatorname{Cov}(X,Y) = \mathbb{E}\,XY - \mathbb{E}\,X\,\mathbb{E}\,Y$. Jeśli $\operatorname{Cov}(X,Y) > 0$, to mówimy, że X, Y są *dodatnio skorelowane*, jeśli $\operatorname{Cov}(X,Y) < 0$ to *ujemnie skorelowane* i *nieskorelowane* jeśli $\operatorname{Cov}(X,Y) = 0$. Przez *współczynnik korelacji* pomiędzy X, Y rozumiemy

$$\operatorname{corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}.$$

Medianą zmienej losowej X jest dowolna liczba is $\zeta_{1/2}$ taka, że

$$\mathbb{P}(X \le \zeta_{1/2}) \ge \tfrac{1}{2}\,, \qquad \mathbb{P}(X \ge \zeta_{1/2}) \ge \tfrac{1}{2}\,.$$

Przez indykator $\mathbf{1}(A) : \Omega \to \mathbb{R}$ rozumiemy zmienną losową

$$\mathbf{1}(A,\omega) = \begin{cases} 1 & \text{jeżeli } \omega \in A, \\ 0 & \text{w przeciwnym razie.} \end{cases}$$

A więc jeśli $A \in \mathcal{F}$ to $\mathbf{1}(A)$ jest zmienną losową i $\mathbb{P}(A) = \mathbb{E}\,\mathbf{1}(A)$. Używa się nastepującej konwencji, że $\mathbb{E}\,[X;A] = \mathbb{E}\,(X\mathbf{1}(A))$ dla zmiennej losowej X i zdarzenia $A \in \mathcal{F}$.

Jeśli chcemy zwrócić uwagę, że średnia, n-ty moment, itd. są zdefiniowane dla zmiennej X lub dystrybunaty F to piszemy $\mu_X, \mu_X^{(n)}, \sigma_X^2, \ldots$ or $\mu_F, \mu_F^{(n)}, \sigma_F^2, \ldots$.

1.3 Niezależność i warunkowanie

Mówimy, że zmienne losowe $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ są *niezależne* jeśli

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{k=1}^{n} \mathbb{P}(X_k \in B_k)$$

dla każdego $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$ lub, równoważnie,

$$\mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n) = \prod_{k=1}^{n} \mathbb{P}(X_k \le x_k)$$

dla wszystkich $x_1, \ldots, x_n \in \mathbb{R}$. Nieskończony ciąg $X_1, X_2, \ldots$ jest ciągiem niezależnych zmiennych losowych, jeśli każdy skończony podciąg ma tą własność. Mówimy, że dwa ciągi $X_1, X_2, \ldots$ i $Y_1, Y_2, \ldots$ są niezależne jeśli

$$\mathbb{P}\Big(\bigcap_{i=1}^{n}\{X_i \le x_i\} \cap \bigcap_{i=1}^{m}\{Y_i \le y_i\}\Big) = \mathbb{P}\Big(\bigcap_{i=1}^{n}\{X_i \le x_i\}\Big)\mathbb{P}\Big(\bigcap_{i=1}^{m}\{Y_i \le y_i\}\Big)$$

dla wszystkicj $n, m \in \mathbb{N}$ i $x_1, \ldots, x_n$, $y_1, \ldots, y_m \in \mathbb{R}$. Dla ciągu $X_1, X_2, \ldots$ niezależnych i jednakowo rozłożonych zmiennych losowych wygodne jest wprowadzenie generycznej zmiennej losowej X, która ma taki sam rozkład. Mówimy, że zmienna losowa jest niezależna od zdarzenie, jeśli jest niezależna od indykatora $\mathbf{1}(A)$ tego zdarzenia.

Niech $A \in \mathcal{F}$. *Warunkowym prawdopodobieństwem* zadarzenia A' pod warunkiem A jest

$$\mathbb{P}(A' \mid A) = \begin{cases} \mathbb{P}(A' \cap A)/\mathbb{P}(A) & \text{jeżeli } \mathbb{P}(A) > 0, \\ 0 & \text{w przeciwnym razie.} \end{cases}$$

Często użyteczne jest wzór na *prawdopodobieństwo całkowite*. Niech $A, A_1, A_2, \ldots \in \mathcal{F}$ będzie ciągiem zdarzeń taki, że $A_i \cap A_j = \emptyset$ dla $i \neq j$ i $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = 1$. Wtedy,

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \mid A_i)\, \mathbb{P}(A_i)\,. \tag{1.3}$$

Warunkową wartością oczekiwaną $\mathbb{E}\,(X \mid A)$ zmiennej losowej X pod warunkiem zdarzenia A jest warunkowa wartością oczekiwana względem rozkładu warunkowego z dystrybunatą $F_{X|A}$, gdzie $F_{X \mid A}(x) = \mathbb{P}(\{X \le x\} \cap A)/\mathbb{P}(A)$. Wtedy wzór na prawdopodobieństwo całkowite można sformułować ogólniej:

$$\mathbb{E}\, X = \sum_{i=1}^{\infty} \mathbb{E}\,(X \mid A_i)\, \mathbb{P}(A_i)\,. \tag{1.4}$$

Ogolnie można zdefiniować dla dwóch zmiennych losowych X, Y warunkową wartość oczekiwaną $\mathbb{E}\,[X|Y]$. Niestety formalna definicja w pełnej ogólności wymaga pojęć teorii miary. Natomiast bez trudności mozna zrozumieć najbardziej przydatne przypadki; gdy zmienna losowa Y jest dyskretna i gdy X, Y mają rozkład ciągly z łączną gęstościa $f(x, y)$.

Lemat 1.1 *Niech X, Y, Z będą zmiennymi losowymi takimi, że odpowiednie wartości oczekiwane, wariancje i kowariancje istnieją skończone. Wtedy*

(i) $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$,
(ii) $\mathrm{Var}[X] = \mathbb{E}[\mathrm{Var}[X|Y]] + \mathrm{Var}[\mathbb{E}[X|Y]]$,
(iii) $\mathrm{Cov}(X,Y) = \mathbb{E}[\mathrm{Cov}(X,Y|Z)] + \mathrm{Cov}(\mathbb{E}[X|Z], \mathbb{E}[Y|Z])$.

Twierdzenie 1.2 *Jesli $X_1, \ldots, X_n$ są niezależnymi zmiennymi losowymi, i ϕ oraz ψ funkcjami niemalejącymi (lub obie nierosnącymi) n zmiennych, to*

$$\mathbb{E}[\phi(X_1, \ldots, X_n)\psi(X_1, \ldots, X_n)] \geq \mathbb{E}[\phi(X_1, \ldots, X_n)]\mathbb{E}[\psi(X_1, \ldots, X_n)]$$

przy założeniu że wartości oczekiwane są skończone.

Dowód Dowód jest indukcyjny względem liczby zmiennych $n = 1, 2, \ldots$. Niech $n = 1$. Wtedy

$$(\psi(x) - \psi(y))(\phi(x) - \phi(y)) \geq 0,$$

skąd

$$\mathbb{E}(\psi(X) - \psi(Y))(\phi(X) - \phi(Y)) \geq 0\,.$$

Teraz rozpisując mamy

$$\mathbb{E}\,\psi(X)\phi(X) - \mathbb{E}\,\psi(X)\phi(Y) - \mathbb{E}\,\psi(Y)\phi(X) + \mathbb{E}\,\psi(Y)\phi(Y) \geq 0\,.$$

Przyjmując, że X i Y są niezależne o jednakowym rozkładzie jak X_1 mamy

$$\mathbb{E}[\phi(X_1)\psi(X_1)] \geq \mathbb{E}[\phi(X_1)]\mathbb{E}[\psi(X_1)]\,.$$

Teraz zakładamy, że twierdzenie jest prawdziwe dla $n-1$. Stąd

$$\begin{aligned}&\mathbb{E}[\phi(X_1, \ldots, X_{n-1}, x_n)\psi(X_1, \ldots, X_{n-1}, x_n)]\\ &\quad \geq \mathbb{E}[\phi(X_1, \ldots, X_{n-1}, x_n)]\mathbb{E}[\psi(X_1, \ldots, X_{n-1}, x_n)],\end{aligned}$$

dla wszystkich x_n. Podstawiając $x_n = X_n$ mamy

$$\begin{aligned}&\mathbb{E}[\phi(X_1, \ldots, X_{n-1}, X_n)\psi(X_1, \ldots, X_{n-1}, X_n)]\\ &\quad \geq \mathbb{E}[\phi(X_1, \ldots, X_{n-1}, X_n)]\mathbb{E}[\psi(X_1, \ldots, X_{n-1}, X_n)].\end{aligned}$$

□

1.4 Transformaty

Niech $I = \{s \in \mathbb{R} : \mathbb{E}\, e^{sX} < \infty\}$. Zauważmy, że I musi być odcinkiem (właściwym) lub prostą półprostą lub nawet punktem $\{0\}$. *Funkcja tworząca momenty* $\hat{m} : I \to \mathbb{R}$ zmiennej losowej X jest zdefiniowana przez $\hat{m}(s) = \mathbb{E}\, e^{sX}$. Zauważmy różnice pomiędzy funkcją tworzącą momenty a *transformatą Laplace-Stieltjesa* $\hat{l}(s) = \mathbb{E}\, e^{-sX} = \int_{-\infty}^{\infty} e^{-sx}\, dF(x)$ zmiennej losowej X lub dystrybuanty F zmiennej X.

Dla zmiennych kratowych z funkcją prawdopodobieństwa $\{p_k,\ k \in \mathbb{N}\}$ wprowadza się *funkcję tworzącą* $\hat{g} : [-1, 1] \to \mathbb{R}$ zdefiniowaną przez $\hat{g}(s) = \sum_{k=0}^{\infty} p_k s^k$.

Niech teraz X będzie dowolną zmienną losową z dystrybuantą F. *Funkcją charakterystyczną* $\hat{\varphi} : \mathbb{R} \to \mathbb{C}$ zmiennej X jest zdefiniowna przez

$$\hat{\varphi}(s) = \mathbb{E}\, e^{isX}. \tag{1.5}$$

W szczególności jesli F jest absolutnie ciągła z gęstością f, to

$$\hat{\varphi}(s) = \int_{-\infty}^{\infty} e^{isx} f(x)\, dx = \int_{-\infty}^{\infty} \cos(sx) f(x)\, dx + i \int_{-\infty}^{\infty} \sin(sx) f(x)\, dx\,.$$

Jeśli X jest kratową zmienną losową z funkcją prawdopodobieństwa $\{p_k\}$, to

$$\hat{\varphi}(s) = \sum_{k=0}^{\infty} e^{isk} p_k\,. \tag{1.6}$$

Jeśli chcemy podkreślić, że transformata jest zmiennej X z dystrybuantą F, to piszemy $\hat{m}_X, \ldots, \hat{\varphi}_X$ lub $\hat{m}_F, \ldots, \hat{\varphi}_F$.

Zauważmy, też formalne związki $\hat{g}(e^s) = \hat{l}(-s) = \hat{m}(s)$ lub $\hat{\varphi}(s) = \hat{g}(e^{is})$. Jednakże delikatnym problemem jest zdecydowanie, gdzie dana transformata jest określona. Na przykład jeśli X jest nieujemna, to $\hat{m}(s)$ jest zawsze dobrze zdefiniowana przynajmiej dla $s \le 0$ podczas gdy $\hat{l}(s)$ jest dobrze zdefiniowana przynajmniej dla $s \ge 0$. Są też przykłady pokazujące, że $\hat{m}(s)$ może być ∞ dla wszystkich $s > 0$, podczas gdy w innych przypadkach, że $\hat{m}(s)$ jest skończone $(-\infty, a)$ dla pewnych $a > 0$.

Ogólnie jest wzajemna jednoznaczność pomiędzy rozkładem a odpowiadającą jemu transformatą, jednakże w każdym wypadku sformułowanie twierdzenia wymaga odpowiedniego wysłowienia. Jeśli n-ty moment zmiennej losowej $|X|$ jest skończony, to n-ta pochodna $\hat{m}^{(n)}(0), \hat{l}^{(n)}(0)$ istnieje jeśli tylko $\hat{m}(s), \hat{l}(s)$ są dobrze zdefiniowane w pewnym otoczeniu $s = 0$. Wtedy mamy

$$\mathbb{E}\, X^n = \hat{m}^{(n)}(0) = (-1)^n \hat{l}^{(n)}(0) = (-i)^n \hat{\varphi}^{(n)}(0)\,. \tag{1.7}$$

Jeśli $\hat{m}(s)$ i $\hat{l}(s)$ są odpowienio dobrze zdefiniowane na $(-\infty, 0]$ i $[0, \infty)$, to trzeba pochodne w (1.7) zastąpić pochodnymi jednostronnymi, t.j.

$$\mathbb{E}\, X^n = \hat{m}^{(n)}(0-) = (-1)^n \hat{l}^{(n)}(0+)\,, \tag{1.8}$$

Jeśli X przyjmuje wartość w podzbiorze $\mathbb{Z}_+$ i jeśli $\mathbb{E}\, X^n < \infty$, to

$$\mathbb{E}\, \left(X(X-1)\dots(X-n+1)\right) = \hat{g}^{(n)}(1-)\,. \tag{1.9}$$

Inną ważną własnością $\hat{m}(s)$ (i również dla $\hat{l}(s)$ i $\hat{\varphi}(s)$) jest własność , że jeśli $X_1, \dots, X_n$ są niezależne to

$$\hat{m}_{X_1+\dots+X_n}(s) = \prod_{k=1}^{n} \hat{m}_{X_k}(s)\,. \tag{1.10}$$

Podobnie, jeśli $X_1, \dots, X_n$ są niezależnymi zmiennymi kratowymi z $\mathbb{Z}_+$ to

$$\hat{g}_{X_1+\dots+X_n}(s) = \prod_{k=1}^{n} \hat{g}_{X_k}(s)\,. \tag{1.11}$$

Ogólnie dla dowolnych funkcji $g_1, \dots, g_n : \mathbb{R} \to \mathbb{R}$ mamy

$$\mathbb{E}\left(\prod_{k=1}^{n} g_k(X_k)\right) = \prod_{k=1}^{n} \mathbb{E}\, g_k(X_k) \tag{1.12}$$

jeśli tylko $X_1, \dots, X_n$ są niezależne.

1.5 Zbieżności

Niech $X, X_1, X_2, \dots$ będą zmiennymi losowymi na $(\Omega, \mathcal{F}, \mathbb{P})$. Mówimy, że ciąg $\{X_n\}$ *zbiega z prawdopdobieństwem 1* do X jeśli

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1.$$

Piszemy wtedy $X_n \overset{\text{a.s.}}{\to} X$. Niech $X, X_1, X_2, \dots$ będą zmiennymi losowymi. Mówimy, że ciąg $\{X_n\}$ *zbiega według prawdopdobieństwa* do X jeśli dla każdego $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$

Piszemy wtedy $X_n \overset{\text{Pr}}{\to} X$.

Niech $X, X_1, X_2, \ldots$ będą zmiennymi losowymi. Mówimy, że ciąg $\{X_n\}$ *zbiega słabo* lub *zbiega według rozkładu* do X jeśli $F_{X_n}(x) \to F_X(x)$ dla wszystkich punktów ciągłości F_X. Piszemy wtedy $X_n \xrightarrow{\mathcal{D}} X$. Ta definicja jest równoważna następującej: $X_n \xrightarrow{\mathcal{D}} X$ wtedy i tylko wtedy gdy

$$\lim_{n\to\infty} \mathbb{E}\, g(X_n) = \mathbb{E}\, g(X) \tag{1.13}$$

dla każdej funkcji ciągłej i ograniczonej $g : \mathbb{R} \to \mathbb{R}$. Ten wariant definicji jest używany do zdefiniowania zbieżności według rozkładu na innych przestrzeniach niż $\mathbb{R}$. Warunkiem dostatecznym dla $X_n \xrightarrow{\mathcal{D}} X$ jest

$$\lim_{n\to\infty} \hat{m}_{X_n}(s) = \hat{m}_X(s) \tag{1.14}$$

dla odcinka otwartego nie tożsamego z $\{0\}$ gdzie $\hat{m}_X(s)$ istnieje. Ze zbieżności z prawdopodobieństwem 1 wynika zbieżność według prawdopodobieństwa, z której wynika zbieżnosc według rozkładu.

Następujące własności są często wykorzystywane. Jesli $X_n \xrightarrow{\mathcal{D}} X$ i $Y_n \xrightarrow{\mathcal{D}} c \in \mathbb{R}$, to

$$X_n Y_n \xrightarrow{\mathcal{D}} cX\,, \qquad X_n + Y_n \xrightarrow{\mathcal{D}} X + c\,. \tag{1.15}$$

2 Elementy procesów stochastycznych

Przez proces stochastyczny $(X(t))_{t\in T}$ z przestrzenią parametrów T nazywamy kolekcję zmiennych losowych. Za T możemy przyjąć np. $T = [0, \infty)$, $T = [a, b]$, itp. W tym rozdziale $T = [0, \infty)$. Zamiast kolekcji zmiennych losowych możemy rozpatrywać kolekcję wektorów losowych z $\mathbb{R}^d$. Przez przyrost procesu $(X(t)_{t\geq 0})$ na odcinku (a, b) rozumiemy $X(b)-X(t)$. Ważne klasy procesów definiuje się przez zadanie odpowiednich własności przyrostów. Na przykład możemy żądać aby na rozłącznych odcinkach przyrosty były niezależne. Zobaczymy to na specjalnym ważny przykładzie ruchu Browna.

2.1 Ruch Browna

Przez standardowy ruch Browna (BM(0,1)) rozumiemy proces stochastyczny $(B(t))_{t\geq 0}$ spełniający warunki

1. B(0) = 0,

2. przyrosty $\mathrm{B}(b_1) - B(a_1), \ldots, B(b_k) - B(a_k)$ są niezależne jeśli tylko odcinki $(a_1, b_1] \ldots, (a_k, b_k]$ są rozłączne $(k = 2, \ldots)$,

3. $B(b) \sim N(0, b)$

4. realizacje $B(t)$ są ciągłe.

Ważną własnościa ruchu Browna jest *samopodobieństwo* z współczynnikiem $\alpha = 1/2$ tj.

$$B(t) =_{\mathrm{d}} t^{1/2} B(1).$$

Definiuje się, dla pewnej klasy procesów X całkę stochastyczną Itô względem ruchu Browna

$$\int_0^T X(t)\, dB(t).$$

W szczególności jeśli funkcja podcałkowa k jest deterministyczna z $\int_0^T k^2(x)\, dx$ to $\int_0^T k(t)\, dB(t)$ ma rozkład normalny $\mathrm{N}(0, \sigma^2)$, gdzie

$$\sigma^2 = \int_0^T k^2(t)\, dt.$$

Korzystając z pojęcia całki stochastycznej Itô definiujemy różniczkę stochastyczną $d\, k(B(t), t)$ następująco. Niech $k(x, t)$ będzie funkcję dwukrotnie ciągle różniczkowalną względem zmiennej x i jednokrotnie względem zmiennej t. Wtedy korzystając z tak zwanego wzoru Itô mamy

$$d\, k(B(t), t) = k_{0,1}(B(t), t)\, dt + k_{1,0}(B(t), t)\, dB(t) + \frac{1}{2} k_{2,0}(B(t), t)\, dt,$$

gdzie $k_{i,0}$ oznacza i-tą pochodną cząstkową k względem x a $k_{0,1}$ pierwszą pochodną cząstkową względem zmiennej t.

3 Rodziny rozkładów

Podamy teraz kilka klas rozkładów dyskretnych i absolutnie ciągłych. Wszystkie są scharakteryzowane przez skończony zbiór parametrów. Podstawowe charakterystyki podane są w tablicy rozkładów w podrozdziale 4.

W zbiorze rozkładów absolutnie ciągłych wyróżniamy dwie podklasy: rozkłady z lekkimi ogonami i ciężkimi ogonami. Rozróznienie to jest bardzo ważne w teorii Monte Carlo.

Mówimy, że rozkład z dystrybuntą $F(x)$ ma wykładniczo ograniczony ogon jeśli dla pewnego $a, b > 0$ mamy $\overline{F}(x) \leq ae^{-bx}$ dla wszystkich $x > 0$. Wtedy mówimy też że rozkład ma *lekki ogon*. Zauważmy, że wtedy $\hat{m}_F(s) < \infty$ dla $0 < s < s_0$, gdzie $s_0 > 0$. W przeciwnym razie mówimy, że rozkład ma *ciężki ogon* a więc wtedy $\hat{m}_F(s) = \infty$ dla wszystkich $s > 0$.

Uwaga Dla dystrybunaty z ciężkim ogonem F mamy

$$\lim_{x\to\infty} e^{sx}\overline{F}(x) = \infty \tag{3.1}$$

dla wszystkich $s > 0$.

3.1 Rozkłady dyskretne

Wśród dyskretnych rozkładów mamy:

- *Rozkład zdegenerowany* δ_a skoncentrowany w $a \in \mathbb{R}$ z $p(x) = 1$ jeśli $x = a$ i $p(x) = 0$ w przeciwnym razie.
- *Rozkład Bernoulliego* $\mathrm{Ber}(p)$ z $p_k = p^k(1-p)^{1-k}$ dla $k = 0, 1;\ 0 < p < 1$: rozkład przyjmujący tylko wartości 0 i 1.
- *Rozkład dwumianowy* $\mathrm{B}(n, p)$ z $p_k = \binom{n}{k}p^k(1-p)^{n-k}$ dla $k = 0, 1 \ldots, n;\ n \in \mathbb{Z}_+, 0 < p < 1$: rozkład sumy n niezależnych i jednakowo rozłożonych zmiennych losowych o rozkładzie $\mathrm{Ber}(p)$. A więc , $\mathrm{B}(n_1, p) * \mathrm{B}(n_2, p) = \mathrm{B}(n_1 + n_2, p)$.
- *Rozkład Poissona* $\mathrm{Poi}(\lambda)$ z $p_k = \mathrm{e}^{-\lambda}\lambda^k/k!$ dla $k = 0, 1 \ldots;\ 0 < \lambda < \infty$: jedna z podstawowych cegiełek na których zbudowany jest rachunek prawdopodobieństwa. Historycznie pojawił się jako graniczny dla rozkładów dwumianowych $\mathrm{B}(n, p)$ gdy $np \to \lambda$ dla $n \to \infty$. Ważna jest własność zamkniętości ze względu na splot, tzn. suma niezależnych zmiennych Poissonowskich ma znowu rozkład Poissona. Formalnie piszemy, że $\mathrm{Poi}(\lambda_1) * \mathrm{Poi}(\lambda_2) = \mathrm{Poi}(\lambda_1 + \lambda_2)$.

- *Rozkład geometryczny* Geo(p) z $p_k = (1-p)p^k$ dla $k = 0, 1, \ldots;\ 0 < p < 1$: rozkład mający dyskretny brak pamięci. To jest $\mathbb{P}(X \geq i+j \mid X \geq j) = \mathbb{P}(X \geq i)$ dla wszystkih $i, j \in \mathbb{N}$ wtedy i tylko wtedy gdy X jest geometrycznie rozłożony.

- *Rozkład obcięty geometryczny* TGeo(p) z $p_k = (1-p)p^{k-1}$ dla $k = 1, 2, \ldots;\ 0 < p < 1$: jeśli $X \sim$Geo(p), to $X|X > 0 \sim$TGeo(p), jest to rozkład do liczby prób Bernoulliego do pierszego sukcesy, jeśli prawdopobieństwo sukcesu wynosi $1-p$.

- *Rozkład ujemny dwumianowy* lub inaczej *rozkład Pascala* NB(α, p) z $p_k = \Gamma(\alpha+k)/(\Gamma(\alpha)\Gamma(k+1))(1-p)^\alpha p^k$ dla $k = 0, 1, \ldots;\ \alpha > 0, 0 < p < 1$. Powyżej piszemy
$$\binom{x}{0} = 1, \qquad \binom{x}{k} = \frac{x(x-1)\ldots(x-k+1)}{k!}\,,$$
dla $x \in \mathbb{R}$, $k = 1, 2, \ldots$. Mamy
$$p_k = \binom{\alpha+k-1}{k}(1-p)^\alpha p^k = \binom{-\alpha}{k}(1-p)^\alpha(-p)^k\,.$$
Co więcej, dla $\alpha = 1, 2, \ldots$, NB(α, p) jest rozkładem sumy α niezależnych i jednakowo rozłożonych Geo(p) zmiennych losowych. To oznacza na przykład, że podklasa ujemnych rozkładów dwumianowych $\{\text{NB}(\alpha, p), \alpha = 1, 2, \ldots\}$ jest zamknięta ze względu na splot t.j. NB(α_1, p)$*$ NB(α_2, p) = NB($\alpha_1 + \alpha_2, p$) jeśli $\alpha_1, \alpha_2 = 1, 2, \ldots$. co więcej poprzedni wzór jest spełniny dla $\alpha_1, \alpha_2 > 0$ and $0 < p < 1$.

- *Rozkład logarytmiczny* Log(p) z $p_k = p^k/(-k\log(1-p))$ for $k = 1, 2, \ldots;\ 0 < p < 1$: pojawia się jako granica obciętych rozkładów dwumianowych.

- *Rozkład (dyskretny) jednostajny* UD(n) z $p_k = n^{-1}$ dla $k = 1, \ldots, n$; $n = 1, 2, \ldots$.

3.2 Rozkłady dyskretne wielowymiarowe

- *Rozkład wielomianowy* M($n, \boldsymbol{a}$), z $p_{\boldsymbol{k}} = \frac{n}{k_1!\ldots k_d!}a_1^{k_1}\ldots a_d^{k_d}$ dla $\boldsymbol{k} = (k_1, \ldots, k_d)$ takiego, że $k_i \in \mathbb{Z}_+$ i $k_1 + \ldots + k_d = n$; zakładamy, że $\boldsymbol{a} = (a_1, \ldots, a_d)$ jest funkcją prawdopodobieństwa; rozkład sumy n niezależnych wektorów losowych przyjmujący wartości $(0, \ldots, 0, 1, 0, \ldots, 0)$ (jedynka na i-tym miejscu) z prawdopodobieństwem a_i.

3.3 Rozkłady absolutnie ciągłe

Wśród rozkładów absolutnie ciągłych mamy:

- *Rozkład normalny* $\mathrm{N}(\mu, \sigma^2)$ z $f(x) = (2\pi\sigma^2)^{-1/2}\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}$ dla wszystkich $x \in \mathbb{R}$; $\mu \in \mathbb{R}, \sigma^2 > 0$: obok rozkładu Poissona druga cegiełka na której opiera się rachunek prawdopodobieństwa. Pojawia się jako granica sum niezalażnych asymptotycznie zaniedbywalnych zmiennych losowych. Rozkłady normalne są zamnięte na branie splotów, tj. $\mathrm{N}(\mu_1, \sigma_1^2) * \mathrm{N}(\mu_2, \sigma_2^2) = \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

- *Rozkład wykładniczy* $\mathrm{Exp}(\lambda)$ z $f(x) = \lambda\mathrm{e}^{-\lambda x}$ dla $x > 0$; $\lambda > 0$: podstawowy rozkład w teorii procesów Markowa ze względu na własność braku pamięci, t.j. $\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t)$ dla wszystkich $s, t \geq 0$ jeśli X jest rozkładem wykładniczym.

- *Rozkład Erlanga* $\mathrm{Erl}(n, \lambda)$ z $f(x) = \lambda^n x^{n-1}\mathrm{e}^{-\lambda x}/(n-1)!$; $n = 1, 2, \ldots$: rozkład sumy n niezależnych i jednakowo rozłożonych $\mathrm{Exp}(\lambda)$ zmiennych losowych. Tak więc, $\mathrm{Erl}(n_1, \lambda) * \mathrm{Erl}(n_2, \lambda) = \mathrm{Erl}(n_1 + n_2, \lambda)$.

- *Rozkład chi kwadrat* $\chi^2(n)$ z $f(x) = (2^{n/2}\Gamma(n/2))^{-1}x^{(n/2)-1}\mathrm{e}^{-x/2}$ jeżeli $x > 0$ i $f(x) = 0$ jeżeli $x \leq 0$; $n = 1, 2, \ldots$: Rozkład sumy n niezależnych i jednakowo rozłożonych zmiennych losowych, które są kwadratami $\mathrm{N}(0, 1)$ zmiennych losowych. To oznacza, że $\chi^2(n_1) * \chi^2(n_2) = \chi^2(n_1 + n_2)$.

- *Rozkład gamma* $\mathrm{Gamma}(a, \lambda)$ z $f(x) = \lambda^a x^{a-1}\mathrm{e}^{-\lambda x}/\Gamma(a)$ for $x \geq 0$; z parametrem kształtu $a > 0$ i parametrem skali $\lambda > 0$: jeśli $a = n \in \mathbb{N}$, to $\mathrm{Gamma}(n, \lambda)$ jest rozkładem $\mathrm{Erl}(n, \lambda)$. Jeśli $\lambda = 1/2$ i $a = n/2$, to $\mathrm{Gamma}(n/2, 1/2)$ jest rozkładem $\chi^2(n)$.

- *Rozkład jednostajny* $\mathrm{U}(a, b)$ z $f(x) = (b-a)^{-1}$ dla $a < x < b$; $-\infty < a < b < \infty$.

- *Rozkład beta* $\mathrm{Beta}(a, b, \eta)$ z

$$f(x) = \frac{x^{a-1}(\eta - x)^{b-1}}{B(a, b)\eta^{a+b-1}}$$

  da $0 < x < \eta$ gdzie $B(a, b)$ oznacza funkcję beta(patrz **[??som.spe.fun??]**); $a, b, \eta > 0$; jeśli $a = b = 1$, to $\mathrm{Beta}(1, 1, \eta) = \mathrm{U}(0, \eta)$.

- *Rozkład odwrotny gaussowski* (*inverse Gaussian distribution*) $\mathrm{IG}(\mu, \lambda)$ z

$$f(x) = \left(\lambda/(2\pi x^3)\right)^{1/2} \exp\left(-\lambda(x-\mu)^2/(2\mu^2 x)\right)$$

dla wszystkich $x > 0;\ \mu \in \mathbb{R}, \lambda > 0$.

- *Rozkład statystyki ekstremalnej* $\mathrm{EV}(\gamma)$ z $F(x) = \exp(-(1+\gamma x)_+^{-1/\gamma})$ dla wszystkich $\gamma \in \mathbb{R}$ gdzie wielkość $-(1+\gamma x)_+^{-1/\gamma}$ jest zdefiniowana jako e^{-x} kiedy $\gamma = 0$. W tym ostatnim przypadku rozkład jest znany jako *Rozkład Gumbela*. Dla $\gamma > 0$ otrzymujemy *rozkład Frécheta* ; rokład ten jest skoncentrowany na półprostej $(-1/\gamma, +\infty)$. W końcu, dla $\gamma < 0$ otrzymujemy *rozkład ekstremalny Weibulla*, skoncentrowany na półprostej $(-\infty, -1/\gamma)$.

3.4 Rozkłady z ciężkimi ogonami

Przykładami absolutnie ciągłych rozkładów, które mają ciężki ogon są podane niżej.

- *Rozkład logarytmiczno normalny* $\mathrm{LN}(a, b)$ z $f(x) = (xb\sqrt{2\pi})^{-1} \exp\left(-(\log x - a)^2/(2b^2)\right)$ dla $x > 0;\ a \in \mathbb{R}, b > 0$; jeśli X jest o rozkładzie $\mathrm{N}(a, b)$ to e^X ma rozkład $\mathrm{LN}(a, b)$.

- *Rozkład Weibulla* $\mathrm{W}(r, c)$ z $f(x) = rcx^{r-1} \exp(-cx^r)$ dla $x > 0$ gdzie parametr kształtu $r > 0$ i parametr skali $c > 0$; $\overline{F}(x) = \exp(-cx^r)$; jeśli $r \geq 1$, to $\mathrm{W}(r, c)$ ma lekki ogon, w przeciwnym razie $\mathrm{W}(r, c)$ ma ciężki.

- *Rozkład Pareto* $\mathrm{Par}(\alpha)$ z $f(x) = \alpha(1/(1+x))^{\alpha+1}$ for $x > 0$; gdzie eksponent $\alpha > 0$; $\overline{F}(x) = (1/(1+x))^{\alpha}$.

- *Rozkład loggamma* $\mathrm{LogGamma}(a, \lambda)$ z $f(x) = \lambda^a/\Gamma(a)(\log x)^{a-1}x^{-\lambda-1}$ dla $x > 1$; $\lambda, a > 0$; jeśli X jest $\Gamma(a, \lambda)$, to e^X ma rozkład $\mathrm{LogGamma}(a, \lambda)$.

- *Rozkład Cauchy'ego* z $f(x) = 1/(\pi(1+x^2))$, gdzie $x \in \mathbb{R}$; rozkład który nie ma średniej z dwustronnie ciężkim ogonem.

3.5 Rozkłady wielowymiarowe

Przykładami absolutnie ciągłych wielowymiarowych rozkładów są podane niżej.

- *Rozkład wielowymiarowy normalny* $\mathrm{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$ ma wektor losowy $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ z wartością oczekiwaną $\mathbb{E}\,\boldsymbol{X} = \boldsymbol{m}$ oraz macierzą kowariancji $\boldsymbol{\Sigma}$, jeśli dowolna kombinacja liniowa $\sum_{j=1}^{n} a_j X_j$ ma jednowymiarowy rozkład normalny. Niech $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^T$. Jeśli $\boldsymbol{Z}$ ma rozkład $\mathrm{N}((0, \ldots, 0), \boldsymbol{I})$, gdzie $\boldsymbol{I}$ jest macierzą jednostkową, to $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{Z}$ ma rozkład normalny $\mathrm{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$.

4 Tablice rozkładów

Następujące tablice podają najważniejsze rozkłady. Dla każdego rozkładu podane są: Średnia, wariancja, współczynnik zmienności zdefiniowany przez $\sqrt{\operatorname{Var} X}/\mathbb{E}\, X$ (w.zm.). Podane są również: funkcja tworząca $\mathbb{E}\, s^N$ (f.t.) w przypadku dyskretnego rozkładu, oraz funkcja tworząca momenty $\mathbb{E}\, \exp(sX)$ (f.t.m.) w przypadku absolutnie ciągłych rozkładów, jeśli tylko istnieją jawne wzory. W przeciwnym razie w tabeli zaznaczamy $*$. Więcej o tych rozkładach można znaleźć w w podrozdziale XI.

rozkład	parametry	funkcja prawdopodobieństwa $\{p_k\}$
$\mathrm{B}(n,p)$	$n=1,2,\ldots$ $0\le p\le 1, q=1-p$	$\binom{n}{k}p^k q^{n-k}$ $k=0,1,\ldots,n$
$\mathrm{Ber}(p)$	$0\le p\le 1, q=1-p$	$p^k q^{1-k}; k=0,1$
$\mathrm{Poi}(\lambda)$	$\lambda>0$	$\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}; k=0,1,\ldots$
$\mathrm{NB}(r,p)$	$r>0, 0<p<1$ $q=1-p$	$\frac{\Gamma(r+k)}{\Gamma(r)k!}q^r p^k$ $k=0,1,\ldots$
$\mathrm{Geo}(p)$	$0<p\le 1, q=1-p$	$qp^k; k=0,1,\ldots$
$\mathrm{Log}(p)$	$0<p<1$ $r=\frac{-1}{\log(1-p)}$	$r\frac{p^k}{k}; k=1,2,\ldots$
$\mathrm{UD}(n)$	$n=1,2,\ldots$	$\frac{1}{n}; k=1,\ldots,n$

Tablica 4.1: Tablica rozkładów dyskretnych

średnia	wariancja	w.zm.	f.t.
np	npq	$\left(\frac{q}{np}\right)^{\frac{1}{2}}$	$(ps+q)^n$
p	pq	$\left(\frac{q}{p}\right)^{\frac{1}{2}}$	$ps+q$
λ	λ	$\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}$	$\exp[\lambda(s-1)]$
$\frac{rp}{q}$	$\frac{rp}{q^2}$	$(rp)^{-\frac{1}{2}}$	$q^r(1-ps)^{-r}$
$\frac{p}{q}$	$\frac{p}{q^2}$	$p^{-\frac{1}{2}}$	$q(1-ps)^{-1}$
$\frac{rp}{1-p}$	$\frac{rp(1-rp)}{(1-p)^2}$	$\left(\frac{1-rp}{rp}\right)^{\frac{1}{2}}$	$\frac{\log(1-ps)}{\log(1-p)}$
$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\left(\frac{n-1}{3(n+1)}\right)^{\frac{1}{2}}$	$\frac{1}{n}\sum_{k=1}^{n} s^k$

Tablica 4.1: Tablica rozkładów dyskretnych (cd.)

rozkład	parametry	f. gęstości $f(x)$
$\mathrm{U}(a,b)$	$-\infty < a < b < \infty$	$\frac{1}{b-a}; x \in (a,b)$
$\mathrm{Gamma}(a,\lambda)$	$a, \lambda > 0$	$\frac{\lambda^a x^{a-1} \mathrm{e}^{-\lambda x}}{\Gamma(a)}; x \geq 0$
$\mathrm{Erl}(n,\lambda)$	$n = 1, 2 \ldots, \lambda > 0$	$\frac{\lambda^n x^{n-1} \mathrm{e}^{-\lambda x}}{(n-1)!}; x \geq 0$
$\mathrm{Exp}(\lambda)$	$\lambda > 0$	$\lambda \mathrm{e}^{-\lambda x}; x \geq 0$
$\mathrm{N}(\mu,\sigma^2)$	$-\infty < \mu < \infty,$ $\sigma > 0$	$\frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma}; x \in \mathbb{R}$
$\chi^2(n)$	$n = 1, 2, \ldots$	$\left(2^{n/2}\Gamma(n/2)\right)^{-1} x^{\frac{n}{2}-1} \mathrm{e}^{-\frac{x}{2}}; x \geq 0$
$\mathrm{LN}(a,b)$	$-\infty < a < \infty, b > 0$ $\omega = \exp(b^2)$	$\frac{\exp\left(-\frac{(\log x - a)^2)}{2b^2}\right)}{xb\sqrt{2\pi}}; x \geq 0$

Tablica 4.2: Tablica rozkładów absolutnie ciągłych

średnia	wariancja	w.zm.	f.t.m.
$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{b-a}{\sqrt{3}(b+a)}$	$\frac{\mathrm{e}^{bs}-\mathrm{e}^{as}}{s(b-a)}$
$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$	$a^{-\frac{1}{2}}$	$\left(\frac{\lambda}{\lambda-s}\right)^a; s<\lambda$
$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$	$n^{-\frac{1}{2}}$	$\left(\frac{\lambda}{\lambda-s}\right)^n; s<\lambda$
$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	1	$\frac{\lambda}{\lambda-s}; s<\lambda$
μ	σ^2	$\frac{\sigma}{\mu}$	$\mathrm{e}^{\mu s+\frac{1}{2}\sigma^2 s^2}$
n	$2n$	$\sqrt{\frac{2}{n}}$	$(1-2s)^{-n/2}; s<1/2$
$\exp\left(a+\frac{b^2}{2}\right)$	$\mathrm{e}^{2a}\omega(\omega-1)$	$(\omega-1)^{\frac{1}{2}}$	$*$

Tablica 4.2: Tablica rozkładów absolutnie ciągłych (c.d.)

rozkład	parametry	f. gęstości $f(x)$
Par(α)	$\alpha > 0$	$\alpha \left(\frac{1}{1+x}\right)^{\alpha+1}; x > 0$
W(r, c)	$r, c > 0$ $\omega_n = \Gamma\left(\frac{r+n}{r}\right)$	$rcx^{r-1}\mathrm{e}^{-cx^r}; x \geq 0$
IG(μ, λ)	$\mu > 0, \lambda > 0$	$\left[\frac{\lambda}{2\pi x^3}\right]^{\frac{1}{2}} \exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right); x > 0$
PME(α)	$\alpha > 1$	$\int_{\frac{\alpha-1}{\alpha}}^{\infty} \alpha \left(\frac{\alpha-1}{\alpha}\right)^{\alpha} y^{-(\alpha+2)}\mathrm{e}^{-x/y}dy$ $x > 0$

Tablica 4.2: Tablica rozkładów absolutnie ciągłych (c.d.)

średnia	wariancja	w.z.	f.t.m.
$\frac{1}{\alpha-1}$ $\alpha>1$	$\frac{\alpha}{(\alpha-1)^2(\alpha-2)}$ $\alpha>2$	$[\frac{\alpha}{\alpha-2}]^{-\frac{1}{2}}$ $\alpha>2$	$*$
$\omega_1 c^{-1/r}$	$(\omega_1-\omega_2^2)c^{-2/r}$	$\left[\frac{\omega_2}{\omega_1^2}-1\right]^{\frac{1}{2}}$	$*$
μ	$\frac{\mu^3}{\lambda}$	$\left(\frac{\mu}{\lambda}\right)^{\frac{1}{2}}$	$\frac{\exp\left(\frac{\lambda}{\mu}\right)}{\exp\left(\sqrt{\frac{\lambda^2}{\mu^2}-2\lambda s}\right)}$
1	$1+\frac{2}{\alpha(\alpha-2)}$	$\sqrt{1+\frac{2}{\alpha(\alpha-2)}}$	$\frac{\int_0^{\frac{\alpha}{\alpha-1}}\frac{x^\alpha}{x-s}dx}{\frac{\alpha^{\alpha-1}}{(\alpha-1)^\alpha}}$ $s<0$

Tablica 4.2: Tablica rozkładów absolutnie ciągłych (cd.)

5 Elementy wnioskowania statystycznego

5.1 Teoria estymacji

W teorii estymacji rozpatruje się $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$, gdzie $\theta \in \Theta$ oraz zmienne losowe $X_1, \ldots, X_n$ zwane *próbą*. Jesli ponadto dla każdego θ zmienne $X_1, \ldots, X_n$ są niezależne o jednakowym rozkładzie (względem $\mathbb{P}_\theta$) to mówimy o *próbie prostej* . Niech $g : \Theta \to \mathbb{R}^d$. *Estymatorem* nazywamy dowolną statystykę $T_n = f(X_1, \ldots, X_n)$, czyli zmienną losową, która jest funkcją od $X_1, \ldots, X_n$. Jesli

$$\mathbb{E}_\theta T_n = g(\theta),$$

to estymator T jest *estymatorem nieobciążonym* funkcji $g(\theta)$, jeśli

$$T_n \to g(\theta)$$

prawie wszędzie $\mathbb{P}_\theta$, to T_n jest *estymatorem mocno zgodnym* jeśli jest zbieżność według rozkładu (względem $\mathbb{P}_\theta$) to mówimy wtedy o *estymatorze zgodnym*.

Niech $X_1, X_2, \ldots, X_n$ będzie próbą prostą. *Średnią próbkową* nazywamy zmienną

$$\hat{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j,$$

natomiast *wariancja próbkowa*

$$\hat{S}_n = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \hat{X}_n)^2.$$

Jeśli istnieje pierwszy moment skończony $m_1 = m_1(\theta) = \mathbb{E}_\theta X_1$ to średnia jest nieobciążonym estymatorem μ_1; jeśli wariancja σ^2 jest skończona to $\hat{S}_n$ jest nieobciążonym estymatorem σ^2; ponadto te estymatory są mocno zgodne.

Do testowania i wyznaczania przedziałów ufności przydatne jest następujące twierdzenie graniczne (patrz np. Asmussen [5]). Przypomnijmy, że α_n jest n-tym momentem cenralnym zmiennej X_1.

Fakt 5.1 *Jeśli* $\mu_4 = \mathbb{E}\, X_1^4 < \infty$ *to*

$$\sqrt{n} \begin{pmatrix} \hat{X}_n - m_1 \\ \hat{S}_n - \sigma^2 \end{pmatrix} \to \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}),$$

gdzie

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & \alpha_3 \\ \alpha_3 & \alpha_4 - \sigma^4 \end{pmatrix}.$$

5.2 Przedziały ufności

Niech $X_1, \ldots, X_n$ będzie próbą prostą na $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$. *Przedziałem ufności na poziomie* α nazywamy odcinek $I = [T_L, T_R]$ spełniający warunki

- jego końce $T_L = T_L(X_1, \ldots, X_n) \leq T_R = T_R(X_1, \ldots, X_n)$ są funkcjami od próby i nie zależą od θ,
- $\mathbb{P}_\theta(T_L \leq \theta \leq T_R) \geq 1 - \alpha$, dla każdego $\theta \in \Theta$.

Oczywiście przedział I ufności jest szczególnym przypadkiem zbioru ufności (confidence region) . Ogólnie I może być zbiorem losowym zdefiniowanym przez próbę. Jest to pojęcie szczególnie przydatne w przypadku estymacji wielowymiarowej. W szczególności możemy rozpatrywać elipsoidy ufności.

Niech $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ będzie próbą prostą z n wektorów losowych o wartości w $\mathbb{R}^d$, z wektorem średniej $\boldsymbol{\mu}$ i miacierzą kowariancji $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1,\ldots,d}$. *Średnią próbkowa* nazywamy wektor

$$\hat{\boldsymbol{X}}_n = \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{X}_j,$$

natomiast próbkową macierzą kowariancji jest

$$\hat{\boldsymbol{\Sigma}}_n = \frac{1}{n-1} \sum_{m=1}^{n} (\boldsymbol{X}_m - \boldsymbol{\mu})^{\mathrm{T}} (\boldsymbol{X}_m - \boldsymbol{\mu}).$$

Średnia próbkowa jest estymatorem niobciążonym wektora wartości oczekiwanej $\boldsymbol{\mu}$, natomiast próbkową macierzą kowariancji jest estymatorem nieobciążnym macierzy kowariancji $\boldsymbol{\Sigma}$. Ponadto estymator $\hat{\boldsymbol{\Sigma}}_n$ jest mocno zgodny oraz

$$(\hat{\boldsymbol{X}}_m - \boldsymbol{\mu})^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_n^{-1} (\hat{\boldsymbol{X}}_m - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} \chi^2(d), \tag{5.2}$$

gdzie $\chi^2(d)$ jest rozkładem chi kwadrat z d stopniami swobody.

5.3 Statystyka Pearsona

Przypuśćmy, że mamy wektor losowy $\boldsymbol{X}_n = (X_{n,1}, \ldots, X_{n,d})$ o rozkładzie wielomianowym $\mathrm{M}(n, \boldsymbol{a})$. Definiujemy statystykę Pearsona

$$C_n(\boldsymbol{a}) = \sum_{j=1}^{d} \frac{(X_{n,j} - n a_j)^2}{n a_j}.$$

Zakładamy, że wszystkie $a_j > 0$. Wtedy (patrz np [44])

$$C_n(\boldsymbol{a}) \xrightarrow{\mathcal{D}} \chi^2(d-1),$$

gdzie $\chi^2(d-1)$ jest rozkładem chi kwadrat z $d-1$ stopniami swobody.

[L'Ecuyer [27] p. 50, [44]].

5.4 Dystrybuanty empiryczne

Dla próby $X_1, \ldots, X_n$ definiujemy *dystrybunatę empiryczną*

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le t), \qquad 0 \le t \le 1\ . \tag{5.3}$$

i statystyką testową

$$D_n = \sup_t |\hat{F}_n(t) - F(t)|\ .$$

Z twierdzenia Gliwienko–Cantelli mamy, że jeśli $X_1, \ldots, X_n$ tworzą próbę prostą z dystrybuantą F to, z prawdopodobieństwem 1, ciąg D_n zmierza do zera. Natomiast, przy założeniu, że F jest funkcją ciągłą, to unormowane zmienne $\sqrt{n}D_n$ zbiegają według rozkładu do rozkładu λ-Kołmogorowa z dystrybuantą

$$K(y) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2y^2}, \qquad y \ge 0\ .$$

[Renyi [36] p.493]

5.5 Wykresy kwantylowe

Coraz większą popularnościa przy weryfikacji czy obserwacje pochodzą z danego rozkładu cieszą się metody graficzne. Na przykład do tego świetnie nadają się wykresy kwantylowe lub inaczej wykres Q-Q . Możemy też przy użyciu tych wykresów weryfikować czy obserwacje z dwóch prób mają ten sam rozkład.

Aby otrzymać wykres kwantylowy należy wykreśleć w kwadracie kwadracie kartezjańskim $\mathbb{R} \times \mathbb{R}$ z jednej strony wartości funkcji kwantylowej dystrybuanty empirycznej utworzonej z obserwacji przeciwko z drugiej strony

teoretycznym wartościom funkcji kwantylowej. Przypomnijmy, że $F^{\leftarrow}(x)$ jest uogólnioną funkcja odwrotną. Jeśli F jest dystrybuantą, to funkcja Q_F zdefiniowana przez $Q_F(y) = F^{\leftarrow}(y)$ nazywa się *funkcją kwantylową* of F. Jednocześnie definiujemu empiryczną wersję funkcji kwantylowej przez rozpatrzenie uogólnionej funkcji od dystrybuanty empirycznej zdefiniowanej w (5.3); to jest $Q_n(y) = Q_{F_n}(y)$ tak że dla próby uporządkowanej $U_{(1)} \leq U_{(2)} \leq \ldots \leq U_{(n)}$ mamy

$$\left\{Q_n(y) = U_{(k)}\right\} = \left\{(k-1)n^{-1} < y \leq kn^{-1}\right\}.$$

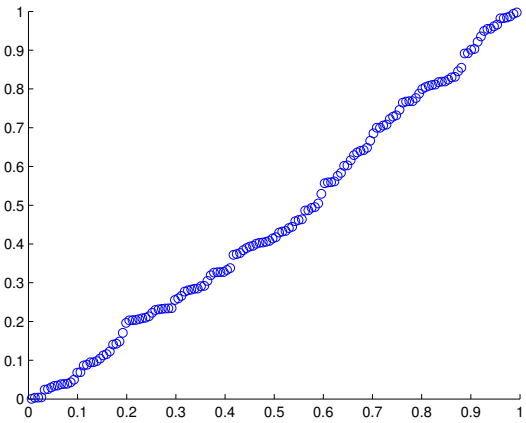
Wykres Q-Q jest więc zbiorem punktów $(Q_F(t), Q_n(t)) : 0 < t < 1$. Z definicji widzimy, że empiryczna funkcja kwantylowa $Q_n(y)$ będzie niemalejącą funkcją schodkową; jedynymi punktami wzrostu są wartości $\{k/n, 1 \leq k \leq n\}$, które tworzą kratę na osi x-ów. A więc wykres Q-Q tworzy też niemalejącą funkcję schodkową, ze skokami odpowiadającymi $t = k/n,\ 1 \leq k \leq n$. Wystarczy więc zrobić wykres w tych punktach tj. wykres $\{(Q_F(k/n), Q_n(k/n)) : k = 1, \ldots, n\}$. Jednakże ze względu na drobny problem, że dla $k = n$ mamy $Q_G(k/n) = \infty$, zalecana jest następujące modyfikacja; wykres należy zrobić w punktach $\{x_k = k/(n+1), 1 \leq k \leq n\}$.

Ogólna filozofia wykresów kwantylowych polega na zaobserwowaniu liniowości wykresu, co jest łatwe do zobaczenia. Na przykład jesli chcemy zweryfikowć, czy obserwacje pochodzą z rozkładu jednostajnego U(0,1). Wtedy oczywiście $Q_F(x) = x$ i robimy wykres punktów $(x_k, U_{(k)})_{k=1,\ldots,n}$. Na rys. 5.1 jest pokazany wykres Q-Q utworzony na podstawie 150 kolejnych rand.

Przykład 5.2 Dla standardowego rozkładu wykładniczego z dystrybuantą $G(x)$ funkcja kwantylowa ma postać $Q_G(y) = -\log(1-y)$ jeśli $0 < y < 1$. Jeśli chcemy porównać wartości próbkowe z tymi od standardowego rozkładu wykładniczego wystarczy rozpatrzyć dwie funkcje kwantylowe. Wykreślamy więc dwie funkcje Q_G and Q_n.

Jeśli nasze obserwacje będą pochodzić z niekoniecznie standardowego rozkładu wykładniczego z parametrem λ tj. z $Q_F(y) = -\lambda^{-1}\log(1-y)$, to spodziewamy się, że będą układały się wzdłuż pewnej prostej. Nachylenie tej prostej będzie λ^{-1}. Jesli nasze obserwacje pochodzą z rozkładu o cięższych ogonach, to możemy spodziewać się wykresu rosnącego szybciej niż prosta, jeśli z lżejszego to wzrost będzie wolniejszy. Na rys. 5.1 jest pokazany wykres Q-Q utworzony na podstawie 150 kolejnych -log(rand)/2.

Ogólnie przez "rozkład teoretyczny" vs "obserwacje" wykres kwantylowy będziemy rozumieli wykres QQ, gdzie na osi x-ów mamy teoretyczną funkcję



Rysunek 5.1: Jednostajny vs obserwacje jednostajne; 150 obserwacji.

kwantylową a na osi y-ków mamy funkcję kwantylową zbudowaną na bazie dystrybuanty empirycznej obserwacji. Możemy mieć też dwi próby. Wtedy będziemy mieli do czynienia "obserwacje I" vs "obserwacje II" wykres kwantylowy.

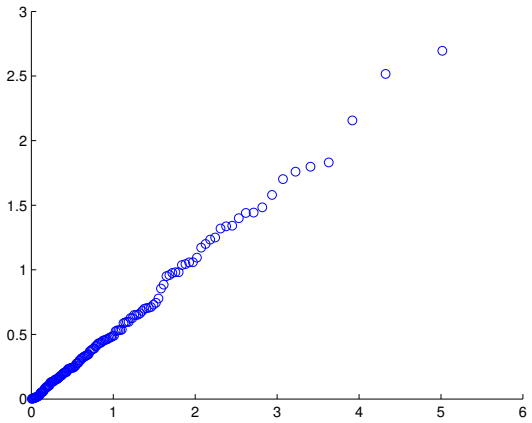
Jeśli obserwacje nie są dobrze zaprezentowane przez rozkład wykładniczy (patrz na przykład rys. 5.3 jak wyglądają na wykresie Q-Q obserwacje pochodzące z Pareto(1.2) ustandaryzowane na średnią 1 przeciwko wykładniczej funkcji kwantylowej), to możemy weryfikować inne rozkłady. Wśród popularnych kandydatów znajdują się rozkład normalny, rozkład logarytmiczno normalny oraz rozkład Weibulla. Krótko omówimy te przypadki osobne.

- *Normalny wykres kwantylowy.* Przypomnijmy, że standardowy rozkład normalny ma dystrybuantę

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-y^2/2}\, dy$$

i niech $\Phi^{-1}(y)$ będzie odpowiadająca mu funkcją kwantylowa. Wtedy standardowy normalny wykres kwantylowy będzie się składał z punktów

$$\left\{ \left(\Phi^{-1}(k/(n+1)), U_{(k)} \right),\ 1 \le k \le n \right\}. \tag{5.4}$$

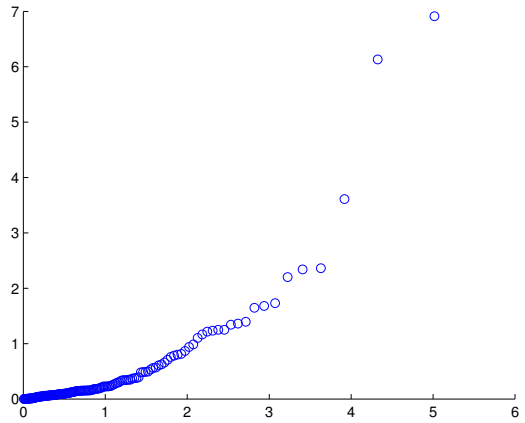


Rysunek 5.2: wykładniczy vs obserwacje wykladnicze; 150 obserwacji.

Zauważmy, że dla ogólnego rozkładu normalnego $N(\mu, \sigma^2)$ funkcja kwantylowa $Q(y) = \mu + \sigma\Phi^{-1}(y)$ (pokazać!). Tak więc jeśli normalny wykres kwantylowy jest w przybliżeniu linią, to jej nachylenie da oszacowanie parametru σ podczas gdy przecięcie w 0 daje oszacowanie μ. Na rys. 5.4 pokazany jest normalny wykres kwantylowy dla obserwacji pochodzących z rozkładu normalnego N(1,2). Proponujemy czytelnikowi odczytać a wykresu, że $m = 1$ i $\sigma = 2$.

- *Logarytmiczno-normalny wykres kwantylowy.* Jest to aparat diagnostyczny do wykrywania logarytmiczno normalengo rozkładu. Wykorzystujemy następującą własność. Zmienna X jest logarytmicnzo-normalna wtedy i tylko wtedy gdy $\log X$ jest normalnie rozłożoną zmienna losową. A więc wykres kwantylowy będzie zadany przez $\{(\Phi^{-1}(k/(n+1)), \log X_{(k)}),\ 1 \leq k \leq n\}$.

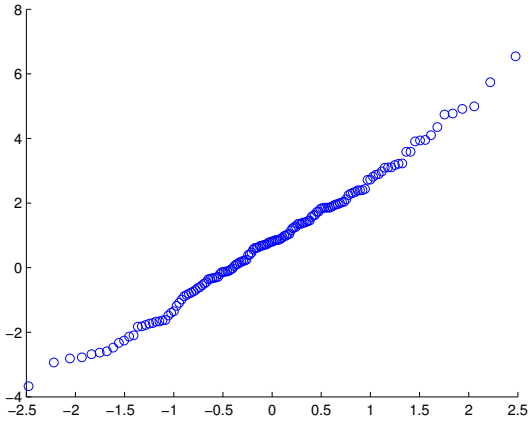
- *Paretowski wykres kwantylowy* jest często wykorzystywanym narzędziem. Przyponijmy, że jeśli U ma rozkład Pareto $\mathrm{Par}(\alpha, c)$ to $\mathbb{P}(U \leq x) = 1 - ((1+x)/c)^{-\alpha}$ dla $x \geq c$, a więc $\log Q(y) = \log c - \alpha^{-1} \log(1-y)$. Paretowski wykres kwantylowy otrzymujemy rysując wykres punktów $\{(-\log(1 - k/(n+1)), \log U_{(k)}),\ 1 \leq k \leq n\}$. Jesli dane pochodzą z rozkładu Pareto to wykres będzie miał kształt prostej przecinającej o na wysokości $\log c$ i nachyleniu α^{-1}. Zauważmy, że użycie Paretowskiego wykresu kwantylowego ma również sens gdy rozkład U jest



Rysunek 5.3: Wykladniczy vs obserwacje Pareto; 150 obserwacji.

Pareto podobny, tj. $\mathbb{P}(U > x) \sim x^{-\alpha} L(x)$, gdzie L jest funkcją wolno zmieniającą się.

- The *Weibullowski wykres kwantylowy* Przypmnijmy, że rozkład ogona w tym wypadku wynosi $\overline{F}(x) = \exp(-cx^r)$. Funkcja kwantylowa jest wtedy $Q(y) = \left(-c^{-1}\log(1-y)\right)^{1/r}$ dla $0 < y < 1$. Jesli weźmiemy logarytm z tego wyrażenia, to znajdujemy, że $\log Q(y) = -r^{-1}\log c + r^{-1}\log(-\log(1-y))$ które automatycznie prowadzi do Weibullowskiego wykresu kwantylowego $\{(\log(-\log(1-k/(n+1))), \log U_{(k)}),\ 1 \le k \le n\}$. Jeśli dane pochodzą z rozkładu Weibulla, to powinniśmy się spodziewać linii prostej o nachyleniu r^{-1} i przecinającej . $-r^{-1}\log c$.



Rysunek 5.4: Normalny vs obserwacje normalne; 150 obserwacji.

5.6 Własność braku pamięci rozkładu wykładniczego

Niektóre algorytmy istotnie wykorzystują tzw. *brak pamięci* rozkładu wykładniczego czasu obsługi i opuszczenie tego założenia będzie wymagało istotnych modyfikacji. Natomiast możemy założyć, że są znane momenty napływania zadań $A_1 < A_2 < \ldots$. Tak jest na przykład gdy zgłoszenia są zgodne z niejednorodnym Procesem Poissona. Tak jest na przykład gdy odstępy między zgłoszeniami są niezależne o jednakowym rozkładzie (niewykładniczym). W każdym zdarzeniu (niech będzi ono w chwili t), gdy dotychczasowy algorytm mówił, że należy generować X i podstawić $t_A = t + X$, należy podstawić najwcześniejszy moment zgłoszenia zadania po chwili t.

Przy konstrukcji algorytmu dla systemów z jednym serwerem wykorzystywaliśmy tak zwaną własność *braku pamięci* rozkładu wykładniczego. Własność można sformułować następująco: dla:

$$\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t), \qquad s, t \geq \ .$$

Łatwo pokazać, że własnośc jest prawdziwa dla zmiennej losowej o rozkładzie wykładniczym $X \sim$Exp(λ). Okazuje się również, że rozkład wykładniczy jest jedynym rozkładem o takiej własności.

Będzie też przydatny następujący lemat.

Lemat 5.3 *Jeśli $X_1, \ldots, X_n$ są niezależnymi zmiennymi losowym oraz $X_i \sim$ Exp(λ_i), to zmienne*

$$T = \min_{i=1,\ldots,n} X_i, \qquad I = \arg \min_{i=1,\ldots,n} X_i$$

są niezależne oraz $T \sim$ Exp($\lambda_1 + \ldots\ldots + \lambda_n$) i $\mathbb{P}(J = j) = \lambda_j/(\lambda_1 + \ldots\ldots + \lambda_n)$.

Dowód Mamy

$$\begin{aligned}
\mathbb{P}(T > t, J = j) &= \mathbb{P}(\min_{i=1,\ldots,n} X_i > t, \min_{i \neq j} X_i > X_j) \\
&= \int_0^\infty \mathbb{P}(\min_{i=1,\ldots,n} X_i > t, \min_{i \neq j} X_i > x) \lambda_j e^{-\lambda_j x} \, dx \\
&= \int_t^\infty \mathbb{P}(\min_{i \neq j} X_i > x) \lambda_j e^{-\lambda_j x} \, dx \\
&= \int_t^\infty e^{-\sum_{i \neq j} \lambda_i x} \lambda_j e^{-\lambda_j x} \, dx \\
&= e^{-\sum_i \lambda_i t} \frac{\lambda_j}{\sum_i \lambda_i}
\end{aligned}$$

6 Procesy kolejkowe

Zajmiemy się symulacją procesów które występują na przykład w zastosowaniach telekomunikacyjnych i przemyslowych. W tym celu będzie nam pomocny język i formalizm tzw. teorii kolejek. Teoria kolejek jest działem procesów stochastycznych. Zajmuje się analizą systemów, do których wchodzą/wpływają zadania/pakiety/itd mające być obsługiwane, i które tworzą kolejkę/wypełniają bufor, i które doznają zwłoki w obsłudze. Zajmiemy sie teraz przeglądem podstawowych modeli i ich rozwiązań.

6.1 Standardowe pojęcia i systemy

Zadania zgłaszają się w momentach $A_1 < A_2 < \ldots$ i są rozmiarów $S_1, S_2, \ldots$. Odstępy między zgłoszeniami oznaczamy przez $\tau_i = A_i - A_{i-1}$. Zadania są obsługiwane według *regulaminów* jak na przykład:

- zgodnie z kolejnościa zgłoszeń (FCFS -First Come First Seved)
- obsługa w odwrotnej kolejności zgłoszeń (LCFS - Last Come First Served)
- w losowej kolejności (SIRO),
- z podziałem czasu (PS), tj jesli w systemie jest n zadań to serwer poświęca każdemy zadaniu $1/n$ swojej mocy.
- najdłuższe zadanie najpierw – LPTF (longest processing time first),
- najkrótsze zadanie najpierw – SPTF (longest processing time first),

Zakładamy tutaj, że zadania są natychmiast obsługiwane jeśli tylko serwer jest pusty, w przeciwnym razie oczekuje w buforze. W przypadku większej liczby serwerów będziemy przyjmować, że są one jednakowo wydolne, to znaczy nie odgrywa roli w przypadku gdy kilka jest wolnych, który będzie obsługiwał zadanie. Bufor może być nieskończony lub z ograniczoną liczba miejsc. W tym ostatnim przypadku, jeśli bufor w momencie zgłoszenia do systemu jest pełny to zadanie jest stracone. Można badać:

- $L(t)$ – liczba zadań w systemie w chwili t,
- $L_q(t)$ – liczba zadań oczekujących w buforze,

- W_n – czas czekania n-tego zadania,
- D_n – czas odpowiedzi dla i-tego zadania (czas od momentu zgłoszenia do momentu wyjścia zadania z systemu)
- $V(t)$ – *załadowanie* (workload) w chwili t, co jest sumą wszystkich zadań które zostały do wykonania (wliczając to co zostało do wyknania dla już obsługiwanych).

W przypadku systemów ze stratami, bada się prawdopodobieństwo, że nadchodzące zadanie jest stracone (blocking probability lub loss probability).

Przez $\bar{l}$, $\bar{w}$, $\bar{d}$, $\bar{v}$ będziemy oznaczać odpowiednio w warunkach stabilności: średnią liczbę zadań w systemie, średni czas czekania, średni czas odpowiedzi, średnie załadowanie.

7 Przegląd systemów kolejkowych

7.1 Systemy z nieskończoną liczbą serwerów

System w którym każde zgłoszenie ma przydzielony natychmiast serwer do obsługi. A więc nie ma czekania. Nieciekawy jest równiez czas odpowiedzi wynoszący tyle ile rozmiar zadania. Będziemy więc pytać się o liczbę $L(t)$ zadań w chwili t. Zadania zgłaszają się w momentach $0 \leq A_1 < A_2 < \ldots$ i ich rozmiary wynoszą odpowiednio $S_1, S_2, \ldots$. A więc i-te zadanie jest w systemie w odcinku $[A_i, A_i + S_i)$. Możemy formalnie zdefiniować

$$L(t) = \sum_{j=1}^{\infty} \mathbf{1}_{[A_i, A_i+S_i)}(t).$$

Jeśli interesujemy się procesem $L(t)$ jedynie w odcinku $(0, t_{\max}]$ to wtedy musimy znaleźć n takie że $A_n \leq t, A_{n+1} > t$; wtedy i-te zadanie jest w systemie w odcinku $[A_i, \min(t_{\max}, A_i + S_i))$. Algorytm na symulację $L(t)$ gdy dane są procedury generowania (A_i) oraz (S_i) jest następujący. Piszemy $A_i = A(i)$ oraz $S_i = S(i)$.

```
l=0;
A=-\log rand/lambda;
while A<=tmax
generate S;
```

```
l=l+min(tmax,A+S)-A;
generate A;
end
lbar=l/tmax
```

M/M/∞ Najłatwiejszym do pełnej analizy jest system gdy odstępy między zgłoszeniami (τ_i) są niezależnymi zmiennymi losowymi o jednakowym rozkładzie wykładniczym z parametrem λ. Wtedy zgłoszenia są zgodne z procesem Poissona z intensywnością λ. Czasy obsług sa niezależne o jednakowym rozkładzie μ. Wtedy mamy, niezaleznie od wielkości $\rho = \lambda/\mu$

- $\lim_{t\to\infty} \mathbb{P}(L(t) = k) = \frac{\rho^k}{k!} e^{-\rho}$ oraz
- $\lim_{t\to\infty} \mathbb{E}\, L(t) = \rho$.

Bardziej interesujące z punkty widzenia symulacji jest istnienie granic z prawdopodobieństwem 1

- $$\lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{1}(L(t) = k) = \frac{\rho^k}{k!} e^{-\rho} \tag{7.5}$$

  oraz

- $$\lim_{t\to\infty} \frac{1}{t} \int_0^t L(t) = \rho \tag{7.6}$$

  .

M/G/∞ Proces zgłoszeń zadań jest taki sam jak w M/M/∞, jednakże rozmiary kolejnych zadań $(S_j)_j$ są niezależne o jednakowym rozkładzie G. Musimy założyć, że $\mathbb{E}\, S_1 = \mu^{-1} < \infty$. W tym systemie jest nieograniczony zasób serwerów a więc żadne zadanie nie czeka i nie ma sensu mówić o czasie czekania który jest zerowy i czasie odpowiedzi, który równy jest rozmiarowi zadania. Wtedy mamy, niezależnie od wielkości $\rho = \lambda \mathbb{E}\, S < \infty$

- $\lim_{t\to\infty} \mathbb{P}(L(t) = k) = \frac{\rho^k}{k!} e^{-\rho}$ oraz
- $$\lim_{t\to\infty} \mathbb{E}\, L(t) = \rho \tag{7.7}$$

  ,

i również są prawdziwe wzory (**[?? eq:erodicM/M/infty.1??]**), (7.6).

Proces $L(t)$ będzie stacjonarny jeśli jako $L(0)$ weźmiemy zmienną losową o rozkładzie Poissona Poi(ρ). Wtedy funkcja kowariancji [1]

$$R(t) = \mathbb{E}\left[(L(s)-\rho)(L(s+t)-\rho)\right] \tag{7.8}$$

$$= \lambda \int_t^\infty (1-G(v))\,dv\ . \tag{7.9}$$

Ponandto znamy wzór na obciążenie jeśli w chwili $t = 0$ nie ma żadnych zadań, tj. $L(0) = 0$:

$$\rho - \mathbb{E}\,L(t) = \lambda \int_t^\infty (1-G(s))\,ds\ .$$

W szczególności dla M/D/∞, tj. gdy rozmiary zadań są deterministyczne mamy $R(t) = \rho(1-\mu t)$ dla $t \le \mu^{-1}$ oraz $R(t) = 0$ dla $t > \mu^{-1}$. [2]

7.2 Systemy ze skończoną liczbą serwerów

M/M/1 Najłatwiejszym do pełnej analizy jest system gdy odstępy między zgłoszeniami (τ_i) są niezależnymi zmiennymi losowymi o jednakowym rozkładzie wykładniczym z parametrem λ. To założenie odpowiada, że zgłoszenia są zgodne z procesem Poissona z intensywanością λ. Czasy obsług są niezależne o jednakowym rozkładzie μ. Wtedy, między innymi, wiadomo że w przypadku $\lambda < \mu$ (lub równoważnie $\rho < 1$, gdzie $\rho = \lambda/\mu$), niezależnie od wyżej wymienionej dyscypliny obsługi:

- $p_k = \lim_{t\to\infty} \mathbb{P}(L(t) = k) = (1-\rho)\rho^k$ $(k = 0, 1, 2, \ldots)$ oraz

$$\bar{l} = \lim_{t\to\infty} \mathbb{E}\,L(t) = \frac{\lambda}{\mu-\lambda} \tag{7.10}$$

  ,

- $\bar{l}_q = \lim_{t\to\infty} \mathbb{E}\,L_q(t) = \frac{\lambda^2}{\mu(\mu-\lambda)}$.

Natomiast w przypadku dyscypliny FCFS mamy

- $\bar{w} = \bar{\lim}_{n\to\infty} \mathbb{E}\,W_n = \frac{\lambda}{\mu(\mu-\lambda)}$,

[1] Cytowanie za Whittem [46]

[2] Napisac procedure generowania relalizacji M/G/infty.

- $\bar{d} = \lim_{n\to\infty} \mathbb{E}\, D_n = \frac{1}{\mu-\lambda}$.

Ponadto mamy z prawdopobieństwem 1 zbieżność

- $\lim_{t\to\infty} \int_0^t L(s)\, ds/t = \frac{\lambda}{\mu-\lambda}$
- $\lim_{t\to\infty} \int_0^t L_q(s)\, ds/t = \frac{\lambda^2}{\mu(\mu-\lambda)}$

a w przypadku dyscypliny FCFS mamy

- $\lim_{k\to\infty} \sum_{j=1}^k W_j/k = \frac{1\lambda}{\mu(\mu-\lambda)}$
- $\lim_{k\to\infty} \sum_{j=1}^k d_j/k = \frac{1}{\mu-\lambda}$

M/M/c Proces zgłoszeń zadań oraz ich rozmiary są takie same jak w M/M/1. Natomiast w tym systemie jest c serwerów obsługujących z tą samą prędkością. Podobnie jak dla M/M/1 możemy pokazać, że, jeśli obługa jest zgodna z kolejnościaą zgłoszeń, to, przy założeniu $\rho < c$

$$\bar{w} = \lim_{n\to\infty} \mathbb{E}\, W_n = \frac{p_W}{\mu(c-\rho)}$$

i mamy również z prawdopdobieństwem 1 zbieżność

$$\lim_{k\to\infty} \sum_{j=1}^k W_j/k = \bar{w}\ .$$

Mamy rówież dla dyscypliny FCFS ale i również pewnej klasy dyscyplin obsługi

$$p_k = \lim_{t\to\infty} \mathbb{P}(L(t) = k) \quad = \quad \begin{cases} \frac{\rho^k}{k!} p_0, & k = 0, \ldots, c \\ \frac{\rho^k}{c! c^{k-c}} p_0, & k > c. \end{cases}$$

gdzie

$$p_0 = \left[\sum_{k=0}^{c} \frac{\rho^k}{k!} + \frac{\rho^{c+1}}{c!(c-\rho)} \right]^{-1} \quad .$$

oraz z prawdopodobieństwem 1

$$\lim_{t\to\infty} \int_0^t L(s)\, ds/t = \ .$$

GI/G/1 Jest to system taki, że $(\tau_k)_k$ ($A_1 = \tau_1, A_2 = \tau_1 + \tau_2, \ldots$) są niezależnymi zmiennymi losowymi o jednakowym rozkładzie F a rozmiary zadań $(S_k)_k$ tworzą ciąg niezależnych zmiennych losowych o jednakowym rozkładzie G. Przypuśćmy, że w chwili $t = 0$ system jest pusty, i pierwsze zadanie przychodzi w momencie A_1. Wtedy oczywiście czas czekania tego zadania wynosi $W_1 = 0$. Dla systemów jednokanałowych prawdziwa jest rekurencja

$$W_{k+1} = (W_k + S_k - \tau_{k+1})_+. \tag{7.11}$$

Można rozpatrywać alternatywne założenia początkowe. Na przykład, przypuścić, że w chwili $A_0 = 0$ mamy zgłoszenie zadania rozmiaru S_0, które będzie musiało czekać przez czas W_0. W takim przypadku rozpatruje się rózne załozenia o rozkładzie W_0, S_0 i A_1, ale zawsze trzeba pamiętać niezależności ażeby poniższe fakty były prawdziwe. Natomiast rekurencja (7.11) jest niezależna od wszlkich założeń probabilistycznych. Przy założeniach jak wyżej W_k tworzy łańcuch Markowa i jeśli $\mathbb{E}\, S_1 < \mathbb{E}\, \tau_1$, to istnieją granice $\lim_{k\to\infty} \mathbb{P}(W_k \le x)$ oraz, jeśli $\mathbb{E}\, S_1^2 < \infty$

$$\bar{w} = \lim_{k\to\infty} \mathbb{E}\, W_k \, .$$

Ponadto z prawdopodobieństwem 1

$$\bar{w} = \lim_{k\to\infty} \sum_{j=1}^{k} W_j/k \, .$$

W przypadku, gdy τ_1 ma rozkład wykładniczy z $\mathbb{E}\, \tau_1 = 1/\lambda$ to

$$\bar{w} = \frac{\lambda \mathbb{E}\, S_1^2}{2(1 - \lambda/\mathbb{E}\, S_1)} \, . \tag{7.12}$$

Wzór (7.12) jest zwany wzorem Pollaczka-Chinczyna. Jeśli S ma rozkład wykładniczy z parametrem μ, to wtedy powyższy wzór pokrywa się z odpowiednim wzorem dla M/M/1.

W szczególności jeśli rozmiary zadań są niezależne o jednakowym rozkładzie wykładniczym, to oznaczamy taki system przez GI/M/1. Jeśli proces zgłoszeń zadań jest procesem Poissona, to wtedy piszemy M/G/1. Jeśli zarówno proces wejściowy jest Poissonowski jak i rozmiary zadań są wykładnicze, to piszemy M/M/1.

GI/G/c Jest to system taki, że odstępy między zgłoszeniami są $(\tau_k)_k$ $(A_1 = \tau_1, A_2 = \tau_1 + \tau_2, \ldots)$ a rozmiary zadań $(S_k)_k$. Jest c serwerów, i zadania oczekujące w buforze do obługi zgodnie z kolejnośćią zgłoszeń. Zamiast rekurencji (7.11) mamy następującą. Oznaczmy $\boldsymbol{V}_n = (V_{n1}, V_{n2}, \ldots, V_{nc})$, gdzie $0 \le V_{n1} \le V_{n2} \le \ldots \le V_{nc}$. Składowe tego wektora reprezentują dla n-go zadania (uporządkowane) załadowanie każdego z serwerów. A więc przychodząc idzie do najmniej załadowanego z załadowaniem V_{n1} i wobec tego jego załadowanie tuż po momencie zgłoszenia wyniesie $V_{n1} + S_n$. Po czasie τ_{n+1} (moment zgłoszenia $n+1$-szego) załadowania wynoszą $V_{n1} + S_n - \tau_{n+1}, V_{n2} - \tau_{n+1}, \ldots, V_{nc} - \tau_{n+1}$. Oczywiście jesli są ujemne to kładziemy 0, tj $(V_{n1} + S_n - \tau_{n+1})_+, (V_{n2} - \tau_{n+1})_+, \ldots, (V_{nc} - \tau_{n+1})_+$. Teraz musimy wyniki uporządkować. Wprowadzamy więc operator $\mathcal{R}(\boldsymbol{x})$, który wektor $\boldsymbol{x}$ z $\mathbb{R}^c$ przedstawia w formie uporządkowaej tj. $x_{(1)} \le \ldots \le x_{(c)}$. A więc dla systemu z c serwerami mamy rekurencję

$$\begin{aligned}&(V_{n+1,1}, V_{n+1,2}, \ldots, V_{n+1,c})\\ &\quad = \quad \mathcal{R}((V_{n1} + S_n - \tau_{n+1})_+, (V_{n2} - \tau_{n+1})_+, \ldots, (V_{nc} - \tau_{n+1})_+ \, .\end{aligned}$$

Przypuśćmy, że w chwili $t = 0$ system jest pusty, i pierwsze zadanie przychodzi w momencie A_1. Wtedy oczywiście czas czekania tego zadania wynosi $V_1 = V_2 = \ldots = V_c = 0$. Rekurencja powyższa jest prawdziwa bez założeń probabilitycznych. Jesli teraz założymy, że ciąg (τ_n) i (S_n) są niezależne i odpowiednio o jednakowym rozkładzie F i G, oraz jeśli $\rho = \frac{\mathbf{E}\, S}{\mathbf{E}\, T} < c$, to system się stabilizuje i istnieje rozkład stacjonarny.

7.3 Systemy ze stratami

M/M/c ze stratami Proces zgłoszeń zadań oraz ich rozmiary są takie same jak w M/M/1. Natomiast w tym systemie jest c serwerów obsługujących z tą samą prędkośćia, ale nie ma bufora. Jeśli przychodzące zadanie zastają wszystkie serwery zajęte, to jest stracone. Można pokazać istnienie granicy

$$p_k = \lim_{t\to\infty} \mathbb{P}(L(t) = k) = \frac{\rho^k/k!}{\sum_{j=0}^{c} \rho^j/j!}, \qquad k = 0, \ldots, c$$

gdzie $\rho = \lambda/\mu$, i teraz ta granica istnieje zawsze niezależnie od wielkości ρ. W szczególności mamy też

$$\lim_{t\to\infty} \frac{\int_0^t \mathbf{1}(L(s) = c)\, ds}{t} = p_c \, .$$

Dla tego systemu szczególnie ciekawą charakterystyką jest tzw. prawdopdobieństwo straty, które możemy w intuicyjny sposób zdefiniować następująco. Niech

$$J_k = \begin{cases} 1 & k\text{-te zadanie jest stracone} \\ 0 & \text{w przeciwnym razie.} \end{cases} \tag{7.13}$$

Wtedy z prawdopodobieństwem 1 istnieje granica

$$p_{\text{block}} \quad = \quad \lim_{k\to\infty} \sum_{j=1}^{k} J_j/k. \tag{7.14}$$

Prawdopodobieństwo p_{block} jest stacjonarnym prawdopdobieństwem, że przychodzące zadanie jest stracone. Odróżnijmy p_{block} od p_c, które jest stacjonarnym prawdopodobieństwem, że system jest pełny, chociaż dla M/M/c ze stratami te prawdopodobieństwa się pokrywają. Tak nie musi być, jeśli zadania zgłaszają się do systemu nie zgodnie z procesem Poissona. Okazuje się, że tzw. wzór Erlanga

$$p_{\text{block}} = \frac{\rho^c/c!}{\sum_{j=0}^{c} \rho^j/j!}$$

jest prawdziwy dla systemów M/G/c ze stratami, gdzie rozmiary kolejnych zadań tworzą ciąg niezaleznych zmiennych losowych z jednakową dystrybuantą G, gdzie $\mu^{-1} = \int_0^\infty x\, dG(x)$ jest wartością oczekiwaną rozmiaru zadania.

8 Błądzenie przypadkowe i technika zamiany miary

8.1 Błądzenie przypadkowe

Zacznijmy od pojęcia *błądzenia przypadkowego*. Niech $\xi_1, \xi_2, \ldots$ będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie. *Błądzeniem przypadkowym* nazywamy ciąg $\{S_n, n = 0, 1, \ldots\}$ zdefiniowany następująco:

$$S_0 = 0, \quad S_n = \xi_1 + \ldots + \xi_n, \quad n = 1, 2, \ldots.$$

Będziemy zakładać, że istnieje i jest skończony pierwszy moment $\mathbb{E}\,\xi_1$. Podamy teraz tzw. twierdzenie ..... Dowód pierwszych dwóch przypadków

jest prostym zastosowaniem prawa wielkich liczb Kołmogorowa. Natomiast dowód trzeciego przypadku jest nietrywialny i można go na przykład znaleźć w książce [39].

Twierdzenie 8.1 *Z prawdopodobieństwem 1 mamy*

$$\lim_{n\to\infty} S_n = \infty, \qquad gdy\ \mathbb{E}\,\xi_1 > 0,$$

$$\lim_{n\to\infty} S_n = -\infty, \qquad gdy\ \mathbb{E}\,\xi_1 < 0,$$

oraz

$$\limsup_{n\to\infty} S_n = \infty, \quad \liminf_{n\to\infty} = -\infty \qquad gdy\ \mathbb{E}\,\xi_1 = 0.$$

Teraz niech $a < 0 < b$ i rozpatrzmy błądzenie $\{S_n\}$ do momentu opuszczenia przez niego przedziału (a, b). Błądzenie może opuścić przedział (a, b) albo przeskakując przez b lub przeskakując przez a. Interesujące jest pytanie o prawdopodobieństwo opuszczenia na przykład przez b. Formalnie definiujemy to prawdopodobieństwo następująco. Niech $\tau = \min\{i \geq 1 : S_i \notin (a, b)\}$, będzie momentem opuszczenia przedziału. Szukane zdarzenie to

$$A = \{S_\tau \geq b, \tau < \infty\}.$$

Oczywiście jesli a, b są skończone, to z twierdzenia 8.1 mamy, że $\mathbb{P}(\tau < \infty) = 1$ ale będzie nas interesował również przypadek, gdy $a = -\infty$. Wtedy oczywiście, również z twierdzenia 8.1, mamy

$$\begin{aligned} \mathbb{P}(\tau < \infty) &= 1 \quad \text{jeśli } \mathbb{E}\,\xi_1 \geq 0, \\ \mathbb{P}(\tau < \infty) &< 1 \quad \text{jeśli } \mathbb{E}\,\xi_1 < 0. \end{aligned}$$

Dla zadania z $a = -\infty$, dalej zwanego zadaniem 1-szym będziemy liczyć $\mathbb{P}(\tau < \infty)$, natomiast dla zadania z a, b skończonymi, będziemy liczyć zadanie 2-go które ma na celu obliczyć $\mathbb{P}(S_\tau \geq b)$.

Zauważmy jeszcze, gdy $\mathbb{P}(\xi_1 = 1) = p$ i $\mathbb{P}(\xi_1 = -1) = q = 1 - p$, to błądzenie przypadkowe generowane przez ciąg $\xi_1, \ldots$ nazywamy *prostym symetrycznym*.

8.2 Technika wykładniczej zamiany miary

Kontynuując rozważania z podrozdziału X.2.1 sformułujemy teraz ważną tożsamość, zwaną tożsamością Walda, prawdziwą dla zmiennych ν będących czasem zatrzymania dla $\{S_n\}$. Niech A będzie zdarzeniem takim, że

$A \cap \{\nu = k\} = \{(\xi_1, \ldots, \xi_k) \in B^k\}$ dla $B^k \subset \mathbb{R}^k$, $k = 0, 1, \ldots$. Oznaczamy to przez $A \in \mathcal{F}_\nu$. W dalszej częśći wykładu będziemy korzystali z $\nu = \tau$, który jest czasem wyjścia błądzenia przypadkowego z (a, b), oraz $A = S_\nu \geq b$ lub $A = \Omega$. Przedstawimy teraz ważną tożsamość, która jest prawdziwa dla dowolnego rozkładu ξ, byleby $\theta \in (\theta_1, \theta_2)$.

Fakt 8.2 (Tożsamość Walda analizy sekwencyjnej) *Jeśli $A \subset \{\tau < \infty\}$, to*

$$\mathbb{P}(A) = \tilde{\mathbb{E}}^{\theta}[\hat{m}^{\nu}(\theta) e^{-\theta(\xi_1 + \ldots + \xi_\nu)}; A]. \tag{8.15}$$

Wniosek 8.3

$$\mathbb{P}(\tau < \infty) = \tilde{\mathbb{E}}^{\theta}[\hat{m}^{\tau}(\theta) e^{-\theta(\xi_1 + \ldots + \xi_\tau)}; \tau < \infty]. \tag{8.16}$$

Przypomnijmy, że $\mathbb{E}\,\xi_1 < 0$. Zbadajmy teraz funkcję $\hat{m}(\theta)$ w przedziale $\theta_1 < \theta < \theta_2$). Ponieważ

$$\frac{d\hat{m}(\theta)}{d\theta}_{|\theta=0} = \mathbb{E}\,\xi_1 < 0,$$

więc funkcja $\hat{m}(\theta)$ w $\theta = 0$ równa się 1 i w tym punkcie maleje. Dalej będziemy przyjmować następujące założenie.

Założenie $\mathbb{E}\,\xi_1 < 0$ i istnieje $\gamma > 0$ dla której $\hat{m}(\gamma) = 1$, oraz

$$\frac{d\hat{m}(\theta)}{d\theta}_{|\theta=\gamma} < \infty.$$

Przy tym złożeniu funkcja $\hat{m}(\theta)$ ma przebieg następujący. W przedziale $[0, \gamma]$ funkcja ściśle maleje od 0 do θ_0 oraz ściśle rośnie od θ_0 do $\gamma + \epsilon$ dla pewnego $\epsilon > 0$. W $\theta = 0, \gamma$ przyjmuje wartość 1.

Lemat 8.4 *Dla zmiennej losowej ξ_1 określonej na przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \tilde{\mathbb{P}}^{\theta})$ funkcja tworząca momenty*

$$\tilde{\mathbb{E}}^{\theta} e^{s\xi_1} = \hat{m}(s + \theta).$$

Stąd

$$\tilde{\mathbb{E}}^{\theta}\xi_1 = \frac{d\hat{m}(\theta + s)}{ds}_{|s=0} > 0, \quad dla \;\; \theta > \theta_0.$$

Dowód

Wniosek 8.5 *Dla $\theta > \theta_0$ mamy $\tilde{\mathbb{P}}^{\theta}(\tau < \infty) = 1$. Stąd dla zadania Y*

$$\mathbb{P}(\tau < \infty) = \mathbb{E}^{\theta}[\hat{m}^{\tau}(\theta)\tilde{e}^{-\theta(\xi_1+\ldots+\xi_\tau)}].$$

W szczególności

$$\mathbb{P}(\tau < \infty) = \tilde{\mathbb{E}}^{\gamma} e^{-\gamma(\xi_1+\ldots+\xi_\tau)}.$$

Natomiast dla zadania 2-go moment zatrzymania τ jest zawsze skończony i dlatego obliczenie $\mathbb{P}(\tau < \infty)$ jest nieciekawe bo wielkość ta równa się 1. Natomiast potrzebne są inne wielkości.

Wniosek 8.6

$$\mathbb{P}(S_\tau \geq b) = \tilde{\mathbb{E}}^{\theta}[\hat{m}^{\nu}(\theta)e^{-\theta(\xi_1+\ldots+\xi_\tau)}; S_\tau \geq b].$$

W szczególności

$$\mathbb{P}(S_\tau \geq b) = \tilde{\mathbb{E}}^{\gamma}[e^{-\gamma(\xi_1+\ldots+\xi_\tau)}; S_\tau \geq b].$$

Jak się okaże, wybór $\theta = \gamma$ jest w pewnym sensie optymalny.

Bibliografia

[1] .

[2] Abramowitz M. & Stegun I. (1965) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards.

[3] Aldous D. (1987) Ultimate instability of exponential back-off protocol for acknowledgment-based transmission control of random access communication channels. *IEEE Trans. on Inform. Theory* **IT-33**: 219–223.

[4] Asmussen S. & Glynn P. (2007) *Stochastic Simulations; Algorithms and Analysis*. Springer, New York.

[5] Asmussen S. (1999) *Stochastic Simulation with a view towards Stochastic Processes*. MaPhySto, University of Aarhus, www.maphysto.dk/cgi-bin/w3-msql /publications/genericpublication.html?publ=89.

[6] Bertsimas D. & Tsitsiklis J. (1993) Simulated annealing. *Statistical Science* **8**: 10–15.

[7] Bratley P., Fox B., and Schrage L. (1987) *A Guide to Simulation*. Springer, New York.

[8] Bremaud P. (1999) *Markov Chains, Gibbs Fields, Monte Carlo Simulations, and Queues*. Springer, New York.

[9] Dagpunar J. (1988) *Principle of Random Variate Generation*. Clarendon Press, Oxford.

[10] Davis M.H.A. (1993) *Markov Models and Optimization*. Chapman & Hall, London.

[11] Diaconis P. Mathematical developments from the analysis of riffle shuffling .

[12] Durret R. (1996) *Probability: Theory and Examples*. Duxbury Press, Belmont.

[13] Feller W. (1987) *Wstęp do rachunku prawdopodobieństwa*, vol. I. PWN, Warszawa.

[14] Fishman G. (1966) *Monte Carlo*. Springer, New York.

[15] Fishman G. (2001) *Discrete-Event Simulation*. Springer, New York.

[16] Glasserman P. (2004) *Monte Carlo Methods in Financial Engineering*. Springer, New York.

[17] Häggström O. (2000) *Finite Markov Chains and Algorithmic Applications*. Chalmers University of Technology, dimacs.rutgers.edu/~dbwilson/exact.html/#haggstrom:course.

[18] Kelly F. (1985) Stochastic models of computer communication systems. *J.R.Statist.Soc.* **47**: 379–395.

[19] Kelly F.P. & MacPhee I. (198?) The number of packets transmitted by collision detect random access schemes. *Ann. Prob.* .

[20] Kingman J. (1992) *Poisson processes*. Oxford University Press, Oxford.

[21] Kingman J. (2002) *Procesy Poissona*. PWN, Warszawa.

[22] Knuth D. (1997) *The Art of Computer Programming – Seminumerical Algorithms*, vol. II. Addison-Wesley, Reading, 3rd ed.

[23] Knuth D. (2002) *Sztuka programowania – Algorytmy seminumeryczne*, vol. II. WNT, Warszawa.

[24] Kolata G. (1990 Jan. 9) In shuffling cards, seven is winning number. *New York Times* .

[25] Kopociński B. (1973) *Zarys teorii odnowy i niezawodności*. PWN, Warszawa.

[26] Kotulski M. (2001) Generatory liczb losowych. *Matematyka Stosowana* **2**: 32–66.

[27] L'Ecuyer P. (1998) Random number generation. In: J.E. Banks (ed.), *Handbook of Simulation*, chap. 4, 93137. Wiley.

[28] L'Ecuyer P. (1999) Good parameters and implementations for combined multiple recursive random generation. *Oper. Res.* **47**: 159–164.

[29] L'Ecuyer P. (2001) Software for uniform random number generation: distinguishing the good and the bad. In: B.P.J.S.D. Medeiros and M.R. eds (eds.), *Proceedings of the 2001 Winter Simulation Conference*, 95–105.

[30] L'Ecuyer P. & Simard R. On the performance of birthday spacing tests with certain families of random number generators .

[31] Levin D., Peres Y., and Wilmer E. *Markov Chains and Mixing Times*.

[32] Lewis P.A.W. & Shedler G. (1979) Simulation of nonhomogeneous poisson process by thinning. *Naval Research Logistic Quartely* **26**: 403–413.

[33] Madras N. (2002) *Lectures on Monte Carlo Methods*. AMS, Providence.

[34] Marsaglia G. (1961) Generating exponential random variables. *Ann. Math. Stat.* .

[35] Niederreiter H. (1992) *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, Philadelphia.

[36] Rényi A. (1970) *Probability Theory*. Akadémia Kiadó, Budapest.

[37] Resing J. (2008) Simulation. Manuscript, Eindhoven University of Technology.

[38] Ripley B. (1987) *Stochastic Simulation*. Wiley, Chichester.

[39] Rolski T., Schmidli H., Schmidt V., and Teugels J. (1999) *Stochastic Processes for Insurance and Finance*. Wiley, Chichester.

[40] Ross S. (1991) *A Course in Simulation*. Macmillan, New York.

[41] Rukhin A.e.a. (2010) *A Statistical Test Suite for Random and Pseudorandom Number Generator for Cryptographic Applications*. National Institute of Standards and Technology.

[42] Savicky P. (2008) A strong nonrandom pattern in matlab default random number generator. Manuscript, Institute of Computer Science, Academy of Science of CR.

[43] Srikant R. and Whitt W. (2008) Simulation run length planning for stochastic loss models. *Proceedings of the 1995 Winter Simulation Conference* **22**: 415–429.

[44] van der Vaart A. (1998) *Asymptotic Statistic*. Cambridge University Press, Cambridge.

[45] Whitt W. (1989) Planing queueing simulation. *Management Science* **35**: 1341–1366.

[46] Whitt W. (1991) The efficiency of one long run versus independent replications in steady-state simulations. *Management Science* **37**: 645–666.

[47] Whitt W. (1992) Asymptotic formulas for markov processes with applications to simulation. *Operations Research* **40**: 279–291.

[48] Wieczorkowski R. & Zieliński R. (1997) *Komputerowe generatory liczb losowych*. WNT, Warszawa.

[49] W.K. G. (2008) Warm-up periods in simulation can be detrimental. *PEIS* **22**: 415–429.

[50] Zieliński R. (1970) *Metody Monte Carlo*. WNT, Warszawa.

[51] Zieliński R. (1972) *Generatory liczb losowych*. WNT, Warszawa.

[52] Zieliński R. i Zieliński W. (1990) *Tablice statystyczne*. PWN, Warszawa.

Skorowidz nazw