

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



Monte Carlo metódy
výpočtu pravdepodobnosti extrémálnych udalostí

(DIPLOMOVÁ PRÁCA)

Bc. Marianna Holosová, 2012

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



Monte Carlo metódy
výpočtu pravdepodobnosti extrémálnych udalostí

(DIPLOMOVÁ PRÁCA)

Vedúci diplomovej práce: doc. Mgr. Radoslav Harman, PhD.

Študijný program: Ekonomická a finančná matematika

Študijný odbor: 1114 Aplikovaná matematika

Bc. Marianna Holosová, 2012



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Marianna Holosová
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský

Názov: Monte Carlo metódy výpočtu pravdepodobnosti extrémnych udalostí
Cieľ: Analýza Monte Carlo metód výpočtu pravdepodobnosti extrémnych udalostí.

Vedúci: doc. Mgr. Radoslav Harman, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Dátum zadania: 13.01.2011

Dátum schválenia: 14.01.2011
prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Čestne prehlasujem, že som diplomovú prácu vypracovala samostatne s využitím teoretických vedomostí a s použitím uvedenej literatúry.

.....
Marianna Holosová

Podakovanie

Predovšetkým sa chcem poďakovať vedúcemu diplomovej práce doc. Mgr. Radoslavovi Harmanovi, PhD., za cenné rady a pripomienky pri tvorbe tejto práce. Osobitne si vážim, že si vo svojom preplnenom rozvrhu vždy našiel čas na konzultácie. Moje poďakovanie patrí tiež môjmu snúbencovi, rodine a priateľom za ich podporu počas písania tejto práce.

Abstrakt

HOLOSOVÁ, Marianna: Monte Carlo metódy výpočtu pravdepodobnosti extrémálnych udalostí [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: doc. Mgr. Radoslav Harman, PhD. , Bratislava, 2012, 69 s.

Cieľom diplomovej práce je aplikovanie metódy váhového výberu na efektívny simulačný odhad pravdepodobnosti zriedkavých udalostí.

Práca popisuje klasické metódy odhadovania pravdepodobnosti udalostí a dôvody, prečo sú tieto metódy neefektívne, ak je simulovaná udalosť zriedkavá. Opisuje metódu váhového výberu, založenú na Radon-Nikodýmovej vete, ktorá znižuje disperziu odhadu pravdepodobnosti zmenením pravdepodobnostného rozdelenia, čo umožňuje nájsť presnejší odhad hľadanej pravdepodobnosti. V práci sa aplikuje exponenciálna zmena miery na simulovanie vybraných špeciálnych prípadov (pravdepodobnosť zbankrotovania, pravdepodobnosť prekročenia bariéry, pravdepodobnosť nadobudnutia extrémálneho stavu Markovovho reťazca, ...) a uvedie sa algoritmická implementácia týchto metód v jazyku R. Metóda váhového výberu sa ukazuje ako presnejšia metóda na odhad pravdepodobnosti zriedkavej udalosti ako klasická metóda odhadu pravdepodobnosti vo všetkých prípadoch uvedených v práci.

Kľúčové slová: Monte Carlo, Zriedkavá udalosť, Metóda váhového výberu, Exponenciálna zmena miery, Momentová vytvárajúca funkcia

Abstract

HOLOSOVÁ, Marianna: Calculating probabilities of extremal events using Monte Carlo methods [Mater thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; supervisor: doc. Mgr. Radoslav Harman, PhD., Bratislava, 2012, 69 p.

The aim of this master thesis is the application of importance sampling method for effective estimation of the probability of rare events.

The thesis describes classical methods for estimating events probabilities and why these methods are ineffective in case of simulated event being rare. It describes the importance of sampling method based on the Radon-Nikodym theorem, which reduces dispersion of estimated probability by changing the probability distribution and thus allows to find a more accurate estimate. Exponential change of measure for simulating selected events (probability of bankruptcy, probability of reaching an extremal state in Markov chain, ...) is applied and algorithmic implementation of these methods in software R is presented. In all cases presented in this thesis, the importance sampling seems to be a more accurate method for estimating probability of a rare event than the classic method of estimating probability.

Keywords: Monte Carlo, Rare event, Importance sampling, Exponential change of measure, Moment generating function

Obsah

Úvod	9
1 Vybrané tvrdenia o spojitých rozdeleniach pravdepodobnosti	11
1.1 Momentové vytvárajúce funkcie	11
1.1.1 Log-normálne rozdelenie	12
1.1.2 Gama rozdelenie	13
1.2 Generovanie náhodných premenných s normálnym rozdelením	13
1.2.1 Metóda inverznej transformácie	14
1.2.2 Zamietacia metóda	14
1.2.3 Box-Mullerov generátor	16
1.2.4 Iné algoritmy na generovanie z normálneho rozdelenia .	17
1.3 Generovanie náhodných premenných s gama rozdelením	17
2 Presnosť Monte Carlo odhadov a intervaly spoľahlivosti	19
2.1 Intervaly spoľahlivosti založené na centrálnej limitnej vete . .	21
2.2 Exaktné intervaly spoľahlivosti	23
3 Metóda váhového výberu	26
3.1 Odhadovanie pravdepodobnosti zriedkavých realizácií normál- neho náhodného vektora	28
3.2 Exponenciálna zmena miery	32
3.3 Odhad pravdepodobnosti prekročenia bariéry typu $S_n > n(\mu + \epsilon)$	35

3.4	Exponenciálna zmena miery pre náhodný proces s diskretným časom	36
3.5	Odhad pravdepodobnosti prekročenia bariéry typu $S_n > x$. .	37
3.6	Odhad pravdepodobnosti záporného zisku 4 roky po sebe . . .	43
3.7	Odhad pravdepodobnosti extrémálnej hodnoty zloženého Poissonovho rozdelenia	44
3.8	Odhad pravdepodobnosti zriedkavých udalostí v Markovovských reťazcoch	47
3.8.1	Hod kockou	49
3.8.2	Ehrenfestov model	52
	Záver	55

Úvod

Monte Carlo simulácie sa používajú na odhadnutie neznámej kvantitu pomocou simulovania náhodného deja v prípadoch, keď nie je možné vypočítať túto kvantitu analyticky. Špeciálne problémy nastávajú, ak sa pomocou Monte Carlo metód snažíme odhadnúť pravdepodobnosť zriedkavej udalosti, napríklad pravdepodobnosť bankrotu spoločnosti, pravdepodobnosť zemetrasenia, pravdepodobnosť zrážky dvoch lietadiel, . . . , ktorých pravdepodobnosť je síce nízka, ale ich následky môžu byť katastrofálne, preto je dôležité túto pravdepodobnosť odhadnúť, čo bude aj našim cieľom v tejto práci. Práca pozostáva z troch kapitol.

V prvej kapitole si priblížime pokročilejšie pojmy týkajúce sa spojitých rozdelení pravdepodobností a možnosti generovania premenných s daným rozdelením, ktoré budeme v práci ďalej používať.

V druhej kapitole sú popísané klasické metódy odhadovania pravdepodobnosti, špeciálna časť je venovaná intervalom spoľahlivosti. Okrem intervalu spoľahlivosti založeného na centrálnej limitnej vete, sú v práci popísané aj exaktné intervaly spoľahlivosti. Interval spoľahlivosti založený na centrálnej limitnej vete budeme v práci naďalej používať napriek nepresnostiam, pretože interval spoľahlivosti založený na Fisherovom-Snedecorovom rozdelení sa dá použiť len pre binomické rozdelenie a na Fishmanovom intervale spoľahlivosti sa neprejaví zníženie smerodajnej odchýlky, čo bude našim cieľom.

V tretej kapitole sa venujeme metóde váhového výberu a jej aplikácii exponenciálnej zmene miery, ktorá je vhodným nástrojom na odhadovanie pravdepodobnosti zriedkavých udalostí. Pri práci tiež popisujeme všeobecné metódy

ako odhadovať extrémálne udalosti, ktoré sa týkajú náhodných procesov s diskretným časom, špeciálne ide o tzv. náhodné prechádzky, teda proces, v ktorom je množina hodnôt spojitá, ale aj procesov, ktorých množina hodnôt je diskretná, ako sú Markovovské reťazce. V prípade náhodnej prechádzky nás môže zaujímať odhadnutie pravdepodobnosti prekročenia bariéry touto náhodnou prechádzkou a v Markovovských reťazcoch pravdepodobnosť, že do určitého počtu krokov sa Markovovský reťazec dostane do nejakého extrémálneho stavu.

Kapitola 1

Vybrané tvrdenia o spojitých rozdeleniach pravdepodobnosti

V tejto kapitole uvidíme vybrané pokročilejšie pojmy a tvrdenia teórie pravdepodobnosti týkajúce sa momentových vytvárajúcich funkcií a spôsobov generovania realizácií náhodných premenných s normálnym rozdelením.

1.1 Momentové vytvárajúce funkcie

V tejto časti budeme postupovať podľa kapitoly 5.7 v publikácii [6].

Momentová vytvárajúca funkcia (m.v.f) premennej X je funkcia $M : \mathbb{R} \rightarrow [0, \infty]$ daná ako $M(t) = E[e^{tX}]$. Ak je X spojitá náhodná premenná s hustotou f , tak platí

$$M(t) = \int_{-\infty}^{\infty} e^{tx} dF(x) = \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$

M.v.f. majú podobné vlastnosti ako pravdepodobnostné vytvárajúce funkcie, napr. ak $M(t) < \infty$ na nejakom otvorenom intervale obsahujúcom nulu, tak

- $E[X] = M'(0)$, $E[X^k] = M^{(k)}(0)$;

- ak X a Y sú nezávislé, tak $M_{X+Y}(t) = M_X(t)M_Y(t)$.

M.v.f. majú nevýhodu, že integrály, ktoré ich definujú, nemusia byť konečné, na rozdiel od charakteristických funkcií, ktorých integrály sú vždy konečné.

Definícia 1.1. *Charakteristická funkcia náhodnej premennej X je funkcia $\phi : \mathbb{R} \rightarrow \mathbb{C}$ definovaná ako*

$$\phi(t) = E[e^{itX}] , kde i = \sqrt{-1}.$$

Poznámka 1.1. *M.v.f. neurčuje jednoznačne distribučnú funkciu. Môžu existovať dve rôzne distribučné funkcie s rovnakou m.v.f.. Postačujúca podmienka pre jednoznačnosť distribučnej funkcie je konečnosť m.v.f. na nejakom okolí bodu $t = 0$.*

M.v.f. normálneho rozdelenia je

$$M(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = e^{\mu t + \frac{\sigma^2 t^2}{2}}.$$

1.1.1 Log-normálne rozdelenie

Log-normálne rozdelenie je spojité pravdepodobnostné rozdelenie náhodných premenných, ktorých logaritmus má normálne rozdelenie. Ak Y má normálne rozdelenie $\mathcal{N}(\mu, \sigma^2)$, tak $X = e^Y$ má log-normálne rozdelenie $L\mathcal{N}(\mu, \sigma^2)$. Hustota log-normálneho rozdelenia je

$$f_X(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, x > 0.$$

Stredná hodnota a variancia log-normálne rozdelenej premennej X je

$$E[X] = e^{\mu + \frac{1}{2}\sigma^2},$$

$$Var[X] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}.$$

Dá sa ukázať, že m.v.f. lognormálneho rozdelenia existuje len na intervale $(-\infty, 0]$.

1.1.2 Gama rozdelenie

Gama rozdelenie je spojité pravdepodobnostné rozdelenie. Hovoríme, že X má gama rozdelenie s parametrami $a > 0$, $b > 0$, ak je X spojitá náhodná premenná s hustotou

$$f(x; a, b) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-\frac{x}{b}} \text{ pre } x \geq 0 \text{ a } a, b > 0.$$

Stredná hodnota a variancia náhodnej premennej X je

$$\begin{aligned} E[X] &= ab, \\ Var[X] &= ab^2. \end{aligned}$$

M.v.f. gama rozdelenia je

$$M(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-\frac{x}{b}} dx = (1 - bt)^{-a} \text{ pre } t < \frac{1}{b}.$$

Exponenciálne rozdelenie je špeciálny prípad gama rozdelenia. Náhodná premenná X má exponenciálne rozdelenie $Exp(\lambda)$ práve vtedy keď má gama rozdelenie $\Gamma(1, \frac{1}{\lambda})$, kde $\lambda > 0$. M.v.f. exponenciálneho rozdelenia je

$$M(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}.$$

1.2 Generovanie náhodných premenných s normálnym rozdelením

Pre generovanie náhodných premenných s normálnym rozdelením je základom generovanie náhodných premenných s rovnomerným rozdelením, ktoré vieme transformovať na realizácie s normálnym rozdelením. Teoreticky je možné transformovať ľubovoľné spojité rozdelenie na normálne rozdelenie, ale generovanie z rovnomerného rozdelenia je najjednoduchšie, a zároveň existujú kvalitné testy, ktoré testujú kvalitu generátorov z rovnomerného rozdelenia (pozri kapitolu 7 v publikácii [6]). V ďalšom texte predpokladáme,

že z rovnomerného rozdelenia na intervale $(0, 1)$ generovať vieme. Podľa nasledujúcej vety na transformáciu stačí poznať kvantilovú funkciu normálneho rozdelenia $\mathcal{N}(0, 1)$.

1.2.1 Metóda inverznej transformácie

Veta 1.1. Veta o inverznej transformácii (pozri kap.1 v publikácii 2)

Nech $F : \mathbb{R} \rightarrow [0, 1]$ je distribučná funkcia a nech ϕ^{-1} je kvantilová funkcia, t.j. $\phi^{-1}(u) = \sup\{x \in \mathbb{R} : F(x) \leq u\}$ pre všetky $u \in (0, 1)$. Nech náhodná premenná U má rozdelenie $R(0, 1)$ (rovnomerné na intervale $(0, 1)$). Potom náhodná premenná $\phi^{-1}(U)$ má distribučnú funkciu F .

Pre normálne rozdelenie sa funkcia ϕ^{-1} nedá napísať v analytickom tvare, preto používame aproximácie.

Možný spôsob približného výpočtu funkčných hodnôt kvantilovej funkcie normálneho rozdelenia je

$$\phi^{-1}(z) \approx \frac{2,30753 + 0,27061c}{1 + 0,99229c + 0,04481c^2}, \text{ kde } c = \sqrt{-2 \ln(z)},$$

ktorý je popísaný v publikácii [8].

Pokiaľ poznáme hodnoty kvantilovej funkcie, tak predchádzajúcu vetu vieme efektívne aplikovať. Ak nevieme počítať kvantilovú funkciu rýchlym algoritmom, tak môžeme použiť „zamietaciu metódu“ (pozri [2]).

1.2.2 Zamietacia metóda

Veta 1.2. Veta o zamietaní

Nech $f, h : R \rightarrow [0, \infty)$ sú dve funkcie hustoty, kladné na intervale $I \subseteq R$ a inde nulové, pričom $c = \sup\{f(x)/h(x) : x \in I\} < \infty$. Nech $Y_1, U_1, Y_2, U_2, \dots$ sú nezávislé náhodné premenné, pričom $Y_1, Y_2, \dots$ majú hustotu h a $U_1, U_2, \dots$ majú rozdelenie $R(0, 1)$. Nech

$$N = \min\{n \in \mathbb{N} : c \cdot h(Y_n)U_n < f(Y_n)\}$$

Potom náhodná premenná Y_N má hustotu f .

Pre použitie vety o zamietaní na generovanie z normálneho rozdelenia $\mathcal{N}(0, 1)$ stačí vhodne zvoliť rozdelenie náhodných premenných $Y_1, Y_2, \dots$, ktoré majú hustotu h . Cauchyho rozdelenie je podobné normálnemu, poznáme jeho hustotu $h(x)$

$$h(x) = \frac{1}{\pi(1+x^2)}, x \in \mathbb{R}$$

a s využitím vety o inverznej transformácii vieme z neho ľahko generovať premenné - kvantilová funkcia Cauchyho rozdelenia $G(U)$ je daná explicitne ako $G(U) = \tan(\pi(U - 1/2))$, kde U má rozdelenie $R(0, 1)$. Hustota normálneho rozdelenia $\mathcal{N}(0, 1)$ je

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Podmienka určujúca index n akceptovanej realizácie má potom tvar

$$c \frac{1}{\pi(1+Y_n^2)} U_n < \frac{1}{\sqrt{2\pi}} e^{-\frac{Y_n^2}{2}},$$

kde $c = \sup\{f(x)/h(x) : x \in I\} = \sqrt{\frac{2\pi}{e}}$, takže podmienka sa dá prepísať ako

$$U_n < \frac{\pi(1+Y_n^2)}{2} e^{\frac{1-Y_n^2}{2}}.$$

Iná možnosť je najprv generovať z pol-normálneho rozdelenia $H\mathcal{N}(0, 1)$, ktoré má hustotu $f(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}$ pre $x \geq 0$, $f(x) = 0$ pre $x < 0$ a za hustotu h zvoliť hustotu $h(x) = e^{-x}$, $x > 0$, čo zodpovedá hustote exponenciálneho rozdelenia $Exp(1)$. V tomto prípade podmienka, ktorá určuje index n akceptovanej realizácie má tvar

$$U_n < \exp(-(Y_n - 1)^2/2).$$

Na vygenerovanie realizácie z normálneho rozdelenia $\mathcal{N}(0, 1)$ stačí prenásobiť realizáciu Y s pravdepodobnosťou $1/2$ hodnotou -1 , resp. 1 , kde Y má hustotu f .

Na generovanie náhodných premenných $X_1, X_2, \dots$ z normálneho rozdelenia $\mathcal{N}(\mu, \sigma^2)$ stačí použiť jednoduchú transformáciu : $X_i = \mu + Y_i\sigma$, kde Y_i je náhodná premenná s rozdelením $\mathcal{N}(0, 1)$.

1.2.3 Box-Mullerov generátor

Nasledujúci spôsob generovania náhodných premenných s normálnym rozdelením je založený na fakte, že dvojrozmerné normalizované normálne rozdelenie $\mathcal{N}_2(0, \mathbf{I})$ má polárne súradnice (R, ψ) s ľahko generovateľným rozdelením: $R \sim \sqrt{\text{Exp}(1/2)}$ a $\psi \sim R(0, 2\pi)$.

Veta 1.3. Box-Mullerov generátor

Nech náhodný vektor $(V_1, V_2)^T$ má rovnomerné rozdelenie na jednotkovej kružnici v rovine, nech $L \sim \text{Exp}(1/2)$ a nech $(V_1, V_2)^T$ a L sú nezávislé. Potom náhodné premenné $X_1 = V_1\sqrt{L}$ a $X_2 = V_2\sqrt{L}$ sú nezávislé, obe s rozdelením $\mathcal{N}(0, 1)$. Špeciálne, ak U, T sú nezávislé s rozdelením $R(0, 1)$, tak náhodné premenné

$$X_1 = \cos(2\pi U)\sqrt{-2\ln(T)}, X_2 = \sin(2\pi U)\sqrt{-2\ln(T)}$$

sú nezávislé, obe s rozdelením $\mathcal{N}(0, 1)$.

Počítaniu sínusov a kosínusov sa môžeme vyhnúť nasledujúcim spôsobom. Náhodný vektor $(V_1, V_2)^T$ vygenerujeme rovnomerne na jednotkovej kružnici, teda $V_1 = \cos(\psi)$ a $V_2 = \sin(\psi)$, kde $\psi \sim R(0, 2\pi)$ tak, že najprv budeme generovať vektor $(W_1, W_2)^T$ rovnomerne na jednotkovom kruhu, a hodnoty V_1, V_2 vypočítame potom ako

$$V_1 = \frac{W_1}{\sqrt{W_1^2 + W_2^2}}, V_2 = \frac{W_2}{\sqrt{W_1^2 + W_2^2}}.$$

Náhodné premenné

$$X_1 = W_1\sqrt{\frac{-2\ln(T)}{W_1^2 + W_2^2}}, X_2 = W_2\sqrt{\frac{-2\ln(T)}{W_1^2 + W_2^2}}$$

sú nezávislé, obe s rozdelením $\mathcal{N}(0, 1)$, čo je efektívnejší spôsob výpočtu, pretože tentoraz stačí vypočítať len odmocninu a logaritmus.

1.2.4 Iné algoritmy na generovanie z normálneho rozdelenia

Ziggurat algoritmus

Patrí do triedy zamietacích algoritmov. Používa sa na generovanie z monotónne klesajúcich pravdepodobnostných rozdelení, ale môže sa použiť aj na generovanie z normálneho rozdelenia tak, že pomocou algoritmu vygenerujeme realizácie z pol-normálneho rozdelenia, a potom ich s pravdepodobnosťou $1/2$ prenásobíme hodnotou 1 , resp. -1 . Myšlienkou metódy je pokrytie cieľovej hustoty pomocou upravenej hustoty, ktorá pozostáva z n horizontálnych blokov rovnakej veľkosti - $n - 1$ obdĺžnikov a jedného bloku, ktorý pozostáva z obdĺžniku a zvyšku cieľovej hustoty (spodný blok). Náhodne vygeneruje bod (x, y) z tejto upravenej hustoty, ak bod nie je súčasťou cieľovej hustoty, tak vygeneruje ďalší bod, ale ak je, tak x je náhodná premenná s požadovanou hustotou. Tento algoritmus je výpočtovo efektívnejší oproti Box-Mullerovmu generátoru, ktorý pre každý pár generovaných hodnôt požaduje počítať logaritmus aj odmocninu.

Okrem spomenutých metód na generovanie z normálneho rozdelenia existujú mnohé ďalšie, napr. Kinderman-Ramage metóda, ktorá má už skôr historický význam vzhľadom na odhalenie jej viacerých nedostatkov.

1.3 Generovanie náhodných premenných s gama rozdelením

Náhodné premenné s gama rozdelením generujeme pomocou náhodných premenných, ktoré majú exponenciálne, resp. beta rozdelenia, preto najprv

musíme vedieť generovať z týchto dvoch rozdelení (pozri [2]).

Generovať náhodné premenné s exponenciálnym rozdelením vieme jednoducho s využitím vety o inverznej transformácii 1.1, pretože vieme efektívne počítať hodnoty kvantilovej funkcie, ktorá je daná explicitne ako

$$G(y) = -\lambda \ln(1 - y) \text{ pre } y \in (0, 1).$$

Potom podľa vety 1.1 platí: Ak $U \sim R(0, 1)$, tak $\lambda \ln(1 - U) \sim \text{Exp}(\lambda)$.

Beta rozdelenie je spojité pravdepodobnostné rozdelenie. Hovoríme, že náhodná premenná X má beta rozdelenie s parametrami $a, b > 0$, ak je X spojitá náhodná premenná s hustotou

$$f(x) = \begin{cases} B^{-1}(a, b)x^{a-1}y^{b-1} & \text{pre } x \in (0, 1), \\ 0 & \text{inak,} \end{cases}$$

kde symbol B označuje beta funkciu. Značíme $X \sim \text{Be}(a, b)$.

Nasledujúca veta hovorí o generovaní realizácií z beta rozdelenia.

Veta 1.4. *Nech $a, b > 0$, nech X_1, X_2 sú nezávislé náhodné premenné s rozdelením $R(0, 1)$. Nech $V = X_1^{1/a}$, $W = X_2^{1/b}$, $Y = \frac{V}{V+W}$ a platí $V + W \leq 1$. Potom náhodná premenná Y má rozdelenie $\text{Be}(a, b)$.*

Gamma rozdelenie je definované v časti 1.1.2 a generovať náhodné premenné s gama rozdelením vieme pomocou nasledujúcej vety.

Veta 1.5. *Nech $a, b > 0$, $n = \lfloor b \rfloor$, $r = b - n$. Nech $Z_0, Z_1, \dots, Z_n, Y$ sú nezávislé náhodné premenné, pričom $Z_0, Z_1, \dots, Z_n \sim \text{Exp}(1)$ a $Y \sim \text{Be}(r, 1 - r)$ ak $r > 0$ a $Y = 0$ ak $r = 0$. Ak $n = 0$, tak $Z = 0$ a ak $n \geq 1$ nech $Z = Z_1 + \dots + Z_n$. Potom náhodná premenná $a^{-1}(Z + Z_0 Y)$ má rozdelenie $\Gamma(a, b)$.*

Výpočtová zložitosť tejto metódy závisí od parametra $n = \lfloor b \rfloor$, so zvyšovaním n sa zvyšuje aj čas potrebný na počítanie. Pokročilejšie metódy majú tú vlastnosť, že výpočtová zložitosť nejde do nekonečna pre $n \rightarrow \infty$. Viaceré takéto metódy sú popísané napríklad v knihe [6].

Kapitola 2

Presnosť Monte Carlo odhadov a intervaly spoľahlivosti

Monte Carlo simulácie sa používajú na odhadnutie neznámej kvantitý z pomocou simulovania náhodného deja, ktorého je z odhadnuteľnou charakteristikou (obvykle je z strednou hodnotou nejakej číselnej charakteristiky simulovaného deja). Vytvorí sa počítačový model tohto deja a po prebehnutí dostatočného množstva simulácií sa môžu dáta spracovať klasickými štatistickými metódami, napríklad určiť priemer a smerodajnú odchýlku. Často však môžeme simulovať veľa odlišných dejov (procesov), pričom z každého z nich je možné odhadnúť hľadanú kvantitu. Tieto procesy sa môžu líšiť presnosťou, s akou umožňujú určiť z .

Predstavme si, že chceme odhadnúť pravdepodobnosť p nastatia udalosti $A : p = P(A)$. Najjednoduchší prístup simulačného odhadovania pravdepodobnosti p spočíva v generovaní n nezávislých realizácií udalostí A , ktoré si označíme ako $A_1, A_2, \dots, A_n$ a zavedieme si náhodnú premennú $Z_i = I\{A_i\}$ - indikátorová funkcia množiny A , ktorá má alternatívne (0 – 1-ové) rozdelenie. Pre strednú hodnotu premennej Z_i platí $E(Z_i) = P(A) = p$ a pre disperziu platí $Var(Z_i) = \sigma^2 = p(1 - p)$.

Priemer náhodných premenných Z_i si označíme ako $\bar{Z}$

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i.$$

Potom stredná hodnota priemeru $\bar{Z}$ je odhadovaná pravdepodobnosť p

$$E[\bar{Z}] = p$$

a disperzia odhadu

$$D[\bar{Z}] = D\left[\frac{1}{n} \sum_{i=1}^n Z_i\right] = \frac{1}{n^2} \sum_{i=1}^n D[Z_i] = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n} = \frac{p(1-p)}{n}.$$

Presnosť odhadu určuje smerodajná odchýlka $\sigma[\bar{Z}]$

$$\sigma[\bar{Z}] = \sqrt{D[\bar{Z}]} = \sqrt{\frac{p(1-p)}{n}}.$$

Predpokladáme, že p je malé, to znamená, že aj variancia $\sigma^2 = p(1-p)$ je malá. Čím je pravdepodobnosť p menšia, tým menšia je aj variancia, to znamená, že absolútna chyba σ^2 je malá pre malé p . Zaujímavejšie je však pozorovať relatívnu chybu - $\sigma[\bar{Z}]/p$, ktorá je vysoká

$$\frac{\sigma[\bar{Z}]}{p} = \frac{\sqrt{p(1-p)}}{p} \sim \frac{1}{p} \rightarrow \infty \text{ pre } p \downarrow 0.$$

Nech $\{A(x)\}$ je trieda zriedkavých udalostí (pod „zriedkavými“ sa obvykle myslia udalosti s pravdepodobnosťou menšou ako 10^{-3}), kde $x \in (0, \infty)$ alebo $x \in \mathbb{N}$. Predpokladajme, že $p(x) = P(A(x)) \rightarrow 0$ pre $x \rightarrow \infty$, pričom x je „parameter“ daného deja a pre každé x je $Z(x)$ nevychýlený odhad pre $p(x)$, teda $E[Z(x)] = p(x)$. Pojmom „algorithmus“ nazvime triedu $\{Z(x)\}$ takých náhodných premenných (pozri [3]).

Hľadáme taký algorithmus, ktorý má ohraničenú relatívnu chybu pre $x \rightarrow \infty$

$$\limsup_{x \rightarrow \infty} \frac{\sqrt{\text{Var}(Z(x))}}{p(x)} < \infty,$$

čo znamená, že pomer štandardnej odchýlky k odhadovanej pravdepodobnosti je menší ako nejaká konečná konštanta pre všetky x . Väčšinou sa používa slabší predpoklad - logaritmická efektivita: $Var(Z(x)) \rightarrow 0$ tak rýchlo, že platí

$$\limsup_{x \rightarrow \infty} \frac{Var(Z(x))}{p(x)^{2-\epsilon}} = 0 \text{ pre nejaké } \epsilon > 0.$$

Podmienku logaritmickkej efektivity je jednoduchšie overiť, preto sa častejšie používa. Pre viacero problémov dokonca neexistujú algoritmy spĺňajúce podmienku ohraničenej relatívnej chyby. My sa v tejto práci nebudeme zaoberať analýzou algoritmov z pohľadu ich efektivity, ale poznamenávame, že metóda exponenciálnej zmeny miery umožňuje skonštruovať logaritmicky efektívne algoritmy, akým je napríklad algoritmus popísaný v časti 3.5.

2.1 Intervaly spoľahlivosti založené na centrálnej limitnej vete

Uvažujme priemer $\bar{Z}$ náhodných premenných $Z_1, \dots, Z_n$ z predchádzajúcej časti. Približné $((1 - \alpha)100\%$ -né) intervaly spoľahlivosti sú

$$\left(\bar{Z} - u_{1-\alpha/2} \sqrt{\frac{\bar{Z}(1 - \bar{Z})}{n}}, \bar{Z} + u_{1-\alpha/2} \sqrt{\frac{\bar{Z}(1 - \bar{Z})}{n}} \right), \quad (2.1)$$

kde $u_{1-\alpha/2}$ je $(1 - \alpha/2)100\%$ -ný kvantil normálneho rozdelenia. To znamená, že presnosť odhadu je daná disperziou a počtom simulácií.

Interval spoľahlivosti daný vzorcom (2.1) je založený na dvoch aproximáciách. Prvou aproximáciou je použitie centrálnej limitnej vety. Označme sumu náhodných premenných Z_i ako X

$$X = \sum_{i=1}^n Z_i \sim Bin(n, p),$$

potom podľa centrálnej limitnej vety pre veľké n platí

$$X \sim N(np, np(1-p)),$$

a teda vieme odhadnúť aj rozdelenie pre $\bar{Z} = \frac{X}{n}$

$$\bar{Z} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

Druhú aproximáciu sme použili pri odhade parametra p pomocou jeho nevychýleného odhadu $\bar{Z} = \frac{X}{n}$, takže disperzia

$$D[\bar{Z}] = \frac{p(1-p)}{n} \doteq \frac{\bar{Z}(1-\bar{Z})}{n}.$$

Poznamenajme, že centrálnu limitnú vetu je možné použiť aj pre iné rozdelenie ako binomické, čiže ak nezávislé náhodné premenné Z_i majú strednú hodnotu p , avšak iné ako alternatívne rozdelenie. V takom prípade nahradíme výraz $\sqrt{\frac{\bar{Z}(1-\bar{Z})}{n}}$ výberovou smerodajnou odchýlkou náhodných premenných $Z_1, \dots, Z_n$.

Interval spoľahlivosti daný vzorcom (2.1) pre odhad pravdepodobnosti zriedkavej udalosti je príliš veľký, pretože σ je oveľa väčšia ako pravdepodobnosť, ktorú sa snažíme odhadnúť. Presnosť odhadu sa dá zvýšiť zväčšením počtu meraní, ale to ide pomaly - napríklad na 10 násobné skrátenie intervalu spoľahlivosti potrebujeme 100-krát viac simulácií a času. Ak uvažujeme udalosti s pravdepodobnosťou p rádu 10^{-9} , ktoré sa vyskytujú napríklad v mnohých telekomunikačných aplikáciách, tak Monte Carlo simulácie sú nielen neefektívne, ale ich použitie nie je vôbec možné. Ďalším problémom je aproximácia pomocou centrálnej limitnej vety kvôli jej nepresnosti pre malé p , čo tiež môže spôsobiť zväčšenie intervalu.

Nepresnosti, ktoré vznikajú aproximáciami môžu spôsobiť, že ľavý kraj intervalu spoľahlivosti bude záporný, ak odhadujeme zriedkavú udalosť. Na druhej strane, ak odhadujeme udalosť, ktorá má vysokú pravdepodobnosť, tak sa môže stať, že pravý kraj intervalu spoľahlivosti bude väčší ako 1.

2.2 Exaktné intervaly spoľahlivosti

Možných intervalov spoľahlivosti pre daný parameter existuje viacero.

Exaktný interval spoľahlivosti pre parameter p binomického rozdelenia môžeme získať využitím kvantilov $F(k_1, k_2; \alpha)$ Fisherovho-Snedecorovho rozdelenia s k_1, k_2 stupňami voľnosti.

$$p_L = \frac{X}{(n - X + 1)F(k_1, k_2; 1 - \frac{\alpha}{2}) + X}, \text{ kde } k_1 = 2(n - X + 1), k_2 = 2X$$

$$p_P = \frac{(X + 1)F(k_1, k_2; 1 - \frac{\alpha}{2})}{n - X + (X + 1)F(k_1, k_2; 1 - \frac{\alpha}{2})}, \text{ kde } k_1 = 2(X + 1), k_2 = 2(n - X).$$

Interval spoľahlivosti (p_L, p_P) sme dostali prepísaním distribučnej funkcie binomického rozdelenia pomocou distribučnej funkcie Fisherovho-Snedecorovho rozdelenia (pozri [1]).

Príklad 2.1. *Pre porovnanie presnosti týchto dvoch intervalov spoľahlivosti uvedieme nasledujúci príklad. Ak pri počte simulácií $n = 1000$ udalosť A nastala práve 5-krát, t.j. $X = 5$, interval spoľahlivosti vypočítaný na základe centrálnej limitnej vety je $(6, 30 \cdot 10^{-4}, 9, 38 \cdot 10^{-3})$ a interval spoľahlivosti vypočítaný pomocou Fisherovho-Snedecorovho rozdelenia je $(1, 63 \cdot 10^{-3}, 1, 16 \cdot 10^{-2})$. Všimnime si ľavé kraje intervalov, druhý je 2,5-krát väčší ako prvý. V prípade nastatia udalosti A len raz, t.j. $X = 1$ je ľavý kraj prvého intervalu záporný $(-9, 60 \cdot 10^{-4})$, kým druhý interval spoľahlivosti vychádza síce malý, ale stále kladný $(2, 50 \cdot 10^{-5})$, z čoho vidíme, že druhý interval spoľahlivosti je podstatne presnejší. (Intervaly boli počítané na hladine významnosti 95%.)*

Uvedieme ešte jeden spôsob výpočtu intervalov spoľahlivosti. Tento prístup má veľkú výhodu v tom, že nie je viazaný na pravdepodobnostné rozdelenie simulovaných náhodných premenných [6].

Veta 2.1. *(Fishman 1991) Nech $Z_1, \dots, Z_n$ označujú nezávislé náhodné premenné nadobúdajúce hodnoty z intervalu $(0, 1)$ so strednou hodnotou $\mu =$*

$E[Z_i] \in (0, 1)$ a nech $\bar{Z}_n = (Z_1 + \dots + Z_n)/n$. Potom pre $s \leq 1$ a $0 < \alpha < 1$, definujeme

$$\rho_1(s, n, \alpha) = \begin{cases} \{t : 0 < t \leq s \leq 1 \text{ a } e^{nw(s-t,t)} = \alpha/2\} & \text{pre } s > 0, \\ 0 & \text{pre } s = 0 \end{cases}$$

a

$$\rho_2(s, n, \alpha) = \begin{cases} \{t : 0 \leq s \leq t < 1 \text{ a } e^{nw(t-s,1-t)} = \alpha/2\} & \text{pre } s < 0, \\ 1 & \text{pre } s = 1, \end{cases}$$

kde $w(\cdot, \cdot)$ je definované ako

$$w(\epsilon, \mu) = (\mu + \epsilon) \ln \left(\frac{\mu}{\mu + \epsilon} \right) + (1 - \mu - \epsilon) \ln \left(\frac{1 - \mu}{1 - \mu - \epsilon} \right).$$

Potom

$$P \left(\rho_1(\bar{Z}_n, n, \alpha) < \mu < \rho_2(\bar{Z}_n, n, \alpha) \right) \geq 1 - \alpha.$$

Funkcia $w(s - t, t)$ je totožná s funkciou $w(t - s, 1 - t)$, preto platí, že $e^{nw(s-t,t)} = e^{nw(t-s,1-t)}$, chápaná ako funkcia premennej t , pretína $\alpha/2$ dvakrát. Prvýkrát na intervale $(0, s)$ a druhýkrát na intervale $(s, 1)$.

Príklad 2.2. (pokračovanie **Príkladu 2.1**) *Interval spoľahlivosti vypočítaný týmto prístupom pre počet pozorovaní $n = 1000$ a počet nastatí udalosti A $X = 5$ je $(1, 10 \cdot 10^{-3}, 1, 37 \cdot 10^{-2})$, pre počet nastatí $X = 1$ je $(9, 24 \cdot 10^{-6}, 6, 56 \cdot 10^{-3})$. Vidíme, že intervaly síce nie sú kratšie, ale sú presnejšie.*

Intervaly spoľahlivosti sme počítali v programe R (pozri [9]) pomocou Brentovho algoritmu - numerická metóda na nájdenie riešenia rovnice s 1 neznámou premennou (funkcia optimize). Táto metóda spája metódu bisekcie, metódu sečníc a kvadratickú interpoláciu. Má spoľahlivosť bisekcie a je rýchla ako menej spoľahlivé metódy.

Je dobré poznamenať, že intervaly spoľahlivosti počítané na základe Fisherovho-Snedecorovho rozdelenia a na základe Fishmanovej vety sú, na

rozdiel od intervalov spoľahlivosti počítaných na základe centrálnej limitnej vety, asymetrické vzhľadom na priemer. Keby boli symetrické, tak pri odhadovaní pravdepodobnosti zriedkavých udalostí by mohol vyjsť ľavý okraj intervalu záporný.

Na skrátenie intervalov spoľahlivosti môžeme zvýšiť n , čo je ale neefektívne. Existujú však metódy, ktoré dokážu skonštruovať náhodnú premennú s tou istou strednou hodnotou, ale s oveľa menšou disperziou, čo je efektívnejší prístup. Medzi najpoužívannejšie metódy patrí metóda váhového výberu (angl. Importance Sampling), metóda kontrolných premenných (angl. Control variates) a stratifikácia (angl. Stratification).

Fishmanov interval spoľahlivosti je zaujímavou alternatívou k intervalom spoľahlivosti založeným na centrálnej limitnej vete, ale len pre binomické rozdelenie. V prípade binomického rozdelenia je odhad disperzie funkciou priemeru $\left(\frac{\bar{Z}(1-\bar{Z})}{n}\right)$. Pokiaľ uvažujeme iné rozdelenie a sme schopní znížiť disperziu, tak toto zníženie sa neprejaví na skrátení Fishmanových intervalov spoľahlivosti. V ďalšom texte nebudeme používať Fishmanove intervaly spoľahlivosti, lebo našim cieľom bude práve znižovanie disperzie.

Kapitola 3

Metóda váhového výberu

Najdôležitejšia pre túto prácu bude metóda váhového výberu, pri ktorej využijeme nasledovnú vetu. Najskôr si potrebujeme zadať absolútnu spojitost mier.

Definícia 3.1. Absolútna spojitost mier

Nech μ a ν sú σ -konečné miery na merateľnom priestore $(\chi, \mathcal{A})$. Miera μ je absolútne spojitá vzhľadom na ν práve vtedy keď pre všetky $A \in \mathcal{A}$ platí

$$\nu(A) = 0 \Rightarrow \mu(A) = 0.$$

Veta 3.1. (Radon-Nikodýmova veta)

Nech μ a ν sú σ -konečné miery na merateľnom priestore $(\chi, \mathcal{A})$ a nech miera μ je absolútne spojitá vzhľadom na ν . Potom existuje nezáporná merateľná funkcia $f : (\chi, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_1)$ taká, že

$$\mu(A) = \int_A f(x) d\nu(x), \quad \forall A \in \mathcal{A}.$$

Funkcia f sa nazýva Radon-Nikodýmova derivácia a môže sa značiť aj $\frac{d\mu}{d\nu}$. Symbol $\mathcal{B}_n$ označuje Borelovské množiny v $\mathbb{R}^n$.

Predstavme si, že chceme vypočítať pravdepodobnosť v tvare $z = E[Z]$. Metóda váhového výberu znižuje disperziu zmenením pravdepodobnostného

rozdelenia P tak, že udalosť, ktorá bola pôvodne zriedkavá, v zmenenej situácii nastáva často, nie zriedkavo. Predpokladajme, že za pravdepodobnostné rozdelenie zvolíme $\tilde{P}$, pre ktoré existuje hustota (alebo Radon-Nikodýmova derivácia) $\tilde{L}$,

$$\tilde{L}(\omega) = \frac{dP}{d\tilde{P}}(\omega) \text{ pre všetky } \omega \in \Omega.$$

Vzhľadom na Radon-Nikodýmovu vetu platí

$$P(A) = \int_A \frac{dP}{d\tilde{P}}(\omega) d\tilde{P}(\omega) \text{ pre všetky udalosti } A.$$

Takže

$$\begin{aligned} E[Z] &= \int_{\Omega} Z(\omega) dP(\omega) = \int_{\Omega} Z(\omega) \frac{dP}{d\tilde{P}}(\omega) d\tilde{P}(\omega) = \\ &= \int_{\Omega} Z(\omega) \tilde{L}(\omega) d\tilde{P}(\omega) = \tilde{E}[Z\tilde{L}], \end{aligned} \quad (3.1)$$

kde $\tilde{E}$ je stredná hodnota pre pravdepodobnostné rozdelenie $\tilde{P}$. To znamená, že na odhadnutie hľadanej hodnoty $z = E[Z]$ môžeme použiť Monte Carlo metódu tak, že generujeme n nezávislých náhodných realizácií $Z_1\tilde{L}_1, \dots, Z_n\tilde{L}_n$ s rozdelením $\tilde{P}$, pomocou priemeru $\bar{Z}$ odhadneme z a použitím výberovej smerodajnej odchýlky a centrálnej limitnej vety vytvoríme interval spoľahlivosti tvaru (2.1). K tomu potrebujeme poznať hodnoty $\tilde{L}_1, \dots, \tilde{L}_n$.

Príklad 3.1. Uvažujme ako pravdepodobnostný priestor priestor $\mathbb{R}^n$, t.j. $\Omega = \mathbb{R}^n$. Na tomto priestore existuje pravdepodobnostné rozdelenie $P_{\mathbf{X}}$ s hustotou f - $P_{\mathbf{X}}(A) = \int_A f(\mathbf{x}) d\mathbf{x}$ a pravdepodobnostné rozdelenie $\tilde{P}_{\mathbf{X}}$ s hustotou $\tilde{f}$ - $\tilde{P}_{\mathbf{X}}(A) = \int_A \tilde{f}(\mathbf{x}) d\mathbf{x}$ ($P_{\mathbf{X}}(A)$ a $\tilde{P}_{\mathbf{X}}(A)$ môžu byť pravdepodobnostné miery indukované na priestore $(\Omega, \mathcal{A})$ náhodným vektorom $\mathbf{X}$). Predpokladáme, že hustoty $f, \tilde{f} > 0$. Radon-Nikodýmova derivácia $\tilde{L}$ je v tomto prípade

$$\tilde{L}(\mathbf{x}) = \frac{dP}{d\tilde{P}}(\mathbf{x}) = \frac{f(\mathbf{x})}{\tilde{f}(\mathbf{x})}.$$

Špeciálne ak $f(\mathbf{x}) = \prod_{i=1}^n f_i(x_i)$, kde $\mathbf{x} = (x_1, \dots, x_n)$, čiže ak $P_{\mathbf{X}}$ zodpovedá vektoru s nezávislými zložkami s hustotou f_i a ak $\tilde{f}(\mathbf{x}) = \prod_{i=1}^n \tilde{f}_i(x_i)$, tak

$$\tilde{L}(\mathbf{x}) = \prod_{i=1}^n \frac{f_i(x_i)}{\tilde{f}_i(x_i)}. \quad (3.2)$$

Nasledujúca veta (pozri kapitolu 5 v [3]) hovorí ako zvoliť hustotu $\tilde{f}$, aby odhad $P(A)$ bol čo najpresnejší.

Veta 3.2. *Nech Z je náhodná premenná na priestore $(\Omega, \mathcal{A})$, pričom $Z(\omega) \neq 0$ pre všetky $\omega \in \Omega$. Nech P a $\tilde{P}$ sú dve pravdepodobnostné miery na $(\Omega, \mathcal{A})$, pričom $P \ll \tilde{P}$, označme $\tilde{L}(\omega) = \frac{dP}{d\tilde{P}}(\omega)$. Nech P^* je taká pravdepodobnostná miera na $(\Omega, \mathcal{A})$, že $P \ll P^*$ a*

$$L^* = \frac{dP}{dP^*} = \frac{E[|Z|]}{|Z|}.$$

Potom $D^(ZL^*) \leq \tilde{D}(Z\tilde{L})$.*

Ak $Z > 0$ všade, tak $D^(ZL^*) = 0$.*

Dôkaz 3.1. *Uvedieme len dôkaz pre prípad $Z > 0$. Platí*

$$E^*[ZL^*] = E^*\left[Z \frac{dP}{dP^*}\right] = \int_{\Omega} Z(\omega) \frac{dP}{dP^*}(\omega) dP^*(\omega) = \int_{\Omega} Z(\omega) dP(\omega) = E[Z].$$

Analogicky $\tilde{E}[Z\tilde{L}] = E[Z]$. Keďže $Z > 0$, tak $Z = |Z|$ a $E[|Z|] = E[Z]$, a teda platí

$$\begin{aligned} D^*[ZL^*] &= E^*[(ZL^*)^2] - (E[Z])^2 = \int_{\Omega} Z^2(\omega) (L^*(\omega))^2 dP^*(\omega) - (E[Z])^2 = \\ &= \int_{\Omega} Z^2(\omega) \frac{E[Z]}{Z(\omega)} \frac{dP(\omega)}{dP^*(\omega)} dP^*(\omega) - (E[Z])^2 = \\ &= \int_{\Omega} Z^2(\omega) \frac{E[Z]}{Z(\omega)} dP(\omega) - (E[Z])^2 = \\ &= E[Z] \int_{\Omega} Z(\omega) dP(\omega) - (E[Z])^2 = (E[Z])^2 - (E[Z])^2 = 0. \end{aligned}$$

Rovnosť $D^(ZL^*) \leq \tilde{D}(Z\tilde{L})$ v prípade $Z > 0$ už plynie triviálne.*

3.1 Odhadovanie pravdepodobnosti zriedkavých realizácií normálneho náhodného vektora

Naším cieľom je odhadnúť pravdepodobnosť udalosti A , že $\mathbf{X} \in M$

$$P(A) = P(\mathbf{X} \in M),$$

kde $\mathbf{X}$ je n -rozmerný normálny náhodný vektor so strednou hodnotou $\mathbf{0}$ a kovariančnou maticou $\mathbf{C}$, t.j. $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{C})$. Nech náhodná premenná $Z : \Omega \rightarrow \mathbb{R}$ je definovaná ako

$$Z(\omega) = \begin{cases} 0 & \text{pre } \mathbf{X}(\omega) \notin M, \\ 1 & \text{pre } \mathbf{X}(\omega) \in M, \end{cases} \quad (3.3)$$

kde $M \in \mathcal{B}_n$ je množina.

Pravdepodobnosť $P(A)$ odhadneme pomocou strednej hodnoty premennej Z pri použití pravdepodobnostného rozdelenia P alebo pomocou strednej hodnoty premennej $\tilde{L}Z$ pri použití pravdepodobnostného rozdelenia $\tilde{P}$, čo môžeme zapísať ako

$$P(A) = E[Z] = \tilde{E}[Z\tilde{L}] = \tilde{E}[Z_N].$$

Ak označíme symbolom f hustotu rozdelenia $\mathcal{N}_n(\mathbf{0}, \mathbf{C})$ a symbolom $\tilde{f}$ inú hustotu nenulovú na množine M , tak potom Z_N má nasledovný tvar

$$Z_N(\omega) = \begin{cases} 0 & \text{pre } \mathbf{X}(\omega) \notin M, \\ \frac{f(\mathbf{X}(\omega))}{\tilde{f}(\mathbf{X}(\omega))} & \text{pre } \mathbf{X}(\omega) \in M. \end{cases} \quad (3.4)$$

Podľa vety 3.2 by ideálna hustota bola taká, ktorá by sa priblížila hustote

$$f^*(\mathbf{x}) = \begin{cases} 0 & \text{pre } \mathbf{x} \notin M, \\ \frac{f(\mathbf{x})}{c} & \text{pre } \mathbf{x} \in M, \end{cases} \quad (3.5)$$

kde c je vhodná konštanta. To znamená, že situácia $\mathbf{X} \notin M$ nastáva pri novom pravdepodobnostnom rozdelení s nulovou pravdepodobnosťou. Vzhľadom k tomu, že plocha pod hustotou je rovná 1, môžeme napísať

$$1 = \int_A f^*(\mathbf{x}) d\mathbf{x} = \int_A \frac{1}{c} f(\mathbf{x}) d\mathbf{x},$$

z čoho dostávame

$$c = \int_A f(\mathbf{x}) d\mathbf{x} = P(A).$$

To znamená, že pravdepodobnosť, ktorú sa snažíme odhadnúť, je rovná práve konštante c . Na zistenie tejto konštanty nám stačí raz vygenerovať premennú Z_N , pretože hustota f^* s pravdepodobnosťou 1 generuje pomer $\frac{f(\mathbf{x})}{f^*(\mathbf{x})}$, ktorý je rovný konštante c . Dostali by sme tak odhad hľadanej pravdepodobnosti s nulovou disperziou analogický odhadu ZL^* z vety 3.2. Problém tohto dokonalého prístupu je v tom, že na generovanie z rozdelenia P^* potrebujeme poznať práve tú konštantu c , ktorú odhadujeme, čiže nepoznáme. Avšak môžeme poznať iné rozdelenie $\tilde{P}$, ktoré je podobné „ideálnemu“ P^* , z ktorého vieme generovať bez znalosti c , a súčasne $\tilde{D}[\tilde{L}Z]$ je oveľa menšia ako $D[Z]$.

Vzhľadom k prudkému poklesu $f(\mathbf{x})$ pre „veľké“ $\mathbf{x}$, väčšina pravdepodobnosti P^* zodpovedajúcej hustote určenej vzťahom (3.5) sa vyskytuje v okolí bodu $\mathbf{x}^*$, ktorý maximalizuje funkciu $f(\mathbf{x})$ na množine M . To naznačuje použitie podobného rozdelenia $\tilde{P}$, ktoré koncentruje väčšinu svojej pravdepodobnosti blízko $\mathbf{x}^*$, ktoré umožňuje ľahký výpočet $\tilde{L}$ a dajú sa z neho ľahko generovať premenné. Ak pôvodná hustota zodpovedá normálnemu rozdeleniu so strednou hodnotou $\mathbf{0}$ a kovariančnou maticou $\mathbf{C}$ a hustota $\tilde{f}$ zodpovedá normálnemu rozdeleniu so strednou hodnotou $\mathbf{x}^*$ a kovariančnou maticou $\mathbf{C}$, Radon-Nikodýmova derivácia $\tilde{L}$ je daná predpisom

$$\begin{aligned}\tilde{L}(\mathbf{x}) &= \frac{f(\mathbf{x})}{\tilde{f}(\mathbf{x})} = \frac{\frac{1}{\sqrt{(2\pi)^n |\mathbf{C}|}} \exp(-\frac{1}{2} \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x})}{\frac{1}{\sqrt{(2\pi)^n |\mathbf{C}|}} \exp(-\frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{x}^*))} \\ &= \exp \left(-\mathbf{x}^{*T} \mathbf{C}^{-1} \mathbf{x} + \frac{1}{2} \mathbf{x}^{*T} \mathbf{C}^{-1} \mathbf{x}^* \right).\end{aligned}\quad (3.6)$$

Uvažujme, že $\mathbf{X}$ je dvojrozmerný normálny náhodný vektor, t.j. $\mathbf{X} \sim N_2(\mathbf{0}, \mathbf{I})$ s hustotou $f(\mathbf{x})$ a množina M je kruh so stredom $(3, 3)$ a polomerom 1, teda $M = \{\mathbf{x} \in R^2 : \|\mathbf{x} - (3, 3)\| \leq 1\}$. Chceme vypočítať pravdepodobnosť p , že vektor $\mathbf{X}$ padne do množiny M , t.j. $p = P(A) = P(\mathbf{X} \in M)$. Situácia je zobrazená na obr. 3.1 vľavo. Využijeme najprv premennú Z , ktorá je zadefinovaná vo vzorci (3.3). Simuláciami v programe R - vygenerovaním n realizácií vektora $\mathbf{X}$ sme pomocou priemeru $\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i$ odhadli pravdepodobnosť $P(A) = E[Z]$ a vypočítali intervaly spoľahlivosti. Z $n = 10^5$ padlo

do množiny M 24 realizácií premennej X , to znamená, že $E[Z] = 2,40 \cdot 10^{-4}$, interval spoľahlivosti bol $(1,40 \cdot 10^{-4}, 3,40 \cdot 10^{-4})$. Po zopakovaní situácie pre rovnaký počet realizácií sme dostali výsledok $E[Z] = 3,20 \cdot 10^{-4}$, interval spoľahlivosti vyšiel $(2,10 \cdot 10^{-4}, 4,30 \cdot 10^{-4})$. Vidíme, že pravdepodobnosť počítaná týmto prístupom má vysokú disperziu, lebo rozdiel v tom, či do množiny M padne 24 alebo 32 realizácií spôsobí viditeľný rozdiel v intervaloch spoľahlivosti. Štandardná odchýlka je $4,90 \cdot 10^{-5}$.

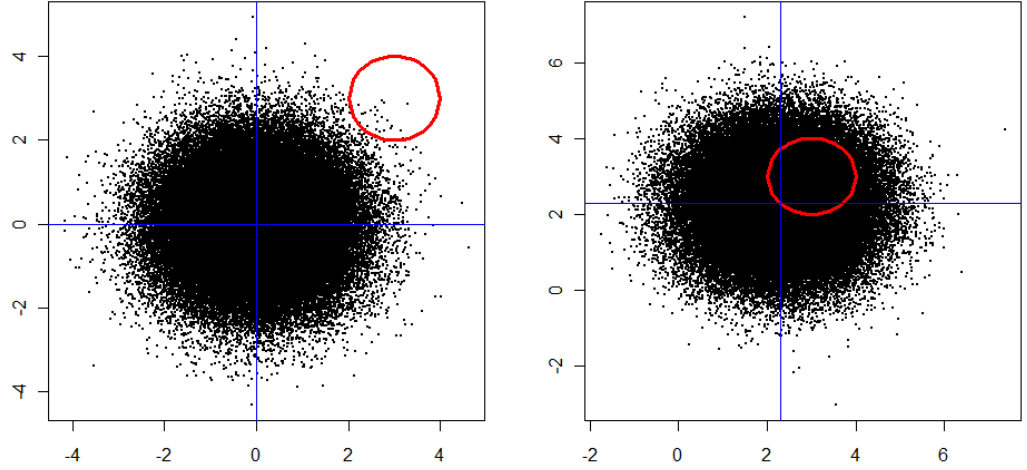
Skúsme na zmenšenie disperzie použiť metódu váhového výberu, premenné budeme generovať z rozdelenia - $\tilde{P}$, pravdepodobnosť udalosti odhadneme ako $P(A) = \tilde{E}[Z_N]$, kde Z_N je definovaná vo vzorci (3.4) a pomer $\frac{f(\mathbf{x})}{\tilde{f}(\mathbf{x})}$ je daný v rovnici (3.6). Teda premenná $Z_N(\mathbf{X})$ je v tomto prípade daná ako

$$Z_N(\omega) = \begin{cases} 0 & \text{pre } \mathbf{X}(\omega) \notin M, \\ \exp(\mathbf{x}^{*T}(\frac{1}{2}\mathbf{x}^* - \mathbf{X}(\omega))) & \text{pre } \mathbf{X}(\omega) \in M, \end{cases}$$

kde $\mathbf{x}^*$ maximalizuje funkciu $f(\mathbf{x})$ na množine M . Generujeme teda náhodné vektory z dvojrozmerného normálneho rozdelenia $\mathbf{X} \sim N_2(\mathbf{x}^*, \mathbf{I})$. Táto situácia je zobrazená na obr. 3.1 vpravo. Zmena oproti obr. vľavo je v tom, že oveľa viac realizácií padne do množiny M .

Pre rovnaký počet simulácií $n = 10^5$ odhadujeme $E[Z_N]$ a dostávame $E[Z_N] = 2,54 \cdot 10^{-4}$ s intervalmi spoľahlivosti $(2,50 \cdot 10^{-4}, 2,58 \cdot 10^{-4})$. Po zopakovaní tejto situácie viackrát dostávame rovnaké výsledky s presnosťou na 4 desatinné miesta, čo nás uisťuje v tvrdení, že pravdepodobnosť počítaná týmto prístupom má menšiu varianciu ako pravdepodobnosť počítaná predchádzajúcim spôsobom. Štandardná odchýlka je $2,22 \cdot 10^{-6}$, čo je naozaj výrazná zmena oproti predchádzajúcemu prístupu, kde štandardná odchýlka bola 22-krát väčšia.

Môžeme ešte vyskúšať obmenu tohto prístupu tak, že za $\mathbf{x}^*$ si zvolíme bod $(3,3)$, teda stred kruhu M , v tomto prípade odhad pravdepodobnosti vyšiel ako $E[Z_n] = 2,54 \cdot 10^{-4}$ s intervalmi spoľahlivosti $(2,50 \cdot 10^{-4}, 2,60 \cdot 10^{-4})$. Štandardná odchýlka v tomto prípade je $2,65 \cdot 10^{-5}$.



Obr. 3.1: Realizácia 10^5 náhodných vektorov z $\mathcal{N}_2(\mathbf{0}, \mathbf{I})$ (ľavá časť obrázku), resp. $\mathcal{N}_2(\mathbf{x}^*, \mathbf{I})$ (pravá časť obrázku). Červená kružnica označuje množinu M .

Pri porovnávaní štandardných odchýliek nám ako najlepšia metóda vychádza metóda váhového výberu, kde za $\mathbf{x}^*$ zvolíme také $\mathbf{x}^*$, ktoré maximalizuje funkciu $f(\mathbf{x})$ na množine M tak, ako to navrhuje veta 3.2.

Analogicky je možné odhadovať pravdepodobnosti aj v prípade, že množina M je iná ako kruh.

Použitím metódy váhového výberu sa podstatne skrátili intervaly spoľahlivosti, pretože sa znížila disperzia odhadovanej pravdepodobnosti.

3.2 Exponenciálna zmena miery

Nech $\mathbf{X}$ je náhodný vektor z rozdelenia s ľahkými chvostami, v zmysle, že platí $E[\exp(\theta^T \mathbf{X})] < \infty$ pre nejaké $\theta \in \mathbb{R}^n$, nech f je hustota tohto náhodného vektora a nech $\kappa(\theta) = \log E[\exp(\theta^T \mathbf{X})]$. Všimnime si, že $E[\exp(\theta^T \mathbf{X})]$ v jednorozmernom prípade zodpovedá m.v.f. v bode $\theta - M(\theta)$. Exponenciálne

naklonenou verziou hustoty f nazývame hustotu $\tilde{f}_\theta$ definovanú vzorcom

$$\tilde{f}_\theta(\mathbf{x}) = \frac{\exp(\theta^T \mathbf{x}) f(\mathbf{x})}{\exp(\kappa(\theta))}. \quad (3.7)$$

Všimnime si, že

$$\tilde{L}_\theta(\mathbf{x}) = \frac{f(\mathbf{x})}{\tilde{f}_\theta(\mathbf{x})} = \exp(-\theta^T \mathbf{x} + \kappa(\theta))$$

je Radon-Nikodýmova derivácia rozdelenia indukovaného hustotou f voči rozdeleniu indukovanému hustotou $\tilde{f}_\theta$.

Pravdepodobnostné rozdelenie dané hustotou $\tilde{f}_\theta$ v mnohých prípadoch patrí do tej istej triedy rozdelení ako f , takže premenné sú ľahko generovateľné, pokiaľ sú ľahko generovateľné premenné z pôvodného rozdelenia. Trieda všetkých rozdelení v tejto forme sa nazýva exponenciálna trieda generovaná rozdelením f .

Ak by nebola splnená podmienka $E[\exp(\theta^T \mathbf{X})] < \infty$, tak $\tilde{f}_\theta(\mathbf{x})$ by sa nedalo definovať. Tento prípad by nastal, ak by rozdelenie náhodného vektora $\mathbf{X}$ malo „ťažké chvosty“, po vynásobení hustoty $f(\mathbf{x})$ výrazom $E[\exp(\theta^T \mathbf{X})]$ by $\tilde{f}$ definovaná vzorcom (3.7) nadobudla hodnotu ∞ , a teda by nebola hustotou. Pokiaľ náhodný vektor $\mathbf{X}$ pochádza z rozdelenia s ľahkými chvostami, tak $E[\exp(\theta^T \mathbf{X})] f(\mathbf{x}) < \infty$ a $\tilde{f}_\theta(\mathbf{x})$, po predelení normalizačnou konštantou $\kappa(\theta)$, nadobúda hodnoty hustoty rozdelenia $\tilde{P}$. Všimnime si, že platí

$$\exp(\kappa(\theta)) = \int_{\mathbb{R}^n} \exp(\theta^T \mathbf{x}) f(\mathbf{x}) d\mathbf{x} = E[\exp(\theta^T \mathbf{X})],$$

a teda

$$\kappa(\theta) = \log E[\exp(\theta^T \mathbf{X})].$$

Takto vieme vypočítať hustotu $\tilde{f}_\theta(\mathbf{x})$ pre rôzne rozdelenia. Pre normálne rozdelenie $\mathcal{N}(\mu, \sigma^2)$, ktoré má hustotu $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$ je nová hustota $\tilde{f}_\theta(x)$ daná ako

$$\tilde{f}_\theta(x) = \frac{\exp(\theta x) f(x)}{\exp(\kappa(\theta))} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x - (\mu + \theta\sigma^2))^2}{2\sigma^2}\right),$$

čo zodpovedá hustote rozdelenia $\mathcal{N}(\mu + \theta\sigma^2, \sigma^2)$.

Exponenciálnu zmenu miery sme už použili v časti 3.1, kde Radon-Nikodýmova derivácia $\tilde{L}$ bola

$$\tilde{L}(\mathbf{x}) = \exp\left(-\mathbf{x}^{*T}\mathbf{C}^{-1}\mathbf{x} + \frac{1}{2}\mathbf{x}^{*T}\mathbf{C}^{-1}\mathbf{x}^*\right),$$

čo zodpovedá zápisu

$$\tilde{L} = \exp(-\theta^T\mathbf{X} + \kappa(\theta)), \text{ kde } \theta^T = \mathbf{x}^{*T}\mathbf{C}^{-1},$$

$$\kappa(\theta) = \frac{1}{2}\mathbf{x}^{*T}\mathbf{C}^{-1}\mathbf{x}^*.$$

Nasledujúci príklad opisuje prípad, kedy má vektor $\mathbf{X}$ nezávislé zložky.

Príklad 3.2. *Nech $X_1, \dots, X_n$ sú nezávislé náhodné premenné s hustotou $f(x)$ a nech metóda váhového výberu zachováva vlastnosť nezávislosti, ale zmení hustotu $f(x)$ na $\tilde{f}_\theta(x) = e^{\theta x - \kappa(\theta)}f(x)$. Potom pomer vierohodnosti, chápaný ako náhodná premenná, má tvar*

$$\tilde{L} = \prod_{i=1}^n \frac{f(X_i)}{\tilde{f}_\theta(X_i)} = \prod_{i=1}^n \frac{f(X_i)}{e^{\theta X_i - \kappa(\theta)}f(X_i)} = e^{-\theta S_n + n\kappa(\theta)},$$

kde $S_n = X_1 + \dots + X_n$. Hodnota Radon-Nikodýmovej derivácie $\tilde{L}$ závisí od S_n , teda nemusíme poznať hodnoty $X_1, \dots, X_n$, stačí poznať ich súčet. To je výhodné, ak uvažujeme taký proces, ktorého náhodné premenné sú súčtom predchádzajúcich - napr. náhodná prechádzka $S_n = \sum_{i=1}^n X_i$, kde $S_0 = 0$ a $X_1, X_2, \dots$ sú nezávislé náhodné premenné.

Exponenciálne naklonenie (angl. exponential tilting) sa používa pri problémoch s náhodnými premennými, ktorých rozdelenia majú ľahké chvosty (to znamená, že $\kappa(\theta)$ existuje pre všetky dostatočne malé $\theta > 0$), napríklad aby suma $S_n = X_1 + \dots + X_n$ dosahovala veľké hodnoty s väčšou pravdepodobnosťou. Ak sa snažíme dosiahnuť hodnoty x_0 alebo väčšie, tak sa snažíme zvoliť θ tak, aby platilo $E_\theta[S_n] = x_0$. Parameter θ je často označovaný ako sedlový bod (pozri kap.5 v knihe 3).

3.3 Odhad pravdepodobnosti prekročenia bariéry typu $S_n > n(\mu + \epsilon)$

Opäť uvažujme náhodnú prechádzku $S_n = X_1 + \dots + X_n$, kde $X_1, X_2, \dots$ sú nezávislé náhodné premenné s rozdelením F so strednou hodnotou μ . Odhadujeme pravdepodobnosť udalosti $A(n) = \{S_n > n(\mu + \epsilon)\}$, kde $\epsilon > 0$ a index zriedkavej udalosti n je v tomto prípade diskretný. Teda odhadujeme pravdepodobnosť, že náhodná prechádzka v konkrétnom čase n prekročí bariéru $n(\mu + \epsilon)$. Takže $p(n) = P(A(n)) \rightarrow 0$ pre $n \rightarrow \infty$, a teda existuje n , pre ktoré je udalosť $A(n)$ zriedkavá.

Predpokladáme, že $X_1, X_2, \dots$ majú normálne rozdelenie $\mathcal{N}(\mu, \sigma^2)$, X_i zodpovedá strate, ktorej poisťovňa čelí v čase i a S_n je celková strata poisťovne v čase n . Bariéra $n(\mu + \epsilon)$ vyjadruje zisk poisťovne v čase n a chceme určiť pravdepodobnosť $p(n)$, že v čase n bude strata poisťovne väčšia ako jej zisk. Zavedieme si premennú $Z(n)$

$$Z(n) = \begin{cases} 1 & \text{ak } S_n > n(\mu + \epsilon), \\ 0 & \text{ak } S_n \leq n(\mu + \epsilon). \end{cases}$$

Pravdepodobnosť $p(n)$ odhadneme ako strednú hodnotu $Z(n)$. Odhad pravdepodobnosti $p(n)$, kde $X_i \sim \mathcal{N}(0, 1)$, $\epsilon = 0,5$ a $n = 60$ pre 10^6 simulácií je $p(60) = 5,30 \cdot 10^{-5}$. Interval spoľahlivosti je $(-1,79 \cdot 10^{-3}, 1,89 \cdot 10^{-3})$, ľavý kraj intervalu spoľahlivosti pre takúto malú pravdepodobnosť vyšiel záporný a pravý kraj pomerne vysoký. Štandardná odchýlka je $7,28 \cdot 10^{-3}$.

Pravdepodobnosť pomocou exponenciálnej zmeny miery odhadneme ako strednú hodnotu náhodnej premennej $Z_{exp}(n)$, kde

$$Z_{exp}(n) = \begin{cases} e^{-\theta S_n + n\kappa(\theta)} & \text{ak } S_n > n(\mu + \epsilon), \\ 0 & \text{ak } S_n \leq n(\mu + \epsilon). \end{cases}$$

Optimálna θ je taká, pre ktorú platí

$$E_\theta[X] = \kappa'(\theta) = \mu + \epsilon. \quad (3.8)$$

Keďže κ' je rastúca funkcia, tak θ je určite kladná.

M.v.f. normálneho rozdelenia je daná ako $\hat{F}[\theta] = \exp(\theta\mu + \frac{1}{2}\sigma^2\theta^2)$, a teda $\kappa(\theta) = \log \hat{F}[\theta] = \theta\mu + \frac{1}{2}\sigma^2\theta^2$, z čoho podľa vzorca (3.8) vieme vypočítať optimálnu θ , platí

$$\begin{aligned}\kappa'(\theta) &= \mu + \epsilon, \\ \mu + \sigma^2\theta &= \mu + \epsilon, \\ \theta &= \frac{\epsilon}{\sigma^2}.\end{aligned}\tag{3.9}$$

Pre rovnaké parametre odhadneme pravdepodobnosť $p(60)$ pomocou exponenciálnej zmeny miery a dostávame výsledok $p(60) = 5,38 \cdot 10^{-5}$ s intervalmi spoľahlivosti $(2,54 \cdot 10^{-5}, 8,22 \cdot 10^{-5})$, ktoré sú podstatne kratšie, pretože štandardná odchýlka $1,12 \cdot 10^{-5}$ sa zmenšila oproti prechádzajúcemu odhadu 65-krát.

3.4 Exponenciálna zmena miery pre náhodný proces s diskretným časom

Budeme uvažovať základný pravdepodobnostný priestor $(\Omega, \mathcal{A}) = (\mathbb{R}^\infty, \mathcal{B}_\infty)$, kde $\mathbb{R}^\infty$ je množina postupností reálnych čísel a $\mathcal{B}_\infty$ sú príslušné borelovské množiny. Uvažujme náhodné premenné X_i , $i = 1, 2, \dots$ na tomto priestore definované ako $X_i(x) = x_i$ pre $x \in \mathbb{R}^\infty$ a $i \in \mathbb{N}$. Postupnosť $\{X_n\}_{n \in \mathbb{N}}$ tvorí náhodný proces s diskretným časom. Nech P a $\tilde{P}$ sú pravdepodobnosti na $(\mathbb{R}^\infty, \mathcal{B}_\infty)$. Symbolom P_n , resp. $\tilde{P}_n$ označme zodpovedajúce rozdelenie náhodného vektora $(X_1, \dots, X_n)'$. Predpokladajme, že P a $\tilde{P}$ sú také, že náhodné premenné $X_1, X_2, \dots$ sú nezávislé a rovnako rozdelené s hustotami f , resp. $\tilde{f}$. Pre jednoduchosť uvažujme, že f aj $\tilde{f}$ sú kladné na celom $\mathbb{R}$.

Naším cieľom je odhadnúť pravdepodobnosť $p = E[Z]$, kde E je operátor strednej hodnoty pre pravdepodobnosť P a Z je nejaká náhodná premenná na $(\mathbb{R}^\infty, \mathcal{B}_\infty)$. Špeciálne, nech

$$Z = g(X_1, \dots, X_\tau)1\{\tau < \infty\},$$

kde τ je merateľná náhodná premenná vzhľadom na $\sigma(X_1, \dots, X_n)$, čiže nastatie, resp. nenastatie udalosti $\{\tau = n\}$ vieme posúdiť z hodnôt $X_1, \dots, X_n$ a $1\{\tau < \infty\}$ je indikátor udalosti $\{\tau < \infty\}$, a teda platí

$$1\{\tau < \infty\} = \begin{cases} 1 & \text{ak } \tau < \infty, \\ 0 & \text{ak } \tau = \infty. \end{cases}$$

Vieme (pozri vzťah (3.2)), že platí $\tilde{L}_n(\mathbf{x}) = \frac{dP_n}{d\tilde{P}_n}(\mathbf{x}) = \prod_{i=1}^n \frac{f(x_i)}{\tilde{f}(x_i)}$, kde $\mathbf{x} = (x_1, \dots, x_n)$. Keďže udalosť $\{\tau < \infty\}$ je disjunktným zjednotením udalostí $\{\tau = n\}$ pre $n = 1, 2, \dots$, platí

$$Z = \sum_{n=1}^{\infty} g(X_1, \dots, X_n) 1\{\tau = n\}.$$

Preto

$$\begin{aligned} E[Z] &= E \left[\sum_{n=1}^{\infty} g(X_1, \dots, X_n) 1\{\tau = n\} \right] = \sum_{n=1}^{\infty} E[g(X_1, \dots, X_n) 1\{\tau = n\}] \\ &= \sum_{n=1}^{\infty} \tilde{E} \left[g(X_1, \dots, X_n) 1\{\tau = n\} \tilde{L}_n \right] = \tilde{E} \left[\sum_{n=1}^{\infty} g(X_1, \dots, X_n) 1\{\tau = n\} \tilde{L}_n \right] \\ &= \tilde{E} \left[\sum_{n=1}^{\infty} Z \tilde{L}_n 1\{\tau = n\} \right] = \tilde{E} [Z \tilde{L}_\tau]. \end{aligned}$$

3.5 Odhad pravdepodobnosti prekročenia bariéry typu $S_n > x$

Pri použití exponenciálnej zmeny miery uvažujeme, že náhodná premenná X má spojité rozdelenie s ľahkými chvostami. Presnejšie, $X_1, X_2, \dots$ sú nezávislé náhodné premenné s distribučnou funkciou F a s m.v.f. $\hat{F}[s] = E[e^{sX}] = \int_{-\infty}^{\infty} e^{sx} F(dx)$. Budeme sa snažiť nájsť také $s > 0$, pre ktoré je $\hat{F}[s]$ konečná.

Exponenciálna zmena miery je definovaná ako

$$f_\theta(x) = \frac{e^{\theta x}}{\hat{F}[\theta]} f(x) = e^{\theta x - \kappa(\theta)} f(x),$$

kde $\kappa(\theta) = \log \hat{F}[\theta]$ je kumulatívna vytvárajúca funkcia (ďalej k.v.f.). Pomer vierohodnosti $L_{n,\theta}$ pre $X_1, \dots, X_n$ je

$$L_{n,\theta} = \left(\frac{dP}{dP_\theta} \right)_n = \prod_{k=1}^n \frac{\hat{F}[\theta]}{e^{\theta X_k}} = e^{-\theta S_n} \hat{F}[\theta]^n = e^{-\theta S_n + n\kappa(\theta)},$$

kde $S_n = X_1 + \dots + X_n$. V tomto prípade Radon-Nikodýmovu deriváciu označujeme $L_{n,\theta}$ (nie $\tilde{L}_n$), pretože sa líši pre rôzne parametre θ . Preto ak Y_n je merateľná náhodná premenná vzhľadom na $\sigma(X_1, \dots, X_n)$, tak platí

$$E[Y_n] = E_\theta[Y_n L_{n,\theta}] = E_\theta[Y_n e^{-\theta S_n + n\kappa(\theta)}].$$

Aby $\hat{F}[s]$ bola konečná je dobré, aby $\hat{F}$ bola strmá (angl. steep) (pozri kap. 6 v [3]). Nech $\theta_{max} = \sup\{\theta : \hat{F}[\theta] < \infty\}$, pre rozdelenie s ľahkými chvostami platí $0 < \theta_{max} \leq \infty$. Strmosť m.v.f. potom znamená, že $\hat{F}[\theta] \uparrow \infty$ pre $\theta \uparrow \theta_{max}$. Príklad m.v.f. $\hat{F}[\theta]$ a k.v.f. $\kappa(\theta)$ rozdelenia F so zápornou strednou hodnotou je na obr. 3.2, kde γ_0 je riešením $\hat{F}'[\theta] = \kappa'(\theta) = 0$, resp. γ_0 je minimum $\hat{F}[\theta]$ a $\kappa(\theta)$ a γ je nenulové riešenie rovnice $\kappa(\gamma) = 0$, resp. $\hat{F}[\gamma] = 1$.

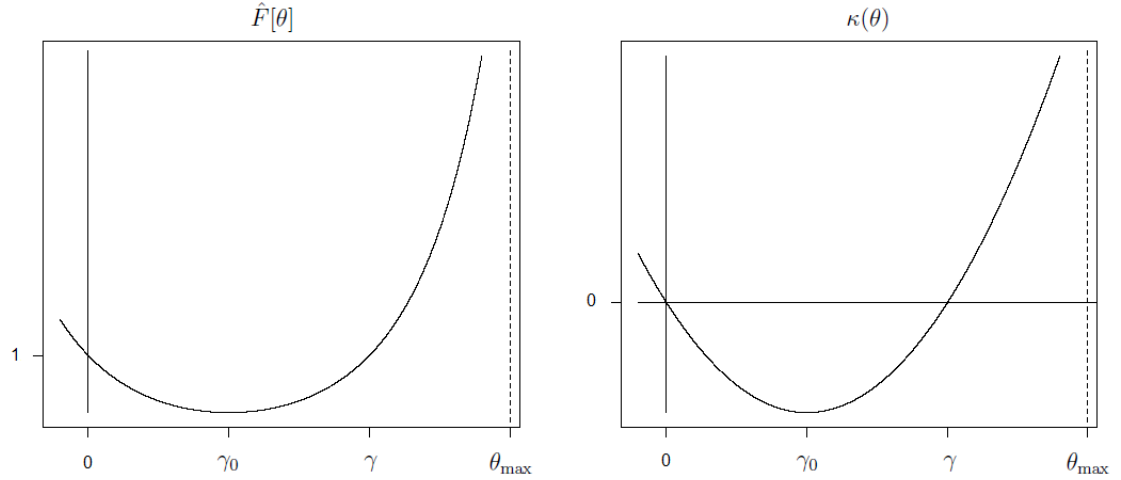
Pre strednú hodnotu μ_θ nového rozdelenia P_θ platí

$$\begin{aligned} \mu_\theta &= E_\theta[X] = \int_{\mathbb{R}} x f_\theta(x) dx = \int_{\mathbb{R}} x e^{\theta x - \kappa(\theta)} f(x) dx = \\ &= \int_{\mathbb{R}} \frac{x e^{\theta x} f(x)}{e^{\kappa(\theta)}} dx = E \left[\frac{X e^{\theta X}}{\hat{F}[\theta]} \right] = \frac{\hat{F}'[\theta]}{\hat{F}[\theta]} = \kappa'(\theta). \end{aligned}$$

Potom $\mu_\theta < 0$ pre $\theta < \gamma_0$, $\mu_\theta > 0$ pre $\theta > \gamma_0$ a $\mu_{\gamma_0} = 0$.

Nech F je taká distribučná funkcia, že ak náhodná premenná X má distribučnú funkciu F , tak $P(X \in (0, \infty]) > 0$ a že $E[X] < 0$. Snažíme sa odhadnúť pravdepodobnosť

$$p(x) = P(\tau_x < \infty), \text{ kde } \tau_x = \inf\{n : S_n > x\},$$



Obr. 3.2: Momentová vytvárajúca funkcia $\tilde{F}[\theta]$ (ľavá časť obrázku) a kumulatívna vytvárajúca funkcia $\kappa(\theta)$ (pravá časť obrázku)

kde x je veľké, a teda $p(x)$ je malé. Vieme posúdiť, či nastala udalosť $\{\tau_x \leq k\}$ len z realizácií náhodných premenných $X_1, \dots, X_k$ a platí $\{\tau_x < \infty\} = \bigcup_{n=1}^{\infty} \{\tau_x = n\}$. Po aplikácii exponenciálnej zmeny miery dostávame podobný výsledok ako v časti 3.4. V tomto prípade je náhodná premenná Z daná ako $Z = 1\{\tau_x < \infty\}$, čo je totožné so zápisom $Z = \sum_{n=1}^{\infty} 1\{\tau_x = n\}$ a pravdepodobnosť $p(x) = P(\tau_x < \infty)$ odhadneme ako strednú hodnotu Z

$$\begin{aligned}
 E[Z] &= E \left[\sum_{n=1}^{\infty} 1\{\tau_x = n\} \right] = \sum_{n=1}^{\infty} E_{\theta} [1\{\tau_x = n\} L_{n,\theta}] \\
 &= E_{\theta} \left[\sum_{n=1}^{\infty} L_{n,\theta} 1\{\tau_x = n\} \right] = E_{\theta} [L_{\tau_x, \theta} 1\{\tau_x < \infty\}] \\
 &= E_{\theta} [e^{-\theta S_{\tau_x} + \tau_x \kappa(\theta)} 1\{\tau_x < \infty\}].
 \end{aligned} \tag{3.10}$$

Je prirodzené voliť parameter θ tak, aby náhodná prechádzka S_n s pravdepodobnosťou 1 prekročila bariéru x , pretože to urýchli proces počítania. Optimálna θ teda musí spĺňať podmienku $P_{\theta}(\tau_x < \infty) = 1$, čo znamená, že náhodná prechádzka S_n vždy prekročí bariéru x , preto musí byť splnená

podmienka $E_\theta[X] \geq 0$, to platí pre $\theta > \gamma_0$. Potom môžeme rovnicu (3.10) prepísať ako

$$p(x) = P(\tau_x < \infty) = E_\theta[L_{\tau_x, \theta}] = E_\theta[e^{-\theta S_{\tau_x} + \tau_x \kappa(\theta)}].$$

Zvyčajne pre optimálnu θ platí $\theta = \gamma$. Predpokladáme, že riešením rovnice $\hat{F}[\gamma] = 1$ je $\gamma > 0$ a $\hat{F}'[\gamma] < \infty$. V tomto prípade platí

$$p(x) = P(\tau_x < \infty) = E_\gamma[e^{-\gamma S_{\tau_x}}].$$

Príklad 3.3. *Predpokladajme, že $S_n = X_1 + \dots + X_n$, kde $X_i \sim \mathcal{N}(-\mu, 1)$ a $\mu > 0$. Čas τ_x bude čas, v ktorom prechádzka S_n nadobudla hodnotu väčšiu ako x . Pre odhadnutie pravdepodobnosti $p(x) = P(\tau_x < \infty)$, čiže pravdepodobnosti, že náhodná prechádzka S_n niekedy vo svojej histórii prekročí x bez použitia exponenciálnej zmeny miery zavedieme náhodnú premennú $Z(x)$, pre ktorú platí*

$$Z(x) = \begin{cases} 1 & \text{ak } \tau_x < \infty, \\ 0 & \text{ak } \tau_x = \infty. \end{cases} \quad (3.11)$$

Je náročné určiť kedy je splnená podmienka $\tau_x = \infty$, pretože ak chceme s istotou vedieť, že náhodná prechádzka S_n nikdy neprekročila bariéru x , je potrebné preskúmať celú históriu náhodnej prechádzky, čiže vygenerovať nekonečne veľa hodnôt X_i , resp. S_i , čo nie je prakticky možné. Preto túto podmienku upravíme na $S_n < B$ pre nejaké $n \in \mathbb{N}$, hranica $B < 0$ je určená tak, aby pre náhodnú prechádzku S_n , ktorá klesne pod B , nastala situácia $S_n > x$ pre nejaké n so zanedbateľnou pravdepodobnosťou. Teda (3.11) môžeme prepísať na

$$Z(x) = \begin{cases} 1 & \text{ak } S_n > x \text{ pre nejaké } n, \\ 0 & \text{ak } S_n < B \text{ pre nejaké } n. \end{cases} \quad (3.12)$$

Náhodná prechádzka S_n pre všetky n s pravdepodobnosťou 1 buď prekročí bariéru x alebo klesne pod hranicu B . Hranicu B vypočítame ako $B = x - d$, kde $d = 22$ je bariéra, ktorú prekročí náhodná prechádzka s driftom $-\mu = -1$ s pravdepodobnosťou 10^{-20} . Bariéru d sme určili pomocou exponenciálnej zmeny miery.

Pre $n = 10^6$ simulácií, náhodná prechádzka s driftom $-\mu = -1$ prekročí bariéru $x = 5$ s pravdepodobnosťou, ktorú sme vypočítali bez použitia exponenciálnej zmeny miery, $p(5) = 1,30 \cdot 10^{-5}$, intervaly spoľahlivosti sú $(5,93 \cdot 10^{-6}, 2,01 \cdot 10^{-5})$ a štandardná odchýlka je $3,60 \cdot 10^{-3}$, pomerne vysoká.

Pravdepodobnosť $p(x)$ s použitím exponenciálnej zmeny miery určíme ako $p(x) = E_\gamma[e^{-\gamma S_{\tau_x}}]$, kde γ je riešením rovnice $\hat{F}[\gamma] = 1$. M.v.f. pre rozdelenie $\mathcal{N}(-\mu, 1)$ je $\hat{F}[s] = \exp(-\mu s + s^2/2)$, to znamená, že γ spĺňa rovnicu $0 = -\mu\gamma + \gamma^2/2$. Predpokladáme, že $\gamma > 0$, preto $\gamma = 2\mu$. Keďže γ je zvolená tak, že $\kappa(\gamma) = 0$, tak platí

$$\begin{aligned}\hat{F}_\gamma[s] &= \int_{\mathbb{R}} e^{sx} f_\gamma(x) dx = \int_{\mathbb{R}} e^{sx} e^{\gamma x - \kappa(\gamma)} f(x) dx = \\ &= \int_{\mathbb{R}} e^{(s+\gamma)x} f(x) dx = \hat{F}[s + \gamma] = \exp(\mu s + s^2/2),\end{aligned}\tag{3.13}$$

čo zodpovedá m.v.f. rozdelenia $\mathcal{N}(\mu, 1)$, ak predpokladáme jednoznačnosť rozdelenia s m.v.f. (3.13). Teda platí rovnosť $p(x) = E_\gamma[e^{-2\mu S_{\tau_x}}]$, kde $S_{\tau_x} = X_1 + \dots + X_{\tau_x}$ a $X_i \sim \mathcal{N}(\mu, 1)$, čo znamená, že náhodná prechádzka má pri novom rozdelení kladný drift a nastáva s pravdepodobnosťou 1. Pri pôvodnom rozdelení mala náhodná prechádzka záporný drift a nastávala zriedka.

Pre rovnaký počet simulácií $n = 10^6$ náhodná prechádzka s driftom $-\mu = -1$ prekročí bariéru $x = 5$ s pravdepodobnosťou $p(5) = 1,46 \cdot 10^{-5}$, ktorej intervaly spoľahlivosti sú $(1,45 \cdot 10^{-5}, 1,46 \cdot 10^{-5})$, štandardná odchýlka je $1,27 \cdot 10^{-5}$, teda 280-krát menšia ako pri odhade, ktorý sme vypočítali bez použitia exponenciálnej zmeny miery a zároveň sa výpočtový čas 3-krát skrátil.

Príklad 3.4. Predpokladajme, že pre krok náhodnej prechádzky platí $X_i = U_i - T_i$, čo je rozdiel dvoch náhodných premenných s exponenciálnym rozdelením, $U \sim \text{Exp}(\delta)$ a $T \sim \text{Exp}(\beta)$, kde $\beta < \delta$. Stredná hodnota náhodnej premennej X_i je záporná, pretože platí

$$E[X_i] = E[U_i] - E[T_i] = 1/\delta - 1/\beta < 0.$$

Pravdepodobnosť $p(x) = P(\tau_x < \infty)$ určíme ako strednú hodnotu premennej $Z(x)$, ktorá je definovaná v (3.12), hranicu B vypočítame ako $B = x - \frac{d}{\delta - \beta}$,

kde $d = 44$ je bariéra, ktorú prekročí náhodná prechádzka s parametrami $\delta = 2, \beta = 1$.

Pre $n = 10^6$ simulácií náhodná prechádzka s parametrami $\delta = 2$ a $\beta = 1$ prekročí bariéru $x = 10$ s pravdepodobnosťou $p(10) = 2,50 \cdot 10^{-5}$, intervaly spoľahlivosti sú $(1,52 \cdot 10^{-5}, 3,48 \cdot 10^{-5})$ a štandardná odchýlka odhadu je $5,00 \cdot 10^{-3}$.

Použitie exponenciálnej zmeny miery vyžaduje, aby sme poznali m.v.f. $\hat{F}[\cdot]$, pretože parameter γ je riešením rovnice $\hat{F}[\gamma] = 1$. M.v.f. náhodnej premennej X vieme vypočítať ako

$$\begin{aligned}\hat{F}_X[s] &= E[e^{sX}] = E[e^{s(U-T)}] = E[e^{sU}] E[e^{-sT}] = \\ &= \frac{\delta}{\delta - s} \frac{\beta}{\beta + s},\end{aligned}$$

teda $\hat{F}[\gamma] = 1$ má pre $\gamma > 0$ jediné riešenie $\gamma = \delta - \beta$. Poznamenajme, že sme využili $E[VW] = E[V]E[W]$ ak V a W sú nezávislé. Potom platí

$$\hat{F}_\gamma[s] = \hat{F}[s + \gamma] = \frac{\beta}{\beta - s} \frac{\delta}{\delta + s}, \quad (3.14)$$

a teda nové rozdelenie je tiež rozdielom dvoch exponenciálnych rozdelení (ak predpokladáme jednoznačnosť rozdelenia s m.v.f. (3.14)), ale s vymenenými parametrami. Pravdepodobnosť $p(x)$ určíme ako $p(x) = E_\gamma[e^{-(\delta-\beta)S_{\tau_x}}]$, kde krok náhodnej prechádzky je $X_i = T_i - U_i$, teda náhodná premenná X má kladnú strednú hodnotu a bariéru x prekročí s pravdepodobnosťou 1. Pre rovnaký počet simulácií $n = 10^6$ a s totožnými hodnotami parametrov $\delta = 2, \beta = 1$ odhadneme pravdepodobnosť ako $p(10) = 2,2686 \cdot 10^{-5}$ s intervalmi spoľahlivosti $(2,2660 \cdot 10^{-5}, 2,2712 \cdot 10^{-5})$. Štandardná odchýlka odhadu je $1,31 \cdot 10^{-5}$, oproti pôvodnému odhadu sa zmenšila 380-krát, čo znamená, že odhad vypočítaný pomocou exponenciálnej zmeny miery je podstatne presnejší.

3.6 Odhad pravdepodobnosti záporného zisku 4 roky po sebe

Využijeme predchádzajúcu teóriu a budeme predpokladať, že $X_1, \dots, X_\tau$ sú nezávislé náhodné premenné s normálnym rozdelením $\mathcal{N}(\mu, 1)$ a ich hodnoty zodpovedajú výškam zisku v rokoch $1, 2, \dots, 10$. Chceme vypočítať pravdepodobnosť p záporného zisku 4 roky po sebe. To znamená, že τ je konštantne 10 rokov a premenná $Z = g(X_1, \dots, X_{10})$ je daná ako

$$Z = \begin{cases} 1 & \text{ak } (\exists i \in \{1, \dots, 7\})(X_i < 0 \wedge X_{i+1} < 0 \wedge X_{i+2} < 0 \wedge X_{i+3} < 0), \\ 0 & \text{inak.} \end{cases}$$

Hľadáme pravdepodobnosť v tvare $p = E[Z]$. Ak $\mu = 1, 2$, tak hľadaná pravdepodobnosť bude veľmi malá. Napríklad pre 10^4 pozorovaní je odhadovaná pravdepodobnosť $9,00 \cdot 10^{-4}$ s intervalmi spoľahlivosti $(3,12 \cdot 10^{-4}, 1,49 \cdot 10^{-3})$. Pri zopakovaní situácie je vysoko pravdepodobné, že dostaneme iný výsledok vzhľadom na vysokú štandardnú odchýlku - $2,99 \cdot 10^{-2}$. Na zníženie disperzie, a teda na spresnenie odhadu vypočítame pravdepodobnosť pomocou exponenciálnej zmeny miery. Tentokrát budeme generovať $X_1, \dots, X_\tau$ z normálneho rozdelenia $\mathcal{N}(\mu^*, 1)$, kde $0 < \mu^* < \mu$ a hľadáme pravdepodobnosť v tvare $p = \tilde{E}[Z\tilde{L}_{10}]$. Nová náhodná premenná $Z_N = Z\tilde{L}_{10}$ je v tomto prípade definovaná ako

$$Z_N = \begin{cases} \tilde{L}_{10} & \text{ak } (\exists i \in \{1, \dots, 7\})(X_i < 0 \wedge X_{i+1} < 0 \wedge X_{i+2} < 0 \wedge X_{i+3} < 0), \\ 0 & \text{inak,} \end{cases}$$

kde

$$\tilde{L}_{10} = \prod_{i=1}^{10} \frac{f(X_i)}{\tilde{f}(X_i)} = \exp \left(5(\mu^{*2} - \mu^2) + (\mu - \mu^*) \sum_{i=1}^{10} X_i \right).$$

Pre $\mu^* = 0.8$ a rovnaký počet pozorovaní je pravdepodobnosť $p = \tilde{E}[Z\tilde{L}_{10}] = 1,10 \cdot 10^{-3}$ s intervalmi spoľahlivosti $(8,02 \cdot 10^{-4}, 1,40 \cdot 10^{-3})$ a štandardnou odchýlkou $1,53 \cdot 10^{-2}$, to znamená, že exponenciálna zmena miery naozaj

zmenšila disperziu odhadu - oproti predchádzajúcemu prístupu sa štandardná odchýlka dvakrát zmenšila.

3.7 Odhad pravdepodobnosti extrémálnej hodnoty zloženého Poissonovho rozdelenia

Nech $X(t) = Y_1 + \dots + Y_{N(t)}$, kde $\{N(t)\}$ je Poissonov proces, v ktorom doby $T_1, T_1 + T_2, \dots$, a Y_i medzi udalosťami majú spoločnú hustotu f . Rozdelenie náhodnej premennej $X(t)$ sa nazýva zložené Poissonovo rozdelenie s parametrami λ a f . Ak chceme nájsť také $\tilde{P}$, aby $\{X(t)\}$ malo opäť zložené Poissonovo rozdelenie s parametrami $\tilde{\lambda}, \tilde{f}$, potom

$$\tilde{L}_t = \frac{e^{-\lambda R(t)}}{e^{-\tilde{\lambda} R(t)}} \prod_{i=1}^{N(t)} \frac{f(y_i) \lambda e^{-\lambda T_i}}{\tilde{f}(y_i) \tilde{\lambda} e^{-\tilde{\lambda} T_i}} = \left(\frac{\lambda}{\tilde{\lambda}} \right)^{N(t)} e^{-(\lambda - \tilde{\lambda})t} \prod_{i=1}^{N(t)} \frac{f(y_i)}{\tilde{f}(y_i)},$$

kde $R(t) = t - T_1 - \dots - T_{N(t)}$ je čas od poslednej udalosti do času t (pozri kap. 5 v [3]).

Príklad 3.5. Zložené Poissonovo rozdelenie sa často používa v poistnej matematike, kde modeluje celkovú výšku škody poisťovne. V zmysle tohto príkladu $Y_1, Y_2, \dots$ zodpovedajú pohľadávkam, ktorým čelí poisťovňa v čase $T_1, T_1 + T_2, \dots$. Doby medzi udalosťami - $T_1, T_2, T_3, \dots$ majú exponenciálne rozdelenie a $N(t)$ vyjadruje počet pohľadávok do času t . To znamená, že $X(t) = Y_1 + \dots + Y_{N(t)}$ je celková suma, ktorú poisťovňa vyplatí do času t .

Chceme vypočítať pravdepodobnosť p , že celková výška škody do času $t = 10$ rokov presiahne hodnotu rezerv $B = 20$, teda $p = P(X(t) > B)$. Predpokladáme, že jednotlivé pohľadávky Y_i majú exponenciálne rozdelenie $\text{Exp}(\lambda_{\text{exp}})$, počet pohľadávok $N(t)$ má Poissonovo rozdelenie $\text{Pois}(\lambda t)$. Zadefinujeme si premennú Z ako

$$Z = \begin{cases} 1 & \text{ak } X(t) > B, \\ 0 & \text{inak.} \end{cases}$$

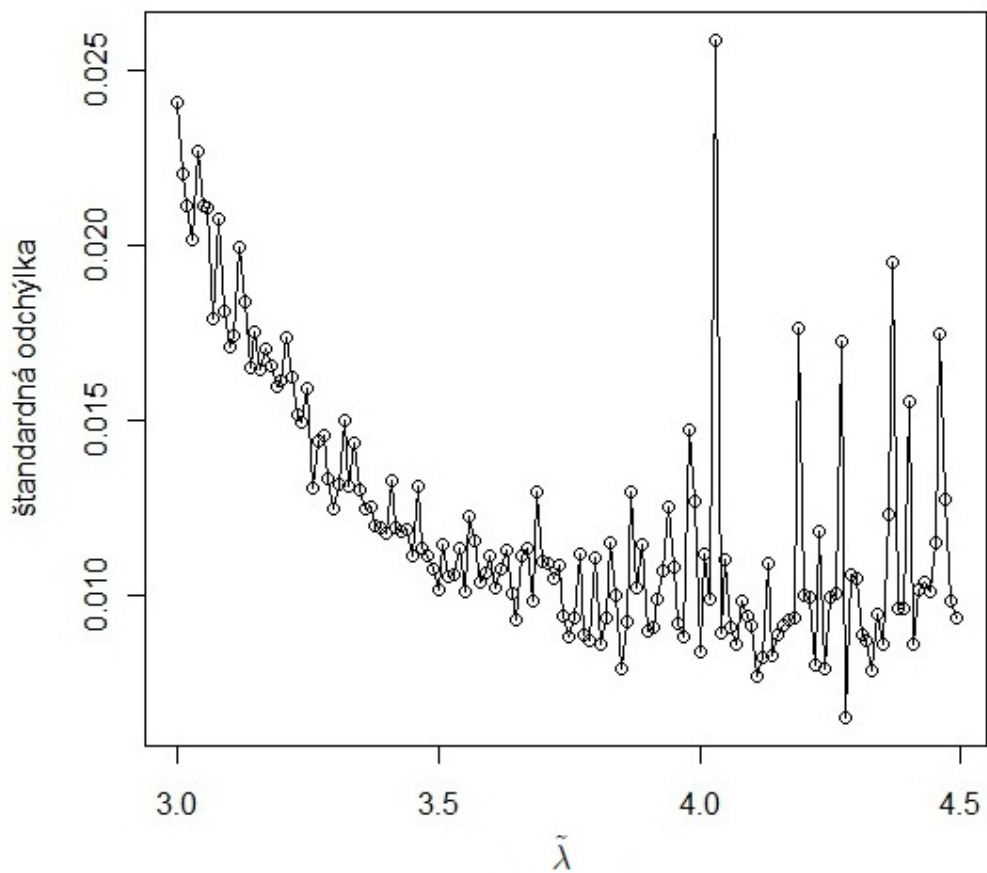
Pravdepodobnosť odhadujeme v tvare $p = E[Z]$. Nech $\lambda_{exp} = \lambda = 3$, teda pohľadávky Y_i majú exponenciálne rozdelenie $Exp(3)$ a počet pohľadávok $N(t)$ má Poissonovo rozdelenie $Pois(30)$, potom pre $n = 10^5$ pozorovaní výška škody presiahla hodnotu rezerv 64-krát, to znamená, že $p = E[Z] = 6,40 \cdot 10^{-4}$ a 95%-né intervaly spoľahlivosti sú $(4,83 \cdot 10^{-4}, 7,97 \cdot 10^{-4})$ so štandardnou odchýlkou $2,53 \cdot 10^{-2}$, ktorá je pomerne vysoká. Na jej zníženie použijeme exponenciálnu zmenu miery. V tomto prípade počet pohľadávok $N(t)$ bude mať Poissonovo rozdelenie $Pois(\tilde{\lambda}t)$ a pravdepodobnosť odhadneme v tvare $p = \tilde{E}[Z_N] = \tilde{E}[Z\tilde{L}_t]$, teda pre náhodnú premennú Z_N platí

$$Z_N = \begin{cases} \tilde{L}_t & \text{ak } X(t) > B, \\ 0 & \text{inak,} \end{cases}$$

kde $\tilde{L}_t = \left(\frac{\lambda}{\tilde{\lambda}}\right)^{N(t)} e^{-(\lambda - \tilde{\lambda})t}$. Na efektívne zníženie disperzie potrebujeme zvoliť vhodné $\tilde{\lambda}$. Je prirodzené požadovať aby $\tilde{\lambda}$ bola väčšia ako λ , pretože potom do času t nastane viac poistných udalostí, a teda častejšie nastane situácia $X(t) > B$, čo je naším cieľom pri použití exponenciálnej zmeny miery. Ak však zvolíme $\tilde{\lambda}$ príliš veľké, tak štandardná odchýlka sa opäť zvýši. Preto sme vypočítali odhad štandardnej odchýlky pre $\tilde{\lambda} = \{3, 3,01, \dots, 4,49, 4,5\}$ a zakreslili sme to do obr. 3.3.

Z grafu vidíme, že so zvyšujúcim $\tilde{\lambda}$ odhad št.odchýlky klesá, no ako každý odhad má svoju varianciu, ktorá sa po čase začne zvyšovať. Najlepšie je preto zvoliť také $\tilde{\lambda}$, pre ktoré sú odhad št.odchýlky aj variancia odhadu pomerne malé. Túto podmienku splňa napr. $\tilde{\lambda} = 3,7$, pre ktorú je št.odchýlka $9,40 \cdot 10^{-3}$, teda takmer 3-krát menšia ako št.odchýlka vypočítaná pre pôvodný odhad. Odhadnutá pravdepodobnosť $p = 5,56 \cdot 10^{-4}$ s intervalmi spoľahlivosti $(4,97 \cdot 10^{-4}, 6,14 \cdot 10^{-4})$.

Priklad 3.6. Opäť uvažujme náhodnú premennú $X = Y_1 + Y_2 + \dots + Y_N$, ktorá má zložené Poissonovo rozdelenie. Náhodné premenné $Y_1, Y_2, \dots, Y_N, N$ sú



Obr. 3.3: Štandardná odchýlka pre rôzne $\tilde{\lambda}$

nezávislé, náhodná premenná N má Poissonovo rozdelenie $\text{Pois}(\lambda)$ a náhodné premenné $Y_1, Y_2, \dots, Y_N$ majú gama rozdelenie $\Gamma(a, b)$ s hustotou f . Chceme odhadnúť pravdepodobnosť $p = P(A(c))$, kde $A(c) = \{X > c\}$ a c je veľké. Pravdepodobnosť p určíme ako strednú hodnotu premennej Z

$$Z = \begin{cases} 1 & \text{ak } X > c, \\ 0 & \text{ak } X \leq c. \end{cases}$$

Pravdepodobnosť $p = E[Z]$ pre $n = 10^6$ simulácií, $\lambda = 30$, $c = 110$,

$a = 4$, $b = 2$ je $1,54 \cdot 10^{-4}$ s intervalmi spoľahlivosti $(1,29 \cdot 10^{-4}, 1,78 \cdot 10^{-4})$ a štandardnou odchýlkou $1,24 \cdot 10^{-3}$.

Po použití exponenciálnej zmeny miery budú premenné $Y_1, Y_2, \dots$ mať rozdelenia s hustotou $f_\theta(y) = \frac{\exp(\theta y) f(y)}{\exp(\kappa(\theta))}$, kde $f(y) = \frac{1}{b^a} \frac{1}{\Gamma(a)} y^{a-1} \exp(-\frac{y}{b})$ pre $y \geq 0$ a $\kappa(\theta) = \ln \hat{F}(y) = -a \ln(1 - b\theta)$, potom platí

$$f_\theta(y) = \frac{1}{\left(\frac{b}{1-b\theta}\right)^a} \frac{1}{\Gamma(a)} y^{a-1} \exp\left(-y \left(\frac{b}{1-b\theta}\right)^{-1}\right),$$

čo zodpovedá predpisu hustoty pre rozdelenie $\Gamma(a, \frac{b}{1-b\theta})$. Takže premenné $Y_1, Y_2, \dots$ majú rozdelenie $\Gamma(a, \frac{b}{1-b\theta})$ a pravdepodobnosť p určíme ako strednú hodnotu premennej Z_{exp}

$$Z_{exp} = \begin{cases} \exp(-\theta x + \kappa(\theta)) & \text{ak } X > c, \\ 0 & \text{ak } X \leq c. \end{cases}$$

Parameter θ sme zvolili $\theta = 0,43$, zvyšné parametre zostali rovnaké a odhad pravdepodobnosti je v tomto prípade $p = 1,55 \cdot 10^{-4}$ s intervalmi spoľahlivosti $(1,54 \cdot 10^{-4}, 1,56 \cdot 10^{-4})$ a štandardnou odchýlkou $5,53 \cdot 10^{-4}$, čo znamená, že štandardná odchýlka sa zmenšila 22-krát.

3.8 Odhad pravdepodobnosti zriedkavých udalostí v Markovských reťazcoch

Markovský proces je taký t -parametrický systém náhodných veličín $\{X_t\}_{t \in I}$, kde I je interval alebo diskrétna množina indexov, pre ktorý platí, že ak je daná hodnota X_s , tak budúce hodnoty X_t pre $t > s$ môžu závisieť iba od X_s , nie však od predošlých hodnôt X_u pre $u < s$ (pozri [7]).

Definícia 3.2. *Markovský reťazec s diskrétnym časom*

Markovský reťazec $\{X(t)\}_{t \in I}$ je Markovský proces s množinou „časov“ $I = \mathbb{N}$ a spočítateľnou množinou stavov S . To znamená, že pre každé $n \in \mathbb{N}$ a $i, j_1, \dots, j_n \in S$ platí:

$$P(X_{n+1} = i | X_1 = j_1, \dots, X_n = j_n) = P(X_{n+1} = i | X_n = j_n).$$

Definícia 3.3. Homogénny Markovovský reťazec

Markovovský reťazec $\{X_t\}_{t \in I}$ nazývame homogénnym, ak pre každé $n \in \mathbb{N}$ a $i, j \in S$ platí

$$P(X_{n+1} = i | X_n = j) = p_{ji},$$

kde p_{ji} nazývame pravdepodobnosťami prechodu.

Predpokladáme, že chceme vypočítať $z = E[g(X_1, \dots, X_\tau)1\{\tau < \infty\}]$, kde τ je čas definovaný v termínoch $\{X_n\}_{n \in \mathbb{N}}$. Ak $\{X_n\}$ je nehomogénny Markovovský reťazec s konečnou množinou stavov a s rozdelením P , tak je potrebné zvoliť také rozdelenie $\tilde{P}$, ktoré zachováva Markovovskú vlastnosť (aby sa zabezpečilo efektívne generovanie trajektórií z rozdelenia $\tilde{P}$).

Je praktické uvažovať takú mieru $\tilde{P}$, pre ktorú platí $\tilde{P}(\tau < \infty) = 1$, potom $z = \tilde{E}[g(X_1, \dots, X_\tau)\tilde{L}_\tau]$.

Nech $\tilde{P}$ je pravdepodobnostné rozdelenie nového homogénneho Markovovského reťazca s pravdepodobnosťami prechodu $\tilde{p}_{ji}$ a nech $\tilde{P}$ má rovnaké počiatočné rozdelenie (rozdelenie náhodnej premennej X_1) ako P , ktorého pravdepodobnosti prechodu sú p_{ji} , tak platí

$$z = \tilde{E}[g(X_1, \dots, X_\tau)\tilde{L}_\tau],$$

dá sa ukázať (pozri kap.5 v [3]), že Radon-Nikodýmova derivácia $\tilde{L}_\tau$ je

$$\tilde{L}_\tau = \prod_{i=2}^{\tau} \frac{p_{X_{i-1}X_i}}{\tilde{p}_{X_{i-1}X_i}},$$

my uvedieme len vlastný popis základnej myšlienky. Ak pri pravdepodobnostom rozdelení P má spojitý náhodný vektor $\mathbf{X}$ hustotu f a pri pravdepodobnostom rozdelení $\tilde{P}$ má $\mathbf{X}$ hustotu $\tilde{f}$, tak v prípade, že náhodný vektor $\mathbf{X}$ nadobudne hodnotu $\mathbf{x}$, Radon-Nikodýmova derivácia nadobudne hodnotu $\tilde{L}(\mathbf{x}) = \frac{f(\mathbf{x})}{\tilde{f}(\mathbf{x})}$. My však uvažujeme diskretný náhodný vektor $\mathbf{X}$, analogický výraz k výrazu $\frac{f(\mathbf{x})}{\tilde{f}(\mathbf{x})}$ je $\frac{P(\mathbf{X}=\mathbf{x})}{\tilde{P}(\mathbf{X}=\mathbf{x})}$, teda

$$\tilde{L}(\mathbf{x}) = \frac{P(\mathbf{X} = \mathbf{x})}{\tilde{P}(\mathbf{X} = \mathbf{x})}.$$

Ak $\mathbf{x}$ má zložky $\mathbf{x} = (j_1, \dots, j_n)$, tak potom

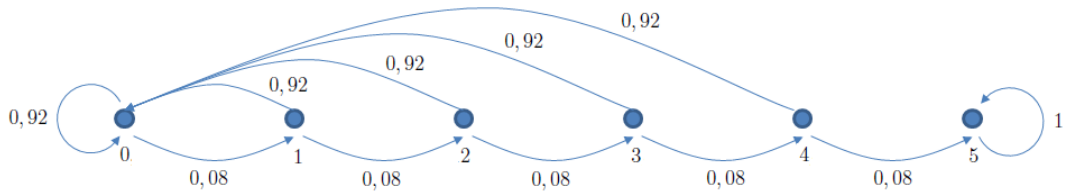
$$\begin{aligned}
P(\mathbf{X} = \mathbf{x}) &= \\
&= P((X_1, X_2, \dots, X_n) = (j_1, j_2, \dots, j_n)) = \\
&= P(X_1 = j_1) \frac{P(X_2 = j_2, X_1 = j_1)}{P(X_1 = j_1)} \dots \frac{P(X_n = j_n, X_1 = j_1, \dots, X_{n-1} = j_{n-1})}{P(X_1 = j_1, \dots, X_{n-1} = j_{n-1})} = \\
&= P(X_1 = j_1) P(X_2 = j_2 | X_1 = j_1) \dots P(X_n = j_n | X_1 = j_1, \dots, X_{n-1} = j_{n-1}) = \\
&= P(X_1 = j_1) P(X_2 = j_2 | X_1 = j_1) \dots P(X_n = j_n | X_{n-1} = j_{n-1}) = \\
&= p_{j_1} p_{j_1 j_2} \dots p_{j_{n-1} j_n}.
\end{aligned}$$

Využili sme vzťah pre podmienenú pravdepodobnosť: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ a Markovovskú vlastnosť. Keďže predpokladáme rovnaké rozdelenie X_1 pri P aj $\tilde{P}$, tak je rozumné očakávať

$$\tilde{L}_\tau = \prod_{i=2}^{\tau} \frac{p_{X_{i-1} X_i}}{\tilde{p}_{X_{i-1} X_i}}.$$

3.8.1 Hod kockou

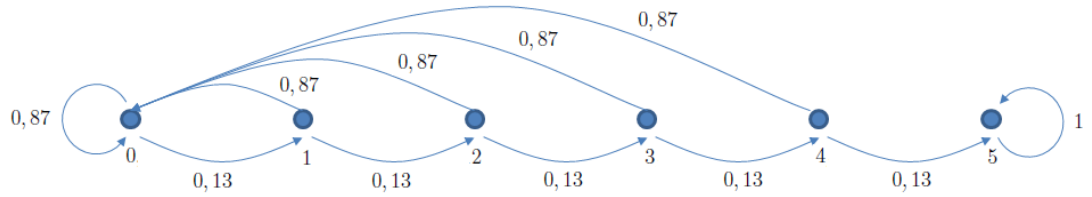
Predstavme si, že máme nevyváženú hraciu kocku, na ktorej 6-ka padá s pravdepodobnosťou $q = 0,08$. Chceme odhadnúť pravdepodobnosť udalosti A - v postupnosti 100 hodov sa na niektorom mieste vyskytne 5 šestiek za sebou, graf prechodových pravdepodobností je zobrazený na obr. 3.8.1.



Obr. 3.4: Graf prechodových pravdepodobností, ak šestka padá s pravdepodobnosťou $q = 0,08$

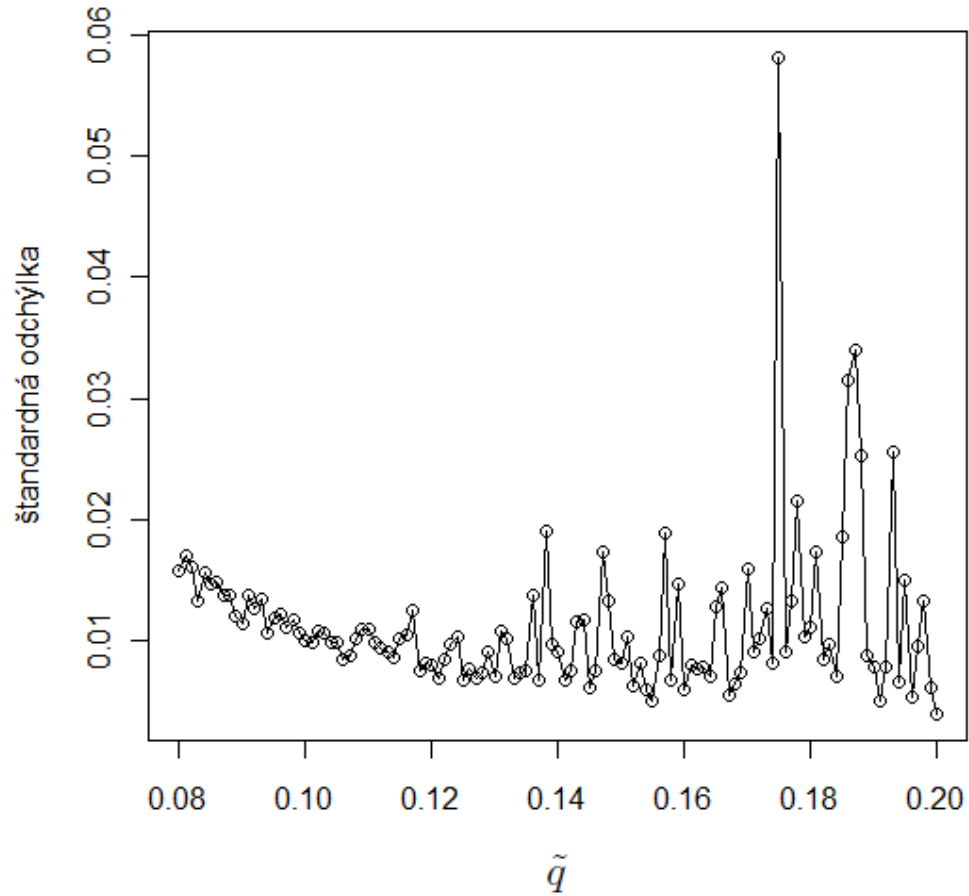
Pravdepodobnosť budeme odhadovať v tvare $p = P(A) = E[Z]$, kde premenná Z je indikátorová funkcia množiny A - $Z_i = I(A_i)$. Pre $n = 10^6$ pozorovaní dostávame odhad $p = 2,98 \cdot 10^{-4}$ s intervalmi spoľahlivosti $(2,64 \cdot 10^{-4}, 3,32 \cdot 10^{-4})$, ktorý má štandardnú odchýlku $1,73 \cdot 10^{-2}$, ktorú sa opäť pokúsime znížiť exponenciálnou zmenou miery.

Pravdepodobnosť budeme odhadovať v tvare $p = \tilde{E}[Z_N] = \tilde{E}[Z\tilde{L}_\tau]$, kde τ je počet hodov potrebných na hodenie 6-ky 5-krát a $\tilde{L}_\tau = \prod_1^\tau \frac{p_{X_{i-1}X_i}}{\tilde{p}_{X_{i-1}X_i}}$, kde prechodové pravdepodobnosti $\tilde{p}$ zvolíme tak, aby zodpovedali analogickému procesu (hádzaniu kockou), v ktorom ale šestka padá s pravdepodobnosťou $\tilde{q}$. Prechodové pravdepodobnosti poznáme - 6-ka padá s pravdepodobnosťou q , resp. $\tilde{q}$ a nepadá s pravdepodobnosťou $1-q$, resp. $1-\tilde{q}$, preto $\tilde{L}_\tau = \frac{q^U(1-q)^{NU}}{\tilde{q}^U(1-\tilde{q})^{NU}}$, kde U je počet hodov 6-ky do nastatia udalosti A a $NU = \tau - U$ je počet hodov čísel rôznych od 6-ky.



Obr. 3.5: Graf prechodových pravdepodobností, ak kocka padá s pravdepodobnosťou $q = 0.13$

Pravdepodobnosť $\tilde{q}$ je pravdepodobnosť s akou padá 6-ka po zmene miery, ktorú volíme prirodzene väčšiu ako pôvodnú pravdepodobnosť q , pretože potom bude sledovaná udalosť nastávať v novom procese častejšie, jej konkrétnu hodnotu opäť určíme z grafu št.odchýlky pre $\tilde{q} = \{0,080, 0,081, \dots, 0,200\}$, ktorý je zobrazený na obr. 3.6.



Obr. 3.6: Štandardná odchýlka pre rôzne $\tilde{q}$

Rovnako ako v predchádzajúcom príklade môžeme pozorovať zníženie št.odchýlky pre zvyšujúce $\tilde{q}$, no zároveň postupný nárast variance odhadu št.odchýlky. Ak zvolíme $\tilde{q} = 0,13$, tak odhadnutá pravdepodobnosť je $2,65 \cdot 10^{-4}$ s intervalmi spoľahlivosti $(2,50 \cdot 10^{-4}, 2,80 \cdot 10^{-4})$ a št.odchýlka $7.71 \cdot 10^{-3}$ sa 2-krát zmenšila.

Všimnime si, že tento problém je podobný ako problém odhadu záporného zisku poisťovne štyrikrát po sebe a aj výsledne zníženie disperzie odhadu je podobné, čiže nie veľmi výrazné. Pre tento typ problémov zrejme nie je

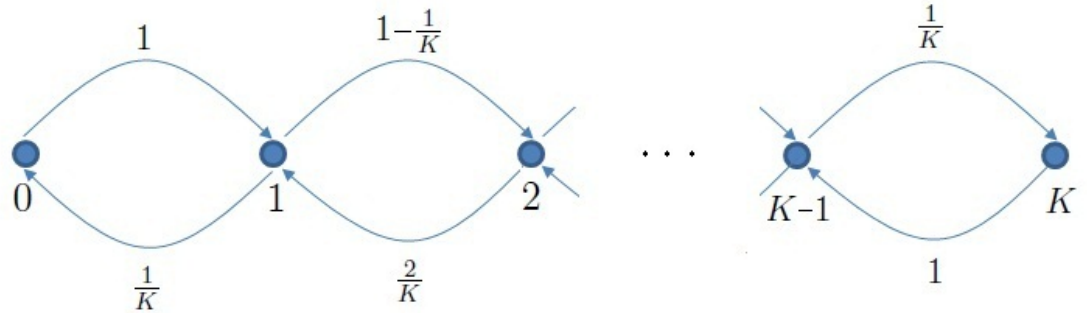
exponenciálna zmena miery až taká prínosná, ako pre iné typy problémov.

3.8.2 Ehrenfestov model

Model, navrhnutý Paulom Ehrenfestom, uvažuje K molekúl v dvoch susediacich kontajneroch (pozri [5]). Kontajnery sú spojené otvorom, cez ktorý prechádzajú molekuly plynu. Predpokladáme, že v každom časovom kroku sa náhodne vybratá molekula presunie z jedného kontajnera do druhého. Nech $X(t)$ je počet molekúl v prvom kontajneri po t časových krokoch, potom $\{X(t)\}$ je Markovovský reťazec s grafom prechodu znázorneným na obr. 3.7 a maticou prechodu $P = (p_{ij})$

$$p_{i,i+1} = 1 - \frac{i}{K}, p_{i,i-1} = \frac{i}{K}, p_{ij} = 0 \text{ ak } |i - j| > 1,$$

pre $0 \leq i \leq K$.



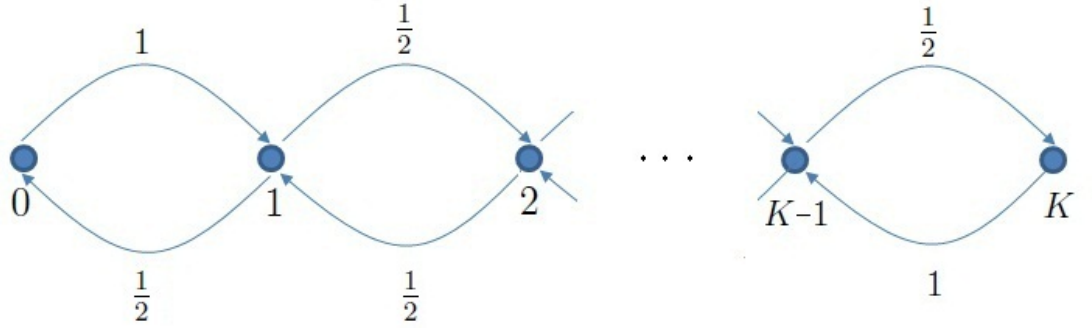
Obr. 3.7: Graf prechodových pravdepodobností pre Ehrenfestov model častíc

Chceme určiť pravdepodobnosť, že do 100 časových krokov nastane situácia, že všetky molekuly sa budú nachádzať len v jednom kontajneri, čiže proces sa dostane do stavu "0", alebo do stavu "K". Táto pravdepodobnosť by sa dala vypočítať aj pomocou pokročilej teórie Markovovských procesov,

ale my to odhadneme simulačne. Uvažujeme $K = 22$ molekúl s počiatočným stavom $X(1) = 11$. Pravdepodobnosť určíme ako strednú hodnotu premennej Z

$$Z = \begin{cases} 1 & \text{ak } X(t) = 0 \text{ alebo } X(t) = K \text{ pre } 0 \leq t \leq 100, \\ 0 & \text{inak.} \end{cases} \quad (3.15)$$

Pre $n = 10^6$ simulácií je odhadnutá pravdepodobnosť $p = 3,70 \cdot 10^{-5}$ s intervalmi spoľahlivosti $(2,51 \cdot 10^{-5}, 4,89 \cdot 10^{-5})$ a štandardnou odchýlkou $6,08 \cdot 10^{-3}$.



Obr. 3.8: Graf prechodových pravdepodobností pre Ehrenfestov model častíc po exponenciálnej zmene miery

Môžeme zvoliť zmenu miery tak, že sa zmení graf prechodových pravdepodobností na taký, ktorý je zobrazený na obr. 3.8, čo znamená, že vyprázdnenie jedného kontajnera nastane častejšie. V tomto prípade určíme pravdepodobnosť p ako strednú hodnotu premennej Z_{exp}

$$Z_{exp} = \begin{cases} \prod_1^\tau \frac{p_{X_{i-1}X_i}}{\tilde{p}_{X_{i-1}X_i}} & \text{ak } X(t) = 0 \text{ alebo } X(t) = K \text{ pre } 0 \leq t \leq 100, \\ 0 & \text{inak,} \end{cases}$$

kde τ označuje časový krok, v ktorom nastalo vyprázdnenie kontajnera. Odhad pravdepodobnosti v tomto prípade vyšiel $p = 3,79 \cdot 10^{-5}$ s intervalmi

spoľahlivosti $(3,77 \cdot 10^{-5}, 3,82 \cdot 10^{-5})$ a štandardnou odchýlkou $1,1 \cdot 10^{-4}$, teda 55-krát menšou ako pri pôvodnom odhade.

Záver

Odhadnúť pravdepodobnosť nejakej udalosti simulačne sa dá jednoducho pomocou klasických metód ako sú výberové štatistiky, či intervaly spoľahlivosti. Problém však nastáva, keď pomocou týchto metód chceme odhadnúť pravdepodobnosť zriedkavej udalosti. V tom prípade vzhľadom k odhadovanej pravdepodobnosti má odhad vysokú smerodajnú odchýlku, čo spôsobí jeho nepresnosť a relatívne široké intervaly spoľahlivosti.

Zvýšiť presnosť odhadu sa dá zvýšením počtu simulácií, ale to je neefektívne. Na spresnenie odhadu o 1 rád potrebujeme 100-krát viac simulácií a času. V mnohých prípadoch oveľa efektívnejší spôsob zvýšenia presnosti je zníženie disperzie, na čo sa používa metóda váhového výberu, v ktorej sa aplikuje exponenciálna zmena miery. Napríklad v príklade 3.4 pri odhade pravdepodobnosti prekročenia bariéry typu $S_n > x$ sa štandardná odchýlka zmenšila 380-krát, keby sme chceli takýto výsledok dosiahnuť zvýšením počtu simulácií, tak by sme museli mať 38000-krát viac simulácií. Okrem spresnenia odhadu má exponenciálna zmena miery aj iné výhody, a to zníženie simuláčného času, napr. v príklade 3.3 je možné overiť, že odhadnutie pravdepodobnosti je po exponenciálnej zmene miery výpočtovo efektívnejšie ako pred ňou.

V práci sú popísané len prípady, kedy náhodné premenné majú rozdelenie s ľahkými chvostami, ktoré sú pre exponenciálnu zmenu miery kľúčové, pretože ak by rozdelenie nemalo ľahké chvosty, tak príslušnú zmenu miery by nebolo možné urobiť. V praxi sa ale často vyskytujú udalosti, ktoré sa nedajú popísať rozdeleniami s ľahkými chvostami. Na prácu by bolo možné

nadviazať analyzovaním metód pre prípady, kedy náhodné premenné majú pravdepodobnostné rozdelenie s ťažkými chvostami.

Literatúra

- [1] LAMOŠ, F., POTOCKÝ, R. : *Pravdepodobnosť a matematická štatistika - Štatistické analýzy*, UNIVERZITA KOMENSKÉHO V BRATISLAVE, 1998
- [2] HARMAN, R. : *Stochastické simulačné metódy*, učebný materiál KAMŠ FMFI UK, <http://www.iam.fmph.uniba.sk/ospm/Harman/sim.pdf>
- [3] ASMUSSEN, S., GLYNN, P.W. : *Stochastic simulation*, SPRINGER, 2007
- [4] BRUNOVSKÝ, P.: *Stochastické metódy operačnej analýzy*, učebný materiál FMFI UK, http://www.iam.fmph.uniba.sk/skripta/brunovsky-smoa/smoa_text_09-10.pdf
- [5] GRIMMETT, G., STIRZAKER, D. : *Probability and Random Processes*, OXFORD UNIVERSITY PRESS, 2001
- [6] FISHMAN, G.S. : *Monte Carlo: Concepts, Algorithmus and Applications*, SPRINGER, 1995
- [7] ŠEVČOVIČ, D., STEHLÍKOVÁ, B., MIKULA, K. : *Analytické a numerické metódy oceňovania finančných derivátov*, STU, 2009
- [8] KALAS, J., PEKÁR, J. : *Simulačné metódy*, MFF UK, 1991

- [9] R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>

Prílohy

Odhadovanie pravdepodobnosti zriedkavých realizácií normálneho náhodného vektora

```
rnorm.2d<-function(n,mu1=0,mu2=0,sd1=1,sd2=1,rho=0)
{
# Generátor z dvojrozmerného normálneho rozdelenia
  X1<-rnorm(n) #X1,X2 su nahodne vektory
  X2<-rnorm(n)
  Y1<-sd1*X1+mu1
  Y2<-sd2*(rho*X1+sqrt(1-rho^2)*X2)+mu2
  rbind(Y1,Y2)
}
#-----
v<-c(3,3)
alfa<-0.05
n<-100000
m<-10
p<-0

for (i in 1:n)
{
X<-rnorm.2d(1)
if (((X[1]-v[1])^2+(X[2]-v[2])^2)<=1)
p=p+1
}
Z<-p/n

del<-qnorm(1-alfa/2)*sqrt(Z*(1-Z)/n)
IS_lavy<-Z-del
IS_pravy<-Z+del
st.odchylka<-sd(Z)
#-----
#exp. zmena miery
#-----
#stredna hodnota rozdelenia, z kt. generujeme je x_max
#x_max maximalizuje funkciu f(x) na množine M
```

```

x_max<-c(v[1]-sin(pi/4),v[2]-sin(pi/4))
C<-matrix( c(1,0,0,1), ncol=2, nrow=2, byrow=TRUE )
for (i in 1:n)
{
X<-rnorm.2d(1,mu1=x_max[1],mu2=-x_max[2])
if (((X[1]-v[1])^2+(X[2]-v[2])^2)<=1)
Z[i]=exp(-x_max%*%C%*%X+1/2*x_max%*%C%*%x_max)
else Z[i]=0
}

del<-qnorm(1-alfa/2)*sd(Z)/sqrt(n)
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
st.odchylka<-sd(Z)
#-----
#stredna hodnota rozdelenia, z kt. generujeme je stred kruhu

x_stred<-c(v[1],v[2])
C<- matrix( c(1,0,0,1), ncol=2, nrow=2, byrow=TRUE )
for (i in 1:n)
{
X<-rnorm.2d(1,mu1=x_stred[1],mu2=-x_stred[2])
if (((X[1]-v[1])^2+(X[2]-v[2])^2)<=1)
Z[i]=exp(-x_stred%*%C%*%X+1/2*x_stred%*%C%*%x_stred)
else Z[i]=0
}

Zp<-mean(Z)
del<-qnorm(1-alfa/2)*sd(Z)/sqrt(n)
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
st.odchylka<-sd(Z)

```

Odhad pravdepodobnosti prekročenia bariéry typu $S_n > n(\mu + \epsilon)$

```

N<-10^6
mu<- 0
sigma<-1
epsilon<-0.5
n<-60
alfa<-0.05
#-----
#odhad pravdepodobnosti prekročenia bariery typu  $S_n > n(\mu + \epsilon)$ 
Z<-rep(0,N)
h<-n*(mu+epsilon)
for (i in 1:N){

```

```

S<-sum(rnorm(n,mu,sigma))
if(S>h)Z[i]<-1
}
del<-qnorm(1-alfa/2)*sqrt( mean(Z)*(1-mean(Z))/n )
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
#-----
#exponential change of measure

theta<-epsilon/(sigma^2)
const<-n*(theta*mu+1/2*sigma^2*theta^2)
h<-n*(mu+epsilon)

Z<-rep(0,N)
for (i in 1:N){
S<-sum(rnorm(n,mu+epsilon,sigma))

if(S>h)Z[i]<-exp(-theta*S+const)
}

del<-qnorm(1-alfa/2)*sd(Z)/sqrt(n)
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)

```

Odhad pravdepodobnosti prekročenia bariéry typu $\{S_n > x\}$

```

#generujeme z N(-mu,1), odhadujem pp, ze nah.prech prekroci barieru x>0
n<-10^6
alfa<-0.05
x<-5
mu<-1

#-----
#ako urcit barieru B, pod ktoru ak klesne nahodna prechadzka S,
#tak so zanedbatelnou pp-ou by prekrocila hranicu x

Z<-rep(0,n)
d<-22/mu
for (i in 1:n){
S<-0
while (S<d){
S<-S+rnorm(1,mu,1)
}
}

```

```

Z[i]<-exp(-2*mu*S)
}
mean(Z)
#-----
#odhad pravdepodobnosti prekrozenia bariery typu S_n>x
#prechadzka S_n je suctom nah.premennych s normalnym rozdelenim

a<-date()
Z<-rep(0,n)
d<-22/mu
B<-x-d
for (i in 1:n){
S<-0
while (S<x & S>B){
S<-S+rnorm(1,-mu,1)
}
if (S>x) Z[i]<-1
if (S<B) Z[i]<-0
}

b<-date()
del<-qnorm(1-alfa/2)*sqrt( mean(Z)*(1-mean(Z))/n )
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
as.numeric(difftime(a,b, tz="", units=x("secs")))
#-----
#Exponential change of measure
a<-date()
Z<-rep(0,n)
for (i in 1:n){
S<-0
while (S<x){
S<-S+rnorm(1,mu,1)
}
Z[i]<-exp(-2*mu*S)
}

del<-qnorm(1-alfa/2)*sd(Z)/sqrt(n)
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
b<-date()
as.numeric(difftime(a,b))
#-----
#odhad pravdepodobnosti prekrozenia bariery typu S_n>x

```

```

#prechadzka S_n=X_1+X_2+... , kde X_i je rozdielom dvoch nah.prem.s exp.rozdelenim

n<-10^6
alfa<-0.05
x<-10
delta<-2
beta<-1

a<-date()
Z<-rep(0,n)
B<-x-d
for (i in 1:n){
  S<-0
  while (S<x & S>B){
    S<-S+rexp(1,delta)-rexp(1,beta)
  }
  if (S>x) Z[i]<-1
  if (S<B) Z[i]<-0
}

del<-qnorm(1-alfa/2)*sqrt( mean(Z)*(1-mean(Z))/n )
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
b<-date()
as.numeric(difftime(a,b))
#-----
#Exponential change of measure
a<-date()
Z<-rep(0,n)
for (i in 1:n){
  S<-0
  while (S<x){
    S<-S+rexp(1,beta)-rexp(1,delta)
  }
  Z[i]<-exp(-(delta-beta)*S)
}

del<-qnorm(1-alfa/2)*sd(Z)/sqrt(n)
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
b<-date()
as.numeric(difftime(a,b))

```


Odhad pravdepodobnosti záporného zisku 4 roky po sebe

```
n<-10000
mu<-1.2
tau<-10
Z<-rep(0,n)

for (j in 1:n){
  A<-0
  x<-rep(0,10)
  x[1]<-rnorm(1,mu,sd)
  for (i in 4:tau){
    x[i]<-rnorm(1,mu,sd)
    if (x[i]<0 & x[i-1]<0 & x[i-2]<0 & x[i-3]<0) {Z[j]<-1}
  }
}

del<-qnorm(1-alfa/2)*sqrt( mean(Z)*(1-mean(Z))/n )
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
#-----

#exp. zmena miery
mu_h<-0.8
Z<-rep(0,n)
for (j in 1:n){
  x<-rnorm(tau,mean=mu_h,sd=1)
  for (i in 4:tau){
    if (x[i]<0 & x[i-1]<0 & x[i-2]<0 & x[i-3]<0){
      Z[j]<-exp((tau/2)*(mu_h^2-mu^2)+(mu-mu_h)*sum(x))
    }
  }
}

del<-qnorm(1-alfa/2)*sd(Z)/sqrt(n)
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
```

Odhad pravdepodobnosti extrémálnej hodnoty zloženého Poissonovho rozdelenia

#odhad pravdepodobnosti, ze vyska skody poistovne presiahne hodnotu rezerv

```
n<-100000
alfa<-0.05
```

```

lambda<-3
lambda_exp<-3

B<-20
t<-10

Z<-rep(0,n)
for (i in 1:n){
N<-rpois(1,lambda*t)
S<-sum(rexp(N,lambda_exp))
if (S>B) Z[i]<-1
}
del<-qnorm(1-alfa/2)*sqrt( mean(Z)*(1-mean(Z))/n )
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
#-----

#exp. zmena miery
lambda_h<-3.7
Z<-rep(0,n)
for (i in 1:n){
N<-rpois(1,lambda_h*t)
S<-sum(rexp(N,lambda_exp))
if (S>B) Z[i]<-(lambda/lambda_h)^N*exp((lambda_h-lambda)*t)
}

del<-qnorm(1-alfa/2)*sd(Z)/sqrt(n)
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
#-----

#obrazok
lambda_h<-seq(1,5,by=0.1)
V<-rep(0,length(lambda_h))
for (j in 1:length(lambda_h)){
Z<-rep(0,n)
for (i in 1:n){
N<-rpois(1,lambda_h[j]*t)
S<-sum(rexp(N,lambda_exp))
if (S>B) Z[i]<-(lambda/lambda_h[j])^N*exp((lambda_h[j]-lambda)*t)
}
V[j]<-sd(Z)
}
plot(lambda_h,V,xlab="lambda",ylab="štandardná odchýlka")
lines(lambda_h,V)
#-----

```

```

#Odhad pravdepodobnosti  $A(c) = \{X > c\}$ 
n<-10^6
alfa<-0.05
lambda<-30

a<-4
b<-2

c<-110

Z<-rep(0,n)
for (i in 1:n){
N<-rpois(1,lambda)
X<-sum(rgamma(N,shape=a,scale=1/b))
if (X>c) Z[i]<-1
}

del<-qnorm(1-alfa/2)*sqrt( mean(Z)*(1-mean(Z))/n )
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
#-----
#exp. zmena miery

theta<-0.043
b_exp<-b/(1-theta*b)

Z<-rep(0,n)
for (i in 1:n){
N<-rpois(1,lambda)
X<-sum(rgamma(N,shape=a,scale=b_exp))
if (X>c) Z[i]<-exp(-theta*X+log(1/(1-b*theta)^a))
}

del<-qnorm(1-alfa/2)*sd(Z)/sqrt(n)
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
#-----
#obrazok
lambda_h<-seq(1,5,by=0.1)
V<-rep(0,length(lambda_h))
for (j in 1:length(lambda_h)){
Z<-rep(0,n)
for (i in 1:n){
N<-rpois(1,lambda_h[j]*t)

```

```

S<-sum(rexp(N,lambda_exp))
if (S>B) Z[i]<-(lambda/lambda_h[j])^N*exp((lambda_h[j]-lambda)*t)
}
V[j]<-sd(Z)
}
plot(lambda_h,V,xlab="lambda",ylab="štandardná odchýlka")
lines(lambda_h,V)

```

Odhad pravdepodobnosti, ze 5 krat za sebou hodime 6ku

#aka je pp., ze 5x za sebou hodim 6ku

```

n<-10^6
N<-100
q<-0.08
alfa<-0.05

Z<-rep(0,n)
for (j in 1:n){
  x<-rbinom(N,1,q)
  for (i in 5:N){
    if (x[i]==1 & x[i-1]==1 & x[i-2]==1 & x[i-3]==1 & x[i-4]==1 ){
      Z[j]<-1
    }
    break
  }
}

del<-qnorm(1-alfa/2)*sqrt( mean(Z)*(1-mean(Z))/n )
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
#-----
#exp. zmena miery
q_h<-0.13
Z<-rep(0,n)
for (j in 1:n){
  x<-rbinom(N,1,q_h)
  for (i in 5:N){
    if ( x[i]==1 & x[i-1]==1 & x[i-2]==1 & x[i-3]==1 & x[i-4]==1 ){
      U<-sum(x[1:i])          #pocet vyskokov
      NU<-i-U #pocet padnuti
      Z[j]<-(q/q_h)^U*((1-q)/(1-q_h))^NU
    }
    break
  }
}

```

```

}
del<-qnorm(1-alfa/2)*sd(Z)/sqrt(n)
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
#-----
#obrazok
q_h<-seq(0.08,0.25,by=0.001)
V<-rep(0,length(q_h))
for (y in 1:length(q_h)){
  Z<-rep(0,n)
  for (j in 1:n){
    x<-rbinom(N,1,q_h[y])
    for (i in 5:N){
      if (x[i]==1 & x[i-1]==1 & x[i-2]==1 & x[i-3]==1 & x[i-4]==1 ){
        U<-sum(x[1:i])      #pocet vysokov
        NU<-i-U #pocet padnuti
        Z[j]<-(q/q_h[y])^U*((1-q)/(1-q_h[y]))^NU
      }
    }
  }
  V[y]<-sd(Z)
}
plot(q_h,V,xlab="q",ylab="štandardná odchýlka")
lines(q_h,V)

```

Ehrenfestov model

```

#funkcia, ktora urcuje presuvanie molekúl z jedného kontajneru do druhého
pp<-function(X,K)
{
  a<-rbinom(1,1,X/K)
  if (a==1) Y=X-1
  if (a==0) Y=X+1
  Y
}
#-----
n<-10^6
N<-100
K<-22 #pocet molekúl v oboch kontajneroch
alfa<-0.05

Z<-rep(0,n)
for (j in 1:n){
  X<-rep(0,N)
  X[1]<-ceiling(K/2) #pociatocny pocet molekúl v 1.kontajneri

```

```

for (i in 2:N){
X[i]<-pp(X[i-1],K)
if (X[i]==0 || X[i]==K){
      Z[j]<-1
break
}
}
}

del<-qnorm(1-alfa/2)*sqrt( mean(Z)*(1-mean(Z))/n )
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)
#-----
#funkcia, ktora vypocita sucin prechodovych pravdepodobnosti
ppp<-function(X,K)
{
m<-length(X)
for (j in 1:(m-1)){
if (X[j]>X[j+1]) p[j]<-X[j]/K
if (X[j]<X[j+1]) p[j]<-(K-X[j])/K
}
prod(p)
}
#-----
Z<-rep(0,n)
for (j in 1:n){
X<-rep(0,N)
X[1]<-ceiling(K/2)
for (i in 2:N){
X[i]<-X[i-1]+2*floor(2*runif(1))-1
if (X[i]==0 || X[i]==K){
      Z[j]<-ppp(X[1:i],K)/((1/2)^(i-1))
break
}
}
}
del<-qnorm(1-alfa/2)*sd(Z)/sqrt(n)
IS_lavy<-mean(Z)-del
IS_pravy<-mean(Z)+del
priemer<-mean(Z)
st.odchylka<-sd(Z)

```