

UNIVERZITA PALACKÉHO V OLOMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA  
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

**DIPLOMOVÁ PRÁCE**  
**MCMC metody**



Vedoucí diplomové práce:  
**RNDr. Karel Hron, Ph.D.**  
Rok odevzdání: 2012

Vypracovala:  
**Bc. Marie Poláková**  
AME, II. ročník

### **Prohlášení**

Prohlašuji, že diplomovou práci na téma MCMC METODY jsem vypracovala samostatně na základě uvedených pramenů a literatury.

V Olomouci dne 11. prosince 2012

.....

## **Poděkování**

Ráda bych na tomto místě poděkovala vedoucímu mé diplomové práce RNDr. Karlu Hronovi, Ph.D. Také bych ráda poděkovala mé rodině za jejich podporu po celou dobu mého studia a Mgr. Blaženě Kratochvílové, Ph.D. za cenné rady při psaní této diplomové práce.

# OBSAH

<b>Úvod .....</b>	<b>5</b>
<b>1. Základní pojmy .....</b>	<b>7</b>
1.1. Markovovy řetězce .....	7
1.2. Bayesovský model .....	14
<b>2. MCMC metody .....</b>	<b>22</b>
2.1. Monte Carlo integrace .....	22
2.2. Skryté Markovovy řetězce .....	34
<b>3. MCMC filtry .....</b>	<b>41</b>
3.1. Importance Sampling .....	42
3.2. Sequential Importance Sampling .....	47
3.3. Resampling .....	47
3.4. Obecný Sequential Monte Carlo filtr .....	54
<b>Závěr .....</b>	<b>55</b>
<b>Příloha.....</b>	<b>56</b>
Příklad 1 - Aposteriorní rozdělení B.....	56
Příklad 7.....	57
Příklad 8.....	57
<b>Literatura .....</b>	<b>58</b>

## Úvod

Markov Chain Monte Carlo (MCMC) metody jsou důležitá třída algoritmů, která nám umožňuje simulovat složité stochastické systémy. V praxi většinou stojíme před problémem, kdy chceme generovat posloupnost vzorků z nějakého složitějšího pravděpodobnostního rozdělení a na základě těchto dat odhadnout parametry tohoto rozdělení (např. střední hodnotu nebo rozptyl). Jednou z možností, jak můžeme těchto výsledků dosáhnout, je využití bayesovského přístupu. Avšak v mnoha případech se setkáváme s problémy, jak bychom měli vhodně určit bayesovské aposteriorní rozdělení. Pro tyto případy vznikly MCMC metody, pomocí nichž vytvoříme Markovův řetězec a po dostatečně velkém počtu kroků rozdělení pravděpodobnosti tohoto řetězce bude konvergovat k našemu cílovému rozdělení. Jednoduše řečeno, snažíme se najít takový Markovův řetězec, jehož stacionární rozdělení se za určitých podmínek bude rovnat našemu cílovému rozdělení. V praxi se MCMC metody používají např.: k výpočtu složitých integrálů, k odhadu stochastické volatility, k řešení optimalizačních problémů pomocí metody simulovaného žihání a nebo k restaurování digitálních obrázků. Je zřejmé, že v posledních letech MCMC algoritmy hrají významnou roli v mnoha oblastech výzkumu – můžeme se s nimi setkat ve statistice, v ekonometrii, ve fyzice, v chemii, v biologii a výpočetní technice.

**Cílem** diplomové práce bude nastudovat teorii Markov Chain Monte Carlo metod a tuto teorii aplikovat na data.

**Teoretická část** práce seznamuje čtenáře s tématem diplomové práce - MCMC metodami. V první kapitole uvádím základní pojmy z oblasti Markovových řetězců, potřebné pro vytvoření MCMC algoritmů, a bayesovského modelu s apriorní hustotou, která nám umožní určit rozdělení pravděpodobnosti na základě známých dat a námi navrženého modelu. V druhé kapitole se věnuji samotným MCMC metodám, kde popisují Monte Carlo integraci, důležitou pro odhad střední hodnoty funkce přes hustotu pravděpodobnosti, a skryté Markovovy řetězce, u kterých si odvodíme rekurzivní vzorec pro apriorní rozdělení. Poslední, třetí kapitola se věnuje metodě Resampling a MCMC filtrům - Importance Sampling, Sequential Importance Sampling a obecnému Sequential Monte Carlo filtru. MCMC filtry používáme v případě, kdy chceme odhadnout data na základě pozorování.

**Praktická část** práce bude součástí teoretické části, přičemž teoretické poznatky budou následně vysvětlovány na praktických příkladech. Pro provedení veškerých potřebných výpočtů budu používat software R. Součástí práce je přiložené CD obsahující skripta k jednotlivým příkladům.

## 1. Základní pojmy

V této kapitole se budu věnovat teoretickým základům Markovových řetězců a bayesovskému modelu, které jsou významné pro pochopení MCMC modelů. V první podkapitole Markovových řetězců vysvětlím definice základních pojmů – náhodný proces, Markovovy řetězce s diskretním stavem, homogenní Markovův řetězec s diskretním stavem a zejména stacionární rozdělení, důležitou vlastnost rozdělení pravděpodobnosti nerozložitelných Markovových řetězců pro  $n \rightarrow \infty$ , které využiji při simulacích rozdělení pravděpodobností MCMC metod. Ve druhé podkapitole za pomoci bayesovského principu objasním a odvodím aposteriorní hustotu, která nám umožní odhadnout parametry neznámého rozdělení pravděpodobnosti za pomoci známého rozdělení pravděpodobnosti.

Podkapitola Markovovy řetězce vychází z literatury [4], [5], [6] a podkapitola bayesovský model vychází ze zdrojů [3], [12].

### 1.1. Markovovy řetězce

V teorii pravděpodobnosti je náhodný proces, neboli stochastický proces chápán jako paralelní pojem deterministického procesu. U deterministického procesu existuje pouze jediná možnost výsledku v čase, naproti tomu stochastický proces využívá neurčitost v budoucím vývoji a k jejímu popisu nám pomáhá rozdělení pravděpodobnosti. I přesto, že máme zadanou počáteční podmínku, musíme počítat s tím, že v případě stochastického procesu nastává více možností, kterými se může proces vydat v závislosti na míře pravděpodobnosti. Posloupnost náhodných veličin, jako nejjednodušší případ náhodného procesu, používám k zavedení Markovových řetězců.

**Definice 1** Náhodný proces je posloupnost náhodných veličin  $\{X_t, t \in T\}$ , které jsou definované na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, \mathbb{P})$ . Pro připomenutí v oboru pravděpodobnosti a matematické statistiky v pravděpodobnostním prostoru chápeme  $\Omega$  jako množinu všech elementárních jevů,  $\mathcal{A}$  jako množinu možných jevů a  $\mathbb{P}$  je potom pravděpodobnost, která definuje zobrazení množiny možných jevů na intervalu  $\langle 0,1 \rangle$ .

Procesy poté můžeme rozdělit podle času nebo stavů. Je-li množina  $T = Z$  nebo  $T = N_0$ , jedná se o náhodný proces s diskrétním časem. Pokud  $T = \langle a, b \rangle$ , kde  $-\infty \leq a < b \leq \infty$ , mluvíme o náhodném procesu se spojitým časem. Dále označíme  $S$  jako množinu hodnot náhodných veličin  $X_t$ . V případě, že náhodné veličiny  $X_t$  nabývají pouze diskrétních hodnot, jedná se o náhodný proces s diskrétními stavy a množina  $S$  je konečná nebo nekonečná spočetná. Pro náhodný proces se spojitými stavy platí, že náhodné veličiny nabývají hodnot z nějakého intervalu a množina  $S$  je nespočetná. Pro účel této práce užívám náhodné procesy s diskrétním časem, ale s libovolnou množinou stavů.

Již víme, co nám vyjadřuje pojem náhodný proces. Proto se mohou dále zabývat speciálním případem náhodného procesu s diskrétními stavy mající markovskou vlastnost.

Uvažujeme pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$  a náhodné veličiny  $X_0, X_1, X_2, \dots$ , které na tomto prostoru nabývají pouze celočíselných hodnot. Necht' máme množinu  $S$  celých čísel  $i$  takových, že  $i \in S$  právě tehdy, když existuje  $n \in N_0$ , pro které platí  $P(X_n = i) > 0$ .

**Definice 2** Markovovým řetězcem s diskrétním časem nazveme náhodný proces  $\{X_n, n \in N_0\}$  v množině stavů  $S$ , jestliže platí:

$$P(X_{n+1} = j / X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+1} = j / X_n = i),$$

pro  $\forall n, n \in \{0, 1, 2, \dots\}$ ,  $\forall i, j, i_0, i_1, \dots, i_{n-1} \in S$  taková, že

$$P(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i) > 0.$$

V předchozím odstavci jsem objasnila náhodnou posloupnost s markovskou vlastností, kde pravděpodobnostní rozdělení budoucího stavu systému v čase  $n+1$ , známe-li stav přítomný  $n$  a minulý  $n-1, n-2, \dots, 0$ , je stejný, jako když známe jen stav přítomný. Jinak řečeno, pravděpodobnost budoucího stavu systému nezávisí na minulosti, ale pouze na přítomnosti.



V případě, že je nadefinovaná podmíněná pravděpodobnost

$$p_{ij}(n, n+1) = P(X_{n+1} = j / X_n = i), i, j \in S,$$

nazývá se pravděpodobností přechodu prvního řádu. Pak pro dané  $i, j \in S$  tato podmíněná pravděpodobnost představuje takovou pravděpodobnost, že se systém v čase  $n+1$  bude nacházet ve stavu  $j$  za podmínky, že se v čase  $n$  nachází ve stavu  $i$ .

Obecně pravděpodobnosti přechodu závisí na časovém parametru  $n$ , jelikož popisují vývoj nějakého stavového systému měnného v čase.

**Definice 3** Homogenní Markovův řetězec s diskrétním časem je taková posloupnost, kde pravděpodobnosti přechodu nezávisí na čase  $n$

$$p_{ij}(n, n+1) = p_{ij}, i, j \in S,$$

$$0 \leq p_{ij} \leq 1, i, j \in S, \quad \sum_{j \in S} p_{ij} = 1, i \in S,$$

tj. pro dané  $i, j \in S$  platí, že pravděpodobnost přechodu nezávisí na čase, ve kterém daný přechod nastává.

Dále uvažujeme pouze homogenní řetězce.

V této chvíli můžeme sestavit matici pravděpodobností přechodu  $\mathbf{P}$ , neboť pro každé  $i \in S$ , pro které existuje  $n \in N_0$ , jsou definovány podmíněné pravděpodobnosti  $P(X_n = i) > 0$ , tj.  $p_{ij} = P(X_{n+1} = j / X_n = i)$  a platí

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} = (p_{ij})_{i,j \in S}.$$

Matice pravděpodobností přechodů je čtvercová a má konečný či nekonečný počet řádků v závislosti na množině stavů  $S$ . Její prvky jsou nezáporná čísla a součet

pravděpodobností přechodu každého řádku je roven jedné. Pak takto určenou matici nazýváme stochastickou maticí.

Pro určení pravděpodobnostního rozdělení homogenního Markovova řetězce v konkrétním stavu a v daném čase potřebujeme matici pravděpodobností přechodu (viz str. 9) a počáteční rozdělení. Potom stav, ve kterém Markovův řetězec začíná, je podmíněn znalostí tzv. počátečního rozdělení

$$p_i^{(0)} = P(X_0 = i) = p^{(0)}(x_0), i \in S,$$

pro které platí  $0 \leq p_i^{(0)} \leq 1, i \in S$  a  $\sum_{i \in S} p_i^{(0)} = 1, i \in S$ , neboť náhodné jevy  $A_i^0 = (X_0 = i), i \in S$  tvoří úplný systém jevů (skupinu jevů).

Máme-li určené počáteční pravděpodobnosti, potom je můžeme zapsat jako jednotlivé složky řádkového vektoru

$$\mathbf{p}^{(0)} = (p_0^{(0)}, p_1^{(0)}, p_2^{(0)}, \dots). \quad (1)$$

Jelikož součet všech těchto složek je roven jedné, pak tento vektor nazýváme tzv. pravděpodobnostním vektorem, který v čase 0 určuje pravděpodobnost toho, v jakém stavu se systém nachází. Jinak řečeno, pravděpodobnostní vektor určuje rozdělení pravděpodobností náhodné veličiny  $X_0$ .

**Definice 4** Pravděpodobnost přechodu Markovova řetězce ze stavu  $i$  do stavu  $j$  po  $n$  krocích (tj. pravděpodobnost přechodu  $n$ -tého řádu) značíme

$$p_{ij}^{(n)} = P(X_{m+n} = j / X_m = i), i, j \in S,$$

je-li  $m \geq 0, n \geq 0$ .

Je evidentní, že pravděpodobnost přechodu po jednom kroku je rovna  $p_{ij}^{(1)} = p_{ij}$ , potom pravděpodobnost přechodu v nulovém kroku můžeme určit následovně

$$p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 1, & \text{je-li } i = j, \\ 0, & \text{je-li } i \neq j. \end{cases}$$

**Definice 5** Absolutní pravděpodobností jednotlivých stavů v čase  $n$  jsou nepodmíněné pravděpodobnosti  $p_i^{(n)} = P(X_n = i), i \in S, n \in \mathbf{N}_0$ , které určují pravděpodobnost toho, že systém se v čase  $n$  nachází ve stavu  $i$ .<sup>1</sup>

Protože náhodné jevy  $A_i^n = (X_n = i), i \in S$  tvoří úplnou skupinu jevů pro každé  $n \in N_0$ , platí pro absolutní pravděpodobnosti vztahy

$$0 \leq p_i^n \leq 1, i \in S, \quad \sum_{i \in S} p_i^n = 1, i \in S.$$

Jak počáteční pravděpodobnosti, tak i absolutní pravděpodobnosti pro libovolné  $n \in N_0$  můžeme zapsat pomocí pravděpodobnostního vektoru

$$\mathbf{p}^{(n)} = (p_0^{(n)}, p_1^{(n)}, p_2^{(n)}, \dots).$$

Uvažujeme-li  $\mathbf{p}^{(n)}$  náhodné veličiny  $X_n$  v čase  $n$ , potom můžeme říct, že počáteční pravděpodobnost určená vztahem (1) je speciálním případem rozdělení pravděpodobnosti  $\mathbf{p}^{(n)}$  pro  $n = 0$ .

**Věta 1** Vztahy platící pro libovolné  $n \in N, j \in S$  jsou následující

$$p_j^{(n+1)} = \sum_{s \in S} p_s^{(n)} p_{sj},$$

$$p_j^{(n)} = \sum_{s \in S} p_s^{(0)} p_{sj}^{(n)}.$$

Nyní si uvedeme pro nás nejdůležitější charakteristiku Markovových řetězců – stacionární rozdělení. Tato vlastnost popisuje chování řetězce ve stabilním stavu v případě, že ho pozorujeme dostatečně dlouhou dobu. Ale abychom jej mohli bez problému určit, potřebujeme znát pojmy, jako jsou přechodný stav, trvalý stav, trvalý

---

<sup>1</sup> Pro upřesnění - absolutní pravděpodobnosti  $p_i^{(n)}$  v čase  $n$  odlišujeme od předešlých pravděpodobností přechodu  $n$ -tého řádu  $p_{ij}^{(n)}$  ze stavu  $i$  do stavu  $j$  pomocí dolních indexů.

stav nulový a nenulový, periodický či neperiodický a také nerozložitelný Markovův řetězec.

Nejprve si zavedeme podmíněnou pravděpodobnost  $f_{ij}^{(n)}$ , která charakterizuje dobu návratu systému do určitého stavu. Neboli říká, že se systém v čase 0 nacházel ve stavu  $i$  a do stavu  $j$  se poprvé dostane právě v  $n$ -tém kroku. Potom pro  $n > 1$  platí

$$f_{ij}^{(n)} = P(X_n = j / X_0 = i, X_r \neq j, 1 \leq r < n),$$

$$f_{ij}^{(0)} = 0, f_{ij}^{(1)} = p_{ij}, i, j \in S.$$

**Definice 6** Řekneme, že stav  $i \in S$  je

1. trvalý, je-li  $f_{ii} = 1$ , tj. pravděpodobnost toho, že systém se v čase 0 nacházel ve stavu  $i$  a někdy se do tohoto stavu  $i$  vrátí, je rovna 1. Dále si určíme střední dobu prvního návratu do stavu  $i$ , která je rovna

$$\mu_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)}. \quad (2)$$

Potom trvalý stav  $i \in S$  nazýváme

- a) nenulový, je-li  $\mu_i < \infty$ , tj. střední doba prvního návratu do stavu  $i$  je konečná a posloupnost (2) tedy konverguje.
  - b) nulový, je-li  $\mu_i = \infty$ , tj. střední doba prvního návrtu do stavu  $i$  je nekonečná a posloupnost (2) diverguje.
2. přechodným, je-li  $f_{ii} < 1$ , tj. existuje kladná pravděpodobnost toho, že se systém do stavu  $i$  již nikdy nevrátí.

**Definice 7** Zavedeme označení  $d_i$  pro největšího společného dělitele čísel  $n \geq 1$ , kde  $p_{ii}^{(n)} > 0$ . Stav  $i \in S$  se nazývá

1. periodický s periodou  $d_i$ , je-li  $d_i > 0$ .
2. neperiodický, je-li  $d_i = 1$ .

Před uvedením posledního pojmu – nerozložitelnost Markovova řetězce, který je nutný ke správnému určení stacionárního rozdělení, potřebujeme znát, co je to uzavřená množina a dosažitelnost z daného stavu.

**Definice 8** Stav  $j \in S$  se nazývá dosažitelný ze stavu  $i \in S$ , existuje-li číslo  $n \geq 0$  takové, že  $p_{ij}^{(n)} > 0$ . Vzájemně dosažitelné stavy budeme označovat za sousledné. Jestliže pro stav  $i \in S$  platí rovnost  $p_{ii} = 1$ , nazýváme jej pohlcujícím (absorpčním) stavem.

**Definice 9** Uzavřená množina je taková neprázdná množina  $C \subset S$ , pro kterou platí

$$p_{ij}^{(n)} = 0 \quad \forall i \in C, \forall j \notin C, \forall n \in N_0,$$

tj. žádný stav vně  $C$  není dosažitelný z žádného stavu uvnitř  $C$ . Uzávěr množiny  $C$  je nejmenší uzavřená množina, která obsahuje množinu  $C$ .

**Definice 10** Řekneme, že Markovův řetězec je nerozložitelný v případě, že neexistuje jiná uzavřená množina stavů, než je množina všech stavů  $S$ . V opačném případě mluvíme o rozložitelném řetězci.

**Definice 11** Uvažujeme homogenní Markovův řetězec  $\{X_n, n \in N\}$  s množinou stavů  $S$  a maticí pravděpodobností přechodu  $P$ . Potom stacionární rozdělení  $\pi$  daného řetězce je rozdělení pravděpodobnosti  $\pi = \{\pi_j, j \in S\}$  na množině  $S$ , tj.  $\pi_j \geq 0, j \in S, \sum_{j \in S} \pi_j = 1$ , pro které platí

$$\pi_j = \sum_{k \in S} \pi_k p_{kj}, \quad j \in S.$$

Použijeme-li vektor  $\pi = (\pi_0, \pi_1, \pi_2, \dots)$ , můžeme předchozí zápis zapsat pomocí následujícího vztahu

$$\pi = \pi P.$$

**Věta 2** V nerozložitelném Markovovém řetězci

1. neexistuje stacionární rozdělení v případě, že jsou všechny jeho stavy přechodné nebo všechny trvalé.
2. existuje stacionární rozdělení a je jediné v případě, že jsou všechny jeho stavy trvalé nenulové.
  - a) Jestliže jsou všechny stavy neperiodické, potom pro stacionární pravděpodobnosti platí

$$\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)} > 0, \quad i, j \in S,$$

$$\pi_j = \lim_{n \rightarrow \infty} p_j^{(n)} > 0, \quad j \in S.$$

- b) Jestliže jsou všechny stavy periodické, platí

$$\pi_j = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)} > 0, \quad i, j \in S,$$

$$\pi_j = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_j^{(k)} > 0, \quad j \in S.$$

## 1.2. Bayesovský model

Abychom získali bayesovský model, musíme vycházet ze základního bayesovského principu.

Předpokládejme, že máme dvě spojité náhodné veličiny A a B definované na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$ , potom podmíněná hustota pravděpodobnosti náhodné veličiny A za podmínky B je definována takto

$$p(A/B) = \frac{p(A \cap B)}{p(B)}, \quad p(B) > 0,$$

sdruženou hustotu pravděpodobnosti  $p(A \cap B)$  vypočítáme jako

$$p(A \cap B) = p(A/B)p(B) \quad (3)$$

a to samé platí i pro

$$p(A \cap B) = p(B/A)p(A). \quad (4)$$

Bayesovo pravidlo, pak vyplývá z rovnosti výše uvedených vztahů (3) a (4), ze které získáme následující rovnici

$$p(B/A) = \frac{p(A/B)p(B)}{p(A)}, p(B) > 0, \quad (5)$$

kde čítec je sdružená hustota pravděpodobnosti  $p(A \cap B)$  a  $p(B)$  je marginální hustota pravděpodobnosti. Tento vzorec platí i pro diskrétní náhodné veličiny, kdy nám stačí pouze nahradit hustotu pravděpodobnosti pravděpodobnostní funkcí.

Předpokládáme, že máme výsledek pokusu  $B$ , který může být vyvolán některou z alternativ  $A_1, A_2, \dots, A_n$ . Počet alternativ je konečný a jednotlivé alternativy se navzájem vylučují, tj. jsou disjunktní a také tvoří úplný systém jevů (neboli množinu všech možných výsledků)

$$A_i \cap A_j = \phi, \quad \bigcup_{i=1}^n A_i = \Omega,$$

$i, j = 1, 2, \dots, n; i \neq j$ ,  $\phi$  je nemožný jev,  $\Omega$  jistý jev,

můžeme psát

$$p(B) = p(B \cap \Omega) = p\left(B \cap \left(\bigcup_{i=1}^n A_i\right)\right) = p\left(\bigcup_{i=1}^n (A_i \cap B)\right) = \sum_{i=1}^n p(A_i \cap B) = \sum_{i=1}^n p(A_i)p(B/A_i).$$

Každá z alternativ (hypotéz) má nějakou pravděpodobnost, je-li znám výsledek pozorování, měření atd. Z matematického hlediska se tedy zajímáme o podmíněnou pravděpodobnost alternativy  $A_k$ ,  $k = 1, 2, \dots, n$ , při známé hodnotě  $B$  zapíšeme Bayesovým vzorcem následovně

$$p(A_k/B) = \frac{p(A_k \cap B)}{p(B)} = \frac{p(A_k)p(B/A_k)}{p(B)} = \frac{p(A_k)p(B/A_k)}{\sum_{i=1}^n p(A_i)p(B/A_i)}.$$

Uvažujeme rozpojitou náhodnou veličinu  $\Theta$ , která může nabývat konečně mnoha hodnot  $\theta = \theta_1, \dots, \theta_m$ . Potom pomocí vhodně zvoleného experimentu chceme získat informaci o těchto hodnotách. Získaná informace je tedy realizace opět rozpojitě veličiny, která reprezentuje výsledky našeho experimentu. Můžeme vidět, že tyto hodnoty se v každém provedeném pokusu liší, což je způsobeno tím, že pozorovaná veličina je svým charakterem náhodná. Tuto náhodnou veličinu pak budeme značit pomocí  $\xi$ . Jednotlivé výsledky (alternativy) tedy nesou informaci o každé možné hodnotě pozorování a můžeme je použít pro odhad skutečné hodnoty.

Jejich vzájemný vztah následně odvodíme pomocí Bayesova vzorce. Předpokládáme-li, že hodnota zjišťované rozpojitě náhodné veličiny je  $\theta$ , pak pomocí této hodnoty dojdeme k výsledku pozorování  $\mathbf{x}$  s danou pravděpodobností  $p(\xi = \mathbf{x}/\Theta = \theta)$ . Jde tedy o určení podmíněné pravděpodobnosti, pro kterou platí

$$p(\xi = \mathbf{x}/\Theta = \theta) = \frac{p(\xi = \mathbf{x} \cap \Theta = \theta)}{p(\Theta = \theta)} = \frac{p(\xi = \mathbf{x})p(\Theta = \theta/\xi = \mathbf{x})}{p(\Theta = \theta)} = \frac{p(\mathbf{x})p(\theta/\mathbf{x})}{p(\theta)}.$$

Ve skutečnosti však stojíme před problémem, kdy známe pouze výsledky experimentu  $\mathbf{x} = x_1, \dots, x_n$ , jejichž chování je popsáno pomocí hodnot  $\theta = \theta_1, \dots, \theta_m$ , přičemž tyto hodnoty jsou neznámé. A my pomocí dat  $\mathbf{x} = x_1, \dots, x_n$  chceme získat nějaké informace o těchto hodnotách  $\theta$ , neboli o hodnotě odhadované veličiny  $\Theta$ . K tomu opět použijeme Bayesův přístup, který nám umožní určit pravděpodobnost  $p(\xi = \mathbf{x}/\Theta = \theta)$  jednotlivých alternativ (hypotéz) náhodné veličiny  $\Theta$ , přičemž známe výsledek experimentu  $\mathbf{x}$

$$p(\Theta = \theta/\xi = \mathbf{x}) = \frac{p(\Theta = \theta \cap \xi = \mathbf{x})}{p(\mathbf{x})} = \frac{p(\Theta = \theta)p(\xi = \mathbf{x}/\Theta = \theta)}{p(\mathbf{x})}.$$

Jelikož chceme na základě realizace  $\mathbf{x}$  získat informace o hodnotách  $\theta$ , budeme dále uvažovat podmíněnou hustotu pravděpodobnosti  $p(\theta/\mathbf{x})$



$$p(\theta/x) = \frac{p(\theta)p(x/\theta)}{p(x)}. \quad (6)$$

V rovnici (6) budeme ignorovat hodnotu ve jmenovateli, jelikož v sobě nezahrnuje žádné informace o hodnotách  $\theta$ . Odtud dostaneme přibližný vztah

$$p(\theta/x) \propto p(x/\theta)p(\theta).^2$$

Potom  $p(\theta/x)$  nazveme aposteriorní hustotou, která je proporcionálním součinem apriorní hustoty  $p(\theta)$  a věrohodnostní funkce  $p(x/\theta)$ . Jinak řečeno, tento vzorec nám umožní určit rozdělení parametrů na základě známých dat a námi navrženého modelu. Apriorní hustota  $p(\theta)$  v sobě zahrnuje všechny informace o parametru  $\theta$  nezávisle na datech  $x$  (nezávisí na výsledku pokusu). Věrohodnostní funkce  $p(x/\theta)$  je již dříve zmíněná podmíněná hustota pravděpodobnosti, pomocí níž generujeme data. Aposteriorní hustota  $p(\theta/x)$  nám dává informaci o parametrech  $\theta$  v závislosti na datech  $x$  (závisí na výsledku pokusu).

**Příklad 1** Uvažujeme apriorní normální rozdělení  $\theta \sim N(\theta_0, \tau_0^2)$  a model vyjadřující závislost  $X$  na  $\theta$ , tj.  $(X/\theta) \sim N(\theta, \sigma^2)$  se známými parametry  $\sigma^2, \theta_0$  a  $\tau_0^2$ . Zprvce nás zajímá aposteriorní rozdělení  $(\theta/X = x) \sim N(\theta_1, \tau_1^2)$ , za předpokladu apriorního rozdělení A a B, které je opět normální a získáme ho z rovnic  $\theta_1 = w\theta_0 + (1-w)x$  a  $\tau_1^{-2} = \tau_0^{-2} + \sigma^{-2}$ , kde  $w = \tau_0^{-2} / (\tau_0^{-2} + \sigma^{-2})$  [11].

Máme následující informace

- Apriorní rozdělení A:  $\theta \sim N(900, 10^2)$
- Apriorní rozdělení B:  $\theta \sim N(900, 100^2)$
- Model:  $(X/\theta) \sim N(\theta, 40^2)$
- Počet realizací pro  $X$  je 200

---

<sup>2</sup> Znaménko  $\propto$  vyjadřuje přímou úměrnost.

Zadruhé chceme vypočítat 95% konfidenční intervaly pro samotný parametr  $\theta$ , ale také pro  $\theta$  jen na základě parametru  $X$ . Zajímá nás, jak moc se od sebe liší rozptyly odhadu parametru  $\theta$  v případě aposteriorního rozdělení A a B.

A. V případě apriorního rozdělení A použijeme následující postup

1. V prvním kroku užitím příkazu „rnorm“ v programu R si vygeneruji jednu hodnotu parametru  $\theta = 896,4695$  [10].

```
> theta0=900
> tau0=10
> theta=rnorm(1,mean=theta0,sd=tau0)
> theta
[1] 896.4695
```

Ve druhém kroku si vygeneruji  $N = 200$  hodnot pro parametr  $X$ , pro který platí  $(X/\theta) \sim N(896,4695, 40^2)$ .

```
> tau=40
> N=200
> Xtheta=rnorm(N,mean=theta,sd=tau)
```

Ve třetím kroku chceme vypočítat váhu  $w = \tau_0^{-2} / (\tau_0^{-2} + \sigma^{-2})$  vyjadřující relativní informaci v apriorním rozdělení. Výsledná hodnota je  $w = 0,9411765$ . Potom můžeme určit  $\theta_1 = w\theta_0 + (1-w)x$  a  $\tau_1^{-2} = \tau_0^{-2} + \sigma^{-2}$ , které potřebujeme pro výpočet aposteriorního rozdělení. Vypočítané hodnoty jsou  $\theta_1 = 899.7638$  a  $\tau_1 = 9.701425$ .

```
> x=mean(Xtheta)
> w=(tau0)^(-2)/((tau0)^(-2)+(tau)^(-2))
[1] 0.9411765
> theta1=w*(theta0)+(1-w)*x
[1] 899.7638
> tau11=(tau0)^(-2)+(tau)^(-2)
[1] 0.010625
> tau1=(tau11)^(-(1/2))
[1] 9.701425
```

V posledním kroku chceme získat aposteriorní rozdělení  $(\theta/X = x) \sim N(\theta_1, \tau_1^2)$ . Po dosazení výsledků ze třetího kroku vygenerujeme 100 hodnot z normálního rozdělení  $(\theta/X, H_A) \sim N(899.7638, 9.701425)$ . Odhadnutá hodnota je  $\theta = 900.2002$ .

```
> thetaX=rnorm(100,mean=(theta1),sd=(tau1))
```

```
> estimatetheta=mean(thetaX)
```

```
> estimatetheta
```

```
[1] 900.2002
```

2. Nyní si zkonstruujeme konfidenční intervaly

- 95% konfidenční interval pro parametr  $\theta$

```
> L=estimatetheta-1.96*(tau1)/((x)^(1/2))
```

```
[1] 898.2987
```

```
> P=estimatetheta+1.96*(tau1)/((x)^(1/2))
```

```
[1] 902.1017
```

- 95% konfidenční interval pro  $\theta$  jen na základě parametru  $X$

```
> X=mean(Xtheta)
```

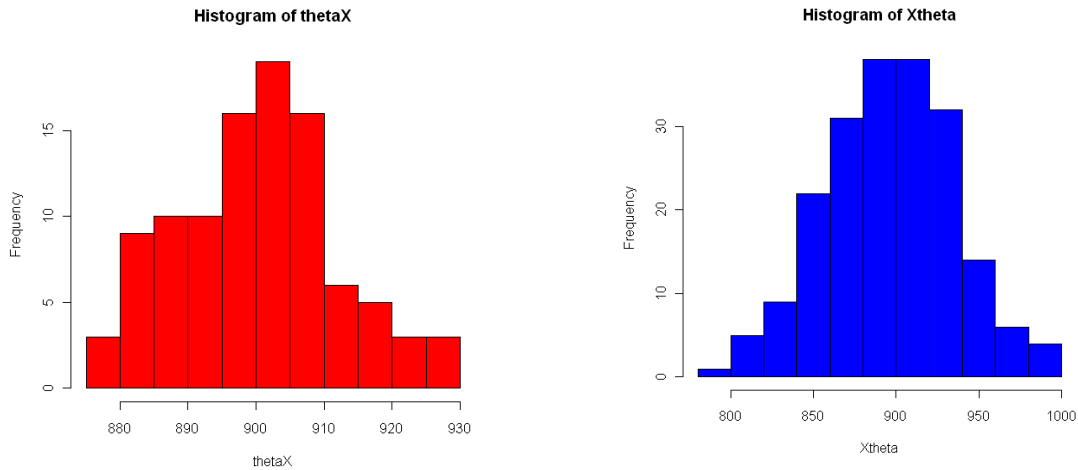
```
> L=X-1.96*tau/((N)^(1/2))
```

```
[1] 890.4412
```

```
> P=X+1.96*tau/((N)^(1/2))
```

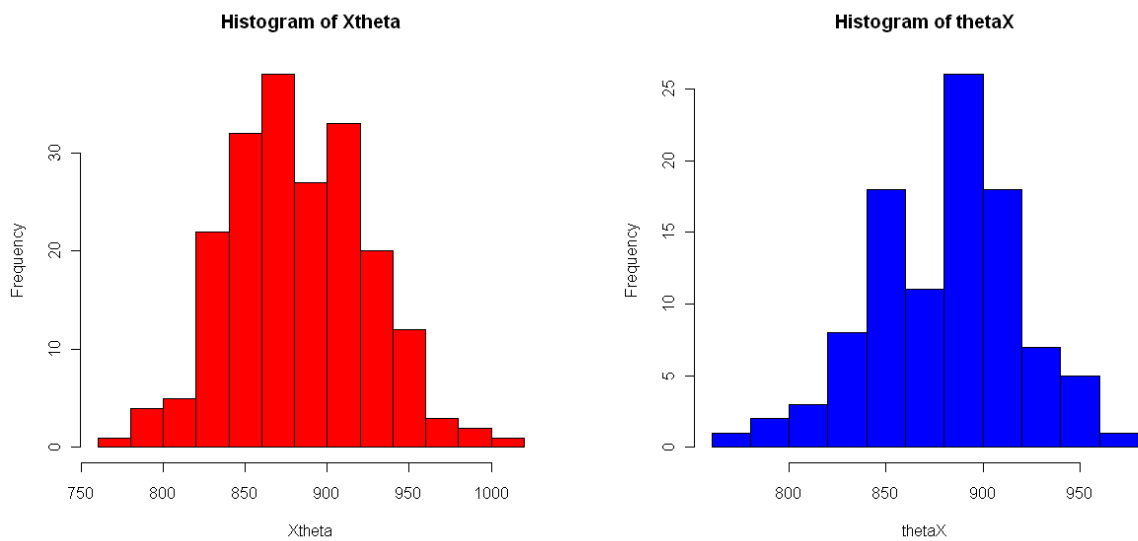
```
[1] 901.5286
```

Na obrázku (Obr.1) pak můžeme vidět, jak moc se od sebe liší rozdělení četností pro parametr  $\theta$ .



**Obr. 1** Rozdělení četností pro samotný parametr  $\theta$  a pro  $\theta$  jen na základě parametru  $X$  v případě A

B. U apriorního rozdělení B použijeme stejný postup, jako v případě apriorního rozdělení A. (viz Příloha). Z vygenerovaných hodnot získáme aposteriorní rozdělení  $(\theta/X, H_B) \sim N(885.0105, 37.13907)$ , které opět bude normální. Odhad  $\theta$  je pak roven hodnotě 880,5503. Grafy 95% konfidenčních intervalů parametru  $\theta$  za předpokladu apriorního rozdělení  $\theta \sim N(900, 100^2)$  znázorňuje (Obr. 2). 95% konfidenční interval pro samotné  $\theta$  je  $(878.1001, 883.0005)$  a pro  $\theta$  jen na základě  $X$  je  $(877.0684, 888.1559)$ .



**Obr. 2** Rozdělení četností pro samotný parametr  $\theta$  a pro  $\theta$  jen na základě parametru  $X$  v případě B

V případě, že uvažujeme nediskrétní marginální věrohodnostní funkci, vyjádříme ji pomocí integrálu následovně

$$p(\mathbf{x}) = \int p(\mathbf{x}/\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}. \quad (7)$$

Pak aposteriorní hustotu  $p(\boldsymbol{\theta}/\mathbf{x})$  získáme využitím marginálního rozdělení  $p(\mathbf{x})$  ze vztahu (6)

$$p(\boldsymbol{\theta}/\mathbf{x}) = \frac{p(\mathbf{x}/\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{x})} = \frac{p(\mathbf{x}/\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{x}/\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}.$$

A jako poslední vztah, který v této podkapitole uvedu, je aposteriorní střední hodnota

$$E(g(\boldsymbol{\theta})/\mathbf{x}) = \int g(\boldsymbol{\theta})p(\boldsymbol{\theta}/\mathbf{x})d\boldsymbol{\theta},$$

kde  $g(\boldsymbol{\theta})$  je funkce, která nás zajímá.

Blíže se tomuto vztahu budu věnovat v Monte Carlo integraci (viz str. 22 - 33).

## 2. MCMC metody

MCMC metody jsou nejvíce využívány v případech, kdy chceme generovat z nějakého pravděpodobnostního rozdělení, které je složité a neumíme z něho generovat přímo. Proto zkonstruujeme Markovův řetězec, jehož stacionární rozdělení je požadované rozdělení. Simulujeme Markovovy řetězce a po dostatečně velkém počtu kroků dostaneme přibližně výběr z daného rozdělení za podmínky, že jsou splněny jisté předpoklady na řetězec, které zaručí, že limitní rozdělení existuje a splývá se stacionárním.

V následující kapitole se budu věnovat Monte Carlo integraci a skrytým Markovovým řetězcům a MCMC metodám.

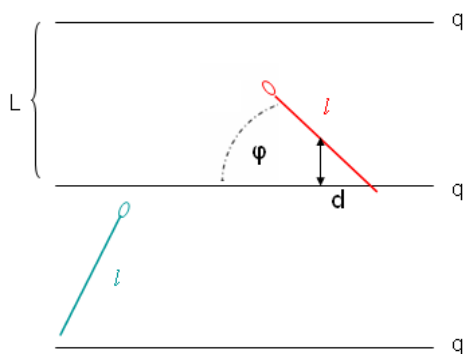
Tato kapitola vychází ze zdrojů [3], [8], [9], [13], [14], [15].

### 2.1. Monte Carlo integrace

Za historicky první příklad vyřešený metodou Monte Carlo se považuje známá Buffonova úloha o jehle z 18. století, jejíž výsledkem je experimentální vyjádření Ludolfova čísla  $\pi$  [1], [2], [3].

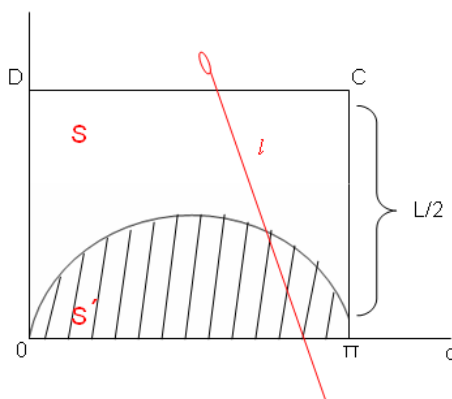
Nechť uvažujeme rovinu a na ní narýsované rovnoběžné přímky od sebe vzdáleny v délce  $L$ . Jaká je pravděpodobnost, že jehla délky  $l < L$ , která je házena náhodně na uvažovanou rovinu, protne některou z přímek?

Je zřejmé, že jehla může protnout jen jednu přímku. Jednotlivé přímky jsou rovnoběžné s osou  $x$ . Poloha jehly je potom dána dvěma charakteristikami – vzdáleností středu jehly od nejbližší přímky  $q$ , kterou označíme  $d$ , a úhlem  $\varphi$  mezi jehlou a přímkou  $q$ . Mohou tedy nastat dvě situace. Jehla může protnout přímku (červená jehla), ale také nemusí (modrá jehla) (viz Obr.3). Za náhodnost se zde považuje stejná pravděpodobnost jakékoliv polohy středu a jakékoliv orientace úhlu zmíněné jehly, kdy tyto dvě nahodilé proměnné jsou vzájemně nezávislé.



**Obr. 3** Vzájemná poloha přímky a jehly

Abychom určili pravděpodobnost toho, zda jehla protne jednu přímku, musíme vycházet z obdélníku  $0 \leq d \leq \frac{L}{2}$  a  $0 \leq \varphi \leq \pi$  na obrázku (Obr. 4). Protnutí, neboli příznivá situace nastane v případě, že bude platit podmínka  $d \leq \frac{l}{2} \sin \varphi$ . Potom pro obdélník  $0\pi CD$ , jehož celková plocha je  $S = \frac{L}{2} \pi$ , potřebujeme vypočítat obsah vyšrafované plochy  $S' = \int_0^{\pi} \frac{l}{2} \sin d \varphi$ , která reprezentuje součet všech příznivých možností, které reprezentují všechny případy protnutí jehly s přímkou.



**Obr. 4** Množina všech možností  $S$  a množina příznivých možností  $S'$

Pravděpodobnost  $p$  tedy odhadneme poměrem příznivých hodů  $m$  (vyšrafované plochy  $S'$ ) a všech celkových hodů jehlou  $n$  (celkové plochy obdélníku  $S$ )

$$p = \frac{m}{n},$$

$$p = \frac{S'}{S} = \frac{\int_0^\pi \frac{l}{2} \sin d\varphi}{\frac{L}{2}\pi} = \frac{\frac{l}{2} \int_0^\pi \sin d\varphi}{\frac{L}{2}\pi} = \frac{\frac{l}{2} [-\cos]_0^\pi}{\frac{L}{2}\pi} = \frac{l[1+1]}{L\pi} = \frac{2l}{L\pi},$$

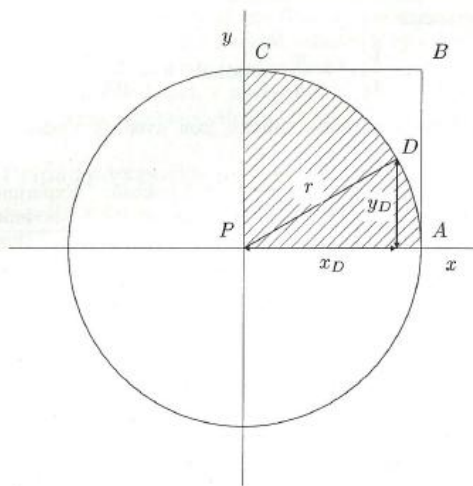
$$\frac{m}{n} \cong \frac{2l}{L\pi},$$

$$\pi \cong \frac{2ln}{Lm}.$$

Nakonec použijeme Bernoulliiovu větu, a tím pro  $n \rightarrow \infty$  bude výraz  $\frac{2ln}{Lm}$  konvergovat

k hodnotě  $\pi$ . V pozdějších letech se zjistilo, že tato metoda není pro přesný výpočet Ludolfova čísla příliš efektivní. To ovšem neznamená, že není dobrou metodou. Tato metoda je využívána například při výpočtu střední hodnoty náhodné veličiny pomocí i někdy příliš složitých integrálů, který za pomoci vhodného algoritmu a velkého počtu pokusů dokážeme snadno odvodit.

**Příklad 2** Nyní se pokusíme o určení hodnoty Ludolfova čísla [3]. Předpokládáme jednotkovou kružnici, která má střed v počátku souřadnicového systému (viz Obr. 5).



**Obr. 5** Jednotková kružnice a jednotkový čtverec

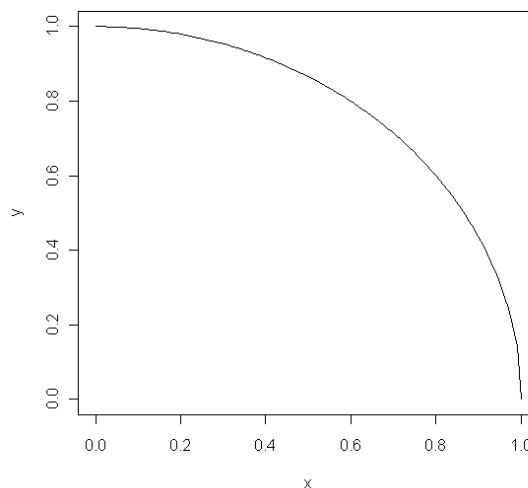
(zdroj: FABIAN, F., KLUBIER., Z. *Metoda Monte Carlo a možnosti jejího uplatnění*)



Dále máme jednotkový čtverec  $PABC$ . Potom vygenerujeme  $N$  náhodných čísel z rovnoměrného rozdělení na intervalu  $(0,1)$ . Z takto vygenerovaných hodnot vytvoříme uspořádané dvojice, které určují souřadnice bodů  $D^i = [x_D^i, y_D^i]$ ,  $i = 1, \dots, N$ . V případě, že platí  $d = \sqrt{(x_D^i)^2 + (y_D^i)^2} \leq 1$ , jedná se o zdařilý pokus. Podobně jako v Buffonově úloze o jehle, hodnotu Ludolfova čísla vypočítáme užitím geometrické pravděpodobnosti. Jednotkový čtverec má obsah  $S = 1$  a zajímá nás obsah vyšrafované oblasti  $S' = \frac{1}{4} \pi r^2$ . K tomu využijeme pravděpodobnost  $p = \frac{S'}{S} = \frac{\frac{1}{4} \pi r^2}{1}$  padnutí bodu  $D^i$  do vyšrafované plochy  $S'$ . Odhad pravděpodobnosti  $p$  získáme pomocí relativní četnosti  $p'$ , která zhodnocuje provedené pokusy. Celkový počet pokusu je  $N = 100000$  a počet pokusů, kdy bod  $D^i$  padne do vyšrafované oblasti  $S'$  je  $m$ , tj. jedná se o příznivý jev a  $p' = \frac{1}{4} \pi r^2 \sim \frac{m}{N}$  a přibližná hodnota Ludolfova čísla je  $\pi \sim \frac{4m}{N}$ .

1. Vykreslíme si jednu čtvrtinu kruhu v jednotkovém čtverci

```
> x=seq(0,1,length=100)
> y=sqrt(1-x^2)
> plot(x,y,type="l")
```



**Obr. 6** Jedna čtvrtina kruhu v jednotkovém čtverci

## 2. Odhadneme hodnotu Ludolfova čísla

```
> set.seed(2565103)
> N=100000
> m=0
> for(i in 1:N){
+ xd=runif(1)
+ yd=runif(1)
+ d=ifelse((xd^2+yd^2)^(1/2)<1,1,0)
+ m=m+d
+ }
> pi=(4*m/N)
[1] 3.14116
```

Metoda Monte Carlo je třída algoritmů, která nám umožňuje simulovat složité systémy založené na mnohokrát opakovaných náhodných pokusech. Tyto algoritmy využívají jak poznatky z teorie pravděpodobnosti, tak i matematické analýzy a výpočetní techniky. Základní myšlenka této metody spočívá ve zjištění vztahu mezi jednotlivými pravděpodobnostními charakteristikami náhodných procesů a veličinami, jež jsou řešením úloh matematické analýzy. Proto na základě vypočtených odhadů jednotlivých statistických charakteristik můžeme například snadno vypočítat určité integrály, určit řešení systému lineárních rovnic, nalézt kořeny rovnic nebo vypočítat hodnoty funkce. Kvalita výsledků získaných metodou Monte Carlo závisí na vhodně zvoleném generátoru náhodných čísel, výpočetním algoritmu, ale také na kontrole přesnosti dosažených výsledků.

Dále se pokusím přiblížit metodu Monte Carlo pomocí slabého zákona velkých čísel a centrální limitní věty [3], [15].

Nejprve si ukážeme, jak probíhá samotná Monte Carlo integrace. Předpokládejme spojitou náhodnou veličinu  $\xi$  definovanou na intervalu  $(a, b)$  hustotou pravděpodobnosti  $p(x)$ . A necht' máme náhodnou veličinu

$$\eta = g(\xi),$$

což je spojitá funkce náhodné veličiny  $\xi$ .

A předpokládáme-li, že existuje střední hodnota náhodné veličiny  $\mu$ , potom platí následující zápis

$$E = E\mu = E_{p(x)}[g(\xi)] = \int_a^b g(\xi)p(x)dx, \quad (8)$$

který znamená, že integrál vyjadřuje očekávanou hodnotu funkce  $g(\xi)$  přes hustotu pravděpodobnosti  $p(x)$ .

Nechť máme náhodnou veličinu  $\xi$  vyjadřující nějakou hodnotu, kterou chceme spočítat. Informace o této hodnotě získáme provedením experimentu, a to tak, že dostaneme nezávislé výsledky jako hodnoty  $\mathbf{x} = x_1, \dots, x_n$  náhodné veličiny  $\xi$ . Je možné napsat

$$y = g(\mathbf{x}),$$

kde  $y$  je posloupnost hodnot náhodné veličiny  $\mu$ . Získali jsme tak realizace náhodné veličiny  $y$ , které jsou opět nezávislé jako hodnoty  $x$  a mají stejné rozdělení pravděpodobnosti.

Na řešení daného problému se můžeme dívat ze dvou hledisek [3]. První přístup vychází ze slabého zákona velkých čísel a druhý přístup používá k výpočtu centrální limitní větu pro nezávislé a stejně rozdělené náhodné veličiny.

1. Na začátku úlohy předpokládáme existenci střední hodnoty náhodné veličiny, tj. platí  $E|g(\xi)| < \infty$  a proto můžeme k řešení dané úlohy použít slabý zákon velkých čísel například s využitím Chinčinyovy věty, která neklade žádné podmínky na rozptyl.

Jestliže máme nezávislé náhodné veličiny  $\mu_1, \dots, \mu_n$  se stejným rozdělením pravděpodobnosti a konečnou střední hodnotou  $E$ , pak pro libovolné  $\varepsilon > 0$  platí

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n \mu_i - E\right| < \varepsilon\right) = 1,$$

$$\text{tj. } \frac{1}{n} \sum_{i=1}^n \mu_i \xrightarrow{P} E \text{ pro } n \rightarrow \infty.$$

Provedeme-li dostatečný počet pokusů  $n$ , pomocí kterých získáme nezávislé výsledky  $x_i$  hledané náhodné veličiny  $\xi$ , potom získáme přibližný stochastický odhad  $\bar{E}$  hledaného integrálu  $E$  (střední hodnoty), který je roven aritmetickému průměru

$$E \approx \bar{E},$$

$$\bar{E} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n g_i.$$

2. Druhý přístup, jak už bylo řečeno, vychází z centrální limitní věty.

Uvažujeme nezávislé náhodné veličiny  $\mu_1, \dots, \mu_n$  se stejným rozdělením pravděpodobnosti, konečnou střední hodnotou  $E$  a konečným nenulovým rozptylem  $\sigma_n^2$  ve tvaru

$$\sigma_n^2 = \sigma(\mu) = \int_a^b [g(x) - E]^2 p(x) dx,$$

potom platí

$$\lim_{n \rightarrow \infty} P \left( \frac{\frac{1}{n} \sum_{i=1}^n \mu_i - E}{\sigma_n / \sqrt{n}} \leq z \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{u^2}{2}} du, \quad (9)$$

tj. aritmetický průměr  $\bar{E} = \frac{1}{n} \sum_{i=1}^n g_i$  má asymptoticky normální rozdělení se střední hodnotou  $E$  a směrodatnou odchylkou  $\sigma_n^*$  definovanou vztahem

$$\sigma_n^* = \frac{\sigma_n}{\sqrt{n}}.$$

Za předpokladu splnění uvedených podmínek ve výrazu (9), můžeme v aplikacích použít vztahu

$$P\left(|E - \bar{E}| < \sigma_n^* t_\alpha\right) = \alpha.$$

Tento vztah jsme získali následující úpravou

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(A < \frac{\bar{E} - E}{\sigma_n / \sqrt{n}} \leq B\right) &= \frac{1}{\sqrt{2\pi}} \int_A^B e^{-\frac{u^2}{2}} du, \\ \lim_{n \rightarrow \infty} P\left(-t_\alpha < \frac{\bar{E} - E}{\sigma_n / \sqrt{n}} \leq t_\alpha\right) &= \frac{1}{\sqrt{2\pi}} \int_{-t_\alpha}^{t_\alpha} e^{-\frac{u^2}{2}} du, \\ \lim_{n \rightarrow \infty} P\left(\left|\frac{\bar{E} - E}{\sigma_n / \sqrt{n}}\right| < t_\alpha\right) &= \frac{2}{\sqrt{2\pi}} \int_0^{t_\alpha} e^{-\frac{u^2}{2}} du, \\ \lim_{n \rightarrow \infty} P\left(|\bar{E} - E| < \sigma_n^* t_\alpha\right) &= 2\Phi(t_\alpha) - 1, \\ P\left(|\bar{E} - E| < \varepsilon\right) &= \alpha. \end{aligned}$$

V prvním kroku jsme dosadili za  $A = -t_\alpha$  a za  $B = t_\alpha$  pro  $t_\alpha > 0$ . Ve druhém kroku jsme integrál upravili do vhodnější podoby. Dále jsme nahradili ve vzorci následující hodnoty  $\sigma_n / \sqrt{n} = \sigma_n^*$  a  $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_\alpha} e^{-\frac{u^2}{2}} du = \Phi(t_\alpha)$ . Výsledný vzorec jsme získali dosazením konečné chyby  $\varepsilon = \sigma_n^* t_\alpha > 0$  a spolehlivosti  $\alpha = 2\Phi(t_\alpha) - 1$ , kterou volíme tak, aby  $\alpha \geq 0,95$ . Potom pomocí vztahu  $\Phi(t_\alpha) = \frac{1 + \alpha}{2}$  pro zvolenou spolehlivost  $\alpha$  můžeme z tabulek pro normální rozdělení určit  $t_\alpha$ .

Uvažujeme-li přípustnou chybu  $\varepsilon = \frac{\sigma_n}{\sqrt{n}} t_\alpha$ ,  $\varepsilon > 0$  a spolehlivost  $\alpha$ , je možné při známém rozptylu  $\sigma_n^2$  určit nutný počet pokusů

$$n = \frac{\sigma_n^2}{\varepsilon^2} t_\alpha^2, \quad (10)$$

můžeme tedy napsat

$$P\left(\left|\bar{E} - E\right| < \varepsilon\right) \sim \alpha,$$

$$E \approx \bar{E}.$$

Pak přibližná hodnota náhodné veličiny  $\bar{E}$  při zvolených hodnotách  $\alpha$  a  $\varepsilon > 0$  je opět dána známým vztahem

$$\bar{E} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n g_i.$$

V praxi se často setkáme s úlohami, ve kterých většinou neznáme rozptyl  $\sigma_n^2$  potřebný pro stanovení dostatečného počtu pokusů  $n$ . Proto se pro určení této veličiny používá iterační postup, kdy si zvolíme počet pokusů  $n = n_0$  a pomocí následujícího vztahu určíme přibližnou hodnotu rozptylu  $\sigma_{n_0}^2$

$$\sigma_{n_0}^2 = \frac{1}{n_0} \sum_{i=1}^{n_0} \left[ g(x_i) - \frac{1}{n_0} \sum_{j=1}^{n_0} g(x_j) \right]^2 \sim \sigma_n^2. \quad (11)$$

Získanou hodnotu  $\sigma_{n_0}^2$  pak dosadíme do vztahu (10) a dostaneme přibližný počet pokusů  $n$ . V případě, že  $n > n_0$  je potřeba provést dodatečné pokusy.

**Příklad 3** Metodou Monte Carlo, použitím vzorce (8), chceme vypočítat integrál

$\int_0^1 (1-x^2)^{3/2} dx = \int_0^1 g(x)p(x)dx$ . Příkazem „runif“ generujeme náhodnou veličinu  $x$  definovanou na intervalu  $(0,1)$ , kdy počet pozorování je roven 1 000 000, tj. generujeme data z pravděpodobnostní hustoty  $p(x)=1$ . Nadefinujeme spojitou funkci  $g(x) = (1-x^2)^{3/2}$  a vypočítáme očekávanou hodnotu integrálu [10]. Pro kontrolu je skutečná hodnota integrálu přibližně rovna  $3\pi/16 \cong 0.5890486225$ . Potom výsledná hodnota se liší o 0,0002135225.

> x=runif(1000000)

```

> gx=(1-(x^2))^(3/2)
> estimate=mean(gx)
> estimate

```

**[1] 0.5888351**

V případě, že daný algoritmus spustím znovu, dostaneme jiný výsledek než v předchozím případě, i když příkazy jsou stejné. Je to způsobeno tím, že opravdu pracujeme s náhodnou veličinou  $x$  a generátorem náhodných čísel. Pokud bychom chtěli zachovat stejné výsledky, neboli zajistit vygenerování stejných hodnot, museli bychom použít příkaz „set.seed“ (viz Příklad 5).

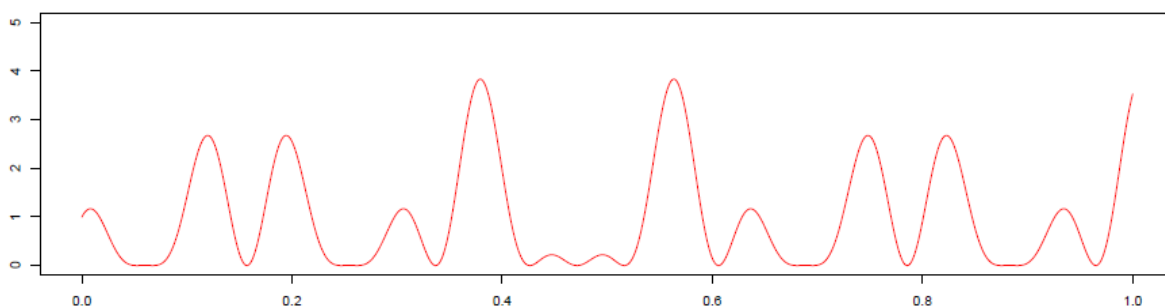
```

> x=runif(1000000)
> gx=(1-(z^2))^(3/2)
> estimate=mean(gx)
> estimate

```

**[1] 0.5893236**

**Příklad 4** Opět se pokusím pomocí metody Monte Carlo vypočítat integrál  $\int_0^1 [\cos(50\theta) + \sin(20\theta)]^2 d\theta \cong 0,965$  [11]. V prvním kroku si vykreslíme graf vyjadřující vztah mezi vygenerovanými hodnoty  $\theta$  na intervalu  $(0,1)$  a funkcí  $h(\theta) = [\cos(50\theta) + \sin(20\theta)]^2$ , kdy pravděpodobnostní hustota normálního rozdělení  $N(0,1)$  je rovna  $p(\theta)=1$ . Potom  $\theta_1, \dots, \theta_n$  jsou nezávislé stejně rozdělené náhodné veličiny opět z  $N(0,1)$ . Výsledná hodnota je o 0,0012057 menší.



**Obr. 7** Vztah mezi hodnotami  $\theta$  a funkcí  $h(\theta)$

```

> pdf(file="Integral4.pdf",width=15,height=10)
> par(mfrow=c(2,1))
> theta1=seq(0,1,length=100000)
> theta2=(cos(50*theta1)+sin(20*theta1))^2
> plot(theta1,theta2,xlab="",ylab="",main="",type="l",col="red",ylim=c(0,5))
> theta=runif(1000000)
> gtheta=(cos(50*theta)+sin(20*theta))^2
> estimate=mean(gtheta)
> estimate
[1] 0.9637943

```

**Příklad 5** Nyní si ukážeme, jak pomocí metody Monte Carlo integrace vypočítáme integrál, jehož integrační meze jsou jiné než  $a=0$  a  $b=1$ . Uvažujme integrál

$\int_0^2 e^{-x} dx = 0,86466472$ , potom očekávanou hodnotu tohoto integrálu vypočítáme pomocí

vzorce  $\hat{I} = (b-a) \frac{\sum f(y_i)}{n}$ , kde  $n$  vyjadřuje počet náhodně vygenerovaných vzorků z normálního rozdělení  $y_1, \dots, y_n \sim R(a,b)$ . Pro zajímavost uvádím, že se tato metoda nazývá přibližná Monte Carlo integrace („crude“) [7].

```

> n=100000
> a=0
> b=2
> set.seed(780257)
> x=runif(m,0,2)
> y=exp(-x)
> odhadI=(b-a)*(sum(y)/n)
> odhadI
[1] 0.8634333

```

Výhoda metody Monte Carlo spočívá v tom, že čím více máme náhodných veličin (neboli vyšší  $n$ ), tím nižší je chyba aproximace. Avšak problémy nelze vyloučit. První problém vzniká v případě, kdy máme vícerozměrnou hustotu pravděpodobnosti  $p(x)$ . Potom je pro nás obtížné z takovéto hustoty získat vzorky. Druhý problém je



samotné získávání vzorků z hustoty pravděpodobnosti  $p(x)$ , jelikož složitost odběru vzorku roste s  $n$ .

Na závěr této podkapitoly bych se chtěla krátce zmínit o nejznámějším algoritmu MCMC metod, a tím je Metropolisův-Hastingsův algoritmus [4], který pro nalezení řešení využívá stacionární rozdělení (viz str. 13 - 14).<sup>3</sup> Tato metoda je založena na nalezení Markovova řetězce  $\{X_n, n \in N_0\}$  se stacionárním rozdělením  $\{\pi_i, i \in S\}$ , který realizujeme pro  $n \rightarrow \infty$ .

Pro snazší porozumění budeme předpokládat pouze diskrétní rozdělení  $\{\pi_i, i \in S\}$  s konečnou množinou stavů  $S$ . Dále uvažujeme prvky  $t_{ij}$  stochastické matice  $\mathbf{T} = (t_{ij})_{i,j \in S}$ , pro které platí, že  $t_{ij} > 0$ , právě když  $t_{ji} > 0$ . Metropolisův-Hastingsův algoritmus přemění matici  $\mathbf{T}$  na matici pravděpodobností přechodu  $\mathbf{P} = (p_{ij})_{i,j \in S}$  Markovova řetězce  $\{X_n, n \in N_0\}$ , který má stacionární rozdělení  $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_n)$  s maticovým zápisem  $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ . Akceptační poměr je určen vztahem

$$a_{ij} = \frac{\pi_j t_{ji}}{\pi_i t_{ij}}, i, j \in S, i \neq j.$$

Pravděpodobnost přechodu ze stavu  $i$  do stavu  $j$  značíme  $t_{ij}$ , potom v případě  $a_{ij} \geq 1$  řetězec přejde ze stavu  $i$  do stavu  $j$ . Pokud  $a_{ij} < 1$  znamená to, že si hodíme falešnou mincí a  $a_{ij}$  bude značit pravděpodobnost padnutí rubu. Padne-li rub, řetězec znovu přejde do stavu  $j$ . Nastane-li opačný případ, řetězec zůstane ve stavu  $i$ . Z tohoto výkladu můžeme formálně zapsat prvky matice  $\mathbf{P}$  vztahem

$$p_{ij} = \begin{cases} t_{ij} & \text{pro } i \neq j, a_{ij} \geq 1, \\ t_{ij} a_{ij} & \text{pro } i \neq j, a_{ij} < 1, \\ t_{ij} + \sum_{k: a_{ik} < 1} t_{ik} (1 - a_{ik}) & \text{pro } i = j. \end{cases}$$

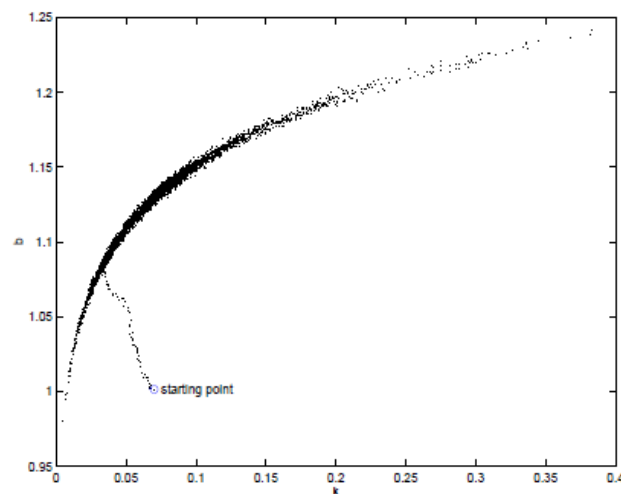
<sup>3</sup> Účel je pouze nastínit využití stacionárního rozdělení v MCMC metodách. Metropolitův Hastingsův algoritmus není pro mou diplomovou práci stěžejní a proto se jím dále v textu nebudu zabývat.

Pro nový řetězec platí vlastnost  $\pi_i p_{ij} = \pi_j p_{ji}$ , proto můžeme psát

$$\sum_{i \in S} \pi_i p_{ij} = \sum_{i \in S} \pi_j p_{ji} = \pi_j \sum_{i \in S} p_{ji} = \pi_j, \forall j \in S.$$

Potom rozdělení pravděpodobnosti  $\{\pi_i, i \in S\}$  je opravdu stacionárním rozdělením Markovova řetězce  $\{X_n, n \in N_0\}$  s maticí pravděpodobnosti přechodu  $\mathbf{P}$ .

Pro představu, zde na obrázku (Obr. 8) ukáži průběh Metropolisova – Hastingsova algoritmu. Graf zobrazuje výběr vzorků z dvourozměrného MCMC řetězce, který vyjadřuje složitý nelineární vztah mezi dvěma parametry. Hodnoty aposterioriho rozdělení začínají v levé části grafu, ale oproti tomu řetězec vychází ze startovacího bodu. Z grafu je pak vidět, že řetězec a aposteriori rozdělení budou mít díky MCMC algoritmu od určitého okamžiku přibližně stejný průběh.

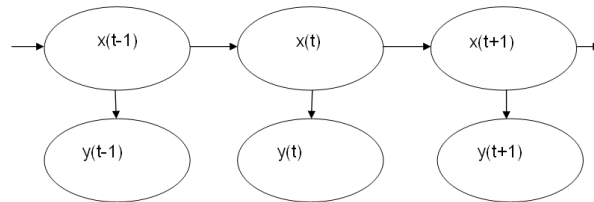


**Obr. 8** Metropolitův – Hastingsův algoritmus  
(zdroj: <https://www.doria.fi/bitstream/handle/10024/36631/isbn9789516976627.pdf>)

## 2.2. Skryté Markovovy řetězce

Abych mohla nadefinovat skryté markovské řetězce, nejprve uvedu velmi jednoduchý příklad, který nám umožní lépe pochopit, čím se dále budu zabývat. Na

následujícím obrázku (Obr. 9) vidíme posloupnost hodnot  $x_t$  a  $y_t$ . Na základě pozorování hodnot  $y_t$  se snažíme odhadnout hodnoty  $x_t$ , které odhadneme pomocí určitých charakteristik.



**Obr. 9** Skrytý Markovův řetězec

**Definice 12** Skrytý Markovův řetězec je určen posloupností  $\{x_t, t \in N\}$ , neboli skrytými stavy procesu a pozorováními  $\{y_t, t \in N\}$ , kde  $t$  je diskretní časový index. Skryté stavy jsou modelovány pomocí počátečního rozdělení Markovova řetězce  $p^{(0)}(x_0)$  a pravděpodobností přechodu  $p(x_t/x_{t-1})$ . Pro pozorování platí, že jsou podmíněně nezávislé<sup>4</sup> vzhledem k danému procesu  $\{x_t, t \in N\}$  a to s pravděpodobností  $p(y_t/x_t)$ . Potom skrytý Markovův řetězec je určen rovnicemi

$$\begin{aligned}
 & p(x_0), \\
 & p(x_t/x_{t-1}), \\
 & p(y_t/x_t),
 \end{aligned}$$

kde  $t \geq 1$ . Označíme jednotlivé stavy  $x_{0:t} \equiv \{x_0, x_1, \dots, x_t\}$  při pozorování  $y_{1:t} \equiv \{y_1, y_2, \dots, y_t\}$  v čase  $t$ . Pro snazší pochopení se dále v práci budu zaměřovat pouze na tento homogenní proces. To znamená, že jak přechodové  $p(x_t/x_{t-1})$ , tak i podmíněně pravděpodobnosti  $p(y_t/x_t)$  jsou nezávislé na časovém indexu  $n$ . Ale v případě, že by nás zajímal nehomogenní proces, lze jej jednoduše odvodit z homogenního procesu.

<sup>4</sup> Uvažujeme tři hodnoty  $a$ ,  $b$  a  $c$ . Pak podmíněnou pravděpodobnost  $a$  za podmínky  $b$  a  $c$  můžeme zapsat jako  $p(a/b,c)$ . V případě, že  $p(a/b,c)$  nezávisí na hodnotě  $b$  použijeme vztah  $p(a/b,c)=p(a/c)$ . Potom řekneme, že hodnota  $a$  je podmíněně nezávislá na hodnotě  $b$ , která je určena hodnotou  $c$ . In: <http://www.cedar.buffalo.edu/~srihari/CSE574/Chap8/8.5-ConditionalIndependence.pdf>

A to pouze za předpokladu, že v celém modelu jsou známy parametry, které nám určují přechodové pravděpodobnosti - parametry mezi hodnotami  $x_t$  a  $y_t$ .

Cílem je získat rekurzivní odhad<sup>5</sup> v čase okrajového rozdělení  $p(x_{0:t}/y_{1:t})$ , včetně jeho charakteristik, mezi které patří marginální, neboli filtrační rozdělení  $p(x_t/y_1, y_2, \dots, y_t) = p(x_t/y_{1:t})$ . V případě, že známe filtrační rozdělení, lze ho použít k výpočtu očekávané hodnoty a bodového odhadu.

Očekávanou hodnotu vypočítáme pomocí metody Monte Carlo

$$I(f_t) = E_{p(x_{0:t}/y_{1:t})}[f_t(x_{0:t})] \cong \int f_t(x_{0:t})p(x_{0:t}/y_{1:t})dx_{0:t}$$

pro funkce  $f_t : X^{t+1} \rightarrow R^{n+t}$ . Mezi vhodné příklady funkcí například patří podmíněná střední hodnota, kde funkce má tvar  $f_t(x_{0:t}) = x_{0:t}$  a nebo podmíněná kovariance  $x_t$  s funkcí  $f_t(x_{0:t}) = x_t x_t - E_{p(x_t/y_{1:t})}[x_t]E_{p(x_t/y_{1:t})}^T[x_t]$ . Vzorec okrajového rozdělení  $p(x_{0:t}/y_{1:t})$  vychází z Bayesovy věty a je daný aposteriorním rozdělením

$$\begin{aligned} p(x_{0:t}/y_{1:t}) &= \frac{p(x_{0:t})p(y_{1:t}/x_{0:t})}{p(y_{1:t})} = \\ &= \frac{p(y_{1:t}/x_{0:t})p(x_{0:t})}{\int p(y_{1:t}/x_{0:t})p(x_{0:t})dx_{0:t}}. \end{aligned}$$

Tuto rovnici jsme získali následujícím způsobem. Pro odvození prvního vzorce vycházíme ze vztahu (5), kde za známé výsledky experimentu  $x$  považujeme posloupnost napozorovaných hodnot  $y_t$  a chceme odhadnout posloupnost hodnot  $x_t$ , namísto neznámých parametrů  $\theta$ . V první rovnici za jmenovatele dosadíme výraz  $p(y_{1:t}) = \int p(y_{1:t}/x_{0:t})p(x_{0:t})dx_{0:t}$ , který jsme opět získali podobným způsobem ze vztahu (6).

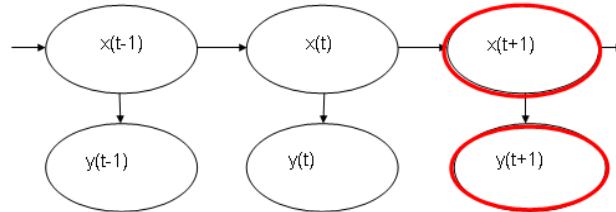
Pro aposteriorní rozdělení můžeme odvodit následující rekurzivní vzorec [14]

---

<sup>5</sup> Rekurzivním odhadem rozumíme aktualizaci parametrů algoritmu na základě nově naměřených dat.

$$\begin{aligned}
p(x_{0:t+1}/y_{1:t+1}) &= \frac{p(x_{0:t+1}, y_{1:t+1})}{p(y_{1:t+1})} = \\
&= \frac{p(x_{t+1}, y_{t+1}/x_{0:t}, y_{1:t})p(x_{0:t}, y_{1:t})}{p(y_{t+1}/y_{1:t})p(y_{1:t})} = \\
&= \frac{p(y_{t+1}/x_{0:t+1}, y_{1:t})p(x_{t+1}/x_{0:t}, y_{1:t})p(x_{0:t}, y_{1:t})}{p(y_{t+1}/y_{1:t})p(y_{1:t})} = \\
&= \frac{p(y_{t+1}/x_{t+1})p(x_{t+1}/x_t)p(x_{0:t}/y_{1:t})p(y_{1:t})}{p(y_{t+1}/y_{1:t})p(y_{1:t})} = \\
&= p(x_{0:t}/y_{1:t}) \frac{p(y_{t+1}/x_{t+1})p(x_{t+1}/x_t)}{p(y_{t+1}/y_{1:t})},
\end{aligned}$$

kde  $p(x_{0:t}/y_{1:t})$  je aposteriorní rozdělení a zlomek vyjadřuje aposteriorní rozdělení v čase  $t+1$  za předpokladu znalosti hodnot  $y_{t+1}$ . Jak můžeme vidět z následujícího obrázku (Obr. 10), pomocí rekurzivního vzorce se snažíme získat hodnoty  $y_{t+1}$  a  $x_{t+1}$  v čase  $t+1$  za předpokladu, že jsme vypočítali všechny předchozí hodnoty  $x_{0:t}$  a  $y_{1:t}$ .



**Obr. 10** Skrytý Markovův řetězec s apriorním rozdělením

Marginální rozdělení  $p(x_t/y_{1:t})$  také splňuje následující rekurzi [14]

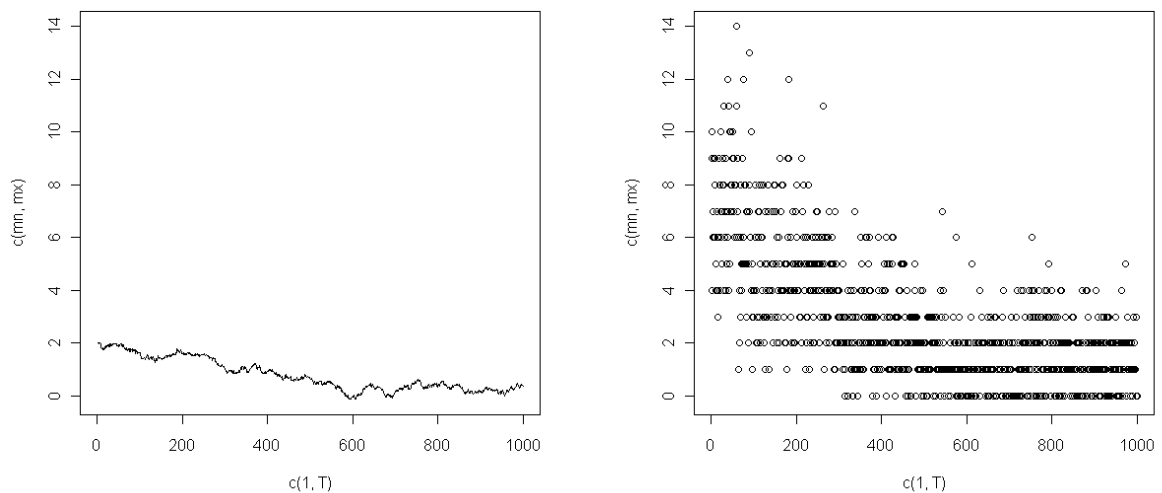
$$\begin{aligned}
p(x_t/y_{1:t}) &= \frac{p(x_t, y_{1:t})}{p(y_{1:t})} = \\
&= \frac{p(y_t/x_t, y_{1:t-1})p(x_t/y_{1:t-1})p(y_{1:t-1})}{p(y_t/y_{1:t-1})p(y_{1:t-1})} = \\
&= \frac{p(y_t/x_t)p(x_t/y_{1:t-1})}{\int p(y_t/x_t)p(x_t/y_{1:t-1})dx_t}.
\end{aligned}$$

Zdá se, že výpočty jednotlivých výrazů jsou jednoduché, ale většinou narazíme na problém při výpočtu  $p(y_{1:t})$ , marginálního rozdělení  $p(x_t/y_{1:t})$ ,  $p(x_t/y_t)$  a  $I(f_t)$ .

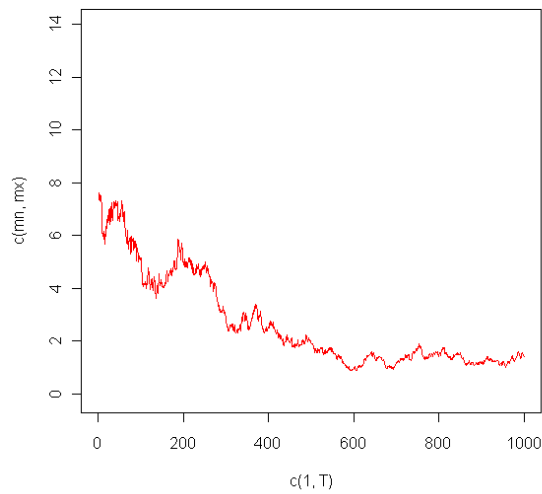
**Příklad 6** Uvažujme náhodnou procházku, která určuje intenzitu pro Poissonovo rozdělení. Což znamená, že  $Y \sim Pois(\exp(X))$ , kde  $X \sim N(0,0.2)$ ,  $T$  vyjadřuje počet pozorování a „set.seed“ je příkaz, který nám umožní zopakovat pokus se stejnými náhodnými čísly [10]. V našem případě nastavíme generátor náhodných čísel tak, aby byl stejný jak pro náhodnou veličinu  $X$ , tak i pro realizace náhodné veličiny  $Y$ .

```
> T=1000
> set.seed(8)
> X=rnorm(T,0,0.2^2)
> X=cumsum(X)+2
> Y=rpois(T,exp(X))
```

První graf (Obr. 11) zobrazuje náhodnou veličinu  $X$  v závislosti na čase od  $1:T$  a bodový graf naopak ukazuje vývoj náhodné veličiny  $Y$  za čas  $T$ . Na dalším grafu (Obr. 12) můžeme vidět vývoj náhodné veličiny  $\exp(X)$  v závislosti na čase  $T$ .



**Obr. 11** Vývoj náhodných veličin  $X$  a  $Y \sim Pois(\exp(X))$

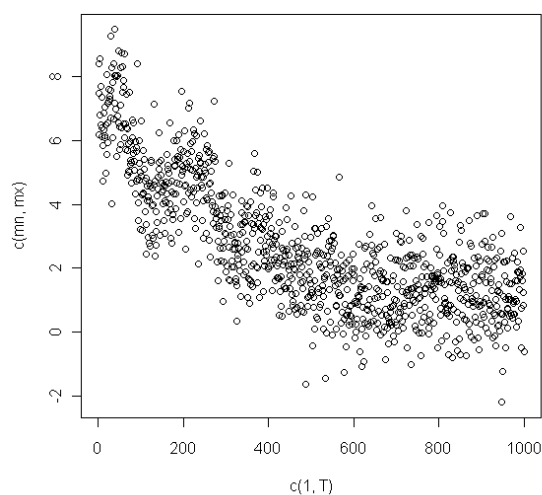


**Obr. 12** Exponenciála náhodné procházky

**Příklad 7** Nyní nás bude zajímat vztah mezi náhodnou procházkou a normálním rozdělením. Normální rozdělení je vymezeno jednotkovým rozptylem a střední hodnotou, která je určena právě danou náhodnou procházkou, tj.  $Y \sim N(\exp(X), 1)$ , kde  $X \sim N(0, 0.2)$ ,  $T$  opět vyjadřuje počet pozorování.

```
> Y=rnorm(T,exp(X),1)
```

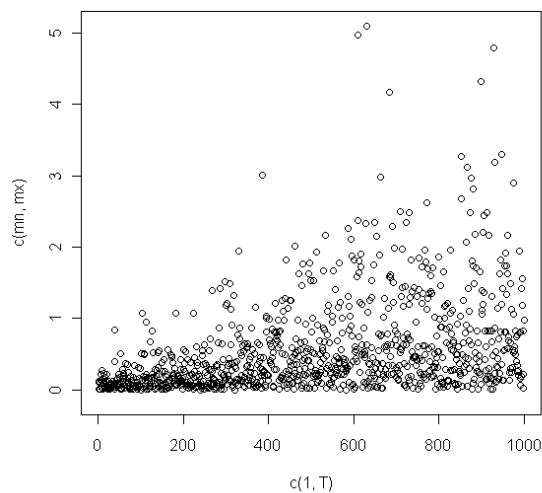
V algoritmu (viz Příloha) se změní hodnota náhodné veličiny  $Y$ , což můžete vidět na grafu (Obr. 13)



**Obr. 13** Vývoj náhodné veličiny  $Y \sim N(\exp(X), 1)$

**Příklad 8** Jako další příklad uvedu náhodnou procházku, která bude vyjadřovat intenzitu pro exponenciální rozdělení, tj.  $Y \sim \text{Exp}(\exp(X), 1)$ , kde  $X \sim N(0, 0.2)$ ,  $T$  opět vyjadřuje počet pozorování (viz Příloha).

```
> Y=rexp(T,exp(X))
```



**Obr. 14** Vývoj náhodné veličiny  $Y \sim \text{Exp}(\exp(X), 1)$



### 3. MCMC filtry

V této kapitole uvedu jednotlivé částicové filtry Importance Sampling, Sequential Importance Sampling a obecný Sequential Monte Carlo filtr. Filtry používáme k odhadnutí původních dat na základě pozorování, například se může jednat o očišťování naměřených dat při experimentu. Dále uvedu metodu Resampling, což už není částicový filtr, ale jedná se o nástroj, který řeší problém vznikající při filtrování.

Kapitola byla vypracována s využitím zdrojů [8], [9], [14], [15].

Předpokládejme, že jsme schopni simulovat  $N$  nezávislých stejně rozdělených náhodných vzorků. Náhodné vzorky  $\{x_{0:t}^i, i = 1, \dots, N\}$  budeme nazývat částicemi určenými dle rozdělení pravděpodobnosti  $p(x_{0:t}/y_{1:t})$ . Empirický odhad tohoto rozdělení má tvar

$$P_N(dx_{0:t}/y_{1:t}) = \frac{1}{N} \sum_{i=1}^N \delta_{x_{0:t}^i}(dx_{0:t}).$$

Míra  $\delta_{x_{0:t}^i}(dx_{0:t})$  z empirického odhadu  $P_N(dx_{0:t}/y_{1:t})$  je Diracova míra vzhledem ke stavu  $x_{0:t}$ .<sup>6</sup> Potom pro  $N$  částic dostaneme odhad  $I(f_t)$  následovně

$$I_N(f_t) = \int f_t(x_{0:t}) P_N(x_{0:t}/y_{1:t}) dx_{0:t} = \int f_t(x_{0:t}) P_N(dx_{0:t}/y_{1:t}) = \frac{1}{N} \sum_{i=1}^N f_t(x_{0:t}^i).$$

---

<sup>6</sup> Necht' množina  $X$  je neprázdná a prvek  $x \in X$ . Diracova míra  $\delta_x$  je definovaná na množině  $X$  (s jakoukoliv  $\sigma$ -algebrou všech podmnožin množiny  $X$ ) vzhledem ke stavu  $x$  následovně:  $\delta_x(A) = 1$  pro  $x \in A$  a  $\delta_x(A) = 0$  pro  $x \notin A$ . Diracova delta, neboli Diracova  $\delta$ -funkce je taková funkce, která v nule nabývá nekonečna a všude jinde má hodnotu nulovou, tj.  $\delta(x) = \infty$  pro  $x = 0$  a  $\delta(x) = 0$  pro  $x \neq 0$ . Pro integrál Diracovy  $\delta$ -funkce platí  $\int_{-\infty}^{\infty} \delta(x) dx = 1$ . In: [http://en.wikipedia.org/wiki/Dirac\\_measure](http://en.wikipedia.org/wiki/Dirac_measure), [http://cs.wikipedia.org/wiki/Delta\\_funkce](http://cs.wikipedia.org/wiki/Delta_funkce).

Tento odhad  $I_N(f_t)$  je nestranný v případě, že aposteriorní rozptyl  $f_t(x_{0t})$  odpovídá vztahu  $\sigma_{f_t}^2 \cong E_{p(x_{0t}/y_{1t})}[f_t^2(x_{0t})] - I^2(f_t) < \infty$  [9]. Potom rozptyl odhadu  $I_N(f_t)$  je roven  $\text{var}(I_N(f_t)) = \frac{\sigma_{f_t}^2}{N}$  a užitím silného zákona velkých čísel platí

$$I_N(f_t) \xrightarrow[N \rightarrow \infty]{s.j.} I(f_t). \quad (12)$$

Ze vztahu (12) je zřejmé, že pro  $N \rightarrow \infty$  rozptyl odhadu  $I_N(f_t)$  konverguje skoro jistě k odhadu  $I(f_t)$ . Navíc, pokud platí  $\sigma_{f_t}^2 < \infty$ , pak dle centrální limitní věty můžeme napsat následující vztah

$$\sqrt{N}[I_N(f_t) - I(f_t)] \xrightarrow[N \rightarrow \infty]{} N(0, \sigma_{f_t}^2),$$

kde  $\Rightarrow$  značí konvergenci dle pravděpodobnosti.

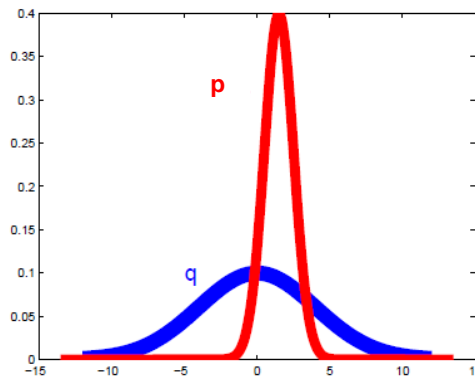
Výhoda této metody spočívá v tom, že z množiny náhodných vzorků  $\{x_{0t}^i, i = 1, \dots, N\}$  lze snadno určit jakoukoliv hodnotu  $I(f_t)$ . Míra konvergence pro tento odhad je nezávislá vzhledem k integrované dimenzi. Oproti tomu, žádná metoda deterministické numerické integrace nemá takovou míru konvergence, pro kterou by platilo, že se zvyšující se dimenzí integrace se konvergence snižuje.

Nevýhodou metody je, že je obvykle nemožné otestovat okrajové rozdělení  $p(x_{0t}/y_{1t})$  v každém čase  $t$ , kdy v rozdělení pravděpodobnosti  $p(x_{0t}/y_{1t})$  se vyskytuje více proměnných s příliš složitými rozděleními a jedinou známou normující konstantou. V oblasti aplikované statistiky se proto zavedli Markov Chain Monte Carlo (MCMC) metody, populární přístupy odběru vzorků ze složitých rozdělení pravděpodobnosti.

### 3.1. Importance Sampling

Metoda Importance Sampling umí odhadnout výsledky simulace v případě, kdy máme více než jedno rozdělení, nebo v případě, kdy je pro nás velmi obtížné získat vzorky z pravděpodobnostní hustoty  $p(x)$ . Proto musíme zavést novou

pravděpodobnostní hustotu  $q(x)$  nazývající se významná funkce. Můžeme ji zvolit tak, že vzorky budeme odebírat např. z Gaussova (neboli normálního) rozdělení. Na obrázku (Obr. 15) vidíme, že střední hodnota pravděpodobnostní hustoty  $q(x)$  má vysoce husté rozmístění a rozptýl se šíří přes nejdůležitější oblasti pravděpodobnostní hustoty  $p(x)$ . Jinak řečeno, snažíme se vygenerovat vzorky z  $q(x)$  tak, aby s co největší pravděpodobností pokrývaly oblast nejvyšších hodnot  $p(x)$ . Metoda Importance Sampling je tedy založena na výběru správné významné funkce  $q(x)$ .



**Obr. 15** Vztah mezi pravděpodobnostními hustotami  $p(x)$  a  $q(x)$ .

(zdroj: <http://web.archive.org/web/20110531150413/http://www.bioass.ac.uk/students/alexm/MCMCintroPresentation.pdf>)

Předpokládejme funkci  $q(x_{0:t}/y_{1:t})$ , která je závislá na hodnotách  $y_{1:t}$  a její doplněk  $p(x_{0:t}/y_{1:t})$ , pak řešíme  $I(f_t)$  následovně

$$I(f_t) = \frac{\int f_t(x_{0:t}) w(x_{0:t}) q(x_{0:t}/y_{1:t}) dx_{0:t}}{\int w(x_{0:t}) q(x_{0:t}/y_{1:t}) dx_{0:t}},$$

kde  $w(x_{0:t})$  je tzv. významná váha

$$w(x_{0:t}) = \frac{p(x_{0:t}/y_{1:t})}{q(x_{0:t}/y_{1:t})}.$$

V případě, že chceme získat Monte Carlo odhad  $I(f_t)$ , nasimulujeme  $N$  nezávislých a stejně rozdělených náhodných veličin  $\{x_{0t}^i, i = 1, \dots, N\}$  dle významné funkce  $q(x_{0t}/y_{1t})$

$$\hat{I}_N(f_t) = \frac{\frac{1}{N} \sum_{i=1}^N f_t(x_{0t}^i) w(x_{0t}^i)}{\frac{1}{N} \sum_{j=1}^N w(x_{0t}^j)} = \sum_{i=1}^N f_t(x_{0t}^i) \tilde{w}_t^i,$$

kde  $\tilde{w}_t^i$  je normovaná významná váha, kterou vypočítáme ze vztahu

$$\tilde{w}_t^i = \frac{w(x_{0t}^i)}{\sum_{j=1}^n w(x_{0t}^j)}.$$

Pro  $N$  jdoucí do nekonečna platí, že odhad  $\hat{I}_N(f_t)$  podle silného zákona velkých čísel konverguje k  $I(f_t)$ .

Tato metoda integrace může být interpretována i jako metoda odběru vzorků. V tomto případě okrajové rozdělení  $p(x_{0t}/y_{1t})$  aproximují pomocí

$$\hat{P}_N(dx_{0t}/y_{1t}) = \sum_{i=1}^N \tilde{w}_t^i \delta_{x_{0t}^i}(dx_{0t}),$$

kde  $\tilde{w}_t^i$  jsou dříve definované normované významné váhy a  $\delta_{x_{0t}^i}(dx_{0t})$  je Diracova míra vzhledem ke stavu  $x_{0t}$ . Pak odhad  $\hat{I}_N(f_t)$  je funkce  $f_t(x_{0t})$  integrovaná s ohledem na empirické opatření  $\hat{P}_N(dx_{0t}/y_{1t})$  a dostaneme

$$\hat{I}_N(f_t) = \int f_t(x_{0t}) \hat{P}_N(dx_{0t}/y_{1t}).$$

Metoda Importance Sampling je metodou obecné Monte Carlo integrace. Její nevýhody spočívají v rekurzivním odhadu, kdy před samotným odhadem pravděpodobnostního rozdělení  $p(x_{0t}/y_{1t})$  potřebujeme znát všechny údaje  $y_{1t}$ . Obecně

také platí, že pokaždé, kdy máme k dispozici nová data  $y_{t+1}$ , musíme přepočítat normované váhy přes výše uvedený algoritmus. Potom výpočetní složitost této operace se zvyšuje časem, což je pro nás nepříjemné.

**Příklad 9** Na tomto příkladu si ukážeme, jak využít metodu Importance Sampling

k hrubému výpočtu integrálu  $\int_0^2 e^{-x} dx = 0,86466472$ , který jsme už dříve vypočítali

Monte Carlo integrací (viz Příklad 5) [7]. Budeme vycházet ze zjednodušeného vzorce

$$\int_a^b f(x) dx = \int_a^b \frac{f(x)}{g(x)} g(x) dx = \int_a^b \frac{f(x)}{g(x)} dGx,$$

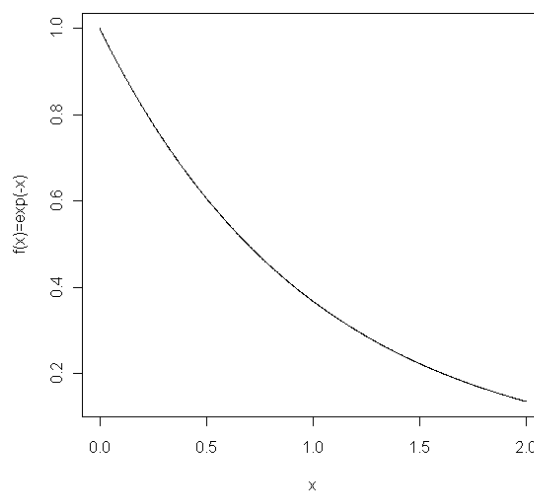
kdy z funkce  $g(x)$  vybereme vzorky a na

základě tohoto výběru odhadneme  $\frac{f(x)}{g(x)}$ . Máme tedy zadané hodnoty  $\lambda = 0.9$ ,  $T = 2$

a funkce  $f(x) = e^{-x}$  a  $g(x) = \lambda e^{-\lambda x} / (1 - e^{-T\lambda})$ .

1. Vygenerujeme hodnoty na intervalu  $(0,2)$  o délce 1 000 hodnot a na základě takto získaných hodnot vypočítáme funkci  $f(x) = e^{-x}$ .

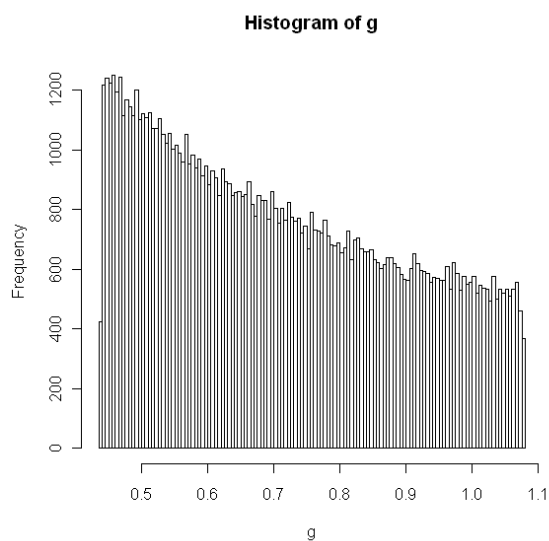
```
> set.seed(286587)
> x=seq(0,2,length=1000)
> y=exp(-x)
> plot(x,y,type="l",xlab="x",ylab="f(x)=exp(-x)")
```



**Obr. 16** Graf funkce  $f(x) = e^{-x}$

2. Ve druhém kroku nás bude zajímat funkce  $g(x) = \lambda e^{-\lambda x} / (1 - e^{-T\lambda})$ , kterou získáme na základě náhodného výběru  $x_1, \dots, x_n \sim R(0,1)$ .

```
> n=100000
> lambda=0.9
> T=2
> x=runif(n,0,1)
> g=lambda*(exp(-lambda*x))/(1-exp(-T*lambda))
> hist(g,100)
```



**Obr. 17** Graf funkce  $g(x) = \lambda e^{-\lambda x} / (1 - e^{-T\lambda})$

3. Odhadneme hodnotu integrálu  $\int_0^2 e^{-x} dx$

```
> f=exp(-x)
> fg=f/g
> I=sum(fg)/n
> I
```

**[1] 0.8824199**

### 3.2. Sequential Importance Sampling

Sequential Importance Sampling na rozdíl od Importance Sampling připouští u významné funkce v čase  $t$  marginální rozdělení v čase  $t-1$ , potom pravděpodobnost hustoty  $q(x_{0:t}/y_{1:t})$  je definována následovně

$$q(x_{0:t}/y_{1:t}) = q(x_{0:t-1}/y_{1:t-1})q(x_t/x_{0:t-1}, y_{1:t}),$$

$$q(x_{0:t}/y_{1:t}) = q(x_0) \prod_{k=1}^t q(x_k/x_{0:k-1}, y_{1:k}).$$

Tato nová funkce umožňuje hodnotit rekurzivně v čase normované váhy

$$\tilde{w}_t^i \propto \tilde{w}_t^i \frac{p(y_t/x_t^i)p(x_t^i/x_{t-1}^i)}{q(x_t^i/x_{0:t-1}^i, y_{1:t})}.$$

Sequential Importance Sampling je atraktivní algoritmus, pomocí něhož získáme jen omezenou verzi významných odběrů vzorků. Je to způsobeno tím, že se zvyšujícím se  $t$  se významné váhy  $\tilde{w}_t^i$  stávají více asymetrickými. Což znamená, že po několika krocích v čase zůstává jen jedna částice, která má nenulové normované váhy a ostatní jsou prakticky nulové. Částice s nenulovými normovanými váhami jsou „částicově neefektivní“, zatímco částice s nulovými váhami jsou „zbytečné“. Z tohoto hlediska nedokáží tyto algoritmy adekvátně reprezentovat okrajové rozdělení. Proto byla zavedena další metoda, která se nazývá Resampling.

### 3.3. Resampling

Hlavní myšlenka metody Resampling je odstranění částic s nízkou normovanou váhou  $\tilde{w}_t^i$  a znásobení částic s velkými normovanými váhami, kdy váhy empirického

rozdělení  $\hat{P}_N(dx_{0:t}/y_{1:t}) = \sum_{i=1}^N \tilde{w}_t^i \delta_{x_{0:t}^i}(dx_{0:t})$  definované v Importance Sampling

nahradíme nenormovanými váhami

$$\bar{P}_N(dx_{0:t}/y_{0:t}) = \frac{1}{N} \sum_{i=1}^N N_t^i \delta_{x_{0:t}}^i(dx_{0:t}),$$

$N_t^i$  je rovno počtu výsledků částice  $x_{0:t}^i$ . Jedná se o celé číslo, pro které platí  $\sum_{i=1}^N N_t^i = N$ . Pokud  $N_t^i = 0$ , částice  $x_{0:t}^i$  „zanikne“. V případě  $N_t^i > 0$  nazýváme částice  $x_{0:t}^i$  „zachovanými“.  $N_t^i$  jsou zvolené tak, že  $\bar{P}_N(dx_{0:t}/y_{1:t})$  se blíží k  $\hat{P}_N(dx_{0:t}/y_{1:t})$  v tom smyslu, že pro jakoukoliv funkci  $f_t$  platí

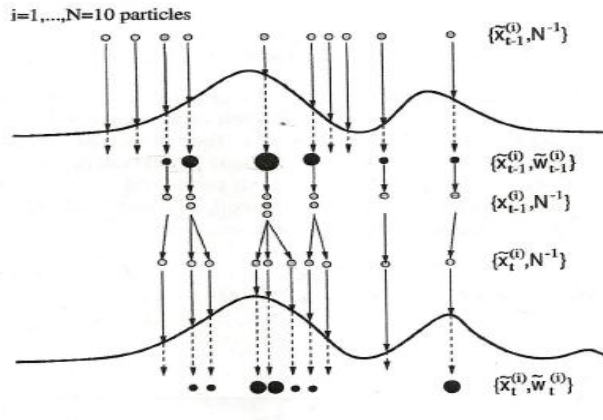
$$\int f_t(x_{0:t}) \bar{P}_N(dx_{0:t}/y_{1:t}) \approx \int f_t(x_{0:t}) \hat{P}_N(dx_{0:t}/y_{1:t}).$$

Po výběrovém kroku jsou „zachované“ částice  $x_{0:t}^i$ ,  $N_t^i > 0$  přibližně rozděleny podle  $p(x_{0:t}/y_{1:t})$ . Existuje mnoho způsobů, jak správně vybrat počet částic  $N_t^i$ . O nejpoužívanějším postupu se zmiňuje Gordon ve své knize z roku 1993. Touto metodou získáme „zachované“ částice tak, že uděláme  $N$  krát výběr z diskrétního rozdělení  $\hat{P}_N(dx_{0:t}/y_{1:t})$ . Což je ekvivalentní k výběru, při kterém získáme  $N_t^i$  výsledků z multinomického rozdělení s parametrem  $\tilde{w}_t^i$ .

Jak to vypadá prakticky si můžeme ukázat na následujícím obrázku (Obr. 18). Uvažujeme deset částic libovolného empirického pravděpodobnostního rozdělení získané filtrem Importance Sampling. Poznamenejme, že Resampling začíná v čase  $t-1$  jako nevyvážené měření  $\{\tilde{x}_{t-1}^i, N^{-1}\}$ , které zajistí aproximaci pravděpodobnosti přechodu  $p(x_{t-1}/y_{1:t-2})$ . Pro každou částici počítáme normované váhy s použitím informací z času  $t-1$ . To má za následek vyvážené měření  $\{\tilde{x}_{t-1}^i, \tilde{w}_{t-1}^i\}$ , které každé částici přiřadí váhu  $\tilde{w}_{t-1}^i$  dle její významnosti. Z obrázku lze rozpoznat, že pátá částice je nejvýznamnější (má největší váhu). První, druhá, sedmá a osmá částice jsou nevýznamné (mají nulovou váhu). Toto vyvážené měření udává aproximaci pravděpodobnosti přechodu  $p(x_{t-1}/y_{1:t-1})$ . Následné převzorkování vybere nejvhodnější částice získané neváženým měřením  $\{\tilde{x}_{t-1}^i, N^{-1}\}$ , které stále aproximují  $p(x_{t-1}/y_{1:t-1})$ . Jinak řečeno, nejvýznamnější pátou částici s největší váhou budeme uvažovat třikrát,



významné částice (čtvrtou a šestou) bereme dvakrát a méně významné částice (třetí, devátou a desátou) uvažujeme jen jednou. S nevýznamnými částicemi, které mají nulovou váhu dále nepracujeme. Samotná predikce je pak představena částicemi, které jsme získali z předcházejícího kroku a vede k měření  $\{\tilde{x}_t^i, N^{-1}\}$  v čase  $t$  aproximující filtrační rozdělání  $p(x_t/y_{1:t-1})$ . Stejný postup můžeme znova aplikovat na získané částice.



**Obr. 18** Resampling

V literatuře se můžeme setkat se třemi nejpoužívanějšími algoritmy metody Resampling

- Systematic Resampling

Uvažujeme částici  $U_1 \sim R\left[0, \frac{1}{N}\right]$  a ostatní částice  $i = 2, \dots, N$  definované

vzorcem  $U_i = U_1 + \frac{i-1}{N}$ , potom za předpokladu  $\sum_{k=1}^0 := 0$  můžeme vypočítat

$$N_t^i = \left\{ U_j : \sum_{k=1}^{i-1} W_t^k \leq U_j \leq \sum_{k=1}^i W_t^k \right\}.$$

**Příklad 10** Uvažujeme  $N = 5$  částic a váhy  $W = \left(\frac{1}{25}, \frac{3}{25}, \frac{12}{25}, \frac{4}{25}, \frac{5}{25}\right)$ .

A naším úkolem je vybrat opět stejný počet vzorků, které jsou na rozdíl od původních částic vybrány. První náhodnou veličinu  $U_1$  si vygenerujeme

v programu R pomocí příkazu „U=runif(1,0,(1/N))“, kdy  $U_1 \sim R\left[0, \frac{1}{5}\right]$  má

rovnoměrné rozdělání [10]. V našem případě dostaneme  $U_1 = 0,138346$ .

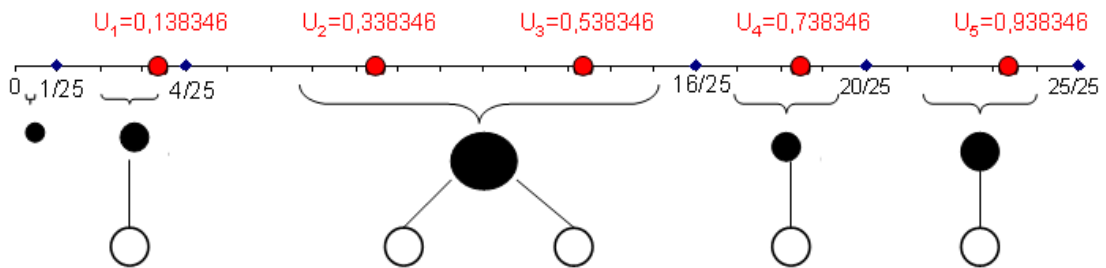
Ostatní částice  $U_i$  vypočítáme dle předchozího vzorce, tzn.  $U_2 = 0,338346$ ,

$U_3 = 0,538346$ ,  $U_4 = 0,738346$  a  $U_5 = 0,938346$ . Jak tedy poznáme, které

částice jsou méně či více významné? Na následujícím obrázku (Obr. 19) vidíme pět intervalů, které jsme získali z vah. Jen pro připomenutí, součet vah musí být roven 1. Potom jednotlivým částicím přiřadíme váhu podle toho, do kterého intervalu padnou. Částici  $U_1$ , patřící do intervalu  $\left(\frac{1}{25}, \frac{4}{25}\right)$ , přiřadíme váhu  $\frac{3}{25}$ .

$U_2, U_3 \in \left(\frac{4}{25}, \frac{16}{25}\right)$  mají váhu  $\frac{12}{25}$ ,  $U_4 \in \left(\frac{16}{25}, \frac{20}{25}\right)$  má váhu  $\frac{4}{25}$  a nakonec

$U_5 \in \left(\frac{20}{25}, \frac{25}{25}\right)$  s váhou  $\frac{5}{25}$ .



**Obr. 19** Systematic Resampling

- Residual Resampling

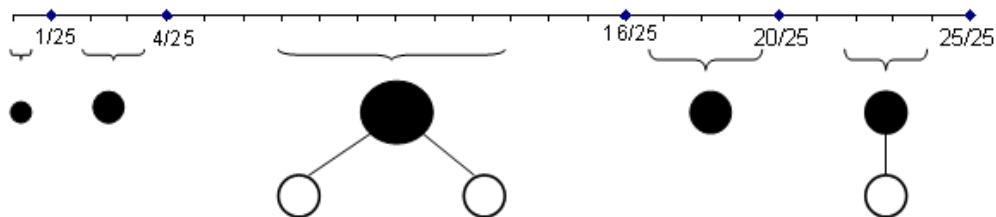
Zvolíme si částice  $\tilde{N}_t^i = \lfloor NW_t^i \rfloor$  a náhodně vybereme částice  $\bar{N}_t^{1:N}$  z multinomického rozdělení s parametry  $(\bar{N}_t^{1:N}, \bar{W}_t^{1:N})$ , které získáme ze vztahu  $N_t^i = \tilde{N}_t^i + \bar{N}_t^i$ .

**Příklad 11** Vycházíme ze stejného zadání jako v případě Systematic Resampling. Zvolíme si vektor vzorků  $\tilde{N}_t^i$  dle předešlého vzorce a obdržíme následující výsledek

$$\tilde{N}_t^i = \left\lfloor 5 * \left( \frac{1}{25}, \frac{3}{25}, \frac{12}{25}, \frac{4}{25}, \frac{5}{25} \right) \right\rfloor = \left\lfloor \left( \frac{5}{25}, \frac{15}{25}, \frac{60}{25}, \frac{20}{25}, \frac{25}{25} \right) \right\rfloor = (0,0,2,0,1).$$

Tím jsme získali dva vzorky, kterým odpovídá váha rovna délce třetího intervalu  $\left(\frac{4}{25}, \frac{16}{25}\right)$ , tzn.  $\frac{12}{25}$  a jeden vzorek, kterému odpovídá váha o délce posledního

intervalu  $\left(\frac{20}{25}, \frac{25}{25}\right)$ , tzn.  $\frac{5}{25}$ .



**Obr. 20** Zobrazení vzorků dle  $\tilde{N}_t^i = (0,0,2,0,1)$

Nyní nám chybí dva vzorky a ty získáme z multinomického rozdělení s parametry  $(2, \bar{W}_t^{1:N})$ . V tomto momentu sem vstupuje náhodnost. K tomu, abychom mohli vygenerovat nové částice, potřebujeme určit vektor vah  $\bar{W}_t^{1:N}$

pro jednotlivé vzorky. Vektor vah je tedy ve tvaru  $\left(\frac{5}{25}, \frac{15}{25}, \frac{2}{25}, \frac{20}{25}, 0\right)$ . Pokud

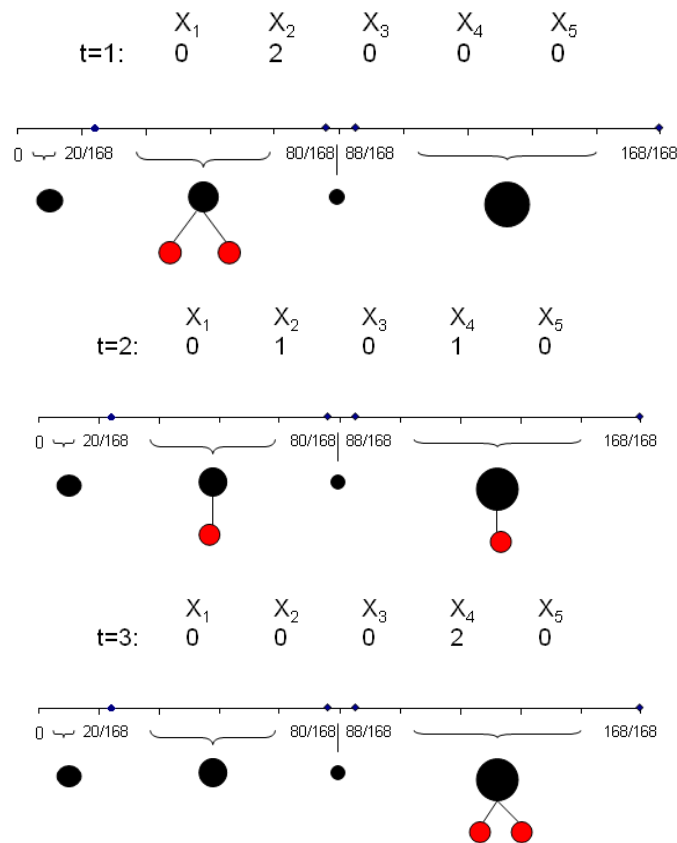
se na něj pozorně podíváme, zjistíme, že uvedené váhy mají ve třetí a páté složce nižší hodnotu oproti původnímu vektoru vzorků  $\tilde{N}_t^i$ . Je to způsobeno tím, že už jsme v prvním kroku Residual Resampling z těchto intervalů částice vybrali. Jelikož součet všech složek vektoru není roven jedné, musíme ho dle známého

vzorce  $v_i = \frac{w_i}{\sum_{i=1}^N w_i}$  znormovat. Dostaneme výsledný vektor

$\left(\frac{20}{168}, \frac{60}{168}, \frac{8}{168}, \frac{80}{168}, 0\right)$  a multinomické rozdělení s těmito parametry

$\left(2, \left(\frac{20}{168}, \frac{60}{168}, \frac{8}{168}, \frac{80}{168}, 0\right)\right)$ . Pokud v programu R použijeme příkaz

„rmultinom“, můžeme například vygenerovat náhodné veličiny v čase, které jsou zobrazeny na následujícím obrázku (Obr. 21) [10].



**Obr. 21** Vygenerované náhodné veličiny z multinomického rozdělení

- Multinomial Resampling

Zde máme vzorky  $N_t^{1:N}$  z multinomického rozdělení s parametry  $(N, W_t^{1:N})$ .

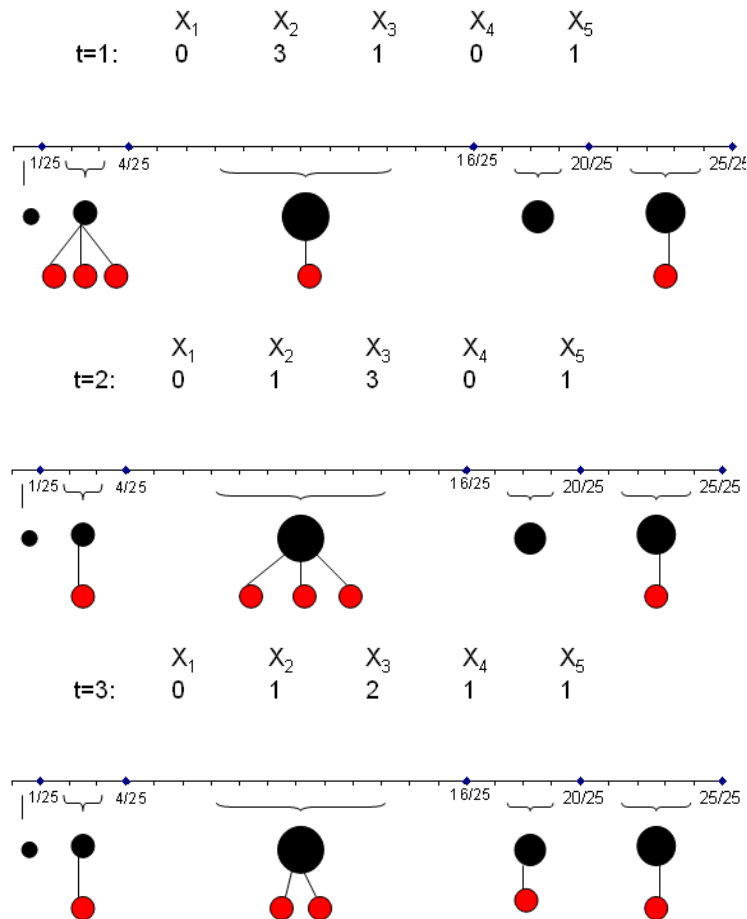
**Příklad 12** Opět uvažujeme  $N = 5$  částic a váhy  $W_t^{1:N} = (W_t^1, \dots, W_t^N)$ .

A abychom získali náhodný vektor  $N_t^{1:N} = (N_t^1, \dots, N_t^N)$ , použijeme

multinomické rozdělení s parametry  $\left(5, \left(\frac{1}{25}, \frac{3}{25}, \frac{12}{25}, \frac{4}{25}, \frac{5}{25}\right)\right)$ . V tomto

případě váhy vyjadřují pravděpodobnosti pro třídy, ze kterých budeme generovat jednotlivé vzorky. Opět využijeme program R a příkaz „rmultinom“ [10]. Jednotlivé výsledky v čase jsou znázorněny na obrázku (Obr. 22).

Například v čase  $t = 2$  obdržíme náhodný vektor vzorků  $N_2^{1:5} = (0, 1, 3, 0, 1)$ .



**Obr. 22** Multinomial Resampling

Metodou Resampling získáme částice přibližně odpovídající cílovému rozdělení  $P_N(dx_{0:t}/y_{1:t})$ , což je zřejmé, protože máme zájem o odhad  $I_N(f_t)$  pro  $N$  částic. Užitím empirického rozdělení  $\hat{P}_N(dx_{0:t}/y_{1:t})$  získáme odhad s nižším rozptylem než je ten, který bychom získali užitím empirického rozdělení  $\bar{P}_N(dx_{0:t}/y_{1:t})$ . Nicméně důležitou výhodou této metody je, že nám umožňuje odstranit částice s nízkou váhou s vysokou pravděpodobností. Tato skutečnost je pro nás velmi užitečná. Nechceme převádět částice s nízkou váhou, ale chceme zaměřit naše výpočetní úsilí s velkou pravděpodobností na oblasti s významnými částicemi. Je zřejmé, že může také nastat i situace, kdy vyřadíme částice s nízkou váhou v čase  $n$ , avšak v čase  $n+1$  mohou mít vysokou váhu. V tom případě je Resampling nešetnou metodou. Před užitím této metody musíme zvážit její výhody i nevýhody. Nicméně tato metoda je z pohledu odhadu prokazatelně prospěšná. Resampling zajistí stabilitu za cenu zvýšeného Monte Carlo rozptylu.

### 3.4. Obecný Sequential Monte Carlo filtr

Metoda Sequential Monte Carlo využívá poznatků z metod Sequential Importance Sampling a Resampling. Jak jsme již zmínili, metoda Resampling efektivně odstraňuje částice s nízkou vahou a znásobuje částice s vysokou vahou. Je to však za cenu okamžitého zavedení některých dalších odchylek. Jestliže částice mají nestandardizované váhy s malým rozptylem, pak převzorkovací krok by mohl být zbytečný. V důsledku toho je v praxi rozumnější použít Resampling pouze v případě, pokud rozptyl nestandardizovaných vah je větší než předem stanovená hranice, tzv. Effective Sample Size (ESS) – efektivní velikost vzorku, která v čase  $n$  vypadá následovně

$$ESS = \left( \sum_{i=1}^N (W_t^i)^2 \right)^{-1}$$

V jednoduchosti je interpretace taková, že výsledky Importance Sampling, které jsou založené na tom, že  $N$  vážených vzorků (pokud jde o odhad rozptylu) se blíží k výsledkům založených na ESS dokonalých vzorcích z cílového rozdělení. ESS nabývá hodnot mezi 1 až  $N$  a převzorkujeme je pouze tehdy, pokud jsou nižší než práh  $N_T$ .

Práh je obvykle roven  $N_T = \frac{N}{2}$ . Za alternativní kritérium lze použít míru neurčitosti

vah  $\{W_t^i\}$ . Její nejvyšší možná hodnota je rovna  $W_t^i = \frac{1}{N}$ . V případě, že míra neurčitosti nedosahuje této hodnoty, provedeme resample krok.

## Závěr

Diplomová práce si kladla za cíl objasnit teorii Markov Chain Monte Carlo (MCMC) metod a tuto teorii aplikovat na data. Nejdříve jsem se věnovala Markovovým řetězcům a bayesovskému modelu. Použila jsem ukázkový příklad, který vyjadřoval vztah mezi apriorním a posteriorním rozdělením. Dále jsem se věnovala MCMC metodám. Velký důraz jsem kladla na Monte Carlo integraci, pomocí které jsem odhadla hodnoty složitých integrálů. V podkapitole skrytých Markovových řetězců jsem vygenerovala náhodnou procházku, na základě které jsem pak určila intenzitu Poissonova, normálního a exponenciálního rozdělení. Poslední kapitola obsahovala MCMC filtry a Resampling metodu, která obsahuje tři nejpoužívanější algoritmy, které jsou následně vysvětleny na praktickém příkladu.

Díky této práci jsem se seznámila s novými metodami, které nejsou v naší zemi velmi rozšířeny. Naučila jsem se pracovat s matematickým programem R. Největší problém mi činila představa aplikovatelnosti metod do reálného života. V práci jsem nevyházela z reálných dat, ale z dat náhodně vygenerovaných, aby byly postupy a výsledky pro čtenáře lépe pochopitelné.

Doufám, že tato práce bude přínosem nejen pro mě, ale i pro další zájemce o MCMC algoritmy.

## Příloha

### Příklad 1 - Aposteriorní rozdělení B

```
> theta0=900
> tau0=100
> theta=rnorm(1,mean=theta0,sd=tau0)
> tau=40
> N=200
> Xtheta=rnorm(N,mean=theta,sd=tau)
> x=mean(Xtheta)
> w=(tau0)^(-2)/((tau0)^(-2)+(tau)^(-2))
[1] 0.137931
> theta1=w*(theta0)+(1-w)*x
[1] 885.0105
> tau11=(tau0)^(-2)+(tau)^(-2)
[1] 0.000725
> tau1=(tau11)^(-(1/2))
[1] 37.13907
> thetaX=rnorm(100,mean=(theta1),sd=(tau1))
> estimatetheta=mean(thetaX)
[1] 880.5503

- 95% konfidenční interval pro theta
> L=estimatetheta-1.96*(tau1)/((x)^(1/2))
> L
[1] 878.1001
> P=estimatetheta+1.96*(tau1)/((x)^(1/2))
> P
[1] 883.0005

- 95% konfidenční interval pro theta/X
> X=mean(Xtheta)
> L=X-1.96*tau/((N)^(1/2))
```



```
> L
```

```
[1] 877.0684
```

```
> P=X+1.96*tau/((N)^(1/2))
```

```
> P
```

```
[1] 888.1559
```

### **Příklad 7**

```
> T=1000
```

```
> set.seed(8)
```

```
> X=rnorm(T,0,0.2^2)
```

```
> X=cumsum(X)+2
```

```
> Y=rnorm(T,exp(X),1)
```

### **Příklad 8**

```
> T=1000
```

```
> set.seed(8)
```

```
> X=rnorm(T,0,0.2^2)
```

```
> X=cumsum(X)+2
```

```
> Y=rexp(T,exp(X))
```

## Literatura

[1] BECKMANN, P. *Historie čísla  $\pi$* . 1. vyd. Praha: Academia, 1998, s. 131-136. ISBN 80-200-0655-9.

[2] BUSLENKO, N. P., ŠREJDER J. A. *Stochastické početní metody*. Praha: Státní nakladatelství technické literatury, 1965, 194 s.

[3] FABIAN, F., KLUBIER., Z. *Metoda Monte Carlo a možnosti jejího uplatnění*. 1. vyd. Praha: Prospektrum, spol.s.r.o., 1998, 148 s. ISBN 80-717-5058-1.

[4] HRON, K., KUNDEROVÁ, P. *Markovovy řetězce a jejich aplikace*. 1. vyd. Olomouc: Univerzita Palackého v Olomouci, 2012, 135 s. ISBN 978-80-244-3132-1.

[5] KUNDEROVÁ, P. *Základy pravděpodobnosti a matematické statistiky*. 1. vyd. Olomouc: Univerzita Palackého, 2004, 186 s. Skripta (Univerzita Palackého). ISBN 80-244-0813-9.

[6] PRÁŠKOVÁ, Z. *Základy náhodných procesů*. 1. vyd. Praha: Karolinum, 2004, 151 s. ISBN 80-246-0971-1.

## Elektronické zdroje

[7] BILKER, W. *Statistical Computing: Numerical Integration II* [online]. [cit. 2012-12-01]. Dostupné z: [http://www.cceb.med.upenn.edu/pages/courses/BSTA670/2009/18\\_NumInt.pdf](http://www.cceb.med.upenn.edu/pages/courses/BSTA670/2009/18_NumInt.pdf)

[8] DOUCET, A., JOHANSEN, A. M. *A Tutorial on Particle Filtering and Smoothing: Fifteen years later* [online]. [cit. 2012-08-16]. Dostupné z: [http://www.cs.ubc.ca/~arnaud/doucet\\_johansen\\_tutorialPF.pdf](http://www.cs.ubc.ca/~arnaud/doucet_johansen_tutorialPF.pdf)

[9] DOUCET, A., de FREITAS N., GORDON, N. *An Introduction to Sequential Monte Carlo Methods* [online]. [cit. 2012-11-24]. Dostupné z: DOUCET, Arnout a A. M. JOHANSEN. *A Tutorial on Particle Filtering and Smoothing: Fifteen years later*

[online]. [cit. 2012-11-24]. Dostupné z: [http://www.cs.ubc.ca/~arnaud/doucet\\_johansen\\_tutorialPF.pdf](http://www.cs.ubc.ca/~arnaud/doucet_johansen_tutorialPF.pdf)

[10] KOMÁREK, A. *Hrátky s R* [online]. [cit. 2012-11-07]. Dostupné z: [http://www.karlin.mff.cuni.cz/~komarek/vyuka/2011\\_12/introR/introR-2011.pdf](http://www.karlin.mff.cuni.cz/~komarek/vyuka/2011_12/introR/introR-2011.pdf)

[11] LOPES, H. F. *Markov Chain Monte Carlo and Sequential Monte Carlo Methods in Stochastic Volatility Models* [online]. [cit. 2012-11-10]. Dostupné z: <http://faculty.chicagobooth.edu/hedibert.lopes/teaching/UPCcourse/UPCcourse-handouts.pdf>

[12] NĚMEC, D. *Bayesiánská analýza* [online]. [cit. 2012-09-24]. Dostupné z: [http://is.muni.cz/el/1456/podzim2011/MPE\\_BAAN/um/27201830/BayesianskaAnalyza.pdf](http://is.muni.cz/el/1456/podzim2011/MPE_BAAN/um/27201830/BayesianskaAnalyza.pdf)

[13] PAWLAS, Z. *MCMC metody (STP 139)* [online]. [cit. 2012-05-06]. Dostupné z: <http://www.karlin.mff.cuni.cz/~pawlas/2007/STP139/mcmc.pdf>

[14] STORDAL, A. S. *Sequential Monte Carlo Method for Bayesian Filtering* [online]. [cit. 2012-11-28]. Dostupné z: [http://www.math.u-bordeaux1.fr/~pdelmora/Masterthesis\\_Stordal.pdf](http://www.math.u-bordeaux1.fr/~pdelmora/Masterthesis_Stordal.pdf)

[15] WALSH, B. *Markov Chain Monte Carlo and Gibbs Sampling* [online]. [cit. 2012-09-24]. Dostupné z: <http://web.mit.edu/~wingated/www/introductions/mcmc-gibbs-intro.pdf>