

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

Katedra aplikovanej matematiky a štatistiky



**Metódy a algoritmy pre modely
stochastickej volatility**

DIPLOMOVÁ PRÁCA

BRATISLAVA 2007

Emília Boháčková

Metódy a algoritmy pre modely stochastickej volatility

DIPLOMOVÁ PRÁCA

Diplomant: Emília Boháčková

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
KATEDRA APLIKOVANEJ MATEMATIKY A ŠTATISTIKY

Matematika
Ekonomická a finančná matematika

Vedúci diplomovej práce: doc. RNDr. Viktor Witkovský, CSc.

BRATISLAVA 2007

Čestne prehlasujem, že som túto diplomovú prácu vypracovala samostatne, len s pomocou nadobudnutých teoretických vedomostí, konzultácií a uvedenej literatúry.

V Bratislave, apríl 2007

Emília Boháčková

Ďakujem vedúcemu diplomovej práce doc. RNDr. Viktorovi Witkovskému, CSc. za jeho odborné vedenie, pripomienky, návrhy a za čas, ktorý mi venoval pri vypracovávaní diplomovej práce.

Abstrakt

Volatilita je kľúčovým faktorom pri určovaní hodnoty aktíva, a teda aj stanovovaní jeho ceny. K jej odhadovaniu existuje viacero prístupov, pričom prevažujú modely časovo závislej volatility, nakoľko lepšie zachytávajú realitu. Medzi ne patria aj modely stochasticej volatility (SV modely), v ktorých volatility popisuje nepozorovateľný náhodný proces. Funkcia vierohodnosti v týchto modeloch sa zväčša nedá analyticky vyjadriť, preto boli vyvinuté špeciálne metódy odhadovania pre SV modely. V diplomovej práci sa z nich zameriame na MCMC (Markov chain Monte Carlo) metódy a na algoritmy, ktoré sa v rámci nich používajú. Jednotlivé algoritmy sú vhodné na odhadovanie rôznych špecifikácií SV modelov, čo sa budeme snažiť využiť pri aplikácii na reálne dáta.

Kľúčové slová: MCMC metódy, stochastická volatilita, Monte Carlo metódy, Markovove reťazce

Abstract

Volatility plays the key role in determining the value of assets and it is crucial in asset pricing. There are several approaches to volatility estimation but the most prevailing ones are models of time-varying volatility because they can better capture the reality. Stochastic volatility models (SV models), that belong to this class of models, describe volatility by the stochastic latent process. Their likelihood function is mostly intractable therefore special estimation methods has been proposed for treating SV models. In our thesis, we focus on MCMC (Markov chain Monte Carlo) methods and on algorithms they are using. Particular algorithms are appropriate to estimating different specifications of SV models and we try to utilize this fact when applying these models on real data.

Keyword: MCMC methods, stochastic volatility, Monte Carlo methods, Markov chains

Obsah

Zoznam skratiek a symbolov	3
Úvod	5
1 Modelovanie volatility	6
1.1 Základné typy modelov	6
1.2 Typické vlastnosti volatility	7
2 SV modely	9
2.1 Diskrétné SV modely	9
2.2 Spojité SV modely	10
2.3 Voľba špecifikácie SV modelu	11
2.4 Konkrétne príklady SV modelov	12
3 Odhadovanie SV modelov	13
3.1 Základný prehľad metód odhadovania pre SV modely	13
3.2 Bayesovská štatistika	14
3.2.1 Použitie bayesovského prístupu v SV modeloch	17
3.3 Monte Carlo výpočet integrálov	19
3.3.1 Použitie MC integrovania v SV modeloch	22
3.4 Bayesovské simulačné techniky	22
3.4.1 Importance Sampling	23
3.4.2 Sampling/Importance Resampling	23
3.4.3 Acceptance-Rejection Sampling	24
3.4.4 Adaptive Rejection Sampling	25
3.5 Markovove reťazce	28
3.5.1 Markovove reťazce s konečnou množinou stavov	28
3.5.2 Markovove reťazce so všeobecnou množinou stavov	32
3.5.3 Konvergenčná teória Markovových reťazcov	36
3.6 Úvod do MCMC metód	37
4 MCMC metódy	39
4.1 Použitie MCMC metód v SV modeloch	39
4.2 Metropolis-Hastingsov algoritmus	40
4.2.1 M-H algoritmus s náhodnou prechádzkou	42
4.2.2 Nezávislý M-H algoritmus	42
4.3 Gibbsov algoritmus	43

4.4	Rozšírenia základných algoritmov	44
4.4.1	Hybridné MCMC metódy	45
4.4.2	Algoritmus s oddialeným zamietnutím	45
4.4.3	Adaptive-Rejection Metropolis Sampling	47
5	Analýza výstupu MCMC metód	49
5.1	Odhad MCSE	50
5.2	Odhad dĺžky burn-in a stanovenie konverencie	51
5.3	Odhad miery zmiešavania	52
6	Praktická aplikácia MCMC metód na odhadovanie SV modelov	53
6.1	Popis dát	54
6.2	Taylorov SV model	54
6.3	SGED SV model	58
	Záver	66
	Literatúra	67
	Príloha 1 - SGED rozdelenie	72
	Príloha 2 - Výsledky praktickej časti	74

Zoznam skratiek a symbolov

ARCH Autoregressive Conditional Heteroskedasticity

ARMS Adaptive Rejection Metropolis Sampling

ARS Adaptive Rejection Sampling

CLT Central Limit Theorem (centrálna limitná veta)

CUSUM Cumulative Sum (kumulatívna suma)

DAX Deutscher Aktienindex (Nemecký akciový index)

DRA Delaying Rejection Algorithm (algoritmus s oddialeným zamietnutím)

EPD Exponential Power Distribution (EPD rozdelenie)

GARCH Generalized Autoregressive Conditional Heteroskedasticity

GED Generalized Error Distribution (GED rozdelenie)

GMM Generalized Method of Moments (zovšeobecnená metóda momentov)

IS Importance Sampling

MC Monte Carlo

MCMC Markov Chain Monte Carlo

MCSE Monte Carlo Standard Error (štandardná chyba aproximácie pomocou MC metódy)

M-H Metropolis-Hastings

QML Quasi Maximum Likelihood (metóda kvázi maximálnej vierohodnosti)

RS Acceptance-Rejection Sampling

RW M-H Random Walk M-H (M-H algoritmus s náhodnou prechádzkou)

SAX Slovenský akciový index

SDR stochastická diferenciálna rovnica

SE Standard Error (štandardná chyba)

SGED Skew GED (zošikmené GED rozdelenie)

SIR Sampling/Importance Resampling

LLN Strong Law of Large Numbers (silný zákon veľkých čísel)

S&P500 Standard and Poor's 500

SV Stochastic Volatility (stochastická volatilita)

SV CJ SV with Correlated Jumps in Returns and Volatility (SV model s korelovanými skokmi vo výnosoch a volatilitě)

SV IJ SV with Independent Jumps in Returns and Volatility (SV model s nezávislými skokmi vo výnosoch a volatilitě)

SV J SV with Jump in Returns (SV model so skokmi vo výnosoch)

\mathcal{B} beta rozdelenie

\mathcal{IG} inverzné gama rozdelenie

\mathcal{N} normálne rozdelenie

\mathcal{U} rovnomerné rozdelenie

Úvod

Volatilita je kľúčovým faktorom vo financiách, hlavne v oblastiach investovania aktív, oceňovania derivátov a risk manažmente. Dlhé roky pretrvával predpoklad o jej konštantnosti, ale nakoniec bola táto hypotéza prekonaná a aj empiricky bolo overené, že volatilita sa vyvíja v čase. To viedlo k rozvoju modelov časovo závislej volatility, medzi inými aj modelov stochastickej volatility. SV modely pritom popisujú volatilitu pomocou nepozorovateľného náhodného procesu. Takýto spôsob vyjadrenia volatility však častokrát vedie k funkcii vierohodnosti, ktorá sa nedá analyticky vyjadriť. Z tohto dôvodu bolo na odhadovanie SV modelov navrhnutých množstvo špeciálnych metód, z ktorých každá má svoje výhody i nevýhody. Hlavne v poslednej dobe sú veľmi rozšírené MCMC metódy založené na bayesovskom prístupe a teórii Markovových reťazcov.

V tejto práci sa zaoberáme práve metódami odhadovania SV modelov, pričom sa zameriame na bayesovské metódy a z nich najmä na MCMC metódy. Popíšeme jednak ich teoretický základ, jednak algoritmy, ktoré využívajú. Keďže v praktickej časti práce aplikujeme niektoré MCMC metódy na odhad dvoch konkrétnych SV modelov, stručne popíšeme aj niektoré jednoduché metódy konvergenčnej diagnostiky MCMC algoritmov.

Práca je rozdelená na šesť častí. V prvej časti definujeme volatilitu, popíšeme jej typické vlastnosti a základné prístupy k jej modelovaniu. Obsahom druhej časti sú SV modely, ich rozdelenie a takisto aj niektoré konkrétne príklady. V tretej časti je zhrnuté teoretické východisko MCMC metód, teda bayesovská štatistika, Monte Carlo integrácia a teória Markovových reťazcov. Tiež sú tu predstavené aj iné simulačné techniky než MCMC metódy. V štvrtej časti sa zaoberáme konkrétnymi MCMC algoritmi, najprv základnými a potom aj ich rozšíreniami. V piatej časti sa nachádza stručný prehľad niektorých metód analýzy výstupu, ktoré patria do konvergenčnej diagnostiky MCMC metód. Nakoniec v šiestej časti aplikujeme niektoré z popísaných MCMC metód na odhadovanie dvoch SV modelov, pričom použijeme dáta akciových indexov.

1 Modelovanie volatility

Ceny aktív na finančnom trhu, napríklad akcií, dlhopisov, úrokových mier a výmenných kurzov, majú prirodzenú tendenciu meniť sa. Pokiaľ sú tieto zmeny veľké, hovoríme, že dané aktíva sú volatilné. Presnejšie, volatilita meria relatívnu zmenu ceny aktíva za daný čas. V ekonomickej teórii financií a finančnej matematike je často definovaná ako (okamžitá) štandardná odchýlka (σ) náhodného Wienerovho procesu, ktorým je modelovaný vývoj ceny daného aktíva. Keďže zachytáva mieru pohybov cien (fluktuácií) aktíva, pre investorov je aj mierou rizika a neistoty spojených s budúcou cenou aktíva: vyššia volatilita znamená vyššie riziko.

Volatilita je preto kľúčovým faktorom pri určovaní hodnoty aktíva, a teda aj stanovovaní jeho ceny. Z toho vyplýva jej dôležitosť tiež v oceňovaní derivátov finančných aktív, v rozhodovaní sa o spotrebe, resp. investíciách, o alokácii aktív, o výbere portfólia a v riadení rizika.

Existuje viacero prístupov k odhadovaniu volatility. Koncepty historickej a implikovanej volatility predpokladajú, že volatilita sa nemení v čase, je konštantná. Hoci správanie sa aktíva sa dá nimi popísať, tento predpoklad je realite veľmi vzdialený. Napríklad Black-Scholesov model nie je schopný vysvetliť známy fenomén volatility smile, ktorý tiež svedčí o tom, že volatilita sa s časom mení. Tento poznatok je dnes považovaný za samozrejímavý a bol experimentálne overený.

1.1 Základné typy modelov

Doteraz bolo navrhnutých množstvo ekonometrických modelov, ktoré sa snažia popísať problém časovo závislej volatility. Jedným z dvoch hlavných prístupov sú pritom GARCH modely, v ktorých je variancia (σ^2) vyjadrená ako funkcia predchádzajúcich pozorovaní ceny aktíva a predchádzajúcich hodnôt variancie. GARCH modely sú zovšeobecnením ARCH modelov, ktoré varianciu vyjadrujú jednoduchšie, len ako funkciu predchádzajúcich cien aktíva. Autormi tejto triedy modelov sú Engle (1982) a Bollerslev (1986). Dnes už existuje množstvo modifikácií základného GARCH modelu, ktoré sú stále široko využívané hlavne kvôli faktu, že na odhad parametrov týchto modelov sa dajú použiť štandardné metódy odhadovania (napr. metóda maximálnej vierohodnosti). Navyše tieto modely sú schopné zachytiť mnohé z typických vlastností volatility.

Druhú skupinu tvoria SV modely, založené na predpoklade, že volatilita sama sleduje latentný (nepozorovateľný) náhodný proces, závislý alebo nezávislý od procesu popisujúceho výnosy aktíva. Môže ísť o diskretný proces (napr. autoregresný) alebo spojitý proces, kedy je volatilita riešením stochastickej diferenciálnej rovnice.

V jednotlivých modeloch boli navrhnuté rôzne typy SDR, ktoré vychádzali jednak z intuície a z toho, aby sa s nimi dalo pomerne jednoducho analyticky narábať, jednak aby čo najlepšie zachytávali (zvolené) empiricky pozorované vlastnosti volatility. Oproti GARCH modelom je odhad parametrov SV modelov neporovnateľne ťažší, pretože nemusí existovať analyticky vyjadriteľná funkcia vierohodnosti. Na odhad týchto modelov bolo preto navrhnutých viacero ekonometrických metód, z ktorých mnohé sú výpočtovo náročné. Avšak vďaka rýchlemu rozvoju výpočtovej techniky hlavne v posledných rokoch sa SV modely stali atraktívnou alternatívou k doteraz používaným GARCH modelom.

1.2 Typické vlastnosti volatility

Rozhodovanie o výbere typu modelu, respektíve o špecifikácii modelu je ovplyvnené empiricky pozorovanými typickými vlastnosťami volatility (*stylized facts*). Schopnosť reprodukovať tieto vlastnosti je žiadúci znak dobrého modelu a práve nesplnenie tejto požiadavky je najčastejším kritériom na zamietnutie modelu. Zvyčajne však nie je snahou zachytiť a vysvetliť všetky typické vlastnosti volatility naraz jedným modelom, ale iba zvolené znaky, na ktoré sa autori zamerajú.

Typické vlastnosti volatility sú dnes už dobre preskúmané a popísané. Sú výsledkom viac ako polstoročného štúdia finančných časových radov zo štatistického hľadiska, pri ktorom bolo zistené, že zmeny cien aktív a finančných nástrojov na rozličných trhoch a za rozličné časové obdobia majú spoločné netriviálne vlastnosti. Týkajú sa samotných časových radov výnosov aktív, ich rozdelenia a tiež ich volatility.

Medzi typické vlastnosti volatility výnosov aktív patria:

- zhlukovanie (*clustering*)

Zhlukovanie volatility znamená, že veľké zmeny ceny aktíva sú často nasledované opäť veľkými zmenami a malé zmeny malými zmenami, pričom zmeny sa chápu v absolútnom vyjadrení (teda ľubovoľného znamienka). Kvantitatívne sa táto vlastnosť prejaví ako kladná, signifikantná a pomaly klesajúca autokorelačná funkcia absolútnej hodnoty výnosov alebo ich štvorcov, hoci výnosy samotné sú nekorelované. Zhlukovanie volatility je úzko späté s leptokurtickým rozdelením výnosov (tzn. rozdelenie má ťažké chvosty), čo zachytávajú aj ARCH modely.

- pretrvávanie (*persistence*)

O volatilitate hovoríme, že je perzistentná alebo tiež, že má vlastnosť dlhej pa-

mäte (*long memory*), ak jej okamžitá úroveň ovplyňuje očakávanú budúcu hodnotu volatility na viac období dopredu, resp. ak trvá istý čas, kým sa volatilita vráti na svoju dlhodobú úroveň. Táto vlastnosť priamo vyplýva zo zhlukovania volatility a zodpovedá pomalému klesaniu autokorelačnej funkcie absolútnej hodnoty výnosov alebo ich štvorcov. Bolo navrhnutých viacero veličín, pomocou ktorých sa meria perzistencia volatility: miera dopredného pretrvávania (*forward persistence*), miera kumulatívneho pretrvávania (*cumulative persistence*), polčas pretrvávania (*half-life of persistence*).

- priťahovanie sa k dlhodobej úrovni (*mean-reversion*)

Zo zhlukovania volatility tiež vyplýva, že obdobie vysokých hodnôt volatility bude nakoniec nasledované “normálnou” úrovňou volatility a takisto po období s nízkou volatilitou nastane nakoniec nárast volatility na “normálnu” úroveň. Vlastnosť mean-reversion teda znamená, že existuje akási dlhodobá ustálená (“normálna”) úroveň volatility, na ktorú by sa volatilita mala nakoniec po rôznych fluktuáciách vrátiť. Z toho dôvodu by všetky predpovede dostatočne vzdialených budúcich hodnôt volatility mali konvergovať k tejto úrovni bez ohľadu na čas, v ktorom boli robené, teda nemali byť ovplyvnené okamžitou hodnotou volatility. Nie je zatiaľ jednoznačné, či ustálená úroveň volatility zostáva konštantná alebo sa môže zmeniť, napríklad následkom nejakej inštitucionálnej zmeny.

- efekt zo zadĺženia alebo pákový efekt (*leverage effect*)

Leverage efekt vyjadruje asymetrický dopad pozitívnych a negatívnych zmien ceny akcií na volatilitu, konkrétne negatívnu koreláciu volatility a výnosov akcií. Je to spôsobené tým, že poklesom ceny akcií vzrastie miera zadĺženosti vlastného imania (*debt-to-equity ratio*) firmy. Dochádza k zvyšovaniu podielu cudzieho kapitálu na celkovom kapitáli (*financial leverage*), čím sa stávajú akcie rizikovejšie, a teda volatilita výnosov vzrastie. Podľa novej teórie naopak správy o rastúcej volatilitě znížia dopyt po akciách kvôli averzii investorov voči riziku. Následný pokles hodnoty akcií spôsobí nárast volatility, čo bude zodpovedať pôvodným správam. Pre volatilitu úrokových mier nebola zatiaľ táto vlastnosť zistená.

- súbežné pohyby a prelievanie (*comovements and spillover*)

Viacere štúdie dokazujú, že zmeny ceny a volatility na jednom medzinárodnom akciovom trhu môžu mať dopad na ceny a volatility na iných medzinárodných

akciových trhoch. Ceny a volatility na rôznych trhoch môžu byť teda navzájom korelované.

- súvis s tokom informácií (*link with information arrivals*)

Podľa niektorých autorov sú výnosy aktív prepojené s tokom informácií. Následne aj volatilita úzko súvisí s objemom obchodu, so zavedením kvót, s udalosťami, ktoré sa dajú predpovedať (napr. oznámenie výšky dividend, uverejnenie makroekonomických ukazovateľov), s uzávierkami trhu a ďalšími relevantnými informáciami.

2 SV modely

V SV modeloch je s časom sa meniaci systém popísaný pomocou dvoch zložiek. Prvou je pozorovateľná premenná, ktorá predstavuje cenu, resp. výnos aktíva a druhou je nepozorovateľná (stavová) premenná, ktorá predstavuje volatilitu, resp. varianciu výnosu aktíva. Cieľom je na základe dát odhadnúť parametre modelu a takisto získať realizácie latentnej premennej, ktorá je pritom časovo závislým parametrom popisujúcim pozorovateľnú premennú.

2.1 Diskrétne SV modely

Diskrétne modely časovo závislej volatility sa dajú všeobecne parametricky vyjadriť ako

$$y_t | z_t \sim \mathcal{P}(\mu_t, \sigma_t^2),$$

kde y_t je výnos aktíva v čase t , z_t je jeho volatilita (resp. spojitá funkcia volatility) v čase t , \mathcal{P} je nejaké štandardné rozdelenie pravdepodobnosti, pričom je často položené $\mu_t = 0$, nakoľko nie je zámerom modelovať túto vlastnosť dát.

V SV modeloch je volatilita z_t chápaná ako nepozorovateľný parameter závislý od svojich predchádzajúcich hodnôt $z_{t-1}, z_{t-2}, \dots, z_{t-k}$, $k \in \mathbb{N}$, preto sa SV modely zaraďujú do triedy *parameter-driven* modelov. Oproti tomu ARCH modely patria medzi *observation-driven* modely, lebo volatilita z_t je vyjadrená ako funkcia predchádzajúcich hodnôt pozorovaných výnosov aktíva $y_{t-1}, y_{t-2}, \dots, y_{t-k}$, $k \in \mathbb{N}$. Ďalším rozdielom je, že v SV modeloch sú prítomné až dva druhy šokov, z ktorých jeden ovplyvňuje latentný parameter. Vďaka tomu, ak pripustíme nenulovú koreláciu medzi šokmi, je možné zachytiť leverage effect, ktorý je vyjadrený práve negatívnou koreláciou medzi týmito šokmi.

Diskrétne SV modely sa dajú pomerne jednoducho zovšeobecniť na viacrozmernej prípad a tiež existuje ich analogická spojitá reprezentácia.

2.2 Spojité SV modely

V spojitom prípade sú dáta považované za realizáciu náhodnej premennej Y_t s hustotou $f(y_t|z_t)$, kde Y_t predstavuje výnos aktíva v čase t a Z_t jeho volatilitu (resp. spojitú funkciu volatility) v čase t . Realizácie volatility Z_t sú generované z hustoty $f(z_t|\theta_Z)$, kde θ_Z reprezentuje parametre latentného stochastického procesu, ktorý volatilita sleduje.

Vo všeobecnosti majú spojité SV modely tvar dvojrozmerného náhodného procesu, ktorý je kombináciou Itôovho a Poissonovho procesu (tzv. *jump-diffusion process* alebo *Poisson-Gaussian process*):

$$dX_t = \mu(X_t, \theta)dt + \sigma(X_t, \theta)dW_t + \xi_t dN_t \quad , \quad (1)$$

kde $X_t = (Y_t, Z_t)'$, $\theta \in \Theta \subseteq \mathbb{R}$ je množina všetkých parametrov modelu, $\mu : \mathbb{R} \times \Theta \rightarrow \mathbb{R}^2$ je funkcia určujúca drift (opäť pre výnos aktíva je často predpoklad nulového driftu), $\sigma : \mathbb{R} \times \Theta \rightarrow \mathbb{R}^{2 \times 2}$ je funkcia definujúca volatility (výnosu aktíva a volatility), $W_t = (W_t^Y, W_t^Z)'$ je dvojrozmerný Brownov pohyb s kovariančnou maticou $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $\xi_t = (\xi_t^Y, \xi_t^Z)'$ sú veľkosti skokov výnosu aktíva a volatility a $N_t = (N_t^Y, N_t^Z)$ je dvojrozmerný Poissonov proces s intenzitami λ_Y a λ_Z .

Pomocou skokov sa niektoré modely snažia zachytiť prudké nečakané zmeny vo vývoji výnosov aktíva a jeho volatility. Do modelu sú zahrnuté, ak uvažujeme aspoň jednu z intenzít Poissonovho procesu rôznu od nuly. Podľa hodnôt intenzít delíme modely so skokmi na:

- jednoduché SV modely (bez skokov): $\lambda_Y = 0, \lambda_Z = 0$
- modely so skokmi vo výnosoch (SVJ modely): $\lambda_Y > 0, \lambda_Z = 0$
- modely so skokmi vo výnosoch aj volatilitě, ktoré sú nezávislé (SVIJ modely): $\lambda_Y > 0, \lambda_Z > 0$
- modely so skokmi vo výnosoch aj volatilitě, pričom skoky vo výnosoch sú podmienené skokmi vo volatilitě (SVCJ modely): $\lambda_Y > 0, \lambda_Z > 0$

Všeobecný spojitý SV model sa dá rozložiť na dve stochastické diferenciálne rovnice, pričom výnos aktíva sa riadi podľa jednej z nich, kým volatilita je riešením druhej SDR. Pokiaľ za aktívum vezmeme akciu, jej výnos je popísaný parciálnou diferenciálnou rovnicou z Black-Scholesovho modelu

$$dS_t = \mu S_t dt + \sigma_t S_t dW_t$$

a špecifikácia SV modelu potom závisí od výberu SDR pre volatilitu σ_t .

Spojité SV model nie je možné priamo odhadovať, takisto realizácie výnosov aktíva poznáme len v izolovaných časových okamžikoch, preto je nutné model najprv diskretizovať. Ak Δ označíme rozdiel časov, v ktorých pozorujeme výnosy aktíva, využitím Eulerovej diskretizácie z (1) dostaneme

$$X_{t+\Delta} = X_t + \mu(X_t, \theta)\Delta + \sigma(X_t, \theta)(W_{t+\Delta} - W_t) + \xi_{t+\Delta}J_{t+\Delta} \quad , \quad (2)$$

pričom Poissonov proces sme aproximovali náhodnou premennou $\xi_t = (\xi_t^Y, \xi_t^Z)$ s Bernoulliho rozdelením, ktorá nadobúda hodnotu 1 s pravdepodobnosťou $\lambda t\Delta$ (nastáva skok) a hodnotu 0 s pravdepodobnosťou $(1 - \lambda)t\Delta$ (nenastáva skok).

Z rovnice (2) vyplýva podmienené rozdelenie výnosov aktíva

$$X_{t+\Delta}|X_t, J_{t+\Delta}, \xi_{t+\Delta} \approx N(X_t + \mu_t\Delta + J_{t+\Delta}\xi_{t+\Delta}, \sigma_t^2\Delta) \quad ,$$

ktoré je kvôli diskretizácii tiež len približné. Je dôležité, aby táto aproximácia nespôsobila významné chyby v odhadoch parametrov modelu. Tieto možné odchýlky však klesajú s rastúcou frekvenciou dát, pričom denné dáta by mali byť už vhodné (dostatočne “husté”) na použitie diskretizovaného SV modelu.

2.3 Volba špecifikácie SV modelu

Doteraz bolo navrhnutých už množstvo rôznych SV modelov, ktoré sa líšia

- vo výbere latentnej premennej: volatilita výnosov σ , variancia výnosov $v = \sigma^2$, log-volatilita výnosov $\log(\sigma)$, log-variancia $\log(v)$ (možnosť jednoduchšej manipulácie s modelom),
- vo výbere funkcií pre drift a volatilitu spojitého procesu $\mu(X_t, \theta)$ a $\sigma(X_t, \theta)$, resp. funkcie popisujúcej závislosť volatility na jej minulých hodnotách v diskretnom prípade (možnosť zachytenia mean-reversion, leverage efektu),
- vo výbere rozdelenia šokov: normálne, t-rozdelenie, (zošikmené) zovšeobecnené exponenciálne (možnosť vysvetlenia empirického leptokurtického rozdelenia výnosov),
- v zahrnutí skokov do modelu (možnosť vysvetlenia veľmi veľkých zmien volatility výnosov aktíva),
- v zahrnutí ďalších vysvetľujúcich premenných (možnosť vysvetlenia ďalších typických vlastností volatility, napr. pretrvávania).

Výber modelu je ovplyvnený nielen zvolenými vlastnosťami volatility, ktoré chceme pomocou neho vysvetliť, ale aj štruktúrou a pôvodom dát. Pritom vždy neplatí, že zložitejší model poskytne lepšie výsledky, prípadne by museli byť vyvážené neúmerne náročnejším odhadovaním parametrov modelu a realizácií volatility.

2.4 Konkrétne príklady SV modelov

Základným a asi najpoužívanejším typom diskrétného SV modelu je model v tvare

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \log(\sigma_{t+1}^2) &= \mu + \phi(\log(\sigma_t^2) - \mu) + \sigma_\eta \eta_t \quad , \end{aligned} \quad (3)$$

kde y_t je výnos aktíva, σ_t je volatilita výnosu, $\epsilon_t, \eta_t \sim IID(0, 1)$ a $cor(\epsilon_t, \eta_t) = \rho$. Log-variancia je popísaná AR(1) procesom, vďaka čomu tento model pripomína GARCH model. Často sa používa aj ekvivalený zápis modelu s vyžitím označenia $h_t = \log(\sigma_t^2)$

$$\begin{aligned} y_t &= \exp(0.5h_t)\epsilon_t \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t. \end{aligned} \quad (4)$$

Ak v (3), resp. (4) predpokladáme, že šoky sú normálne rozdelené a nezávislé, teda $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$ a $\rho = 0$, ide o Taylorov model, pre normalitu η_t nazývaný tiež lognormálny SV model. Výhodou tohto modelu je jeho jednoduchosť, preto sa často používa ako základný model na porovnanie s inými zložitejšími špecifikáciami.

Spomedzi spojitých SV modelov je veľmi populárny Hestonov model, ktorý má tvar

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{V_t} S_t dW_t^{(1)} \\ dV_t &= \kappa(\theta - V_t) dt + \sigma_V \sqrt{V_t} dW_t^{(2)} \quad , \end{aligned} \quad (5)$$

kde S_t je cena akcie, V_t je variancia ceny akcie a $W_t^{(1)}, W_t^{(2)}$ sú Brownove pohyby s koreláciou ρ (tzn. $dW_t^{(1)} dW_t^{(2)} = \rho dt$). Variancia sleduje mean-reverting proces, ktorý pôvodne použili Cox, Ingersoll a Ross (1985) na modelovanie krátkodobých úrokových mier. Obľúbenosť tohto modelu vyplýva z jeho dvoch hlavných vlastností: pripúšťa závislosť medzi výnosmi akcie a volatilitou (dá sa ním teda zachytiť leverage efekt) a vyplýva z neho čiastočne analytická rovnica na oceňovanie opcí.

V rámci modelov so skokmi SVCJ model zahŕňa pre isté hodnoty parametrov aj jednoduchý SV model a SVJ model. Dá sa zapísať nasledovne:

$$\begin{aligned} d \log S_t &= \mu dt + \sqrt{V_t} dW_t^Y + \xi^Y dN_t \\ dV_t &= \kappa(\theta - V_t) dt + \sigma_V \sqrt{V_t} dW_t^V + \xi^V dN_t \quad , \end{aligned} \quad (6)$$

kde S_t je cena akcie, V_t je jej variancia, W_t^Y, W_t^V sú Brownove pohyby s koreláciou ρ , N_t je Poissonov proces s intenzitou λ a ξ^Y, ξ^V sú veľkosti skokov, pričom

$\xi^V \sim \exp(\mu^V)$ a $\xi^Y | \xi^V \sim \mathcal{N}(\mu^Y + \rho^J \xi^V, (\sigma^Y)^2)$. Takto špecifikovaný model použili Asgharian a Bengtsson (2006).

Zaujímavý SV model navrhli Cappuccio, Lubian a Raggi (2004), ktorí vychádzajú z Taylorovho SV modelu, ale predpokladajú, že šoky v rovnici pre výnosy akcie majú SGED rozdelenie (pozri Prílohu 1):

$$\begin{aligned} y_t &= \beta \exp(0.5h_t)\epsilon_t \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t \quad , \end{aligned} \tag{7}$$

kde y_t sú výnosy akcie, $h_t = \log(\sigma_t^2)$ je log-variancia výnosov akcie, $\epsilon_t \sim SGED(\lambda, \nu)$, pričom $E(\epsilon_t) = 0$ a $Var(\epsilon_t) = 1$, a $\eta_t \sim \mathcal{N}(0, 1)$. Autori tento model pomenovali Skew-GED SV model a snažili sa ním o zachytenie asymetrie a ťažkých chvostov podmieneného rozdelenia výnosov akcie.

3 Odhadovanie SV modelov

Podobne ako u väčšiny nelineárnych modelov s latentnými premennými ani v SV modeloch nie je možné explicitne vyjadriť funkciu vierohodnosti, čo znemožňuje použitie klasických odhadovacích metód. Ďalej je nutné si uvedomiť, že nejde len o odhady parametrov modelu. Zväčša je dôležitejšie získať odhad práve latentnej premennej, čiže realizácie volatility, prípadne aj jej predikcie. Pritom dimenzia tejto nepozorovanej premennej je veľmi veľká, kvôli čomu aj štandardné simulačné metódy zlyhávajú. Preto boli SV modely v porovnaní s (G)ARCH modelmi dlho nezaujímavé. V posledných rokoch však došlo k výraznému rozvoju metód odhadovania vhodných pre SV modely, často spojených s použitím už dostupnej rýchlej výpočtovej techniky.

3.1 Základný prehľad metód odhadovania pre SV modely

Metódy štatistickej inferencie pre SV modely je možné rozdeliť do dvoch základných skupín. Ide o metódy využívajúce

- štatistické vlastnosti dát (momenty rozdelenia, funkcia vierohodnosti)
- simulačné techniky.

Do prvej triedy metód sa zaraďuje GMM a QML. GMM je pomerne jednoduchá metóda odhadovania vychádzajúca z porovnania teoretických a empirických momentov výnosov. Takto sa dajú získať odhady parametrov SV modelu, nie však realizácie

volatility. Ďalšou nevýhodou tejto metódy je fakt, že získané odhady nie sú invariantné vzhľadom na špecifikáciu modelu. V metóde QML sa zasa uvažuje o stavovej reprezentácii (*state-space representation*) SV modelu, kde sa využíva Kalmanov filter na získanie jednak kvázi funkcie vierohodnosti, a teda odhadov parametrov SV modelu, jednak odhadov latentnej volatility. Ide zároveň o vyhladzovanie (*smoothing*) a filtrovanie (*filtering*), teda výstupom sú aj predikcie volatility na jeden časový krok dopredu. Čo sa týka realizácie, QML je pomerne jednoduchá a dá sa rozšíriť aj na viacrozmerné SV modely. Podobne ako GMM, je však neefektívna, pretože využíva len kvázi funkciu vierohodnosti.

Metódy patriace do druhej skupiny sa ďalej členia podľa toho, či vychádzajú z klasického (frekventistického) prístupu alebo z bayesovského prístupu k pravdepodobnosti. V rámci klasických simulačných metód ide o metódy založené na pomocných modeloch (Simulated Method of Moments, Efficient Method of Moments a Indirect Inference). Ich základnou myšlienkou je nájsť také hodnoty parametrov, pre ktoré nasimulované dáta zodpovedajú pozorovaným dátam v tom zmysle, že obe skupiny dát budú dávať tie isté odhady v pomocnom modeli. Nástrojom je teda zvolený pomocný model, jednoduchý na odhadovanie, pričom je využitá jednoduchosť simulácií v SV modeli. Tieto metódy sú dosť výpočtovo náročné a navyše ich výstupom nie sú realizácie volatility. Bayesovské simulačné metódy slúžia ako prostriedok na použitie MC metód pri numerickom výpočte integrálov, ktorými je definovaná či už funkcia vierohodnosti, alebo aposteriórna združená hustota parametrov a volatility (pozri časti 3.2 a 3.3). Patria k nim Importance Sampling, Sampling/Importance Resampling, Acceptance-Rejection Sampling (pozri časť 3.4) a Markov Chain Monte Carlo metódy, ktoré podrobnejšie popíšeme v častiach 3.6 a 4.

3.2 Bayesovská štatistika

Rozdiel medzi klasickým (frekventistickým) a bayesovským prístupom k štatistike je v úplnej zmene uchopenia problému odhadovania parametrov nejakého štatistického (ekonometrického) modelu. V klasickej štatistike sa predpokladá, že poznáme správnu štruktúru (špecifikáciu) modelu a že chceme odhadnúť skutočnú hodnotu parametrov tohto modelu. Pritom je dôležitý práve pojem “skutočná hodnota parametrov”. Klasická štatistika totiž vychádza z predpokladu, že istá skutočná hodnota parametrov existuje, iba je neznáma. Tvrdenie, že nejaký parameter leží s pravdepodobnosťou α medzi hodnotami c_1 a c_2 , nemá zmysel. Namiesto toho treba konštatovať, že interval $[c_1, c_2]$ s pravdepodobnosťou α obsahuje skutočnú hodnotu parametra. Dáta a štatistiky na nich založené (ako napr. hranice intervalu

spoľahlivosti) sú považované za náhodné, zatiaľ čo parametre sú fixné.

Z bayesovského pohľadu neexistuje nič také, ako skutočná hodnota parametrov θ . Všetko čo máme, sú dáta, z praktického hľadiska chápané ako fixné hodnoty. Ďalej môžeme mať nejakú apriórnu predstavu o očakávaných hodnotách parametrov, ktorá sa dá zhrnúť do apriórnej hustoty parametrov $p(\theta)$ (*prior density*). Jej parametre sa nazývajú hyperparametre. Aposteriórna hustota parametrov $p(\theta|\text{dáta})$ (*posterior density*) je potom tvorená kombináciou apriórnej hustoty parametrov a funkcie vierohodnosti pozorovaných dát $p(\text{dáta}|\theta)$ (*likelihood function*), čo sa dá vyjadriť

$$p(\theta|\text{dáta}) = \frac{p(\text{dáta}|\theta)p(\theta)}{\int p(\text{dáta}|\theta)p(\theta)d\theta} \propto p(\text{dáta}|\theta)p(\theta).$$

Táto aposteriórna hustota je jediný kvantitatívny a pravdepodobnostný opis toho, čo vieme o parametroch modelu. Postačuje však na účely analýzy parametrov, oproti klasickému prístupu sa dá tiež použiť na ďalšiu analýzu celého modelu (napr. v teórii rozhodovania) a na mnohé rozšírenia (napr. optimálny hedging portfólia pri zachovaní neistoty v parametroch modelu).

Kritickou a často aj frekventistami kritizovanou časťou bayesovskej analýzy nejakého modelu je voľba apriórnej hustoty parametrov. Jej výber nie je jednoznačný, ale závisí od presvedčenia štatistika, ktorý analýzu vykonáva, o dôležitosti informácie obsiahnutej v dátach. Ak dáta (čiže funkcia vierohodnosti) obsahujú podstatnú informáciu, aposteriórna hustota parametrov bude závisieť od apriórnej hustoty len v malej miere, bude teda relatívne stabilná vzhľadom na voľbu apriórnej hustoty parametrov. To znamená, že platí istá asymptotická vlastnosť bayesovského prístupu: dostatočné množstvo dát (vtedy zahŕňajú dostatočnú informáciu) dominuje nad ľubovoľnou apriórnu hustotou parametrov (okrem patologických príkladov, napr. nevlastná apriórna hustota).

Typ apriórneho rozdelenia parametrov sa zvyčajne volí tak, aby uľahčil odvodenie aposteriórnej hustoty. Často sa používajú konjugované apriórne rozdelenia, ktoré majú vlastnosť, že pre danú funkciu vierohodnosti dávajú aposteriórne rozdelenie, ktoré patrí do rovnakej triedy rozdelení ako apriórne rozdelenie.

Konjugované apriórne hustoty sú známe pre množstvo používaných funkcií vierohodnosti (pozri Tabuľka 1), ale sú obmedzené v tvaroch a niekedy ich nemožno použiť v konkrétnych modeloch. Flexibilita konjugovaných apriórnych hustôt sa dá čiastočne zlepšiť použitím ich zmesi (*mixture*), čo vedie k aposteriórnej kombinácii upravených konjugovaných zložiek.

Pri voľbe apriórnej hustoty parametrov, ktorá má vyjadrovať naše presvedčenie o ich možných hodnotách, je dôležitejšia stredná hodnota a variancia (resp. jej prevrátená hodnota (*precision*)) ako tvar rozdelenia. Veľmi často sa používa nein-

funkcia vierohodnosti (príslušné rozdelenie)	apriórne rozdelenie
binomické	beta
multinomické	Dirichletovo
Poissonovo	gama
normálne	
μ neznáma, σ^2 známa	normálne
μ známa, σ^2 neznáma	inverzné chí-kvadrát
viacrozmerné normálne	
μ neznáma, \mathbf{V} známa	viacrozmerné normálne
μ známa, \mathbf{V} neznáma	inverzné Wishartovo

Tabuľka 1: Konjugované apriórne rozdelenia pre bežne používané funkcie vierohodnosti (Walsh (2002))

formatívna apriórna hustota (*flat, diffuse prior*) tvaru

$$p(\theta) = \begin{cases} k = \frac{1}{b-a} & a \leq \theta \leq b \\ 0 & \text{inak} \end{cases}.$$

Ide vlastne o rovnomerné rozdelenie na intervale (a, b) vyjadrujúce predpoklad, že hodnota parametra bude niekde medzi a a b , ale pritom žiadnu z hodnôt medzi nimi nenadobudne s väčšou pravdepodobnosťou než ostatné. Aposteriórna hustota v tomto prípade nadobudne tvar

$$p(\theta|\text{dáta}) = C p(\text{dáta}|\theta), \quad C = \text{konšt.}$$

teda

$$p(\theta|\text{dáta}) \propto p(\text{dáta}|\theta).$$

V prípade, že $a = -\infty$ alebo $b = \infty$, sa takáto apriórna hustota nazýva nevlastná (*improper prior*), pretože po zintegrovaní na \mathbb{R} nedáva 1, ale ∞ . Nevlastné apriórne hustoty môžu spôsobiť problémy v inferencii (hlavne pri testovaní hypotéz a rozhodovaní o výbere modelu), preto ich treba používať opatrne.

Neinformatívne apriórne hustoty sa volia hlavne preto, aby sa minimalizoval ich potenciálny vplyv na aposteriórnu hustotu. Samozrejme, ak máme odôvodnené presvedčenie, že parametre s vyššou pravdepodobnosťou nadobudnú konkrétne hodnoty, môžeme použiť aj (viac) informatívne apriórne hustoty. Ako bolo spomenuté už vyššie pri dostatočnom množstve dát by informácia zahrnutá v dátach mala prevažovať.

Napriek výhodám bayesovského prístupu zaujímavým a využívaným sa stal až v poslednom desaťročí 20. storočia. Dôvodov pomalej adaptácie tohto alternatívneho štatistického prístupu je viacero. Prvým a najdôležitejším je zmena chápania podstaty problému potrebná na prechod od frekvenčnej k bayesovskej štatistike. Po storočiach považovania parametrov za neznáme pevné kvantily Bayes prezentoval parameter ako náhodnú premennú s príslušným rozdelením, čo pre väčšinu štatistikov a ekonometrov nebolo jednoduché prijať. Ďalším problémom bola výpočtová zložitosť spojená s odvádzaním aposteriorných rozdelení. Iba v špeciálnych prípadoch je aposteriórna hustota vyjadriteľná analyticky. Pretože ide o viacrozmerné integrály, najčastejšie je nutné počítať ich numericky. Tieto ťažkosti boli odstránené aplikáciou MC metód na výpočet integrálov. Simulačné metódy slúžiace na aproximáciu integrálov boli už skôr využívané vo fyzike, v ekonometrii došlo k ich rozvoju až v 70. a 80. rokoch 20. storočia. V posledných rokoch vďaka výkonným počítačom zažívajú ďalší prudký rozvoj, pričom výskum sa nezameriava len na vývoj nových metód, ale aj na zlepšovanie, rozvoj a možnosti novej aplikácie tých starších.

3.2.1 Použitie bayesovského prístupu v SV modeloch

Pre využitie bayesovského prístupu v SV modeloch je nutné rozlišovať nasledujúce zložky modelu: funkciu vierohodnosti, hustotu rozdelenia stavovej premennej popisujúcu jej vývoj a hustotu rozdelenia parametrov. V modeli považujeme za danú množinu pozorovaní $Y = \{Y_t\}_{t=1}^T$ a snažíme sa charakterizovať parametre θ a stavovú premennú $X = \{X_t\}_{t=1}^T$, teda potrebujeme určiť aposteriórnu hustotu $p(\theta, X|Y)$. Podľa Bayesovej vety sa dá rozložiť

$$p(\theta, X|Y) \propto p(Y|X, \theta)p(X|\theta)p(\theta),$$

kde $p(Y|X, \theta)$ je funkcia vierohodnosti, $p(X|\theta)$ je hustota rozdelenia stavovej premennej daného špecifikáciou modelu a $p(\theta)$ je hustota apriórneho rozdelenia parametrov modelu zachytávajúceho naše presvedčenie alebo informácie (nevyplývajúce z dát) o parametroch (často napríklad kladnosť niektorého parametra vyplývajúca z jeho ekonomickej interpretácie).

Hustota $p(Y|X, \theta)$, ktorú sme nazvali funkciou vierohodnosti, je v skutočnosti rozšírená funkcia vierohodnosti (*(data-)augmented likelihood*), pretože vyjadruje pravdepodobnosti podmienené nielen na parametroch, ale aj na stavovej premennej. Funkcia vierohodnosti, ako ju poznáme z klasickej štatistiky, je vlastne marginálna funkcia vierohodnosti, ktorú dostaneme zintegrovaním rozšírenej funkcie vierohod-

nosti cez stavovú premennú

$$p(Y|\theta) = \int p(Y, X|\theta)dX = \int p(Y|X, \theta)p(X|\theta)dX.$$

Vo väčšine SV modelov vo financiách sa $p(Y|\theta)$ nedá analyticky vyjadriť, čo znemožňuje použitie metódy maximálnej vierohodnosti. Oproti tomu bayesovské simulačné metódy narábajú s rozšírenou funkciou vierohodnosti $p(Y|X, \theta)$, ktorá je vďaka podmieňovaniu aj stavovou premennou zvyčajne vyjadriteľná v uzavretom tvare.

Iným užitočným rozkladom aposteriórnej hustoty je

$$p(\theta, X|Y) \propto p(\theta|X, Y)p(X|\theta, Y). \quad (8)$$

Vychádza z Hammersley-Cliffordovej vety, podľa ktorej združené rozdelenie pravdepodobnosti sa dá jednoznačne popísať množinou jej úplných podmienených rozdelení. Výhoda tohto rozkladu spočíva v tom, že je jednoduchšie charakterizovať a simulovať z dvoch podmienených rozdelení ako zo združeného rozdelenia vyššej dimenzie. Aj tieto podmienené rozdelenia sa dajú ďalej rozložiť na ich podmienené rozdelenia, čo môže opäť zjednodušiť simulácie. Ak teda predpokladáme, že vektor parametrov θ sa dá rozložiť na k zložiek (jedno- alebo viacrozmerých), t.j. $\theta = (\theta_1, \dots, \theta_k)$, potom rozdelenie prislúchajúce k $p(\theta|X, Y)$ je určené množinou podmienených rozdelení s hustotami

$$\begin{aligned} & p(\theta_1|\theta_2, \theta_3, \dots, \theta_k, X, Y) \\ & p(\theta_2|\theta_1, \theta_3, \dots, \theta_k, X, Y) \\ & \vdots \\ & p(\theta_k|\theta_1, \theta_2, \dots, \theta_{k-1}, X, Y). \end{aligned}$$

Podobne združené rozdelenie stavovej premennej $p(X|\theta, Y)$ je charakterizované množinou podmienených rozdelení $p(X_t|X_{-t}, \theta, Y)$ pre $t = 1, \dots, T$, kde X_{-t} predstavuje prvky X okrem X_t .

Čo sa týka voľby apriórnych rozdelení parametrov s príslušnými hustotami $p(\theta)$, ak je to možné, najčastejšie sa používajú konjugované rozdelenia, aby sa zjednodušil výpočet aposteriórnych hustôt. V rámci SV modelov si treba dať pozor na používanie neinformatívnych apriórnych rozdelení, pretože nevlastné apriórne hustoty by mohli viesť k výpočtovým problémom a spôsobiť, že aposteriórna hustota by bola takisto nevlastná.

Na získanie odhadov jednotlivých parametrov modelu sa využívajú marginálne aposteriórne hustoty

$$p(\theta_i|Y) = \int p(\theta, X|Y)dX d\theta_i = \int p(\theta_i, \theta_{(-i)}, X|Y)dX d\theta_i,$$

kde θ_i je i -ty prvok vektora parametrov θ a $\theta_{(-i)}$ označuje zvyšné parametre modelu. Za odhad parametra θ_i sa štandardne považuje priemer (prípadne medián) tejto marginálnej aposteriórnej hustoty.

Odhadovanie stavovej premennej je zložitejšie, rozlišujeme tri druhy odhadov. Každý je založený na príslušnej marginálnej aposteriórnej hustote, ktoré sa líšia v množstve informácií, ktorým podmieňujeme:

- vyhladzovanie (*smoothing*) $p(X_t|Y_1, \dots, Y_T), t = 1, \dots, T$
- filtrovanie (*filtering*) $p(X_t|Y_1, \dots, Y_t), t = 1, \dots, T$
- predikcie (*forecasting*) $p(X_{t+1}|Y_1, \dots, Y_t), t = 1, \dots, T$.

Odhadovanie realizácií stavovej premennej na základe všetkých dát, teda vyhladzovanie, je statický problém, rieši sa jednorázovo s použitím všetkých dát Y

$$p(X_t|Y) = \int p(\theta, X|Y) d\theta dX_{-t},$$

kde X_{-t} predstavuje všetky realizácie stavovej premennej okrem realizácie v čase t . Odhad realizácie volatility X_t v čase t bude opäť daný priemerom (resp. mediánom) marginálnej aposteriórnej hustoty $p(X_t|Y)$.

Zvyšné dva typy odhadovaní sú sekvenčné problémy, kedy odhad stavovej premennej (jej realizácie v čase t) je založený len na súbežných dátach (do času t). Metódy na riešenie sekvenčného odhadovania sa nazývajú filtračné, patrí medzi ne napr. *particle filter* a *practical filter*.

Na získanie odhadov parametrov modelu a realizácií stavovej premennej, je teda nutné poznať aposteriórnu združenú hustotu $p(\theta, X|Y)$ a vedieť ju integrovať. Pretože jej tvar býva zložitý (čo platí vo väčšine modelov) a navyše integrály sú viac-rozmerné, na ich výpočet musíme použiť MC metódu integrovania.

3.3 Monte Carlo výpočet integrálov

Monte Carlo metódy predstavujú široko využívanú triedu algoritmov na simulovanie správania sa rozličných fyzikálnych a matematických systémov. Od ostatných simulačných metód sa odlišujú využívaním náhodných čísel, teda sú v istom zmysle stochastické. Kvôli veľkému počtu opakovaní výpočtov a krokov týchto algoritmov, sú to metódy viac-menej predurčené na realizáciu pomocou počítačov.

MC metódy sú pomenované po rovnomennom meste v Monaku, známom pre svoje kasína, kde sa v hazardných hrách (napr. v rulete) stále opakujú udalosti so známymi pravdepodobnosťami. Hoci poznáme viacero izolovaných a nerozvinutých aplikácií týchto metód už v skorších rokoch, moderné využitie MC metód sa datuje

od 40. rokov 20. storočia, kedy boli použité počas vývoja atómovej bomby (Ulam, Metropolis, von Neumann, Fermi).

Monte Carlo metódy sú založené na simulovaní celého systému (všetkých neznámych parametrov) v čase. Ide pritom o veľké množstvo simulácií, ktoré predstavujú možné budúce realizácie systému. Výsledkom nie sú jednotlivé hodnoty, ale pravdepodobnostné rozdelenie, na základe ktorého je potom možné získať odhady parametrov systému (napr. bodový odhad sa aproximuje výberovým priemerom simulácií).

Aby sme mohli použiť MC metódy, nemusíme poznať rovnice (obyčajné alebo diferenciálne) popisujúce systém. Stačí, aby bolo známe rozdelenie pravdepodobnosti (resp. jeho hustota) parametrov systému. Ďalšou veľkou výhodou MC metód je ich podstata - simulovanie realizácií systému. Teda na odhadovanie systému nepotrebuje pozorovať žiadne dáta ako výsledky experimentov, ktoré sú niekedy komplikované, časovo alebo finančne náročné, poprípade aj nemožné (napr. systémy s nepozorovateľnými stavovými premennými). MC metódy podstatne rozšírili triedu modelov, ktoré sa dajú riešiť (odhadovať).

V matematike sa využívajú MC metódy na riešenie rôznych problémov, najčastejšou aplikáciou je však Monte Carlo výpočet integrálov (*MC integration*). Táto metóda je určená na numerický výpočet integrálov pomocou náhodných (resp. pseudonáhodných) čísel, teda ide o algoritmus na získanie približnej hodnoty konečných integrálov, zvyčajne aj viacrozmerných. Klasické numerické metódy vyčíslujú podintegrálnu funkciu na pravidelnej mriežke, metóda MC integrovania volí body, v ktorých je podintegrálna funkcia (resp. jej násobok) vyhodnocovaná, náhodne a nezávisle. Na to je potrebné vhodne rozložiť podintegrálnu funkciu $f(x)$ na súčin hustoty nejakého rozdelenia $p(x)$, ozn. cieľová hustota, a inej funkcie $g(x)$. Cieľovú hustotu následne aproximuje histogram simulácií zo zvoleného rozdelenia a hodnotu integrálu výberový priemer funkčných hodnôt (funkcie $g(x)$) týchto simulácií:

$$I = \int_{\Omega} f(x)dx = \int_{\Omega} g(x)p(x)dx = E_p [g(X)] \approx \frac{1}{N} \sum_{n=1}^N g(x_n) = \hat{I}_N, \quad (9)$$

kde I je integrál, ktorý chceme vypočítať a x_1, \dots, x_N je náhodný výber z rozdelenia s hustotou $p(x)$ na oblasti Ω ($\int_{\Omega} p(x)dx = 1$). Najjednoduchšou voľbou, no nie vždy najlepšou (z pohľadu veľkosti variancie odhadu), je rovnomerné rozdelenie na oblasti Ω . Pomocou rozkladu podintegrálnej funkcie sa dajú približne počítať nielen momenty zvoleného rozdelenia s hustotou $p(x)$, ale aj samotné pravdepodobnosti, ak za $g(x)$ zvolíme indikátor množiny.

Pretože \hat{I}_N je výberový priemer náhodnej premenej $g(X)$ s realizáciami $g(x_n)$, $n = 1, \dots, N$ a strednou hodnotou $E(g(X)) = I$, podľa zákona veľkých čísel s

počtom simulácií idúcim do nekonečna konverguje skoro všade k svojej strednej hodnote I

$$\hat{I}_N \xrightarrow{\text{s.v.}} I, \quad N \rightarrow \infty.$$

Navyše, ak $\text{Var}(g(X)) = \sigma^2$ je konečná, podľa centrálnej limitnej vety je \hat{I}_N asymptoticky normálne rozdelený so strednou hodnotou I a varianciou $\frac{\sigma^2}{N}$

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

Štandardná chyba takejto aproximácie integrálu (9) sa nazýva MCSE a má tvar

$$\text{MCSE} = \text{SE}(\hat{I}_N) = \frac{\sigma}{\sqrt{N}},$$

kde asymptotickú varianciu σ^2 z CLT môžeme odhadnúť ako výberovú varianciu náhodnej premennej $g(X)$

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{n=1}^N (g(x_n) - \hat{I}_N)^2.$$

Vo všeobecnosti sú výhodami MC integrovania

- konvergencia rádu $O(N^{-1/2})$ bez ohľadu na dimenziu integrálu a hladkosť podintegrálnej funkcie,
- jednoduchosť - algoritmy vykonávajú iba dva typy operácií (generovanie náhodných hodnôt, výpočet funkčných hodnôt),
- všeobecnosť - dá sa použiť aj na oblastiach, kde by klasické kvadratury boli problematické,
- lepšie zvládanie singularít podintegrálnych funkcií.

Tieto výhodné vlastnosti MC integrovania sú však viazané na takú hustotu $p(x)$, z ktorej generovanie náhodného výberu je štandardná a nenáročná úloha ¹. V mnohých prípadoch (pri využívaní MC integrovania v rozličných oblastiach) kvôli zložitému tvaru cieľovej hustoty nie sú možné priame simulácie. Vtedy treba siahnuť po niektorej zo simulačných techník založených na využití pomocných hustôt, z ktorých sa, naopak, dá simulovať ľahko. Príklady bayesovských simulačných techník sú uvedené v časti 3.4.

¹Pod výrazmi “simulovať z hustoty”, “generovať náhodný výber z hustoty” rozumieme robenie náhodného výberu z rozdelenia, ku ktorému daná hustota prislúcha. Podobne “náhodný výber z hustoty” znamená náhodný výber z rozdelenia s danou hustotou.

3.3.1 Použitie MC integrovania v SV modeloch

Vráťme sa teraz späť k SV modelom a označeniam z časti 3.2.1. Potom aplikáciou MC metódy výpočtu integrálov pre ľubovoľnú integrovateľnú funkciu $f(\theta, X)$ dostaneme

$$\begin{aligned}\hat{f}(\theta, X) &= E_p[f(\theta, X)|Y] = \int f(\theta, X)p(\theta, X|Y)dXd\theta \approx \\ &\approx \frac{1}{N} \sum_{n=1}^N f(\theta^{(n)}, X^{(n)}),\end{aligned}\tag{10}$$

kde $(\theta^{(n)}, X^{(n)})$, $n = 1, \dots, N$ je náhodný výber zo združenej aposteriórnej hustoty $p(\theta, X|Y)$.

Ak vhodne zvolíme $f(\theta, X) = \theta_i$, $i \in \{1, \dots, k\}$, z (10) dostaneme bodový odhad jednotlivých parametrov modelu

$$\hat{\theta}_i = E_p[\theta_i|Y] \approx \frac{1}{N} \sum_{n=1}^N \theta_i^{(n)} \quad i = 1, \dots, k$$

alebo pri $f(\theta, X) = X_t$, $t \in \{1, \dots, T\}$, (10) dáva odhad realizácií stavovej premennej v jednotlivých časoch t

$$\hat{X}_t = E_p[X_t|Y] \approx \frac{1}{N} \sum_{n=1}^N X_t^{(n)} \quad t = 1, \dots, T.$$

Jediné, čo k týmto odhadom potrebujeme, je teda náhodný výber zo združenej aposteriórnej hustoty. SV modely však patria medzi prípady, kedy priame simulácie kvôli tvaru $p(\theta, X|Y)$ nie sú možné, preto musíme využiť niektorú zo simulačných techník.

3.4 Bayesovské simulačné techniky

Tieto metódy sa používajú na generovanie náhodného výberu z hustôt, z ktorých nevieme simulovať priamo. Najčastejšie ide o aposteriórne združené hustoty potrebné na výpočet integrálov MC metódou, preto ich nazývame bayesovské simulačné techniky. Kvôli zjednodušeniu zápisu však nebudeme používať “bayesovské” označenia funkcií z časti 3.2, ale zachováme označenie cieľovej hustoty $p(x)$ ako v časti 3.3.

Spoločným znakom bayesovských simulačných techník je, že sa vyhnú simulovaniu z cieľovej hustoty tak, že náhodný výber urobia z pomocnej hustoty $q(x)$, ozn. inštrumentálna hustota (*proposal density*), z ktorej sa dá simulovať priamo. Následne tento výber nejakým spôsobom upravia, aby predstavoval výber z cieľovej hustoty.

3.4.1 Importance Sampling

V IS metóde volíme inštrumentálnu hustotu $q(x)$ tak, aby jej nosič zahrňal nosič cieľovej hustoty $p(x)$ a vhodne upravíme integrál I :

$$I = \int g(x)p(x)dx = \int g(x)\frac{p(x)}{q(x)}q(x)dx = \int g(x)w(x)q(x)dx = E_q [g(X)w(X)],$$

kde $w(x)$ je váhová funkcia (*importance weight function*) zachytávajúca vzťah cieľovej a inštrumentálnej hustoty, ktorý sa prejaví aj v odhade integrálu

$$I \approx \frac{1}{N} \sum_{n=1}^N g(x_n)w(x_n) = \hat{I}_N,$$

kde x_1, \dots, x_N je náhodný výber z rozdelenia s hustotou $q(x)$. Vyššie váhy, a teda aj význam v priemere, budú mať hodnoty z náhodného výberu, ktoré by sa vyskytovali menej často pri simulovaní z inštrumentálnej hustoty, no častejšie pri simulovaní z cieľovej hustoty. Odhad je konzistentný a nevychýlený s varianciou

$$Var(\hat{I}_N) = \frac{1}{N} Var [f(X)w(X)] = \frac{1}{N} \left(\int_{\Omega} f^2(x)w(x)dx - I^2 \right).$$

Kritériom na výber vhodnej inštrumentálnej hustoty je minimalizácia variancie odhadu. Preto, aby IS metóda bola efektívna, inštrumentálna hustota by mala tvarom čo najlepšie aproximovať cieľovú hustotu. V takom prípade budú váhy $w(x_i)$ blízke k 1 a variancia odhadu bude malá (variancia bude minimálna, ak inštrumentálna hustota bude proporcionálna k cieľovej hustote). Táto požiadavka je často nedosiahnuteľná, hlavne ak ide o chvosty cieľovej hustoty. Vždy si treba dať pozor na to, aby pomocná hustota mala ťažšie chvosty ako cieľová, inak môže váhová funkcia nadobúdať veľmi veľké hodnoty. Navyše stredná hodnota váhovej funkcie $E(w(X))$ musí byť konečná.

Importance Sampling metóda nie je výpočtovo náročná a netreba pri nej riešiť ani problémy s konvergenciou. Je však prakticky nepoužiteľná, ak ide o integrály väčších dimenzií.

3.4.2 Sampling/Importance Resampling

Rubin (1988) zaviedol SIR metódu, ktorá na generovanie potrebného výberu x_1, \dots, x_N využíva váhy z výberu získaného pomocou IS metódy. Postup pri metóde SIR je nasledovný:

1. vygenerujeme náhodný výber x_1^*, \dots, x_R^* z pomocnej hustoty $q(x)$ a vypočítame IS váhy $w_n = w(x_n^*) = \frac{p(x_n^*)}{q(x_n^*)}$ pre všetky $n = 1, \dots, R$

2. znovu urobíme náhodný výber x_1, \dots, x_M , $M \ll R$, ale z vektora hodnôt vygenerovaného v kroku 1 s príslušnými pravdepodobnosťami $p_{SIR}(x_n^*) = \frac{w_n}{\sum_{n=1}^R w_n}$, $n = 1, \dots, R$.

SIR metóda bude opäť relatívne efektívna vtedy, ak bude pomocná hustota dobre aproximovať tú cieľovú, pretože iba vtedy budú mať pravdepodobnosti $p_{SIR}(x_n^*)$ malú varianciu. V opačnom prípade, ak budú pravdepodobnosti $p_{SIR}(x_n^*)$ príliš rozdielne, treba urobiť výber v prvom kroku s veľkým rozsahom. Následne je možné časť z neho, t.j. x_n^* s malými váhami, zanedbať.

3.4.3 Acceptance-Rejection Sampling

RS metóda je založená na simulovaní náhodných hodnôt z inštrumentálnej hustoty, z ktorých však len niektoré prijmemo do (konečného) náhodného výberu. Aby sme mohli použiť RS metódu, inštrumentálna hustota musí byť tzv. (hornou) obálkovou funkciou cieľovej hustoty (*envelope function*), teda potrebujeme poznať konštantu c takú, že

$$p(x) < cq(x) \quad \text{pre } \forall x \in \mathcal{D}, \quad \text{pričom } c < \infty,$$

kde \mathcal{D} je nosič hustoty $p(x)$.

Algoritmus RS vygeneruje hodnotu x_n náhodného výberu nasledovne:

1. vygeneruj $y \sim q(x)$ a $u \sim \mathcal{U}(0, 1)$
2. ak $u < \frac{p(y)}{cq(y)}$, prijmi y a nastav $x_n = y$
inak zamietni y a opakuj krok 1

Číslo $\alpha_{RS} = \frac{p(y)}{cq(y)}$ z kroku 2 sa nazýva pravdepodobnosť prijatia (*acceptance probability*). Ak by sme sa na dvojicu $[y, cq(y)u]$ pozerali ako na bod v rovine, bude ležať pod funkciou $cq(x)$. Pri rozhodovaní o prijatí, resp. zamietnutí zisťujeme, či leží aj pod funkciou $p(x)$: ak áno, prijmemo ho, inak ho zamietneme. Tým je zabezpečené, že x -ové súradnice prijatých bodov, teda x_1, \dots, x_N , budú náhodným výberom z hustoty $p(x)$.

Podobne ako u IS metódy, výhodou RS metódy je jej jednoduchosť, ale na získanie jednej hodnoty x_n náhodného výberu z cieľovej hustoty treba vygenerovať v priemere až c kandidátov z inštrumentálnej hustoty (miera prijatia (*acceptance rate*) = $\frac{1}{c}$). Problémom tohto algoritmu teda nie je len potreba poznať hodnotu c , aby $q(x)$ bola obálkou hustoty $p(x)$. Ak inštrumentálna hustota neaproximuje dobre cieľovú hustotu, bude $c \gg 1$ (frekvencia zamietnutia bude vysoká), čo znamená, že bude potrebných veľa výpočtov funkčných hodnôt hustoty $p(x)$. V mnohých

praktických problémoch má $p(x)$ komplikovaný tvar, ktorý spôsobuje výpočtovú náročnosť jej vyhodnocovania, preto ani táto metóda nie je vhodná na výpočet (hlavne viacrozmerných) integrálov.

3.4.4 Adaptive Rejection Sampling

Metódu ARS navrhli Gilks a Wild (1992), resp. Gilks (1992) ako nástroj na generovanie náhodného výberu z jednorozmernej cieľovej hustoty $p(x)$, ktorá však musí spĺňať predpoklad logaritmickkej konkávnosti. Ten algoritmus využíva na skonštruovanie jej hornej obáľkovej funkcie, ktorá je použitá ako inštrumentálna hustota $q(x)$. Vylepšenie oproti RS je v znížení počtu vyhodnocovaní $p(x)$ vďaka priebežnému “zlepšovaniu” $q(x)$. Po každom zamietnutí kandidáta vygenerovaného z $q(x)$ totiž získanú informáciu o $p(x)$ zahrnieme do predpisu $q(x)$ tak, že c bude monotónne klesať. Inštrumentálna hustota bude postupne konvergovať k cieľovej hustote, pričom sa jej bude dotýkať práve v zamietnutých kandidátskych bodoch.

Rozšírená je aj verzia algoritmu ARS, ktorá navyše využíva dolnú obáľkovú funkciu cieľovej hustoty (*squeezing function*). Ďalej pri konštrukcii obáľkovej funkcie (funkcií) môžeme buď použiť deriváciu logaritmu cieľovej hustoty (*derivative based ARS*), alebo sa jej naopak vyhnúť (*derivative free ARS*). My popíšeme prípad algoritmu s konštrukciou len hornej obáľkovej funkcie a bez využitia derivácie (vtedy totiž netreba, aby bola cieľová hustota spojitely diferencovateľná).

Predpokladajme teda, že cieľová hustota $p(x)$ je na svojom nosiči \mathcal{D} logaritmicky konkávna (*log-concave*), čiže jej logaritmus je konkávna funkcia

$$\frac{\ln f(b) - \ln f(a)}{b - a} > \frac{\ln f(c) - \ln f(b)}{c - b} \quad \forall a, b, c \in \mathcal{D} \text{ také, že } a < b < c.$$

Nech $S_n = \{x_i ; i = 0, \dots, n + 1\}$ je súčasná množina bodov usporiadaných vzostupne, pričom x_0 a x_{n+1} môžu byť aj nekonečné limity oblasti \mathcal{D} . Ďalej nech $L_{i,j}(x; S_n)$ pre $1 \leq i \leq j \leq n$ predstavuje priamku prechádzajúcu bodmi $[x_i, \ln p(x_i)]$ a $[x_j, \ln p(x_j)]$, pre ostatné (i, j) nech $L_{i,j}(x; S_n)$ nie je definovaná.

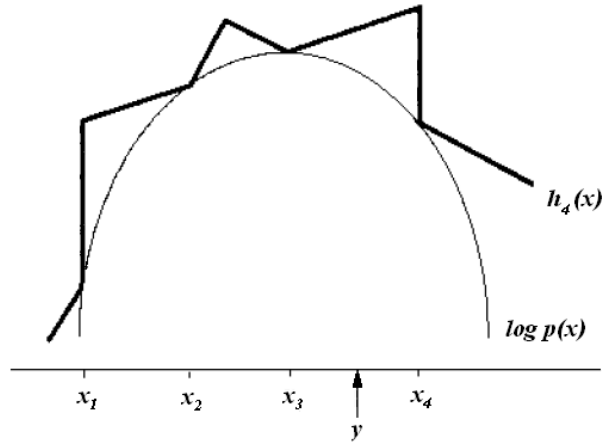
Definujeme po častiach lineárnu funkciu $h_n(x)$ (*hull function*):

$$h_n(x) = \min [L_{i-1,i}(x; S_n), L_{i+1,i+2}(x; S_n)], \quad x_i \leq x < x_{i+1}, \quad ^2 \quad (11)$$

ktorá je obáľkovou funkciou funkcie $\ln p(x)$ (teda $\ln p(x) \leq h_n(x) \quad \forall x \in \mathcal{D}$), čo je dôsledkom predpokladanej logaritmickkej konkávnosti cieľovej hustoty $p(x)$. Inštrumentálna hustota bude potom daná predpisom

$$q_n(x) = \frac{1}{c_n} \exp h_n(x), \quad (12)$$

²ak b nie je definované, potom $\min\{a, b\} = \min\{b, a\} = a$



Obr. 1: Grafické znázornenie ARS metódy, y je vygenerované z $h_4(x)$

kde

$$c_n = \int \exp h_n(x) dx. \quad (13)$$

Je to po častiach hustota exponenciálneho rozdelenia, a teda sa z nej dá simulovať priamo. Ďalšou výhodnou vlastnosťou $q_n(x)$ je, že je aktualizovaná vždy po výpočte funkčnej hodnoty hustoty $p(x)$.

ARS algoritmus vygeneruje hodnotu x_A z cieľovej hustoty nasledovne:

1. inicializuj n a S_n
2. vygeneruj $y \sim q_n(\cdot)$
3. vygeneruj $u \sim \mathcal{U}(0, 1)$
4. ak $u > \frac{p(y)}{\exp h_n(y)}$, potom

ZAMIETNUTIE:

nastav $S_{n+1} = S_n \cup \{y\}$ a preusporiadaj body v S_{n+1} vzostupne
zvyš n o 1 a choď späť na krok 2

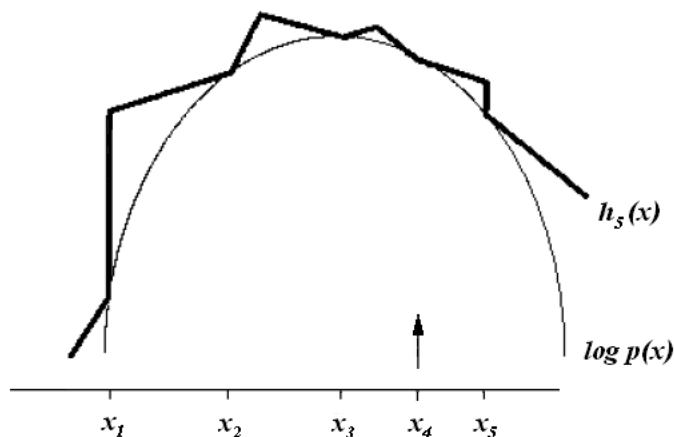
inak

PRIJATIE:

nastav $x_A = y$.

5. vráť x_A .

Po každom zamietnutí kandidáta y v kroku 4 sa zvýši počet dotykových bodov funkcií $\ln p(x)$ a $h_n(x)$ o jeden, čím sa zmenší c_n a zníži sa pravdepodobnosť



Obr. 2: Grafické znázornenie ARS metódy, zamietnutý bod y je zahrnutý do množiny S_5 , na základe ktorej je zostrojená aktualizovaná funkcia $h_5(x)$

zamietnutia nového kandidáta v ďalšej iterácii. Opakovaním krokov 2-5 ARS algoritmu dostaneme náhodný výber z cieľovej hustoty $p(x)$, pričom pravdepodobnosť zamietania bude stále klesať vďaka postupne sa zlepšujúcej funkcii $h_n(x)$.

Pri neohraničenej oblasti \mathcal{D} by mohli nastať komplikácie, preto ak \mathcal{D} je neohraničená zľava, štartovacie body v S_n treba zvoliť tak, aby smernica $L_{1,2}(x; S_n)$ bola kladná. Podobne, ak je \mathcal{D} neohraničená sprava, musí byť splnené, že $L_{n-1,n}(x; S_n)$ bude mať zápornú smernicu.

Priemerný počet iterácií potrebných na prijatie jednej hodnoty závisí na voľbe štartovacích bodov S_n a na predpise cieľovej hustoty $p(x)$. Vo všeobecnosti umiestnenie štartovacích bodov nemá na efektívnosť algoritmu veľký vplyv, dôležitý je hlavne ich počet. Malo by to byť číslo aspoň 3. Celkovo je ARS metóda veľmi efektívna, no použiteľná len na generovanie náhodného výberu z jednorozmernej logaritmicke konkávnej hustoty. Aj keď väčšina štandardných jednorozmerných rozdelení a mnohé úplné podmienené rozdelenia z používaných bayesovských modelov túto vlastnosť spĺňajú, vznikla potreba narábania aj s logaritmicke nekonkávnymi funkciami hustôt. Zovšeobecnením ARS metódy bola vytvorená ARMS metóda, pomocou ktorej je možné generovať náhodný výber zo všeobecnej (aj logaritmicke nekonkávnej) jednorozmernej cieľovej hustoty (pozri časť 4.4.3).

Metódy, ktoré sme uviedli, volia inštrumentálnu hustotu tak, aby vhodne aproximovala cieľovú hustotu a zároveň, aby sa z nej dalo priamo simulovať. Ak má cieľová hustota veľkú dimenziu alebo zložitý tvar, splniť obe požadované vlastnosti inštrumentálnej hustoty často nie je možné. Vtedy môžeme použiť dômyselnejšie

simulačné metódy založené na teórii Markovových reťazcov, a to MCMC metódy.

3.5 Markovove reťazce

V tejto časti uvedieme základnú teóriu Markovových reťazcov, na ktorej sú založené MCMC metódy. Pôjde o podstatné definície a vlastnosti, pričom najprv sa budeme zaoberať Markovovými reťazcami s konečnou množinou stavov (*finite state space Markov chains*), následne reťazcami so všeobecnou množinou stavov (*general state space Markov chains*) a nakoniec stručne zhrnieme konvergenčnú teóriu.

Pozn. 1 Pod Markovovým reťazcom budeme rozumieť Markovov reťazec s diskrétnym časom (*discrete-time Markov chain*).

3.5.1 Markovove reťazce s konečnou množinou stavov

Definícia 3.5.1 Markovov reťazec $\{X_t\}_{t \in T}$, $T = \mathbb{Z}_+$, je náhodný proces nadobúdajúci hodnoty z množiny stavov (state space) \mathcal{X} ³ a spĺňajúci Markovovu vlastnosť (tiež “*memoryless*” property)

$$\begin{aligned} P(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \\ = P(X_{n+1} = x_{n+1} | X_n = x_n) \quad \forall n \in T, \forall x_j \in \mathcal{X}. \end{aligned}$$

Ak $X_t = x_j$, hovoríme, že reťazec v čase t je v stave x_j .

Markovova vlastnosť znamená, že jediná informácia z minulých nadobudnutých hodnôt reťazca, ktorú potrebujeme na predikciu nasledujúcej hodnoty, je len terajší stav reťazca. Inak sa dá povedať, že pri danom súčasnom stave reťazca sú minulé a budúce stavy nezávislé.

Jednotlivé reťazce sú definované ich začiatočným pravdepodobnostným rozdelením stavov $p(0)$ a pravdepodobnosťami prechodu (*transition probabilities*) $P(X_{n+1}|X_n)$. Rozdelenie (stavov) reťazca v čase $n \in T$ je dané vektorom

$$p(n) = (p_n(x), x \in \mathcal{X}), \quad \text{kde } p_n(x) = P(X_n = x) \quad \forall x \in \mathcal{X}.$$

Na toto rozdelenie sa dá pozeráť aj ako na marginálne rozdelenie náhodnej premennej X_n . Pravdepodobnosti prechodu $P(X_{n+1} = x_j | X_n = x_i)$ vyjadrujú pravdepodobnosť toho, že reťazec v jednom časovom kroku prejde zo stavu x_i do stavu x_j . Tieto pravdepodobnosti zasa zodpovedajú podmienenému rozdeleniu $X_{n+1}|X_n$.

³Tu teda uvažujeme o \mathcal{X} s konečným počtom prvkov.

Definícia 3.5.2 Markovov reťazec $\{X_t\}_{t \in T}$ sa nazýva *stacionárny alebo homogénny* (v čase), ak pravdepodobnosti prechodu nezávisia od času, t.j.

$$p_{ij} = P(X_{n+k+1} = x_j | X_{n+k} = x_i) = P(X_{n+1} = x_j | X_n = x_i) \quad \forall x_i, x_j \in \mathcal{X}, \forall k \in \mathbb{N}, \forall n \in T.$$

Odtiaľ budeme uvažovať len homogénne Markovovove reťazce. V tomto prípade môžeme p_{ij} usporiadať do matice (pravdepodobností) prechodu (*transition probability matrix*) $\mathbb{P} = (p_{ij})_{x_i, x_j \in \mathcal{X}}$. Je to stochastická matica, tzn. štvorcová matica, ktorej prvky sú nezáporné čísla a súčty po riadkoch dávajú 1.

Pod pravdepodobnosťami prechodu k -teho rádu rozumieme pravdepodobnosti prechodu reťazca zo stavu x_i do stavu x_j za k časových krokov

$$p_{ij}(k) = P(X_{n+k} = x_j | X_n = x_i), \quad x_i, x_j \in \mathcal{X}, k \in \mathbb{N}, n \in T.$$

Takisto ich možno zapísať do matice (pravdepodobností) prechodu k -teho rádu

$$\mathbb{P}(k) = (p_{ij}(k))_{x_i, x_j \in \mathcal{X}}.$$

Veta 3.5.1 Pre homogénny Markovov reťazec $\{X_t\}_{t \in T}$ platí

$$\mathbb{P}(n+m) = \mathbb{P}(n)\mathbb{P}(m) \quad \forall n, m \in T \quad (14)$$

$$\mathbb{P}(n) = \mathbb{P}^n \quad \forall n \in T \quad (15)$$

$$p(n) = p(0)\mathbb{P}(n) \quad \forall n \in T. \quad (16)$$

Rovnice (14) z vety (3.5.1) sa nazývajú Chapman-Kolmogorovove rovnice. Táto veta hovorí, že pokiaľ poznáme začiatočné pravdepodobnostné rozdelenie a maticu pravdepodobností prechodu, poznáme správanie sa reťazca vo všetkých časoch z T .

Definícia 3.5.3 Pravdepodobnostné rozdelenie $\pi = (\pi_j)_{x_j \in \mathcal{X}}$ sa nazýva *limitné* (aj *rovnovážne alebo stabilizované*) rozdelenie reťazca, ak

$$\lim_{n \rightarrow \infty} p(n) = \pi.$$

Teda ak π existuje, potom bez ohľadu na to, aké bolo začiatočné pravdepodobnostné rozdelenie $p(0)$, Markovov reťazec bude konvergovať (s $n \rightarrow \infty$) k náhodnej premennej X_n s limitným pravdepodobnostným rozdelením π . Navyše toto limitné rozdelenie bude jediné a bude spĺňať $\pi\mathbb{P} = \pi$.

Definícia 3.5.4 Vektor $v = (v_j)_{x_j \in \mathcal{X}}$, ktorý spĺňa

$$v_j \geq 0 \quad \forall x_j \in \mathcal{X} \quad (17)$$

$$\sum_{x_j \in \mathcal{X}} v_j = 1 \quad (18)$$

$$v\mathbb{P} = v \quad (19)$$

sa nazýva *stacionárne (alebo invariantné) rozdelenie reťazca určeného maticou prechodu* \mathbb{P} .

Pozn. 2 Rovnice (19) sa nazývajú aj *globálne rovnice rovnováhy (general balance)*.

Z definície (3.5.4) je zrejmé, že ak raz dosiahneme stacionárne rozdelenie (prípadne v ňom začneme), t.j. $v = p(m)$ pre nejaké $m \in T$, už sa nebude s časom meniť, teda rozdelenie náhodných premenných v nasledujúcich časoch X_n bude tiež v , tzn. $p(n) = v$ pre $\forall n > m$. Stacionárne rozdelenie je teda invariantné vzhľadom na maticu prechodu. Markovov reťazec s konečným počtom stavov má vždy aspoň jedno stacionárne rozdelenie.

Limitné rozdelenie je nutne stacionárnym rozdelením a ak existuje, je aj jediným stacionárnym rozdelením. O limitnom rozdelení reťazca hovorí aj nasledujúca veta.

Veta 3.5.2 *Ak existujú*

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \pi_j \quad \forall x_i, x_j \in \mathcal{X},$$

potom existujú aj

$$\lim_{n \rightarrow \infty} p_n(x_j) = \pi_j \quad \forall x_j \in \mathcal{X}.$$

Limitné rozdelenie teda existuje, ak matica prechodu konverguje v čase a bude dané ľubovoľným riadkom práve tejto limitnej matice prechodu

$$\mathbb{P}(n) \longrightarrow \begin{pmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{pmatrix}, \quad n \rightarrow \infty.$$

Naopak, stacionárne rozdelenie nemusí byť aj limitným rozdelením, teda môže existovať, aj keď matica prechodu nekonverguje v čase. Prakticky je ľahšie nájsť stacionárne ako limitné rozdelenie reťazca. Preto v snahe nájsť limitné rozdelenie reťazca nájdeme stacionárne rozdelenie, o ktorom musíme dokázať, že bude aj hľadaným limitným rozdelením.

Definícia 3.5.5 *Markovov reťazec, pre ktorý existuje práve jedno jeho stacionárne rozdelenie, nazývame ergodický (ergodic). V opačnom prípade nazveme reťazec neergodický (non-ergodic).*

Pre existenciu ergodického Markovovho reťazca existuje viacero postačujúcich podmienok. Jednou z nich je nasledujúca veta.

Veta 3.5.3 (Markovova) Ak existuje $n \in \mathbb{N}$ také, že všetky prvky matice \mathbb{P}^n sú kladné (t.j. nenulové), potom Markovov reťazec s maticou pravdepodobností prechodu \mathbb{P} je ergodický.

Pre Markovov reťazec s konečnou množinou stavov je ergodickosť ekvivalentná s dvojicou iných vlastností, a to neredukovateľnosť a aperiodickosť.

Definícia 3.5.6 Markovov reťazec $\{X_t\}_{t \in T}$ s konečnou množinou stavov \mathcal{X} sa nazýva neredukovateľný (irreducible) alebo nerozložiteľný (indecomposable), ak

$$\forall x_i, x_j \in \mathcal{X}, x_i \neq x_j \quad \exists k \in \mathbb{N} : \quad p_{ij}(k) > 0.$$

Pre neredukovateľný reťazec teda platí, že všetky stavy sú dosiahnuteľné za konečný počet krokov, čiže reťazec sa môže dostať z ľubovoľného stavu do všetkých ostatných stavov konečným počtom krokov.

Definícia 3.5.7 Markovov reťazec $\{X_t\}_{t \in T}$ s konečnou množinou stavov \mathcal{X} sa nazýva aperiodický, ak

$$\forall x_i \in \mathcal{X} \quad NSD(k > 1 : p_{ii}(k) > 0) = 1.$$

Pozn. 3 $NSD(k_1, k_2, \dots)$ označuje najväčší spoločný deliteľ čísel k_1, k_2, \dots

Markovov reťazec je aperiodický, ak žiaden stav z \mathcal{X} nemá periódu, teda neexistujú časti množiny stavov, z ktorých by reťazec nadobúdal hodnoty iba v pravidelných časových intervaloch.

Veta 3.5.4 (Ergodická) Pre ľubovoľný neredukovateľný a aperiodický Markovov reťazec $\{X_t\}_{t \in T}$ existuje také rozdelenie reťazca $\pi = (\pi_j)_{x_j \in \mathcal{X}}$, že platí

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \pi_j \quad \forall x_i, x_j \in \mathcal{X}.$$

Z tejto vety a vety (3.5.2) vyplýva, že π bude limitné a jediné stacionárne rozdelenie stavov reťazca, teda reťazec je ergodický. Preto aj viacerí autori definujú ergodický Markovov reťazec ako reťazec, ktorý je neredukovateľný a aperiodický. Z neredukovateľnosti vyplýva existencia a jednoznačnosť stacionárneho rozdelenia, zatiaľčo aperiodicita zabezpečí, že toto rozdelenie bude aj limitným rozdelením.

Vlastnosti Markovovho reťazca nutne na existenciu jeho limitného rozdelenia, však nehovoria nič o tvare tohto rozdelenia. Preto ak by sme chceli zistiť, či je nejaké rozdelenie limitným rozdelením reťazca, overíme najprv postačujúcu podmienku existencie jediného stacionárneho rozdelenia, ktorou je reverzibilita reťazca, a následne ukážeme, že je aj limitným rozdelením. Takýto prístup bude potrebný v MCMC metódach.

Definícia 3.5.8 *Markovov reťazec sa nazýva reverzibilný (v čase) (time-reversible), ak existuje také rozdelenie $\pi = (\pi_j)_{x_j \in \mathcal{X}}$, že vzhľadom naň matica prechodu $\mathbb{P} = (p_{ij})_{x_i, x_j \in \mathcal{X}}$ reťazca spĺňa podmienky reverzibility*

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall x_i, x_j \in \mathcal{X}. \quad (20)$$

Pozn. 4 *Rovnice (20) sa nazývajú aj lokálne rovnice rovnováhy (detailed balance).*

Tak ako stacionarita znamenala rovnováhu s časom idúcim dopredu, reverzibilita znamená obojsmernú rovnováhu - ak raz dosiahne reverzibilný Markovov reťazec rovnovážne rozdelenie, pravdepodobnosť prechodu do nejakého stavu bude rovnaká pre čas idúci dopredu i dozadu. Ľahko sa overí, že po zosumovaní oboch strán rovníc (20) cez $x_i \in \mathcal{X}$ dostaneme globálne rovnice rovnováhy, $\pi \mathbb{P} = \pi$, tzn. π bude stacionárne rozdelenie reťazca. Reverzibilita je teda silnejšia vlastnosť ako stacionarita, no pri overovaní podmienok reverzibility sa vyhneme sumovaniu cez množinu stavov \mathcal{X} . V praxi mávajú diagonálne prvky \mathbb{P} často zložitý tvar, preto ďalšou výhodou je, že tieto podmienky stačí overovať len pre $x_i \neq x_j$.

3.5.2 Markovove reťazce so všeobecnou množinou stavov

Napriek počiatkovej snahe vyhradiť termín Markovov reťazec len pre náhodný proces s konečnou (resp. spočítateľnou) množinou stavov (s diskretným alebo spojitým časovým parametrom), aplikácie ukázali nutnosť rozšíriť definíciu aj na reťazec so všeobecnou množinou stavov. My budeme opäť uvažovať iba Markovove reťazce s diskretným časom.

V tejto časti bude teda množina stavov \mathcal{X} ľubovoľná (všeobecná) množina, na ktorej existuje σ -algebra, ozn. $\mathcal{F}(\mathcal{X})$, a $T = \mathbb{Z}_+$. Markovova vlastnosť, ktorú musí Markovov reťazec spĺňať, má tvar

$$\begin{aligned} P(X_{n+1} \in A | X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \\ = P(X_{n+1} \in A | X_n = x_n), \quad \forall n \in T, \forall A \in \mathcal{F}(\mathcal{X}), \forall x_j \in \mathcal{X}. \end{aligned}$$

Reťazec je daný

- začiatočným rozdelením určeným pravdepodobnostnou mierou μ na $\mathcal{F}(\mathcal{X})$
- prechodovým (pravdepodobnostným) jadrom (*transition (probability) kernel*)

$$P = \{P(x, A), x \in \mathcal{X}, A \in \mathcal{F}(\mathcal{X})\},$$

pre ktoré platí

1. pre všetky $A \in \mathcal{F}(\mathcal{X})$ je $P(\cdot, A)$ nezáporná merateľná funkcia na \mathcal{X}
2. pre každé $x \in \mathcal{X}$ je $P(x, \cdot)$ pravdepodobnostná miera na $\mathcal{F}(\mathcal{X})$,

nasledovne

$$\begin{aligned} P_\mu(X_0 \in A_0, X_1 \in A_1, \dots, X_{n+1} \in A_{n+1}) &= \\ &= \int_{x_0 \in A_0} \int_{x_1 \in A_1} \dots \int_{x_n \in A_n} \mu(dx_0) P(x_0, dx_1) \dots P(x_n, A_{n+1}), \quad n \in \mathbb{N}. \end{aligned} \quad (21)$$

Výraz (21) definuje tzv. konečno-rozmerné rozdelenie Markovovho reťazca a ak platí pre všetky $n \in \mathbb{N}$, Markovov reťazec sa nazýva homogénny s počiatočným rozdelením μ a prechodovým jadrom $P(x, A)$.

Prechodové jadro je zovšeobecnením pravdepodobností prechodu Markovových reťazcov s konečným počtom stavov

$$P(x, A) = P(X_{n+1} \in A | X_n = x), \quad \forall x \in \mathcal{X}, A \in \mathcal{F}(\mathcal{X}).$$

Podobne môžeme definovať aj prechodové jadro k -teho rádu (*k-step probability transition kernel*) $\{P^k(x, A), x \in \mathcal{X}, A \in \mathcal{F}(\mathcal{X})\}$, kde

$$P^0(x, A) = \delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

$$P^k(x, A) = \int_{\mathcal{X}} P(x, dy) P^{k-1}(y, A) = P(X_{n+k} \in A | X_n = x).$$

Veta 3.5.5 *Pre ľubovoľné m také, že $0 \leq m \leq n$, platí*

$$P^n(x, A) = \int_{\mathcal{X}} P^m(x, dy) P^{n-m}(y, A), \quad x \in \mathcal{X}, A \in \mathcal{F}(\mathcal{X}). \quad (22)$$

Rovnice (22) sú Chapman-Kolmogorovove rovnice pre homogénny Markovov reťazec so všeobecnou množinou stavov.

Ťažšie je definovať aj limitnú pravdepodobnostnú mieru reťazca, potrebujeme k tomu určiť spôsob konvergencie dvoch pravdepodobnostných mier.

Definícia 3.5.9 *Nech μ a ν sú pravdepodobnostné miery na tom istom pravdepodobnostnom priestore (Ω, \mathcal{F}) . Potom celková variačná vzdialenosť (total variation distance) je daná*

$$\|\mu(\cdot) - \nu(\cdot)\| = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|.$$

Limitnou pravdepodobnostnou mierou Markovovho reťazca potom rozumieme takú pravdepodobnostnú mieru π , pre ktorú platí

$$\lim_{n \rightarrow \infty} \|P^n(x, \cdot) - \pi(\cdot)\| = 0 \quad \text{pre } \forall x \in \mathcal{X}.$$

Definícia 3.5.10 Pravdepodobnostná miera π na $(\mathcal{X}, \mathcal{F}(\mathcal{X}))$ s vlastnosťou

$$\pi(A) = \int_{\mathcal{X}} \pi(dx)P(x, A) \quad \forall A \in \mathcal{F}(\mathcal{X})$$

sa nazýva *stacionárna alebo invariantná miera Markovovho reťazca s prechodovým jadrom $P(x, A)$* .

Opäť platí, že limitná miera je stacionárnou mierou Markovovho reťazca.

Definícia 3.5.11 Čas prvého návratu (first return time) do množiny A definujeme

$$\tau_A = \inf \{n \geq 1; X_n \in A\}.$$

Definícia 3.5.12 Markovov reťazec sa nazýva ϕ -neredukovateľný (ϕ -irreducible), ak existuje miera ϕ na $(\mathcal{X}, \mathcal{F}(\mathcal{X}))$ taká, že pre $\forall A \in \mathcal{F}(\mathcal{X})$ platí

$$\phi(A) > 0 \Rightarrow P(\tau_A < \infty) > 0 \quad \forall x \in \mathcal{X}.$$

Analogicky ako u Markovových reťazcov s konečnou množinou stavov, ϕ -neredukovateľnosť znamená, že reťazec začínajúci v ľubovoľnom bode $x \in \mathcal{X}$ nadobudne hodnotu z množiny A nulovej miery (vzhľadom na ϕ) s nenulovou pravdepodobnosťou. Dá sa to vyjadriť aj tak, že pre každú takú množinu A existuje $n \geq 1 : P^n(x, A) > 0$.

Platí, že ak je Markovov reťazec s limitnou mierou π ϕ -neredukovateľný, potom π je jediná stacionárna miera a reťazec je aj π -neredukovateľný.

Definícia 3.5.13 ϕ -neredukovateľný Markovov reťazec je *periodický*, ak existujú disjunktné množiny $A_1, A_2, \dots, A_d \subset \mathcal{X}, d \geq 2$ (nazývané *periodický rozklad*) také, že

$$\begin{aligned} P(x, A_{i+1}) &= 1 & \forall x \in A_i, \quad 1 \leq i \leq d-1 \\ P(x, A_1) &= 1 & \forall x \in A_d \\ \phi(A_1) &> 0. \end{aligned}$$

Ak $d = 1$, Markovov reťazec sa nazýva *aperiodický*.

Veta 3.5.6 Pre ϕ -neredukovateľný aperiodický Markovov reťazec so stacionárnou pravdepodobnostnou mierou π platí

$$\lim_{n \rightarrow \infty} \|P^n(x, \cdot) - \pi(\cdot)\| = 0 \quad \text{pre } \pi\text{-s.v. } x \in \mathcal{X}.$$

V praxi často konvergencia pre skoro všetky x vzhľadom na mieru π nestačí, pretože reťazec začína v konkrétnom bode, čo je množina nulovej miery, a teda konvergencia prechodového jadra k π by nebola zaručená.

Definícia 3.5.14 Množina A sa nazýva Harris-rekurentná, ak

$$P(\{X_t\}_{t \in T} \in A \text{ nekonečne často} \mid X_0 = x) = 1 \quad \forall x \in A.$$

ϕ -neredukovateľný reťazec $\{X_t\}_{t \in T}$ sa nazýva Harris-rekurentný, ak všetky množiny $A \in \mathcal{F}(\mathcal{X})$ také, že $\phi(A) > 0$, sú Harris-rekurentné.

Pozn. 5 $\{\{X_t\}_{t \in T} \in A \text{ nekonečne často}\} := \bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} \{X_k \in A\}$

Harris-rekurentnosť je pre Markovove reťazce s konečnou množinou stavov ekvivalentná s neredukovateľnosťou.

Definícia 3.5.15 Harris-rekurentný aperiodický Markovov reťazec sa nazýva ergodický.

Veta 3.5.7 Pre ergodický Markovov reťazec so stacionárnou pravdepodobnostnou mierou π platí

$$\lim_{n \rightarrow \infty} \|P^n(x, \cdot) - \pi(\cdot)\| = 0 \quad \forall x \in \mathcal{X}.$$

Ak teda Markovov reťazec spĺňa silnejšiu podmienku (ako je ϕ -neredukovateľnosť), a to Harris-rekurentnosť, prechodové jadro konverguje k π pre ľubovoľný začiatkový bod $x \in \mathcal{X}$.

Na overenie stacionarity nejakej pravdepodobnostnej miery je tak isto možné použiť reverzibilitu reťazca.

Definícia 3.5.16 Markovov reťazec sa nazýva reverzibilný, ak existuje taká pravdepodobnostná miera π , že jeho prechodové jadro spĺňa

$$\int_A \pi(dx)P(x, B) = \int_B \pi(dx)P(x, A) \quad \forall A, B \in \mathcal{F}(\mathcal{X}). \quad (23)$$

Pozn. 6 Podmienka (23) sa niekedy skrátene zapisuje v tvare

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx), \quad (24)$$

ktorý je analógiou lokálnych rovníc rovnováhy, tu však ide o pravdepodobnosti množín a nie stavov z \mathcal{X} .

Ak v (23) položíme $A = \mathcal{X}$, dostaneme

$$\int_{\mathcal{X}} \pi(dx)P(x, B) = \int_B \pi(dx) = \pi(B) \quad \forall B \in \mathcal{X},$$

t.j. z reverzibility reťazca vyplýva stacionarita π .

3.5.3 Konvergenčná teória Markovových reťazcov

Pri splnení podmienok konvergenzie platí pre homogénne Markovove reťazce obdoba silného zákona veľkých čísel a centrálnej limitnej vety. Pre Markovove reťazce s konečnou množinou stavov sú týmito podmienkami nereducovateľnosť a aperiodicita (t.j. ergodicita) reťazca. SLLN a CLT majú tvar uvedený nižšie.

Aby boli spomenuté tvrdenia platné pre ľubovoľnú začiatočnú pravdepodobnostnú mieru Markovovho reťazca so všeobecnou množinou stavov, nestačí iba ϕ -nereducovateľnosť a aperiodicita reťazca, pretože konvergenca by nemusela byť zaručená, ak by reťazec začínal v nejakom bode - množine s nulovou mierou (podobne ako pri podmienkach existencie limitnej pravdepodobnostnej miery). Označme

$$\mu = E_\pi[g(X)] = \int_{\mathcal{X}} g(x)\pi(dx)$$

$$\hat{\mu}_N = \frac{1}{N} \sum_{n=1}^N g(x^{(n)}),$$

kde $x^{(1)}, \dots, x^{(N)}$ sú realizácie Markovovho reťazca $\{X_t\}_{t \in T}$.

SLLN platí pre Harris-rekurentný Markovov proces so stacionárnou pravdepodobnostnou mierou π a ľubovoľnú funkciu $g(x)$ takú, že $E_\pi[|g(x)|] < \infty$, t.j.

$$\hat{\mu}_N \xrightarrow{s.v.} \mu, \quad \text{ak } N \rightarrow \infty. \quad (25)$$

Pozn. 7 $\hat{\mu}_N$ sa nazýva aj *ergodický priemer*.

Pre CLT sú potrebné ešte silnejšie predpoklady, okrem Harris-rekurentnosti reťazca a existencie stacionárnej pravdepodobnostnej miery π musí byť Markovov reťazec navyše buď geometricky ergodický, alebo V-rovnomerne ergodický (vždy s príslušnými podmienkami na funkciu $g(x)$). Potom platí

$$\sqrt{N} \frac{(\hat{\mu}_n - \mu)}{\sqrt{\sigma_N^2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \quad \text{ak } N \rightarrow \infty, \quad (26)$$

kde

$$\sigma_N^2 = \frac{1}{N} \sum_{k=1}^N \text{Var}_\pi[g(X_k)] + \frac{1}{N} \sum_{k=1}^N \sum_{l=1}^N \text{Cov}[g(X_k), g(X_l)] = \gamma_0 + 2 \sum_{k=1}^{N-1} \frac{N-k}{N} \gamma_k.$$

Teda

$$\sqrt{N} (\hat{\mu}_N - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2), \quad \text{ak } N \rightarrow \infty,$$

kde

$$\sigma^2 = \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k. \quad (27)$$

Definície týchto silnejších predpokladov konvergenzie sú pomerne zložité (pozri Meyn a Tweedie (1993)), preto ich nebudeme uvádzať. Ich splnenie je dokázané pre rôzne typy úloh a konkrétnych MCMC metód, preto v jednotlivých aplikáciách sa zväčša overuje už len ergodicita reťazca a podmienky reverzibility.

3.6 Úvod do MCMC metód

Podobne ako simulačné techniky (popísané v časti 3.4), aj Markov Chain Monte Carlo (MCMC) metódy sa využívajú na získanie “náhodného výberu” (tiež nazývaného aj MCMC výber) z cieľového rozdelenia, z ktorého nie je možné simulovať priamo. V skutočnosti však nejde o náhodný výber, teda realizácie nezávislých rovnako rozdelených náhodných premenných s daným rozdelením, ale o realizácie vhodného Markovovho reťazca. Vhodný reťazec bude taký, ktorého limitným rozdelením bude práve cieľové rozdelenie. Myšlienku MCMC simulácií prvý použil Metropolis (1953) ako metódu na efektívne simulovanie energetických úrovní atómov v kryštalickej štruktúre. Hastings (1970) ju neskôr zovšeobecnil a prispôbil na využitie v štatistických úlohách.

Cieľom MCMC metód je teda zostrojiť také prechodové jadro, aby ním daný Markovov reťazec mal množinu stavov zhodnú s nosičom cieľovej hustoty a aby cieľové rozdelenie bolo jeho limitným rozdelením. Ide o obrátený problém ako v klasickom prístupe teórie Markovových reťazcov, kde na základe znalosti prechodového jadra hľadáme limitné rozdelenie.

Nech $p(x)$, $x \in \mathcal{X}$ (\mathcal{X} je často otvorená podmnožina \mathbb{R}^d a hustoty sa berú vzhľadom na Lebesgueovu mieru), je cieľová hustota, v ktorej netreba poznať normovací koeficient. Potom príslušné cieľové rozdelenie $\pi(\cdot)$ je dané

$$\pi(A) = \frac{\int_A p(x) dx}{\int_{\mathcal{X}} p(x) dx}.$$

Na skonštruovanie prechodového jadra $P(x, \cdot)$ sa využíva pomocné prechodové jadro $Q(x, \cdot)$, dané inštrumentálnou hustotou $q(x, y)$, s vlastnosťami

1. $Q(x, \cdot)$ je rozdelenie, z ktorého vieme simulovať,
2. pre $\forall x, y \in \mathcal{X}$ vieme vyčíslieť hodnotu $q(x, y)$.

Výber inštrumentálnej hustoty $q(x, y)$ nutne nesúvisí s cieľovou hustotou $p(x)$, v podstate môžeme zvoliť ľubovoľnú hustotu, ktorá spĺňa vlastnosti 1 a 2.

Pri súčasnom stave x , v ktorom sa Markovov reťazec nachádza, vygenerujeme z $Q(x, \cdot)$ kandidáta y na nasledujúci stav a prijmeme ho s pravdepodobnosťou $\alpha(x, y)$ (*acceptance probability*). V takom prípade sa reťazec pohne zo stavu x do stavu

y . Inak ho zamietneme a reťazec zostane v stave x . Výsledné prechodové jadro pre ľubovoľnú množinu $A \subseteq \mathcal{X}$ je potom dané súčtom pravdepodobnosti pohnutia sa x a pravdepodobnosti zotrvania v x

$$P(x, A) = \int_A q(x, y)\alpha(x, y)dy + I_A(x) \left(1 - \int_{\mathcal{X}} q(x, y)\alpha(x, y)dy \right), \quad (28)$$

kde $I_A(x)$ je indikátor množiny A a $q(x, y)\alpha(x, y)$ je hustota zodpovedajúca pravdepodobnosti prijatia stavu y (a teda pohnutia sa z x).

Pravdepodobnosť prijatia $\alpha(x, y)$ (ktorá vplýva na konštrukciu prechodového jadra) definujeme tak, aby boli splnené podmienky reverzibility cieľového rozdelenia $\pi(\cdot)$ vzhľadom na prechodové jadro $P(x, \cdot)$. Tie zabezpečia, že dané rozdelenie bude stacionárnym rozdelením Markovovho reťazca $\{X_t\}_{t \in T}$. Na to, aby bolo skutočne aj limitným rozdelením, musí reťazec spĺňať aj podmienky konverencie (v zmysle celkovej variačnej vzdialenosti $P^n(x, \cdot)$ a $\pi(\cdot)$, t.j. rozdelenie $X_n \approx \pi$), čiže ϕ -neredukovateľnosť a aperiodickosť, resp. Harris-rekurentnosť a aperiodickosť.

V takom prípade platí pre Markovov reťazec SLLN, tzn. charakteristiky cieľového rozdelenia môžeme vypočítať Monte Carlo metódou presne ako s použitím náhodného výberu z $p(x)$

$$I = \int g(x)p(x)dx = \int g(x)d\pi = E_{\pi}[g(X)] = \mu \approx \frac{1}{N} \sum_{n=1}^N g(x^{(n)}) = \hat{\mu}_N, \quad (29)$$

kde $x^{(1)}, \dots, x^{(N)}$ sú realizácie Markovovho reťazca.

Dnes už existuje množstvo algoritmov na zostrojenie prechodového jadra s požadovanými vlastnosťami, vychádzajú však z dvoch základných algoritmov - Metropolis-Hastingsovho a Gibbsovho. MCMC algoritmy sa dajú použiť na generovanie MCMC výberu v podstate z ľubovoľného (aj veľmi komplikovaného mnohorozmerného) rozdelenia. Jediná podmienka na toto rozdelenie je, aby bolo možné vyčíslieť jeho hustotu (nie nutne normovanú) v ľubovoľnom bode z \mathcal{X} .

Stacionarita cieľového rozdelenia pre Markovov reťazec je teda zabezpečená konštrukciou prechodového jadra, ktorá je vždy súčasťou MCMC algoritmu. V jednotlivých aplikáciách týchto metód však treba zvlášť overiť splnenie podmienok konverencie. Aj keď to často nie je problém, neexistuje žiaden všeobecný postup. Mnohí autori sformulovali kritériá, ktoré zabezpečia ϕ -neredukovateľnosť a aperiodickosť, ale len pre špecifické problémy na základe typu algoritmu, tvaru a dimenzie cieľovej hustoty a typu množiny stavov reťazca. V praktických implementáciách MCMC metód so všeobecnou množinou stavov nie je s ϕ -neredukovateľnosťou zväčša problém. Navyše často implikuje aj Harris-rekurentnosť. Problémy však môžu napríklad nastať, ak by mala inštrumentálna hustota ľahšie chvosty ako cieľová hustota. Taktiež aperiodicita je väčšinou splnená.

Ďalšou žiadúcou vlastnosťou Markovovho reťazca, ktorého realizácie získavame MCMC metódou, je geometrická ergodicita, pretože potom platí CLT a na základe odhadu odhadu asymptotickej variancie pre ergodický priemer je možné zistiť kvalitu odhadov charakteristík cieľového rozdelenia. Opäť však neexistuje všeobecný postup na overenie tejto vlastnosti reťazca a je dokázaná (pomocou overenia minorizačnej alebo driftovej podmienky) len pre rôzne konkrétne algoritmy a tvary cieľových hustôt. Podobne je to v oblasti určovania kvantitatívnych mier konvergenzie, kde sa tiež mnohí autori snažili odvodiť hranice pre čas a počet iterácií potrebných na konvergenciu - no opäť len v špecifických prípadoch úloh. Otázna zostala aj použiteľnosť týchto výsledkov v praxi.

V aplikáciách MCMC metód je častejšie dôležitejšia praktická konvergenzia, ktorá sa dá určiť pomocou konvergenčnej diagnostiky (pozri časť 5). To však nič nemení na tom, že pri používaní MCMC metód treba poznať teoretické predpoklady, na ktorých sú založené a snažiť sa ich v čo najväčšom rozsahu aj analyticky overiť. Táto snaha by mala byť podporená aj faktom, že všetky MCMC algoritmy sú vykonávané na počítači, preto nasimulovaný Markovov reťazec je vždy len aproximáciou teoreticky zostrojeného reťazca. Nástroje konvergenčnej diagnostiky teda v skutočnosti pracujú len s realizáciami približného, a nie teoretického, reťazca.

4 MCMC metódy

4.1 Použitie MCMC metód v SV modeloch

Ako už bolo spomenuté, SV modely patria medzi prípady, kedy priame simulácie z aposteriórnej hustoty $p(\theta, X|Y)$ nie sú možné, pretože často nejde o hustotu žiadneho štandardného rozdelenia. Použitie MCMC metód je potrebné najmä vtedy, ak kvôli veľkej dimenzii, prípadne zložitému tvaru $p(\theta, X|Y)$ sa nedajú použiť iné bayesovské simulačné techniky.

Široká využiteľnosť MCMC metód na simulovanie v podstate z ľubovoľnej cieľovej aposteriórnej združenej hustoty vyplýva z faktu, že tieto metódy umožňujú jej postupný rozklad (podľa (8)) na úplné podmienené hustoty menších dimenzií. V extrémnom prípade namiesto simulovania z $(T + k)$ -rozmernej $p(\theta, X|Y)$, MCMC algoritmy na získanie potrebných realizácií Markovovho reťazca použijú $T + k$ jednorozmerných úplných podmienených hustôt ($p(\theta_i|\theta_{-i}, X, Y)$, $i = 1, \dots, k$ a $p(X_t|X_{-t}, \theta, Y)$, $t = 1 \dots, T$). Aj rozšírenú funkciu vierohodnosti a hustotu rozdelenia stavovej premennej možno zjednodušiť rozkladom vyplývajúcim z predpokladu

markovovskosti

$$p(Y|X, \theta) = p(Y_1, \dots, Y_T|X, \theta) = \prod_{t=1}^T p(Y_t|Y_{t-1}, X_{t-1}, \theta)$$

$$p(X|\theta) = \prod_{t=1}^T p(X_t|X_{t-1}, \theta).$$

Ďalšou výhodou je, že netreba poznať normovací koeficient v cieľovej hustote $p(\theta, X|Y)$, ktorý vzhľadom na jej komplikovaný tvar zväčša ani nie je možné vypočítať.

Medzi dve základné MCMC metódy patrí Metropolis-Hastingsov algoritmus a Gibbsov algoritmus. Ostatné MCMC metódy sú buď ich rôznymi obmenami, alebo kombináciami (tzv. hybridné MCMC metódy).

4.2 Metropolis-Hastingsov algoritmus

Nech $p(x)$ je jednorozmerná cieľová hustota s nosičom \mathcal{X} a $\pi(\cdot)$ k nej prislúchajúce rozdelenie. Zvolíme štartovací bod $x^{(0)} \in \mathcal{X}$ a prechodové jadro $Q(x, \cdot)$ s hustotou $q(x, y)$. Hodnota $Q(x, dy)$ vyjadruje pravdepodobnosť prechodu zo stavu x do stavu y , teda pravdepodobnosť, že z rozdelenia $Q(x, \cdot)$ vygenerujeme bod y , ak predchádzajúci vybratý bod bol x .

Bod $x^{(n+1)} \in \mathcal{X}$ dostaneme v $(n+1)$ -vej iterácii v týchto krokoch:

1. vygeneruj kandidáta (na nasledujúci $((n+1)$ -vý) stav) $y \sim Q(x^{(n)}, \cdot)$,
2. vypočítaj pravdepodobnosť prijatia

$$\alpha(x^{(n)}, y) = \begin{cases} \min \left\{ 1, \frac{p(y)q(y, x^{(n)})}{p(x^{(n)})q(x^{(n)}, y)} \right\} & p(x^{(n)})q(x^{(n)}, y) > 0 \\ 1 & p(x^{(n)})q(x^{(n)}, y) = 0 \end{cases} \quad (30)$$

3. s pravdepodobnosťou $\alpha(x^{(n)}, y)$ prijmi y a priraď $x^{(n+1)} = y$,
inak zamietni y a priraď $x^{(n+1)} = x^{(n)}$

Prakticky potrebujeme v kroku 3 vygenerovať $u \sim \mathcal{U}(0, 1)$ a y prijme, ak $u < \alpha(x^{(n)}, y)$. V opačnom prípade ho zamietneme.

Výber kandidáta y je náhodný a závisí len od posledného vygenerovaného bodu $x^{(n)}$, preto opakovaním krokov 1 - 3 dostávame realizácie Markovovho reťazca. Pri výpočte pravdepodobnosti prijatia by sa normovací koeficient cieľovej hustoty vykrátil, preto $p(x)$ môže byť aj nenormovaná hustota. Jediná požiadavka na ňu je, aby sa dala vyčísliť v ľubovoľnom bode $x \in \mathcal{X}$.

Platí, že M-H algoritmus generuje realizácie Markovovho reťazca so stacionárnym rozdelením π . Ako dôkaz potrebujeme overiť podmienky reverzibility (24) rozdelenia π vzhľadom na prechodové jadro tohto reťazca, ktoré je dané vzťahom (28). Pre prvú časť prechodového jadra vyjadrujúcu pravdepodobnosť pohnutia sa z x dostaneme

$$\begin{aligned}\pi(dx)Q(x, dy)\alpha(x, y) &= [p(x)dx][q(x, y)\alpha(x, y)dy] = \\ &= p(x)q(x, y) \min \left\{ 1, \frac{p(y)q(y, x)}{p(x)q(x, y)} \right\} dx dy = \\ &= \min \{ p(x)q(x, y), p(y)q(y, x) \} dx dy = \\ &= p(y)q(y, x) \min \left\{ 1, \frac{p(x)q(x, y)}{p(y)q(y, x)} \right\} dx dy = \\ &= [p(y)dy][q(y, x)\alpha(y, x)dx] = \pi(dy)Q(y, dx)\alpha(y, x)\end{aligned}$$

a pre druhú časť prechodového jadra vyjadrujúcu pravdepodobnosť zotrvania v x je zrejmé

$$p(x)dx I_A(x) \left(1 - \int_{\mathcal{X}} q(x, y)\alpha(x, y)dy \right) = p(y)dy I_A(y) \left(1 - \int_{\mathcal{X}} q(y, x)\alpha(y, x)dx \right).$$

π -neredukovateľnosť reťazca zabezpečí podmienka na inštrumentálnu hustotu

$$p(x) > 0 \Rightarrow q(x, y) > 0 \quad \forall x \in \mathcal{X},$$

tzn. nosič inštrumentálnej hustoty, musí zahŕňať nosič \mathcal{X} cieľovej hustoty. Aperiodicita reťazca bude splnená, ak hustoty $p(x)$, $q(x, y)$ budú spojité (je pritom dôležité, že $p(x^{(0)}) > 0$, t.j. štartovací bod $x^{(0)}$ volíme z \mathcal{X}). Pri splnení týchto podmienok, menovateľ pravdepodobnosti prijatia bude rovný 0 s nulovou pravdepodobnosťou.

Z π -neredukovateľnosti pre M-H algoritmus vyplýva aj Harris-rekurentnosť a ak je \mathcal{X} ohraničená v \mathbb{R} , pomocou minorizačnej alebo driftovej podmienky sa dá ukázať aj geometrická ergodicita.

Okrem toho, že inštrumentálna hustota musí spĺňať podmienky potrebné na zabezpečenie stacionarity cieľového rozdelenia a konvergenie reťazca, je jej výber obmedzený len tým, aby sa dala vyčíslíť v ľubovoľných bodoch x, y z jej nosiča a aby sme z nej vedeli priamo simulovať, čo ponecháva značnú ľubovôľu pri jej voľbe. Práve vďaka tomu môžeme voliť typ $q(x, y)$, resp. jej škálovanie, tak, aby reťazec skonvergoval čo najrýchlejšie alebo aby bol M-H algoritmus výpočtovo čo najjednoduchší.

Výber vhodnej inštrumentálnej hustoty pritom záleží na type úlohy, v ktorej algoritmus používame a na tvare cieľovej hustoty. Existujú niektoré všeobecné tvary inštrumentálnej hustoty, pri ktorých nastávajú špeciálne prípady M-H algoritmu.

4.2.1 M-H algoritmus s náhodnou prechádzkou

Ako vidno aj z názvu tejto metódy, algoritmus generuje kandidátov na nasledujúci stav podľa modelu náhodnej prechádzky

$$y = x^{(n)} + z, \quad z \sim F(\cdot),$$

čo zodpovedá inštrumentálnej hustote tvaru

$$q(x, y) = f(y - x) = f(z),$$

kde f je hustota rozdelenia F a pravdepodobnosti prijatia

$$\alpha(x^{(n)}, y) = \min \left\{ 1, \frac{p(y)f(x^{(n)} - y)}{p(x^{(n)})f(y - x^{(n)})} \right\}.$$

Existuje viacero možností pre voľbu F , môže ísť napr. o rovnomerné, normálne alebo t -rozdelenie.

Ak je inštrumentálna hustota navyše symetrická, t.j.

$$q(x, y) = f(y - x) = f(x - y) = q(y, x),$$

pravdepodobnosť prijatia sa zjednoduší na tvar

$$\alpha(x^{(n)}, y) = \min \left\{ 1, \frac{p(y)}{p(x^{(n)})} \right\}.$$

Práve takýto algoritmus navrhli Metropolis et al. (1953), preto sa niekedy nazýva aj Metropolisov algoritmus.

Zmenami variancie inštrumentálnej hustoty v RW M-H algoritme môžeme dosiahnuť rýchlejšiu konvergenciu alebo vyššiu mieru zmiešavania reťazca, ide teda o ladiaci parameter (*tuning parameter*).

4.2.2 Nezávislý M-H algoritmus

Hastings (1970) zovšeobecnil Metropolisov algoritmus voľbou inštrumentálnej hustoty v tvare

$$q(x, y) = q(y).$$

To znamená, že kandidátov na nasledujúci stav bude algoritmus vyberať nezávisle od posledného stavu reťazca. Závislosť dvoch po sebe idúcich stavov sa však zachová kvôli tvaru pravdepodobnosti prijatia

$$\alpha(x^{(n)}, y) = \min \left\{ 1, \frac{p(y)q(x^{(n)})}{p(x^{(n)})q(y)} \right\} = \min \left\{ 1, \frac{p(y)/q(y)}{p(x^{(n)})/q(x^{(n)})} \right\} = \min \left\{ 1, \frac{w(y)}{w(x^{(n)})} \right\}.$$

Tvarom $\alpha(x^{(n)}, y)$ sa nezávislý M-H algoritmus podobá na bayesovské simulačné techniky IS a RS. Takisto bude efektívny vtedy, ak inštrumentálna hustota bude dobre aproximovať cieľovú hustotu.

4.3 Gibbsov algoritmus

Tento algoritmus sa používa, ak je cieľová hustota r -rozmerná, $r \geq 2$, a teda potrebujeme realizácie Markovovho reťazca $\{X_t\}$, ktoré budú r -rozmernými vektormi. Stratégia v Gibbsovom algoritme je rozdeliť náhodný vektor X_t na k ($1 < k \leq r$) blokov a prechodové jadro definovať ako súčin úplných podmienených hustôt jednotlivých blokov vektora. Nasledujúci stav reťazca dostaneme potom postupným generovaním jednotlivých blokov z týchto úplných podmienených hustôt, pričom podmieňovať budeme vždy naposledy vygenerovanými hodnotami ostatných blokov. Výhoda Gibbsovho algoritmu je v tom, že úplné podmienené hustoty majú menšiu dimenziu a jednoduchší tvar, ako cieľová hustota $p(x)$.

Ak zvolíme štartovací vektor $x^{(0)} = (x_1^{(0)}, \dots, x_k^{(0)})$, vektor $x^{(n+1)} = (x_1^{(n+1)}, \dots, x_k^{(n+1)})$ dostaneme v $(n+1)$ -vej iterácii nasledovne:

1. vygeneruj $x_1^{(n+1)} \sim p(x_1 | x_2^{(n)}, x_3^{(n)}, \dots, x_k^{(n)})$
2. vygeneruj $x_2^{(n+1)} \sim p(x_2 | x_1^{(n+1)}, x_3^{(n)}, \dots, x_k^{(n)})$
3. vygeneruj $x_3^{(n+1)} \sim p(x_3 | x_1^{(n+1)}, x_2^{(n+1)}, x_4^{(n)}, \dots, x_k^{(n)})$
- ⋮
4. vygeneruj $x_k^{(n+1)} \sim p(x_k | x_1^{(n+1)}, x_2^{(n+1)}, \dots, x_{k-1}^{(n+1)})$.

$(n+1)$ -vý vektor MCMC výberu získame teda prechodom cez $k-1$ “medzihodnoty”

$$\begin{aligned} & (x_1^{(n+1)}, x_2^{(n)}, \dots, x_k^{(n)}) \\ & (x_1^{(n+1)}, x_2^{(n+1)}, x_3^{(n)}, \dots, x_k^{(n)}) \\ & \vdots \\ & (x_1^{(n+1)}, \dots, x_{k-1}^{(n+1)}, x_k^{(n)}), \end{aligned}$$

ktoré závisia len od predchádzajúceho vygenerovaného vektora $x^{(n)}$, preto Gibbsov algoritmus simuluje Markovov reťazec s prechodovým jadrom daným hustotou

$$p_G(x, y) = \prod_{i=1}^k p(y_i | y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_k). \quad (31)$$

Gibbsov algoritmus sa dá chápať ako špeciálny prípad M-H algoritmu, ktorým budeme generovať i -tu zložku vektora $x^{(n+1)}$. Inštrumentálna hustota bude mať tvar

$$q(x, y) = \begin{cases} p(y_i | x_{-i}) & y_{-i} = x_{-i} \\ 0 & \text{inak} \end{cases} \quad i = 1, \dots, k,$$

kde $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$. Z takto definovanej $q(x, y)$ vyplýva príslušná pravdepodobnosť prijatia

$$\begin{aligned}\alpha(x, y) &= \frac{p(y)q(y, x)}{p(x)q(x, y)} = \frac{p(y)p(x_i|y_{-i})}{p(x)p(y_i|x_{-i})} = \frac{p(y)/p(y_i|x_{-i})}{p(x)/p(x_i|y_{-i})} = \\ &= \frac{p(y)/p(y_i|y_{-i})}{p(x)/p(x_i|x_{-i})} = \frac{p(y_{-i})}{p(x_{-i})} = 1,\end{aligned}$$

pretože $y_{-i} = x_{-i}$ a pre podmienenú pravdepodobnosť platí

$$p(x_i|x_{-i}) = \frac{p(x_i, x_{-i})}{p(x_{-i})} = \frac{p(x)}{p(x_{-i})}.$$

V každom kroku jediným možným prechodom je teda prechod do stavu y , ktorý má všetky okrem i -tej zložky totožné s predchádzajúcim stavom x . Ak zvolíme

$$x = \left(x_1^{(n+1)}, \dots, x_{i-1}^{(n+1)}, x_i^{(n)}, \dots, x_k^{(n)} \right),$$

potom nasledujúci stav bude

$$y = \left(x_1^{(n+1)}, \dots, x_i^{(n+1)}, x_{i+1}^{(n)}, \dots, x_k^{(n)} \right),$$

čo je práve i -ta “medzihodnota” z Gibbsovho algoritmu.

Niektoré podmienky konvergenie pre Gibbsov algoritmus sa dajú oproti podmienkam všeobecnejšieho M-H algoritmu zjednodušiť, pričom vo väčšine aplikácií sú ľahko splniteľné. Čo sa týka rozdelenia vektora na bloky, odporúča sa vytvárať ich čo najmenej a hlavne zložky vektora s vysokou koreláciou by mali byť v tom istom bloku, inak by mohla byť konvergencia veľmi pomalá.

Gibbsov algoritmus je založený na predpoklade, že vieme generovať hodnoty z jednotlivých úplných podmienených hustôt $p(y_i|x_{-i})$. Niektoré bloky môžu byť jedno- alebo dvojrozmerné a z ich príslušných úplných podmienených hustôt sa dá simulovať priamo. Inak musíme využiť iné simulačné techniky, napr. ARS alebo aj M-H algoritmus (v takom prípade už ide o hybridný MCMC algoritmus, pozri časť 4.4).

4.4 Rozšírenia základných algoritmov

Kvôli potrebám v praxi vzniklo množstvo rozšírení základných algoritmov. Mnohé sú založené na kombinovaní prechodových jadier základných MCMC algoritmov v rámci simulovania jediného Markovovho reťazca, aby sa dosiahla lepšia alebo rýchlejšia konvergencia. Nazývajú sa hybridné MCMC metódy. Ďalšie typy MCMC metód sa snažia riešiť problém nevyhnutnosti vytvárať pre špecifické úlohy zvláštne algoritmy. Pomocou adaptívnych stratégií sa tieto MCMC algoritmy snažia prispôbiť jednotlivým modelom (najmä typom cieľových hustôt), ktoré odhadujú, preto sa nazývajú adaptívne MCMC metódy.

4.4.1 Hybridné MCMC metódy

Hybridné MCMC metódy kombinujú vhodné prechodové jadrá náhodných procesov, pričom vytvoria buď ich zmes (*mixture*), alebo cyklus (*cycle*) tak, aby výsledné prechodové jadro definovalo Markovov reťazec s požadovaným stacionárnym rozdelením. Pritom pri použití jednotlivých prechodových jadier osobitne by výsledkom simulovania nemusela byť nutne realizácia potrebného Markovovho reťazca, ako je tomu pri ich skombinovaní. V MCMC metóde tvorenej zmesou prechodových jadier, sa v každom kroku jednotlivé jadrá vyberajú na základe vopred špecifikovaných pravdepodobností. V metódach založených na cykle prechodových jadier sa jadrá používajú pravidelne v stanovenom poradí.

Aj samotný Gibbsov algoritmus je v podstate cyklom prechodových jadier (ozn. aj Gibbsove prechodové jadrá), ktorými sú úplne podmienené rozdelenia blokov náhodného vektora s cieľovým rozdelením. Výsledné prechodové jadro je dané hustotou (31). Keďže jednotlivé prechodové jadrá v Gibbsovom algoritme sú špeciálnymi prípadmi Metropolis-Hastingsových prechodových jadier, je prirodzené, že v Gibbsovom algoritme môžeme použiť aj krok s iným M-H prechodovým jadrom. Vtedy dostávame algoritmus známy pod názvom Metropolis-within-Gibbs algoritmus. Výhodné je ho použiť vtedy, ak nevieme priamo simulovať z úplných podmienených pravdepodobností niektorých blokov náhodného vektora z Gibbsovho algoritmu.

Podobne, ako v Gibbsovom algoritme používame cyklus Gibbsových prechodových jadier, existuje MCMC metóda, ktorá využíva cyklus jednorozmerných M-H prechodových jadier. Nazýva sa One-Variable-at-a-Time M-H algoritmus a umožňuje vlastne M-H simulácie z viacrozmerného cieľového rozdelenia.

Pri používaní hybridných, rovnako ako základných, MCMC metód treba dať pozor, aby výsledné prechodové jadro malo požadované stacionárne rozdelenie a takisto, aby simulovaný Markovov reťazec spĺňal podmienky konvergencie. Kvôli zložitejšej konštrukcii týchto algoritmov je overenie potrebných vlastností reťazca komplikovanejšie.

4.4.2 Algoritmus s oddialeným zamietnutím

DRA algoritmus patrí medzi adaptívne MCMC metódy a je vlastne upravenou verziou M-H algoritmu. Jeho cieľom je pritom znížiť počet zamietnutí, ktoré v klasickom M-H algoritme môžu viesť k menej efektívnym ergodickým odhadom, nakoľko Markovov reťazec pri zamietnutí zostáva v predchádzajúcom stave. V DRA algoritme však po zamietnutí kandidáta nepriradíme do nasledujúceho stavu ten predchádzajúci, ale vygenerujeme ďalšieho kandidáta, ktorého ako v základnom

M-H algoritme prijme s istou pravdepodobnosťou. Táto pravdepodobnosť prijatia bude daná upraveným predpisom tak, aby cieľové rozdelenie bolo stacionárnym rozdelením simulovaného Markovovho reťazca. Ak aj tohto kandidáta zamietneme, postup generovania ďalšieho kandidáta môžeme opakovať buď kým kandidáta v niektorom z krokov neprijmeme, alebo kým po (napr. aj náhodne) určenom počte krokov proces neukončíme (vtedy nasledujúci stav reťazca bude zhodný s predchádzajúcim stavom).

Nech π je cieľové rozdelenie s hustotou $p(x)$ s nosičom $\mathcal{X} \subset \mathbb{R}^d$ a nech posledná nasimulovaná realizácia Markovovho reťazca je x . Ako v základnom M-H algoritme z inštrumentálnej hustoty $q_1(x, \cdot)$ vygenerujeme kandidáta na nasledujúci stav, y_1 (*first stage proposal*) a vypočítame pravdepodobnosť prijatia

$$\alpha_1(x, y_1) = \min \left\{ 1, \frac{p(y_1)q_1(y_1, x)}{p(x)q_1(x, y_1)} \right\}.$$

Ak y_1 zamietneme, v druhom stupni z novej inštrumentálnej hustoty $q_2(x, y_1, \cdot)$, ktorá môže závisieť nielen od súčasného stavu reťazca x , ale aj od zamietnutého kandidáta y_1 , vygenerujeme ďalšieho kandidáta na nasledujúci stav, y_2 (*second stage proposal*). Pravdepodobnosť prijatia kandidáta z druhého stupňa je definovaná tak, aby pre Markovov reťazec platili podmienky reverzibility (24), a to

$$\alpha_2(x, y_1, y_2) = \min \left\{ 1, \frac{p(y_2)q_1(y_2, y_1)q_2(y_2, y_1, x)[1 - \alpha_1(y_2, y_1)]}{p(x)q_1(x, y_1)q_2(x, y_1, y_2)[1 - \alpha_1(x, y_1)]} \right\}.$$

Po zamietnutí aj kandidáta z druhého stupňa, môžeme tento postup opakovať, pričom pravdepodobnosť prijatia v i -tom stupni bude mať tvar

$$\begin{aligned} \alpha_i(x, y_1, \dots, y_i) &= \\ &= \min \left\{ 1, \frac{p(y_i)q_1(y_i, y_{i-1})q_2(y_i, y_{i-1}, y_{i-2}) \dots q_i(y_i, y_{i-1}, \dots, y_1, x)}{p(x)q_1(x, y_1)q_2(x, y_1, y_2) \dots q_i(x, y_1, \dots, y_i)} \right. \\ &\quad \left. \frac{[1 - \alpha_1(y_i, y_{i-1})][1 - \alpha_2(y_i, y_{i-1}, y_{i-2})] \dots [1 - \alpha_i(y_i, \dots, y_1)]}{[1 - \alpha_1(x, y_1)][1 - \alpha_2(x, y_1, y_2)] \dots [1 - \alpha_i(x, y_1, \dots, y_{i-1})]} \right\}. \end{aligned} \quad (32)$$

Pretože pravdepodobnosti prijatia v každom stupni sú konštruované tak, aby zachovávali podmienky reverzibility, proces oddialenia zamietnutia môže byť prerušený v ľubovoľnom stupni. Je dokázané, že ergodický priemer realizácií Markovovho reťazca zostrojených pomocou DRA algoritmu je efektívnejší (Tierney, Mira (1999)) ako pri realizáciách získaných základným M-H algoritmom.

Je pravda, že opakovaním stále ďalších stupňov DRA algoritmu sa zvyšuje jeho výpočtová zložitosť oproti základnému M-H algoritmu. Treba si však uvedomiť, že získané hodnoty cieľovej hustoty v zamietnutých kandidátoch zo skorších stupňov nie sú zbytočné. Informáciu z nich využijeme pri výbere kandidáta v neskorších

stupňoch, a to stále v rámci jednej iterácie, teda pred určením nasledovného stavu reťazca. V M-H algoritme by tým došlo k porušeniu Markovovej vlastnosti reťazca, kdežto v DRA algoritme nám práve takýto spôsob umožňuje dočasné prispôsobenia (zlepšenia z pohľadu kvality generovania kandidátov) inštrumentálnej hustoty.

4.4.3 Adaptive-Rejection Metropolis Sampling

Metóda ARMS patrí tak isto do adaptívnych MCMC metód, konkrétne je adaptívnym zovšeobecnením špeciálneho prípadu nezávislého M-H algoritmu využívajúceho RS na simulovanie z inštrumentálnej hustoty. ARMS metóda kombinuje výhodný spôsob ARS metódy (časť 3.4.4) na skonštruovanie inštrumentálnej hustoty a M-H algoritmus tak, že sa dá pomocou nej simulovať aj z logaritmicky nekonkávnej hustoty. ARMS metódu navrhli Gilks, Best a Tan (1995) ako nástroj na simulovanie z ľubovoľnej jednorozmernej úplnej podmienenej hustoty v rámci Gibbsovo algoritmu.

Nech $p(x)$ je úplná podmienená hustota (môže byť aj nenormovaná), z ktorej posledná nasimulovaná hodnota je x_P . Pomocou ARMS algoritmu chceme z $p(x)$ vygenerovať novú hodnotu x_M . Oproti ARS algoritmu sa zmení len predpis funkcie $h_n(x)$:

$$h_n(x) = \max[L_{i,i+1}(x, S_n), \min L_{i-1,i}(x, S_n), L_{i+1,i+2}(x, S_n)], \quad 4 \quad (33)$$

$$x_i \leq x \leq x_{i+1},$$

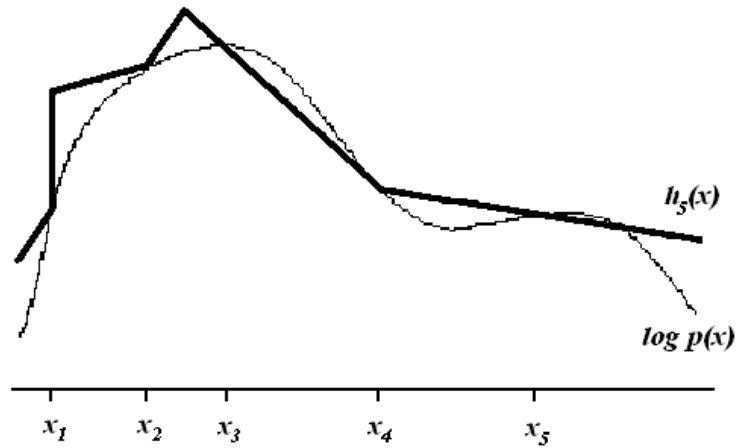
súčasná množina bodov S_n , priamky $L_{i,j}(x, S_n)$ a inštrumentálna hustota $q_n(x)$ zostanú definované rovnako ako v ARS algoritme (pozri časť 3.4.4). Začiatkové body x_i z S_n však musia byť zvolené nezávisle od x_P . Funkcia $h_n(x)$ už vo všeobecnosti nebude obáľkovou funkciou logaritmu cieľovej hustoty, čo je znázornené na obrázku (3).

ARMS algoritmus potom vygeneruje bod x_M v nasledujúcich krokoch:

1. inicializuj n a S_n nezávisle od x_P
2. vygeneruj $y \sim q_n(\cdot)$
3. vygeneruj $u \sim \mathcal{U}(0, 1)$
4. ak $u > \frac{p(y)}{\exp h_n(y)}$, potom

ARS ZAMIETNUTIE:

nastav $S_{n+1} = S_n \cup \{y\}$ a preusporiadaj body v S_{n+1} vzostupne
zvýš n o 1 a choď späť na krok 2



Obr. 3: Grafické znázornenie ARMS metódy, $h_5(x)$ nie je obáľkovou funkciou $\log p(x)$

inak

ARS PRIJATIE:

nastav $x_A = y$.

5. vygeneruj $u \sim \mathcal{U}(0, 1)$

6. ak $u > \min \left\{ 1, \frac{p(x_A) \min\{p(x_P), \exp h_n(x_P)\}}{p(x_P) \min\{p(x_A), \exp h_n(x_A)\}} \right\}$, potom

M-H ZAMIETNUTIE:

nastav $x_M = x_P$

inak

M-H PRIJATIE:

nastav $x_M = x_A$.

7. vráť x_M .

Ak by $p(x)$ bola logaritmicky konkávna, ARMS algoritmus by sa zredukoval na ARS algoritmus, pretože h_n z (33) by nadobudla tvar (11) (teda tvar h_n z ARS) a následne v kroku 6 by vždy nastalo prijatie.

Dobrá voľba štartovacích bodov v S_n je dôležitá z viacerých dôvodov. Fakt, že sú zvolené nezávisle od x_P , zabezpečí stacionaritu cieľového rozdelenia v rámci Gibbsovho algoritmu. Ďalej vhodné štartovacie body môžu znížiť pravdepodobnosť zamietnutia v kroku 6, pričom na ich výber sa dá využiť funkcia $\exp h_N(x)$ (N je hodnota n , pri ktorej sme prešli na krok 5) z predchádzajúcej iterácie Gibbsovho algoritmu. Nakoniec, aby sa ARMS metóda dala použiť aj na neohraničenom nosiči hustoty $p(x)$, štartovacie body musíme voliť opäť podľa pravidiel opísaných v ARS metóde.

5 Analýza výstupu MCMC metód

Hoci sú mnohé z teoretických výsledkov pre konvergenciu MCMC metód sľubné, zväčša sú založené na veľmi prácnych výpočtoch, ktoré sa musia vykonávať pre jednotlivé implementácie MCMC algoritmov zvlášť. Navyše teoretické postupy nie sú všeobecné a ani ich závery nie sú častokrát jednoducho využiteľné v praxi. Preto pre praktické použitie MCMC metód je dôležitejšie rozpoznať praktickú konvergenciu, na čo sa využívajú rôzne metódy konvergenčnej diagnostiky.

Praktická konvergencia zahŕňa

- stanovenie počtu simulácií N_0 , ktoré nebudeme zahŕňať do ergodického priemeru, aby sme minimalizovali vplyv zvolenej štartovacej hodnoty Markovovho reťazca (táto prechodná fáza reťazca sa nazýva *burn-in period*);
- stanovenie celkovej dĺžky N simulovaného reťazca, aby sme získali MCMC výber zo stacionárneho rozdelenia reťazca;
- zistenie miery zmiešavania reťazca (mixing), ktorá hovorí o tom, nakoľko nasimulované hodnoty pokrývajú celú množinu stavov.

Konvergenčná diagnostika predstavuje metódy (formy štatistickej analýzy) na stanovenie praktickej konvergenencie. Nezaručí však stanovenie konvergenencie spoľahlivo a vždy, preto iba o ňu sa opierať nemôžeme. Napríklad v prípade, ak je miera zmiešavania reťazca malá (slow mixing), t.j. hodnoty nasimulovaného reťazca nepochádzajú z celej množiny stavov, ale len z nejakých jej častí, mnohé diagnostické nástroje zlyhajú. Avšak vo väčšine prípadov nám konvergenčná diagnostika poskytne vcelku dobrú predstavu o nasimulovanom reťazci.

Výber vhodných diagnostických nástrojov závisí od konkrétneho problému, v ktorom MCMC metódy používame. Kritériami pre výber by mali byť

- aplikovateľnosť nástroja pre použitý MCMC algoritmus,
- výpočtová náročnosť a celková zložitosť použitia nástroja,
- počet nasimulovaných reťazcov, ktoré nástroj používa,
- typ výstupu, ktorého konvergenciu nástroj monitoruje (celé rozdelenie alebo len skalárne funkcie jeho parametrov),
- interpretovateľnosť výsledkov nástroja (nakoľko je nutná subjektívna interpretácia alebo skúsenosť používateľa).

Samozrejme, že každá metóda konvergenčnej diagnostiky má svoje výhody i nevýhody, takže výber bude ovplyvnený aj subjektívnym hodnotením dôležitosti tohto ktorého kritéria.

Medzi najjednoduchšie nástroje konvergenčnej diagnostiky patria metódy nazývané analýza výstupu (*output analysis*). Ide o (niekedy aj neformálne) metódy založené na monitorovaní vybraného výstupu zo samotného nasimulovaného Markovovho reťazca. Tieto nástroje patria medzi najľahšie implementovateľné, pretože nevyžadujú žiadne dodatočné simulácie, a dajú sa použiť pre ľubovoľný MCMC algoritmus, pretože používajú iba hodnoty nasimulovaného reťazca.

5.1 Odhad MCSE

Štandardná chyba MC aproximácie integrálu (29), teda ergodického priemeru, je vyjadrená pomocou asymptotickej variancie σ^2 (27) z CLT

$$MCSE = SE(\hat{\mu}_N) = \frac{\sigma}{\sqrt{N}} = \frac{1}{\sqrt{N}} \left(\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \right)^{\frac{1}{2}}$$

a určuje presnosť tohto bodového odhadu. Na odhadovanie σ^2 bolo navrhnutých viacero metód, z ktorých kvôli jej jednoduchosti uvedieme metódu skupinových priemerov (*batch means method*).

V metóde skupinových priemerov nasimulovaný Markovov reťazec dĺžky N s hodnotami $x^{(1)}, \dots, x^{(N)}$ rozdelíme na m skupín (*batches*) rovnakej veľkosti (prípadne okrem poslednej) $b = \lfloor \frac{N}{m} \rfloor$ a pre každú definujeme priemer (*batch mean*)

$$\hat{\mu}_{N,k} = \frac{1}{b} \sum_{i=(k-1)b+1}^{kb} g(x^{(i)}), \quad k = 1, \dots, m-1$$

$$\hat{\mu}_{N,m} = \frac{1}{N - b(m-1)} \sum_{i=(m-1)b+1}^N g(x^{(i)}).$$

Podľa (určitého znenia) CLT sú tieto skupinové priemery $\hat{\mu}_{N,k}$ asymptoticky nezávislé a rovnako rozdelené s rozdelením $\mathcal{N}(\mu, \frac{\sigma^2}{b})$. Z toho vyplýva, že σ^2 môžeme odhadnúť ako

$$\hat{\sigma}^2 = \frac{b}{m-1} \sum_{k=1}^m (\hat{\mu}_{N,k} - \hat{\mu}_N)^2. \quad (34)$$

Otázkou pri tejto metóde zostáva voľba dĺžky jednotlivých skupín b . Musí spĺňať $1 \ll b \ll N$, pričom sa odporúča, aby počet skupín m nebol príliš veľký (podľa niektorých autorov, aby nepresiahol 30).

Odhad asymptotickej variancie metódou skupinových priemerov nie je vo všeobecnosti konzistentný, ak však definujeme

$$b_N = \lfloor N^\alpha \rfloor, \quad m_N = \left\lfloor \frac{N}{b} \right\rfloor,$$

kde $\alpha \in (0, 1)$ (často sa volí $\alpha = 0.5$), dostaneme konzistentný odhad σ^2 . Následne jeho použitím získame asymptoticky správny interval spoľahlivosti pre ergodický priemer $\hat{\mu}_N$ s polomerom

$$t_{m_N-1} \frac{\hat{\sigma}}{\sqrt{N}},$$

kde t_{m_N-1} je príslušný kvantil Studentovho t-rozdelenia s $m_N - 1$ stupňami voľnosti.

5.2 Odhad dĺžky burn-in a stanovenie konvergenencie

Najjednoduchšou, no naozaj neformálnou, metódou na odhad dĺžky burn-in je zobrazenie trajektórie (čiže nasimulovaných hodnôt) reťazca v závislosti od času (tzv. *trace plot*). Ak nebude na grafe zrejмый trend, dĺžka burn-in bude čas, od ktorého už budú realizácie reťazca relatívne ustálené. Do grafu je možné zobrazíť aj priebežný priemer (teda priemer všetkých realizácií do času, v ktorom ho zobrazujeme), potom sa nazýva aj *running mean plot*. Ak sa nasimulovaný reťazec zjavne neustálil na nejakej hodnote, konvergenca ešte nebola dosiahnutá. Samozrejme nejde o žiadne spoľahlivé metódy na určenie konvergenencie, dávajú však (prvú) hrubú predstavu o nasimulovanom reťazci.

Jednoduchým a formálnym prístupom na určenie celkového počtu simulácií N , a teda na detekciu dosiahnutia konvergenencie, je metóda založená na výpočte MCSE pomocou konzistentnej metódy skupinových priemerov (popísaná v časti 5.1). N stanovíme tak, aby príslušný intervalový odhad ergodického priemeru bol čo najpresnejší (teda aby mal čo najmenší polomer). Zvolíme si hodnotu ϵ , ktorá bude reprezentovať už dostatočnú presnosť, a N budeme zväčšovať (čiže simulovať ďalšie hodnoty reťazca), až kým nebude splnená podmienka

$$t_{m_N-1} \frac{\hat{\sigma}}{\sqrt{N}} + p(N) \leq \epsilon, \quad (35)$$

kde $p(N) = \epsilon I(n < n^*)$ a $n^* > 0, n^* \in \mathbb{Z}_+$ je fixné. Použitie indikátora $p(N)$ má zabezpečiť, aby simulácie neboli ukončené príliš skoro, pretože pre malé N nie je odhad $\hat{\sigma}$ dobrý. Z podmienok (geometrická ergodicita Markovovho reťazca), ktoré zabezpečia konzistentosť odhadu $\hat{\sigma}^2$, tiež vyplýva, že s pravdepodobnosťou 1 nájdeme dostatočne veľké N v konečnom čase, teda že konvergenca bude dosiahnutá za konečný počet simulácií. Podmienku (35) však treba overovať rozumne často, aby sa podstatne nezvýšila výpočtová náročnosť MCMC algoritmu.

5.3 Odhad miery zmiešavania

Ďalším užitočným grafom je graf odhadnutých autokorelácií (ako funkcie operátora posunu v čase), ktorý indikuje mieru zmiešavania reťazca. Pomaly klesajúci priebeh naznačuje malú mieru zmiešavania, to znamená, že hodnoty neboli simulované z celej množiny stavov a treba v simuláciách pokračovať. Opäť ide o formálne nepodloženú a subjektívne interpretovateľnú metódu.

CUSUM metóda (Yu a Mykland (1998), Brooks (1996)) je graficko-štatistická metóda, ktorá dáva kvalitatívnu mieru zmiešavania využívajúc priebeh kumulatívnych súm nejakej skalárnej štatistiky, ktorú odhadujeme z realizácií $x^{(1)}, \dots, x^{(N)}$ jedného nasimulovaného reťazca. Postup zostrojenia priebehu kumulatívnych súm (*CUSUM path plot*) je nasledovný:

1. vypočítame ergodický priemer štatistiky $g(\cdot)$

$$\hat{\mu}_N = \frac{1}{N - N_0} \sum_{j=N_0+1}^N g(x^{(j)}),$$

kde N_0 dĺžka je burn-in,

2. vypočítame kumulatívne (čiastkové) sumy

$$\hat{S}_t = \sum_{j=N_0+1}^t (g(x^{(j)}) - \hat{\mu}_N), \quad t = N_0 + 1, \dots, N,$$

3. zobrazíme $\{\hat{S}_t\}$ vzhľadom na t pre $t = N_0 + 1, \dots, N$ a spojíme susedné body úsečkami.

O rýchlosti zmiešavania potom svedčí hladkosť výsledného grafu kumulatívnych súm, a to: hladký priebeh indikuje malú mieru zmiešavania, kým “strapatý” (“*hairly*”) priebeh hovorí o veľkej miere zmiešavania.

Nedostatok metódy - subjektívnosť určovania hladkosti priebehu kumulatívnych súm - sa snažili autori odstrániť stanovením referenčného grafu kumulatívnych súm. Ten je založený na náhodnom výbere z normálneho rozdelenia so strednou hodnotou a varianciou, ktoré sa rovnajú príslušným výberovým momentom realizácií Markovovho reťazca. Dobrá miera zmiešavania potom zodpovedá veľkej podobnosti referenčného CUSUM grafu a grafu kumulatívnych súm zostrojeného na základe hodnôt Markovovho reťazca.

Brooks (1996) nakoniec matematicky definoval “strapatosť” pomocou

$$d_j = \begin{cases} 1 & \text{ak } \hat{S}_{j-1} > \hat{S}_j \text{ a } \hat{S}_j < \hat{S}_{j+1} \text{ alebo } \hat{S}_{j-1} < \hat{S}_j \text{ a } \hat{S}_j > \hat{S}_{j+1} \\ 0 & \text{inak} \end{cases},$$

$j = N_0 + 1, \dots, N - 1$. Potom

$$D_t = \frac{1}{t - N_0 - 1} \sum_{j=N_0+1}^{t-1} d_j \quad N_0 + 2 \leq t \leq N$$

nadobúda hodnoty medzi 0 a 1, pričom 0 znamená úplne hladký priebeh a 1 znamená maximálne "strapatý" priebeh. d_j má hodnotu 1, ak je bod (j, \hat{S}_j) lokálnym extrémom priebehu kumulatívnych súm, teda $\sum d_j$ vlastne vyjadruje koľkokrát trajektória $g(x^j)$ pretne $\hat{\mu}_N$. Dá sa ukázať, že D_t je realizáciou binomického rozdelenia so strednou hodnotou 0.5 a varianciou $\frac{1}{4(t-N_0-1)}$, preto podľa zákona veľkých čísel pre dostatočne veľké t budú D_t normálne rozdelené a konvergenciu môžeme konštatovať, ak aspoň $100(1 - \frac{\alpha}{2})\%$ z nich leží v intervaloch

$$\left(0.5 - u_{\frac{\alpha}{2}} \sqrt{\frac{1}{4(t - N_0 - 1)}}, 0.5 + u_{\frac{\alpha}{2}} \sqrt{\frac{1}{4(t - N_0 - 1)}} \right),$$

kde $u_{\frac{\alpha}{2}}$ je príslušný kvantil $\mathcal{N}\left(0.5, \frac{1}{4(t-N_0-1)}\right)$. Hranice týchto intervalov sú určené za predpokladu, že d_j sú nezávislé a rovnako rozdelené, preto ich treba brať len približne.

Pomocou štatistiky D_t sa dá určiť aj dĺžka N_0 burn-in. Ak máme N nasimulovaných hodnôt Markovovho reťazca, postup je nasledovný:

- zvolíme $n = \frac{N}{20}$, $n_0 = \frac{n}{2}$ a na základe týchto prvých n hodnôt (pričom za dĺžku burn-in považujeme n_0) vypočítame hodnotu D_n ,
- predchádzajúci výpočet opakujeme pre prvých $2n, 4n, \dots, N$ realizácií reťazca,
- zobrazíme priebeh hodnôt D_n, D_{2n}, \dots, D_N a čas, od ktorého sa táto postupnosť ustáli na nejakej hodnote, označíme t^* .

Dĺžku burn-in môžeme následne odhadnúť ako

$$\hat{N}_0 = \frac{t^*}{2}.$$

6 Praktická aplikácia MCMC metód na odhadovanie SV modelov

V tejto časti práce ukážeme praktické použitie MCMC metód pri odhadovaní SV modelov. Konkrétne pôjde o dva typy SV modelov - Taylorov (log-normálny) SV model a SGED SV model, pričom prvý z nich budeme odhadovať dvoma rôznymi metódami. V každom z problémov použijeme tri sady dát, a to výnosy akciových indexov S&P500, DAX a SAX. Na získané výsledky nakoniec aplikujeme popísané metódy analýzy výstupu.

6.1 Popis dát

Na praktickú analýzu sme zvolili spomenuté tri akciové indexy. Ako zástupcu Ameriky sme vybrali S&P500, ktorý je u odborníkov považovaný za štandardné meradlo výkonnosti amerického akciového trhu, z európskych indexov sme siahli po DAX-e, ktorý je zasa vedúcim nemeckým akciovým indexom. SAX sme zvolili preto, lebo sme sa chceli aplikovať SV modely aj na slovenské dáta. Zdrojom dát boli internetové stránky <http://finance.yahoo.com/> a <http://www.bsse.sk/>.

	Taylorov SV model	SGED SV model
S&P500	30. 4. 2003 - 20. 4. 2007	25. 4. 2005 - 20. 4. 2007
DAX	23. 5. 2003 - 20. 4. 2007	9. 5. 2005 - 20. 4. 2007
SAX	5. 3. 2003 - 20. 4. 2007	24. 3. 2005 - 20. 4. 2007

Tabuľka 2: Rozsah použitých dát

Pre Taylorov SV model sme použili 1001 a pre SGED SV model 501 hodnôt S_t každého z indexov za roky uvedené v Tabuľke 2. Výnosy Y_t indexov sme získali štandardnou logaritmicou transformáciou

$$Y_t = \log S_{t+1} - \log S_t = \log \frac{S_{t+1}}{S_t},$$

čím sa zmenšil rozsah dát o 1. Deskriptívne charakteristiky takto upravených dát sú uvedené v Tabuľke 3. Hodnoty šikmosti a špicatosti naznačujú, že marginálne rozdelenia výnosov indexov sú mierne ľavostranne zošikmené a majú ťažšie chvosty ako normálne rozdelenie.

Na simulovanie, odhad parametrov modelov a realizácií volatility, ako aj na analýzu výstupu sme používali MATLAB 7.0.4, v ktorom sme naprogramovali všetky potrebné funkcie.

6.2 Taylorov SV model

Pri voľbe parametrizácie modelu a spôsobe jeho odhadovania sme vychádzali z článku Jacquier, Polson a Rossi (1994). Uvažujeme teda model v tvare:

$$y_t = \sqrt{v_t} \epsilon_t$$
$$\log v_{t+1} = \alpha + \beta \log v_t + \sigma_H \eta_t,$$

pričom $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$ sú nezávislé. Ďalej označíme výnosy indexov $Y = \{y_t\}_{t=1}^T$, varianciu výnosov $V = \{v_t\}_{t=1}^T$ a parametre modelu $\theta = \{\alpha, \beta, \sigma_H^2\}$, kde $T = 1000$.

	Taylorov SV model			SGED SV model		
	S&P500	DAX	SAX	S&P500	DAX	SAX
počet pozorovaní	1000	1000	1000	500	500	500
priemer	0.0482	0.0956	0.0924	0.0489	0.1074	-0.0248
medián	0.0857	0.1263	0.0589	0.0874	0.1686	0.0094
minimum	-3.5343	-3.5162	-5.0268	-3.5343	-3.4633	-4.2268
maximum	2.2138	4.2530	4.0738	2.1336	2.6051	4.0738
štand. odchýlka	0.6994	1.0029	1.0611	0.6397	0.9097	0.9422
šikmost'	-0.1574	-0.2308	-0.2732	-0.3174	-0.4089	-0.3886
špicatosť	3.9038	4.0208	6.3346	5.3203	3.9260	7.3922

Tabuľka 3: Deskriptívne charakteristiky výnosov

S využitím rozkladu (8) môžeme združenú aposteriórnu hustotu vyjadriť v tvare

$$\begin{aligned}
p(V, \theta|Y) &\propto p(V|\theta, Y)p(\theta|V, Y) \\
&\propto p(V|\theta, Y)p(\alpha, \beta|\sigma_H^2, V, Y)p(\sigma_H^2|\alpha, \beta, V, Y).
\end{aligned}$$

Simulovanie parametrov modelu

Rovnica pre log-varianciu je vlastne jednorozmerná lineárna regresia, preto ak zvolíme štandardné apriórne rozdelenia parametrov

$$\begin{aligned}
\alpha, \beta &\sim \mathcal{N}(a, A) \\
\sigma_H^2 &\sim \mathcal{IG}(b, B),
\end{aligned}$$

ich úplné podmienené rozdelenia budú mať tvar

$$\begin{aligned}
\alpha, \beta|\sigma_H^2, V, Y &\sim \mathcal{N}(a^*, A^*) \\
\sigma_H^2|\alpha, \beta, V, Y &\sim \mathcal{IG}(b^*, B^*).
\end{aligned}$$

Upravené hyperparametre sú dané vzťahmi

$$\begin{aligned}
a^* &= A^*(A^{-1}a + \sigma_H^{-2}W'Z), \quad A^* = (A^{-1} + \sigma_H^{-2}W'W)^{-1} \\
b^* &= b + \frac{T}{2}, \quad B^* = B + (Z - W\theta_{-\sigma_H^2})'(Z - W\theta_{-\sigma_H^2}),
\end{aligned}$$

kde

$$\begin{aligned}
Z &= (\log v_2, \dots, \log v_T)' \\
W &= \begin{pmatrix} 1 & \log v_1 \\ \vdots & \\ 1 & \log v_{T-1} \end{pmatrix} \\
\theta_{-\sigma_H^2} &= (\alpha, \beta)'.
\end{aligned}$$

Simulovanie realizácií variancie

Podmienená aposteriórna hustota pre varianciu sa dá rozložiť na jednorozmerné úplné podmienené hustoty nasledovne

$$p(V|\theta, Y) \propto p(Y|\theta, V)p(V|\theta) \propto \prod_{t=1}^T p(v_t|v_{t-1}, v_{t+1}, \theta, Y),$$

pričom pre $t = 1, \dots, T$

$$\begin{aligned} p(v_t) &\stackrel{\text{ozn.}}{=} p(v_t|v_{t-1}, v_{t+1}, \theta, Y) \propto p(y_t|v_t, \theta)p(v_{t+1}|v_t, \theta)p(v_t|v_{t-1}, \theta) \propto \\ &\propto v_t^{-\frac{1}{2}} \exp\left(-\frac{y_t^2}{2v_t}\right) \times \exp\left(-\frac{(\log v_t - \alpha - \beta \log v_{t-1})^2}{2\sigma_H^2}\right) v_t^{-1} \times \\ &\quad \times \exp\left(-\frac{(\log v_{t+1} - \alpha - \beta \log v_t)^2}{2\sigma_H^2}\right) = \\ &= v_t^{-\frac{1}{2}} \exp\left(-\frac{y_t^2}{2v_t}\right) v_t^{-1} \exp\left(-\frac{(\log v_t - \mu_t)^2}{2\sigma^2}\right), \end{aligned}$$

kde

$$\mu_t = \frac{\alpha(1 - \beta) + \beta(\log v_{t-1} + \log v_{t+1})}{1 + \beta^2}, \quad \sigma^2 = \frac{\sigma_H^2}{1 + \beta^2}.$$

Potrebné v_0 a v_{T+1} získame z rovnice pre varianciu takto

$$\begin{aligned} \log v_0 &= \alpha + \beta \log v_1 + \sigma_H \eta_0, & \eta_0 &\sim \mathcal{N}(0, 1) \\ \log v_{T+1} &= \alpha + \beta \log v_T + \sigma_H \eta_T, & \eta_T &\sim \mathcal{N}(0, 1). \end{aligned}$$

Aj keď hodnoty variancie budeme simulovať po jednom, ani tieto jednorozmerné úplné podmienené hustoty nie sú hustotami žiadneho štandardného rozdelenia, preto použijeme RS metódu kombinovanú s M-H algoritmom. Za inštrumentálnu hustotu $q(x)$ pre RS metódu sme zvolili hustotu rozdelenia $\mathcal{IG}(d, D_t)$ s parametrami danými

$$\begin{aligned} d &= 0.5 + \frac{1 - 2 \exp(-\sigma^2)}{1 - \exp(-\sigma^2)} \\ D_t &= (d - 1) \exp(\mu_t + 0.5\sigma^2) + 0.5y_t^2, \end{aligned}$$

pretože dobre aproximuje cieľovú hustotu. Keďže RS doplníme aj M-H krokom, konštantu c nemusíme voliť tak, aby inštrumentálna hustota bola obáľkovou funkciou cieľovej hustoty, preto

$$c = 1.1 \frac{p(m)}{q(m)},$$

kde $m = \frac{D_t}{d+1}$ je modus zvoleného inverzného gama rozdelenia. Kandidáta y prijatého v RS algoritme, prijmeme za nový stav reťazca, ak ho prijmeme aj v M-H kroku, ktorý zodpovedá nezávislému M-H algoritmu s inštrumentálnou hustotou

$$f(x) \propto \min \{p(x), cq(x)\}.$$

V opačnom prípade je nový stav rovný predchádzajúcemu.

Druhý spôsob, ktorý sme použili na simulovanie hodnôt variancie je ARS metóda, pretože jednorozmerné podmienené hustoty $p(x)$ sú logaritmicky konkávne funkcie (podľa Geweke (1994)).

MCMC výber $\left\{ \alpha^{(j)}, \beta^{(j)}, \sigma_H^{2(j)}, V^{(j)} \right\}_{j=1}^N$ z $p(V, \theta | Y)$ získame teda nasledujúcim hybridným MCMC algoritmom:

1. z úplných podmienených rozdelení simulujeme jednotlivé parametre
 - (a) $(\alpha^{(j)}, \beta^{(j)})$ z $p(\alpha, \beta | \sigma_H^{2(j-1)}, V^{(j-1)}, Y)$ priamo
 - (b) $\sigma_H^{2(j)}$ z $p(\sigma_H^2 | \alpha^{(j)}, \beta^{(j)}, V^{(j-1)}, Y)$ priamo
2. z úplných podmienených rozdelení simulujeme po jednej zložky vektora variancií

$$v_t^{(j)} \text{ z } p(v_t | v_{t-1}^{(j)}, v_{t+1}^{(j-1)}, \theta^{(j)}, y_t), \quad t = 1, \dots, T$$

buď pomocou kombinácie RS metódy a M-H algoritmu alebo pomocou ARS metódy.

Hodnoty hyperparametrov apriórnych rozdelení parametrov modelu sme zvolili

$$a = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad A = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix},$$

$$b = 2.5, \quad B = 0.025.$$

Štartovacie hodnoty parametrov modelu sme vygenerovali z ich apriórnych rozdelení a štartovacie hodnoty variancie z $\mathcal{IG}(2.5, 0.25)$.

Výsledky

Celkovo sme urobili $N = 5000$ simulácií, z čoho $N_0 = 1000$ predstavovalo burn-in. Výstupom simulácií sú trajektórie jednotlivých parametrov modelu a realizácií variancie. Bodové odhady sme z nich dostali ako výberový priemer. Ďalej pre všetky bodové odhady sme vypočítali odhady MCSE, štandardných odchýlok a 95%-ných intervalov spoľahlivosti. Tieto hodnoty sú zhrnuté v Tabuľke 5 v Prílohe 2.

Pomocou priebehov trajektórií s priebežnými priermi sme overili, že dĺžka burn-in by mala byť dostatočná, hlavne pri použití ARS metódy. Čo sa týka miery zmiešavania, na základe grafov autokorelácií, kumulatívnych súm a mier “strapastosti” môžeme konštatovať, že lepšia bola dosiahnutá použitím ARS metódy. Pritom sme skúmali mieru zmiešavania reťazca len na podmnožine množiny stavov prislúchajúcej parametrom SV modelu. Najlepšia bola dosiahnutá pre parameter β , a to

pre všetky tri sady dát a obe metódy odhadovania. V Prílohe 2 sú na ilustráciu zobrazené uvedené grafy pre parametre modelu odhadovaného metódou ARS pre akciový index SAX.

Nakoniec na obrázkoch 4 a 5 sú zobrazené odhadnuté realizácie variancie výnosov jednotlivých indexov v porovnaní so samotnými výnosmi. Vidno, že prudkým výkyvom hodnôt výnosov zväčša zodpovedajú vyššie hodnoty odhadnutých variancií.

Z pohľadu kvality odhadov i konvergencie sa nám celkovo javila lepšie použitá metóda ARS, avšak za cenu dlhšieho trvania simulácií (cca. 2,5-násobne dlhšie oproti kombinácií RS metódy a M-H algoritmu).

6.3 SGED SV model

Model sme zostavili a odhadovali na základe článkov Cappuccio, Lubian a Raggi (2004) a Kim, Shephard a Chib (1998). Model má teda tvar:

$$y_t = \beta \exp\left(\frac{h_t}{2}\right) \epsilon_t$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_H \eta_t,$$

pričom $\epsilon_t \sim SGED(\lambda, \nu)$ a $\eta_t \sim \mathcal{N}(0, 1)$ sú nezávislé, $h_t = \log v_t$, $\beta = \exp(\mu/2)$. Prepokladáme, že

$$E[\epsilon_t] = 0, \quad Var[\epsilon_t] = 1$$

$$h_1 \sim \mathcal{N}\left(\mu, \frac{\sigma_H^2}{1 - \phi^2}\right).$$

Ďalej označíme výnosy indexov $Y = \{y_t\}_{t=1}^T$, log-varianciu výnosov $H = \{h_t\}_{t=1}^T$ a parametre modelu $\theta = \{\mu, \phi, \sigma_H^2, \lambda, \nu\}$, kde $T = 500$.

Opäť s využitím rozkladu (8) vyjadríme združenú aposteriórnu hustotu nasledovne

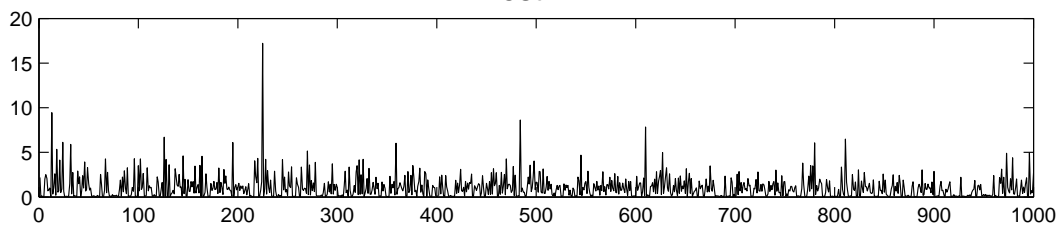
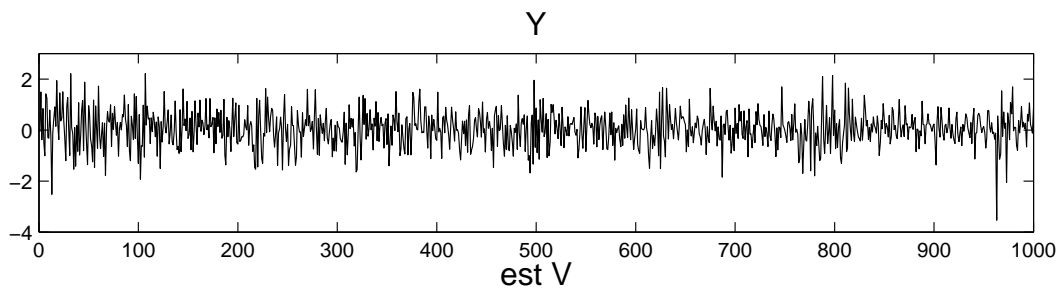
$$\begin{aligned} p(H, \theta|Y) &\propto p(H|\theta, Y)p(\theta|H, Y) \propto \\ &\propto p(H|\theta, Y)p(\mu|\phi, \sigma_H^2, \lambda, \nu, H, Y)p(\phi|\mu, \sigma_H^2, \lambda, \nu, H, Y) \times \\ &\times p(\sigma_H^2|\mu, \phi, \lambda, \nu, H, Y)p(\lambda|\mu, \phi, \sigma_H^2, \nu, H, Y)p(\nu|\mu, \phi, \sigma_H^2, \lambda, H, Y). \end{aligned}$$

Simulovanie parametrov modelu

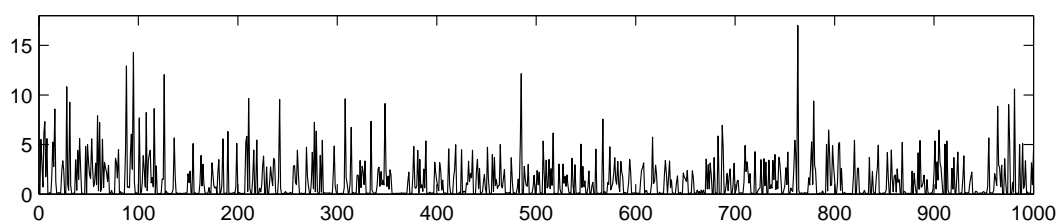
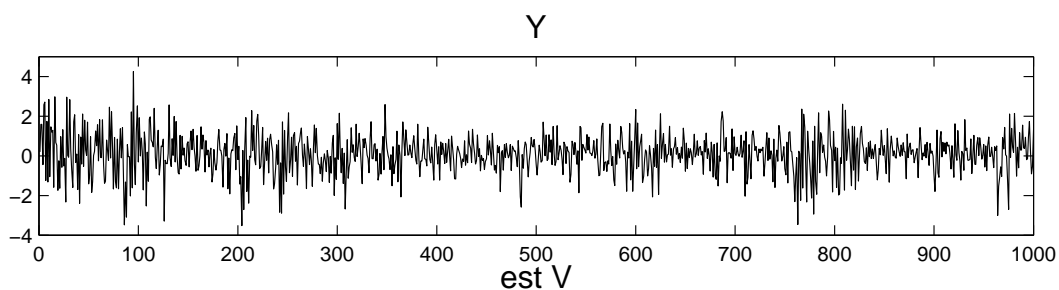
Ak označíme jednotlivé parametre θ_i , úplné (aposteriórne) podmienené hustoty parametrov majú tvar

$$p(\theta_i|\theta_{-i}, H, Y) \propto p(\theta_i)p(h_1|\theta_{-i}) \prod_{t=1}^{T-1} p(h_{t+1}|h_t, \theta_{-i}) \quad \text{pre } \mu, \phi \text{ a } \sigma_H^2$$

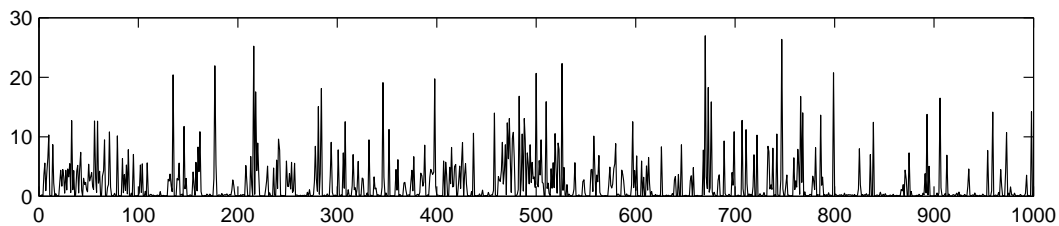
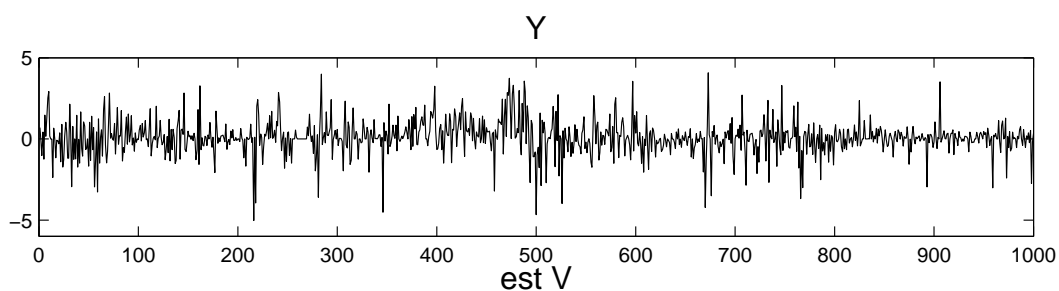
$$p(\theta_i|\theta_{-i}, H, Y) \propto p(\theta_i) \prod_{t=1}^{T-1} p(y_t|h_t, \theta_{-i}) \quad \text{pre } \lambda \text{ a } \nu.$$



(a) S&P500

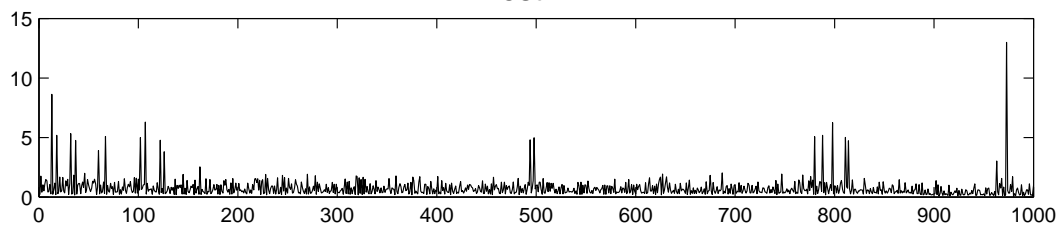
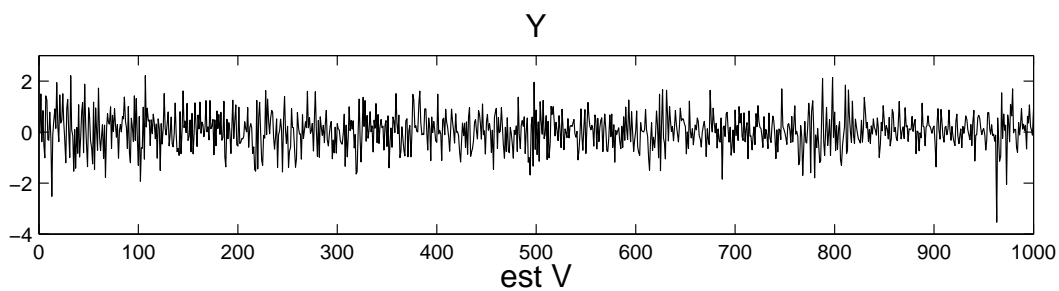


(b) DAX

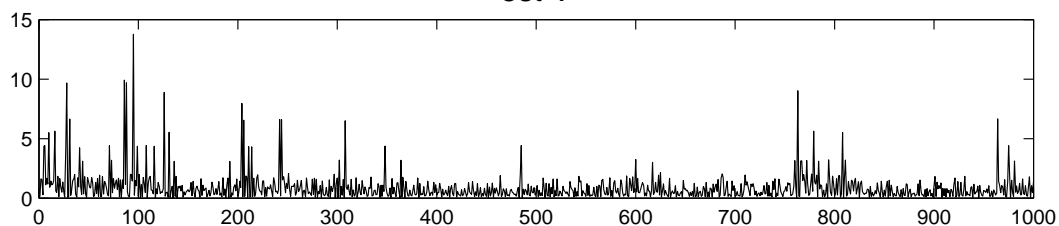
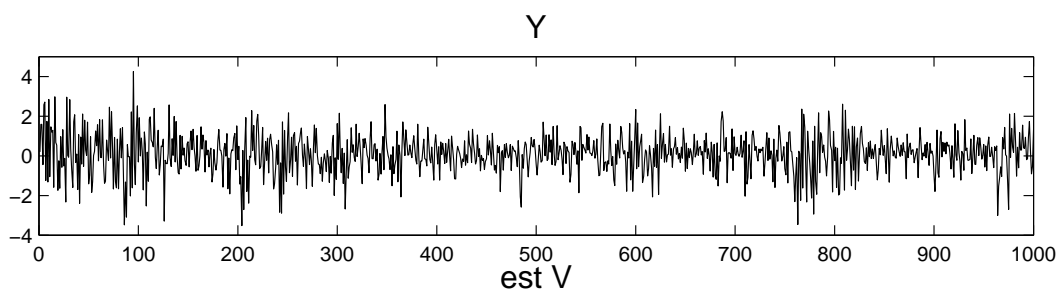


(c) SAX

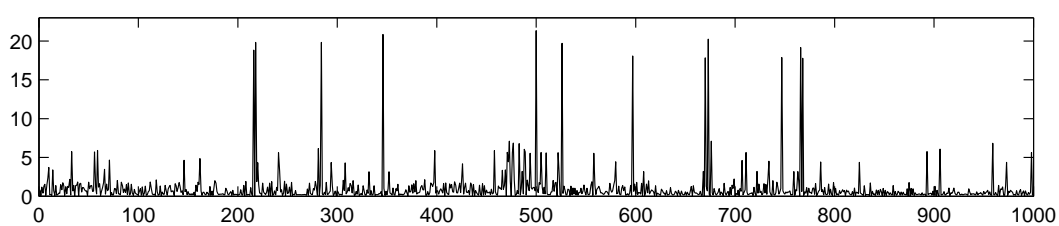
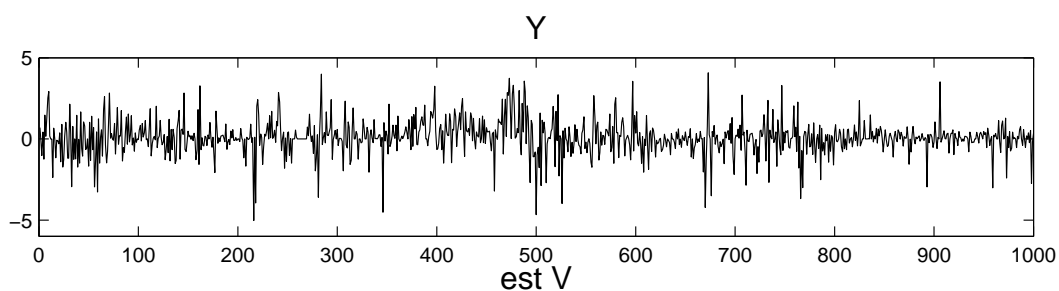
Obr. 4: Odhadnuté realizácie variance v Taylorovom SV modeli pomocou RS M-H



(a) S&P500



(b) DAX



(c) SAX

Obr. 5: Odhadnuté realizácie variancie v Talorovnom SV modeli pomocou ARS

Simulovanie μ

Pre parameter μ uvažujeme neinformatívnu apriórnu hustotu, z ktorej preň vyplýva tvar úplného podmieneného rozdelenia

$$\mu|\phi, \sigma_H^2, \lambda, \nu, H, Y \sim \mathcal{N}(c^*, C^*),$$

s hyperparametrami danými

$$c^* = C^* \left[\frac{1 - \phi^2}{\sigma_H^2} h_1 + \frac{1 - \phi}{\sigma_H^2} \sum_{t=1}^{T-1} (h_{t+1} - \phi h_t) \right], \quad C^* = \frac{\sigma_H^2}{T(1 - \phi)^2}.$$

Simulovanie ϕ

Apriórnu hustotu pre parameter ϕ špecifikujeme

$$p(\phi) \propto \left(\frac{1 + \phi}{2} \right)^{b-1} \left(\frac{1 - \phi}{2} \right)^{B-1}, \quad b, B > 0.5,$$

pričom prakticky nasimulujeme $\phi^* \sim \mathcal{B}(b, B)$ a položíme $\phi = 2\phi^* - 1$. Z úplnej podmienenej hustoty

$$p(\phi|\mu, \sigma_H^2, \lambda, \nu, H, Y) \propto \left(\frac{1 + \phi}{2} \right)^{b-1} \left(\frac{1 - \phi}{2} \right)^{B-1} (1 - \phi^2)^{0.5} \exp \left(-\frac{(h_1 - \mu)^2(1 - \phi^2)}{2\sigma_H^2} \right) \times \\ \times \exp \left(-\frac{1}{2\sigma_H^2} \sum_{t=1}^{T-1} [h_{t+1} - \mu - \phi(h_t - \mu)]^2 \right)$$

sa nedá simulovať priamo, ale je logaritmicky konkávna, preto použijeme ARS metódu.

Simulovanie σ_H^2

Pre parameter σ_H^2 zvolíme konjugovanú apriórnu hustotu rozdelenia $\mathcal{IG}(a, A)$, teda úplná podmienená hustota bude hustota rozdelenia $\mathcal{IG}(a^*, A^*)$ s hyperparametrami upravenými nasledovne

$$a^* = a + \frac{T}{2}, \quad A^* = A + \frac{1}{2} \left((h_1 - \mu)^2(1 - \phi^2) + \sum_{t=1}^{T-1} [h_{t+1} - \mu - \phi(h_t - \mu)]^2 \right).$$

Simulovanie λ a ν

Aby šoky ϵ_t v rovnici pre výnosy mali strednú hodnotu 0 a varianciu 1, zavedieme

$$\epsilon_t = \lambda_2 X + \lambda_1,$$

kde $X \sim \mathcal{SGED}(\lambda, \nu)$, $\lambda_1 = -E(X)/\sqrt{\text{Var}(X)}$ a $\lambda_2 = 1/\sqrt{\text{Var}(X)}$.

Keďže za apriórne rozdelenia parametrov λ a ν zvolíme

$$\lambda \sim \mathcal{U}(\lambda_L, \lambda_H), \quad \nu \sim \mathcal{U}(\nu_L, \nu_H),$$

úplné podmienené hustoty týchto parametrov majú tvar

$$p(\lambda|\mu, \phi, \sigma_H^2, \nu, H, Y) \propto \prod_{t=1}^T \frac{1}{\lambda_2} \exp\left(-\frac{1}{2}|z_t|^\nu\right) \left[1 + \operatorname{sgn}(\lambda z_t) \gamma\left(\frac{1}{2}|\lambda z_t|^\nu, \frac{1}{\nu}\right)\right]$$

$$p(\nu|\mu, \phi, \sigma_H^2, \lambda, H, Y) \propto \prod_{t=1}^T \frac{\nu}{2^{1+\frac{1}{\nu}} \Gamma(\frac{1}{\nu}) \lambda_2} \exp\left(-\frac{1}{2}|z_t|^\nu\right) \left[1 + \operatorname{sgn}(\lambda z_t) \gamma\left(\frac{1}{2}|\lambda z_t|^\nu, \frac{1}{\nu}\right)\right],$$

pričom

$$z_t = \frac{\exp\left(-\frac{h_t}{2}\right) y_t - \lambda_1}{\lambda_2},$$

$\Gamma(a) = \int_0^\infty e^{-t} t^{a-1} dt$ je gama funkcia, $\gamma(x, a) = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt$ je neúplná gama funkcia. Na simulovanie z týchto hustôt využijeme ARMS metódu, lebo nejde o logaritmicky konkávne funkcie.

Simulovanie realizácií log-variance

Podmienená a posteriori hustota variance $p(H|\theta, Y)$ sa dá znova rozložiť na jednorozmerné úplné podmienené hustoty tvaru

$$p(h_t|h_{-t}, \theta, Y) \propto p(y_t|h_t, \theta) p(h_{t+1}|h_t, \theta) p(h_t|h_{t-1}, \theta) \propto$$

$$\propto \exp\left(-\frac{h_t}{2}\right) \exp\left(-\frac{1}{2}|z_t|^\nu\right) \left[1 + \operatorname{sgn}(\lambda z_t) \gamma\left(\frac{1}{2}|\lambda z_t|^\nu, \frac{1}{\nu}\right)\right] \times$$

$$\times \exp\left(-\frac{1}{2\sigma_H^2} [h_{t+1} - \mu - \phi(h_t - \mu)]^2\right)$$

$$\times \exp\left(-\frac{1}{2\sigma_H^2} [h_t - \mu - \phi(h_{t-1} - \mu)]^2\right), \quad t = 2, \dots, T-1$$

a pre krajné zložky vektora variancií platí

$$p(h_1|h_{-1}, \theta, Y) \propto p(y_1|h_1, \theta) p(h_2|h_1, \theta) p(h_1|\theta),$$

$$p(h_T|h_{-T}, \theta, Y) \propto p(y_T|h_T, \theta) p(h_T|h_{T-1}, \theta).$$

Z predpisov týchto hustôt je zrejmé, že priame simulácie nie sú možné, namiesto toho použijeme dvojstupňový DRA algoritmus. V prvom stupni pôjde o nezávislý M-H algoritmus s inštrumentálnou hustotou q_1 danou rozdelením $\mathcal{N}(d, D_t)$ s parametrami

$$d = \frac{\sigma_H^2}{1 + \phi^2}, \quad D_t = h_t^* + 0.5d[y_t^2 \exp(-h_t^*) - 1],$$

kde

$$h_t^* = \mu + \frac{\phi[h_{t-1} + h_{t+1} - 2\mu]}{1 + \phi^2},$$

kým v druhom stupni zvolíme inštrumentálnu hustotu z RW M-H algoritmu danú rozdelením $\mathcal{N}(0, D_t)$.

MCMC výber $\left\{\mu^{(j)}, \phi^{(j)}, \sigma_H^{2(j)}, \lambda^{(j)}, \nu^{(j)}, H^{(j)}\right\}_{j=1}^N$ z $p(H, \theta|Y)$ aj v tomto modeli dostaneme hybridným MCMC algoritmom:

1. z úplných podmienených rozdelení simulujeme po jednej zložky vektora log-variancií

$$h_t^{(j)} \text{ z } p(h_t|h_{t-1}^{(j)}, h_{t+1}^{(j-1)}, \theta^{(j)}, y_t), \quad t = 1, \dots, T$$

pomocou DRA algoritmu

2. z úplných podmienených rozdelení simulujeme jednotlivé parametre

(a) $\sigma_H^{2(j)}$ z $p(\sigma_H^2|\phi^{(j-1)}, \mu^{(j-1)}, \nu^{(j-1)}, \lambda^{(j-1)}, H^{(j)})$ priamo

(b) $\phi^{(j)}$ z $p(\phi|\sigma_H^{2(j)}, \mu^{(j-1)}, \nu^{(j-1)}, \lambda^{(j-1)}, H^{(j)})$ pomocou ARS metódy

(c) $\mu^{(j)}$ z $p(\mu|\sigma_H^{2(j)}, \phi^{(j)}, \nu^{(j-1)}, \lambda^{(j-1)}, H^{(j)})$ priamo

(d) $\nu^{(j)}$ z $p(\nu|\sigma_H^{2(j)}, \phi^{(j)}, \mu^{(j)}, \lambda^{(j-1)}, H^{(j)})$ pomocou ARMS metódy

(e) $\lambda^{(j)}$ z $p(\lambda|\sigma_H^{2(j)}, \phi^{(j)}, \mu^{(j)}, \nu^{(j)}, H^{(j)})$ pomocou ARMS metódy.

Hodnoty hyperparametrov apriórnych rozdelení parametrov modelu sme zvolili

$$a = 2, 5, \quad A = 0.025,$$

$$b = 20, \quad B = 1.5$$

$$\lambda_L = -3 \quad \lambda_H = 3$$

$$\nu_L = 1 \quad \nu_H = 3.$$

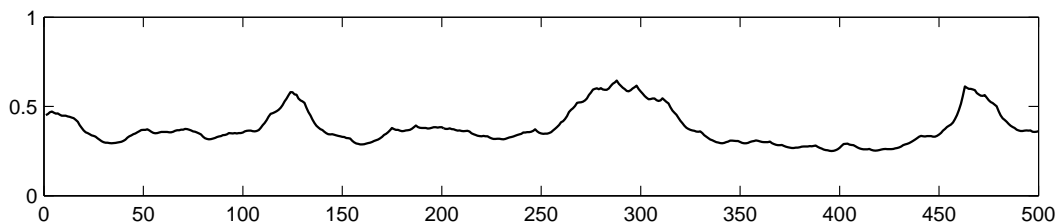
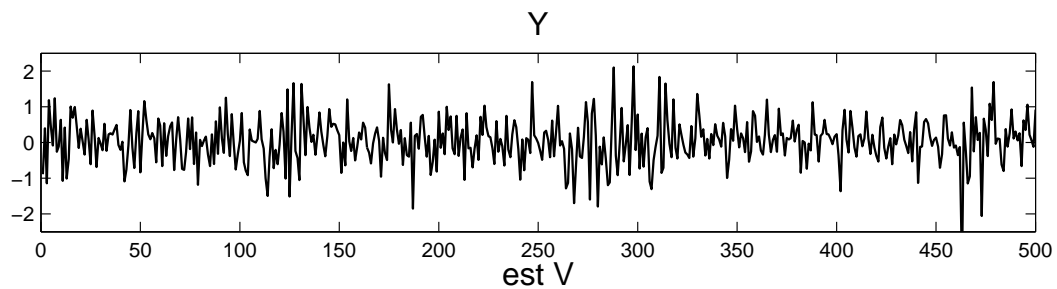
Štartovacie hodnoty μ sme vygenerovali z $\mathcal{N}(0, 20)$, štartovacie hodnoty ostatných parametrov modelu z ich apriórnych rozdelení a štartovacie hodnoty variancie z $\mathcal{N}\left(\mu^{(0)}, \frac{\sigma_H^{2(0)}}{1-\phi^{(0)2}}\right)$.

Výsledky

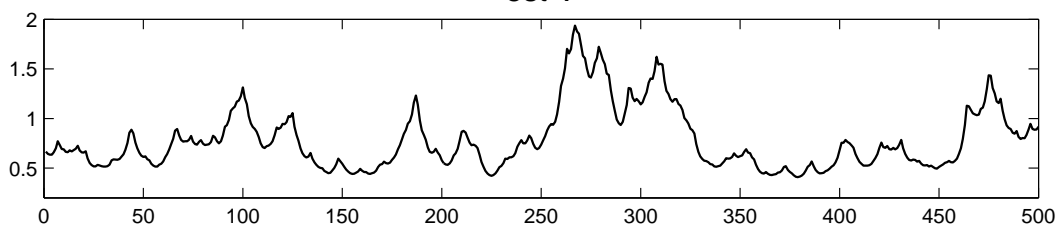
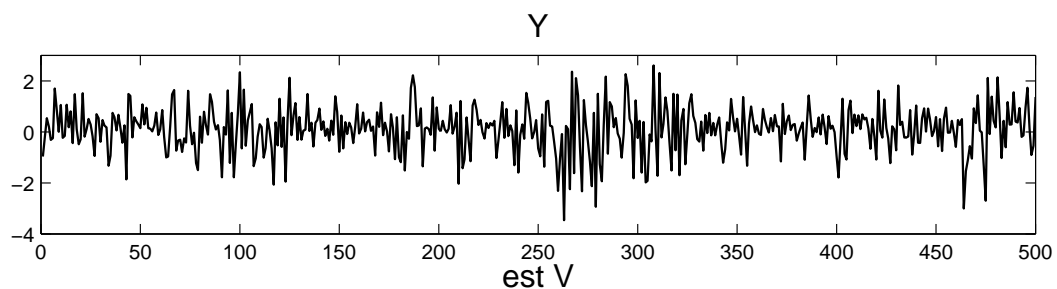
V tomto modeli sme tiež urobili celkovo $N = 5000$ simulácií, z čoho $N_0 = 1000$ bola dĺžka burn-in. Bodové odhady jednotlivých parametrov modelu aj realizácií log-variancie sme získali opäť ako výberový priemer z ich trajektórií, ktoré boli výstupom modelu. V Tabuľke 4 v Prílohe 2 sú uvedené aj odhady ich MCSE, štandardných odchýlok a 95%-ných intervalov spoľahlivosti. Najzaujímavejšie parametre SGED SV modelu sú λ a ν , pretože hovoria o asymetrii a chvostoch rozdelenia výnosov. Výsledky potvrdzujú, že rozdelenia výnosov všetkých troch indexov sú ľavostranne zošikmené a pre S&P500 a SAX majú ťažké chvosty. Použitie SGED rozdelenia pre šoky v rovnici výnosov bolo teda adekvátne.

Odhadnuté realizácie variancie takisto pomerne dobre odrážajú fluktuácie výnosov indexov, čo je zachytené v grafoch na obrázku 6.

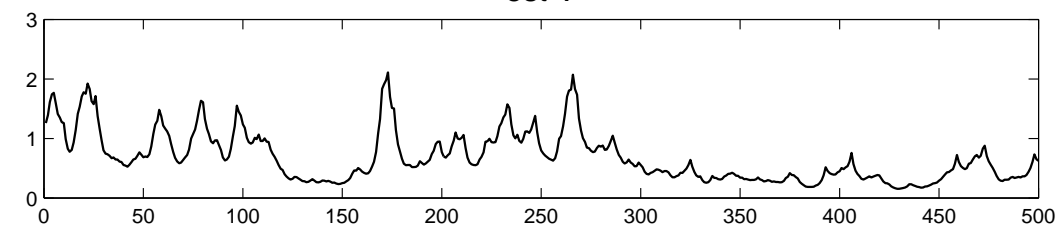
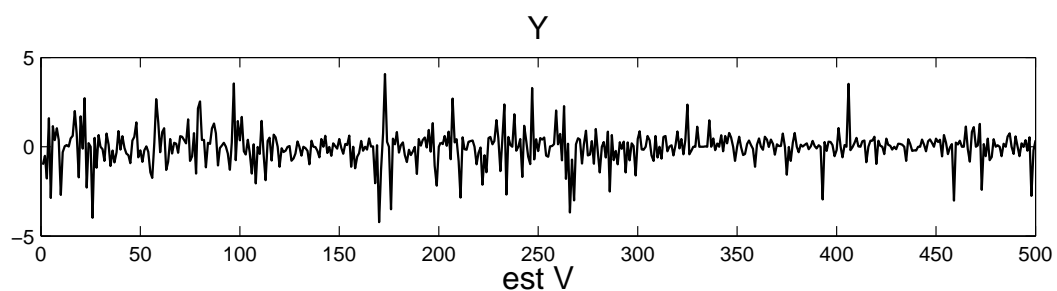
Z výsledkov analýzy výstupu, v rámci ktorej sme použili už spomenuté metódy, vyplýva pomerne malá miera zmiešavania v podstate pre všetky parametre modelu.



(a) S&P500



(b) DAX



(c) SAX

Obr. 6: Odhadnuté realizácie variancie v SGED SV modeli

Ako príklad uvádzame v Prílohe 2 grafy pre parametre λ a ν akciového indexu S&P500. Znamená to teda, že by bolo potrebné zvýšiť celkový počet simulácií a aj dĺžku burn-in. Nemožno však konštatovať, že hybridná MCMC metóda, ktorú sme použili v tomto modeli, je menej efektívna než metóda z Taylorovho SV modelu, pretože SGED SV model je zložitejší (kvôli čomu sme v ňom volili aj menší rozsah použitých dát).

Záver

V tejto práci sme sa zaoberali metódami odhadovania SV modelov. Z množstva modelov, ktoré boli na tento účel navrhnuté sme sa zamerali na MCMC metódy kvôli ich širokej využiteľnosti a aj využívanosti. Dôležitou súčasťou práce je popis teoretického pozadia týchto metód, teda bayesovský prístup, Monte Carlo integrovanie a teória Markovových reťazcov. Prvé dve témy súvisia viac s praktickým použitím MCMC metód, zatiaľ čo na teórii Markovových reťazcov je založená ich konvergencia. Považujeme za veľmi dôležité oboznámiť sa aj s touto stránkou MCMC metód, aby mohli byť správne používané. Ďalej sme popísali konkrétne algoritmy, ktoré MCMC metódy využívajú, a tiež ich niektoré rozšírenia. S ohľadom na praktické použitie sme sa venovali aj niekoľkým metódam konvergenčnej diagnostiky.

V praktickej časti práce sme demonštrovali použitie MCMC metód. Aplikovali sme ich pritom na dva konkrétne SV modely, Taylorov SV model a SGED SV model. Použili sme dáta troch akciových indexov: S&P500, DAX a SAX, konkrétne ich výnosy. V každom z modelov sme využívali MCMC algoritmy tak, aby bolo odhadovanie správne (algoritmy sme volili podľa tvaru cieľových hustôt), a pritom efektívne (použili sme aj adaptívne algoritmy). Na výsledky odhadovania sme aplikovali aj jednoduché metódy konvergenčnej diagnostiky, ktoré pri SGED SV modeli svedčili o slabej konvergencii nasimulovaného reťazca, ktorá by sa mohla zlepšiť buď vyšším počtom celkových simulácií alebo nejakou zmenou metódy odhadovania. Pre Taylorov SV model sme z pohľadu konvergenzie konštatovali celkom dobré výsledky, samozrejme s ohľadom na jednoduchosť metód analýzy výstupu.

Čo sa týka rozšírenia práce, dobrým námetom je skúmanie ďalších dvoch spôsobov odhadovania volatility v SV modeloch, a to filtrovanie a predikcie, nakoľko my sme sa venovali len vyhladzovaniu.

Literatúra

- [1] ANDERSON, E.C. Monte Carlo Methods and Importance Sampling. Lecture notes for Statistical Genetics, 1999, University of Washington, Department of Biology. Dostupné na internete: <<http://ib.berkeley.edu/labs/slatkin/eriq/classes.htm>>, Statistics 578C.
- [2] ANDRIEU, C., DE FREITAS, N., DOUCET, A., JORDAN, M. I. An Introduction to MCMC for Machine Learning. In *Machine Learning*, 2003, vol. 50, no. 1-2, p. 5-43. Dostupné na internete: <<http://www.springerlink.com/content/1573-0565/>>.
- [3] ASGHARIAN, H., BENGTSSON, CH. Jump Spillover in International Equity Markets. In *Journal of Financial Econometrics*, 2006, vol. 4, no. 2, p. 167-203. Dostupné na internete: <<http://www.fma.org/Siena/Papers/>>, 330350.
- [4] BILLIO, M., SARTORE, D. Stochastic volatility models: A survey with application to option pricing and Value at Risk. In *Applied Quantitative Methods for Trading and Investment*, 2003, chap. 8, p. 239-291. Dostupné na internete: <<http://www.greta.it/en/wp.htm>>, Working papers, 2002, n. 0210.
- [5] BOLLERSLEV, T. Generalized Autoregressive Conditional Heteroskedasticity. In *Journal of Econometrics*, 1986, vol. 31, no. 3, p. 307-327.
- [6] BROOKS, S.P. Markov Chain Monte Carlo Method and Its Application. In *The Statistician*, 1998, vol. 47, no. 1, p. 69-100. Dostupné na internete: <<http://www.jstor.org/journals/00390526.html>>.
- [7] BROOKS, S.P. Quantitative Convergence Diagnosis for MCMC via CUSUMS. In *Statistics and Computing*, 1998, vol. 8, no. 3, p. 267-274. Dostupné na internete: <<http://citeseer.ist.psu.edu/104448.html>>.
- [8] BROOKS, S.P., ROBERTS, G.O. Assessing Convergence of Markov Chain Monte Carlo Algorithms. In *Statistics and Computing*, 1998, vol. 8, no. 4, p. 319-335. Dostupné na internete: <<http://www.statslab.cam.ac.uk/~steve/mypapers/bror95a.html>>.
- [9] BROTO, C., RUIZ, E. Estimation Methods for Stochastic Volatility Models: A Survey. In *Journal of Economic Surveys*, 2004, vol. 18, no. 5, p. 613-649. Dostupné na internete: <<http://ideas.repec.org/p/cte/wsrepe/ws025414.html>>.

- [10] CAPPUCCIO, N., LUBIAN, D., RAGGI, D. MCMC Bayesian Estimation of a Skew-GED Stochastic Volatility Model. In *Studies in Nonlinear Dynamics & Econometrics*, 2004, vol. 8, no. 2, p. 1211-1211. Dostupné na internete: <<http://ssrn.com/abstract=445640>>.
- [11] CONT, R. Empirical properties of asset returns: stylized facts and statistical issues. In *Quantitative Finance*, 2001, vol. 1, no. 2, p. 223-236. Dostupné na internete: <<http://citeseer.ist.psu.edu/cont01empirical.html>>.
- [12] COX, J.C., INGERSOLL, J.E., ROSS, S.A. A Theory of the Term Structure of Interest Rates. In *Econometrica*, 1985, vol. 53, no. 2, p. 385-407. Dostupné na internete: <<http://links.jstor.org/journals/00129682.html>>.
- [13] ENGLE, R.F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. In *Econometrica*, 1982, vol. 50, no. 4, p. 987-1007. Dostupné na internete: <<http://links.jstor.org/journals/00129682.html>>.
- [14] ENGLE, R.F., PATTON, A.J. What good is a volatility model?. In *Quantitative Finance*, 2001, vol. 1, no. 2, p. 237-245. Dostupné na internete: <<http://citeseer.ist.psu.edu/416228.html>>.
- [15] FLEGAL, J.M., HARAN, M., JONES, G.L. Markov Chain Monte Carlo: Can We Trust the Third Significant Figure? Draft, 2007. Dostupné na internete: <<http://www.arxiv.org/abs/math/0703746>>.
- [16] GEYER, Ch.J. Markov Chain Monte Carlo Lecture Notes. Lecture notes, 1998 (revised 2005), University of Minnesota, School of Statistics. Dostupné na internete: <<http://www.stat.umn.edu/geyer/8931/>>, Notes, n1998.
- [17] GEWEKE, J. [Bayesian Analysis of Stochastic Volatility Models]: Comment. In *Journal of Business & Economic Statistics*, 1994, vol. 12, no. 4, p. 397-399. Dostupné na internete: <<http://links.jstor.org/journals/07350015.html>>.
- [18] GHYSELS, E., HARVEY, A., RENAULT, E. Stochastic volatility. Working Paper, 1996, Cahiers de recherche 9613, Université de Montréal, Département de sciences économiques. Dostupné na internete: <<http://www.cirano.qc.ca>>, Scientific Publications>, November 1995, 95s-49.
- [19] GILKS, W.R. Derivative-free Adaptive Rejection Sampling for Gibbs Sampling. In *Bayesian Statistics 4*, 1992, p. 641-649.

- [20] GILKS, W.R., WILD, P. Adaptive Rejection Sampling for Gibbs Sampling. In *Applied Statistics (Journal of the Royal Statistical Society: Series C)*, 1992, vol. 41, no. 2, p. 337-348. Dostupné na internete: <<http://www.jstor.org/journals/00359254.html>>.
- [21] GILKS, W.R., BEST, N.G., TAN, K.K.C. Adaptive Rejection Metropolis Sampling within Gibbs Sampling. In *Applied Statistics (Journal of the Royal Statistical Society: Series C)*, 1995, vol. 44, no. 4, p. 455-472. Dostupné na internete: <<http://www.jstor.org/journals/00359254.html>>.
- [22] HASTINGS, W.K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. In *Biometrika*, 1970, vol. 57, no. 1, p. 97-109. Dostupné na internete: <<http://links.jstor.org/journals/00063444.html>>.
- [23] JACQUIER, E., POLSON, N., ROSSI, P. Bayesian Analysis of Stochastic Volatility Models. In *Journal of Business & Economic Statistics*, 1994, vol. 12, no. 4, p. 371-417. Dostupné na internete: <<http://citeseer.ist.psu.edu/jacquier94bayesian.html>>.
- [24] JOHANNES, M., POLSON, N. MCMC Methods for Continuous-Time Financial Econometrics. In HANSEN, L., AIT-SAHALIA, Y. *Handbook of Financial Econometrics*. Elsevier Science, to appear 2007. Dostupné na internete: <<http://home.uchicago.edu/~lhansen/>>.
- [25] KIM, S., SHEPHARD, N., CHIB, S. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. In *The Review of Economic Studies*, 1998, vol. 65, no. 3, p. 361-393. Dostupné na internete: <<http://links.jstor.org/journals/00346527.html>>.
- [26] METROPOLIS, N. et al. Equation of State Calculations by Fast Computing Machines. In *Journal of Chemical Physics*, 1953, vol. 21, no. 6, p. 1087-1092. Dostupné na internete: <<http://jcp.aip.org/jcp/>>.
- [27] MEYN, S.P., TWEEDIE, R.L. Markov Chains and Stochastic Stability. London:Springer-Verlag, 1993. 548 p. Dostupné na internete: <<http://probability.ca/MT/>>.
- [28] MIRRA, A. On Metropolis-Hastings Algorithms with Delayed Rejection. In *Metron - International Journal of Statistics*, 2001, vol. 59, no. 3-4, p. 231-241. Dostupné na internete: <<http://citeseer.ist.psu.edu/517259.html>>.

- [29] MØLLER, J. Notes on Markov Chain Monte Carlo Methods. Lecture notes, 1999, Aalborg University, Department of Mathematical Sciences. Dostupné na internete: <<http://www.math.aau.dk/~jm/>>, Recent publications, 1999.
- [30] OSTATNÍKOVÁ, B. Stochastic volatility modeling. Diplomová práca, 2006, Univerzita Komenského, Fakulta matematiky, fyziky a informatiky.
- [31] PÁZMAN, A. Bayesovská štatistika. Bratislava: Univerzita Komenského, 2003. 98 s.
- [32] PEŠKO, Š. Operačná analýza II. Prednášky pre IRS, 2000-2001, Žilinská univerzita, Fakulta riadenia a informatiky. Dostupné na internete: <<http://frcatel.fri.utc.sk/users/pesko/ipubl.html>>, A2.
- [33] PEŠKO, Š. Teória hromadnej obsluhy. Prednášky pre AM, 2001, Žilinská univerzita, Fakulta riadenia a informatiky. Dostupné na internete: <<http://frcatel.fri.utc.sk/users/pesko/ipubl.html>>, A3.
- [34] POIRIER, D.J., TOBIAS, J.L. Bayesian Econometrics. In MILLS, T.C., PATTERSON, K. *Palgrave Handbook of Econometrics, Volume 1: Theoretical Econometrics*. Palgrave Macmillan, 2006, p. 841-870. Dostupné na internete: <<http://www.econ.iastate.edu/faculty/tobias/>>, Selected Publications.
- [35] ROBERTS, G.O., ROSENTHAL, J.S. General State Space Markov Chains and MCMC Algorithms. In *Probability Surveys*, 2004, vol. 1, p. 20-71. Dostupné na internete: <<http://www.i-journals.org/ps/viewarticle.php?id=15>>.
- [36] RUBIN, D.B. Using the SIR Algorithm to Simulate Posterior Distributions. In BERNARDO, J.M., DeGROOT, M.H., LINDLEY, D.V., SMITH A.F.M. *Bayesian Statistics 3*. Oxford: Oxford University Press, 1988, p. 395-402.
- [37] TIERNEY, L., MIRRA, A. Some Adaptive Monte Carlo Methods for Bayesian Inference. In *Statistics in Medicine*, 1999, vol. 18, no. 17-18, p. 2507-2515. Dostupné na internete: <<http://citeseer.ist.psu.edu/141391.html>>.
- [38] WALSH, B. Introduction to Bayesian analysis. Lecture notes for Postgraduate course in Evolutionary Quantitative Genetics 2006, 2002 University of Aarhus. Dostupné na internete: <<http://nitro.biosci.arizona.edu/workshops/Aarhus2006/Aarhus-06.html>>, Lecture 1: Basic statistical machinery, Advanced notes for various statistics topics, Bayesian methods.

- [39] Yu, B., MYKLAND, P. Looking at Markov Samplers through CUSUM Path Plots: a Simple Diagnostic Idea. In *Statistics and Computing*, 1998, vol. 8, no. 3, p. 275-286. Dostupné na internete: <<http://citeseer.ist.psu.edu/81956.html>>.
- [40] ZHENG, S.J. Adaptive MCMC: Background, Theory and Simulations. Research report, 2006, University of Toronto, Department of Statistics. Dostupné na internete: <<http://probability.ca/jeff/grad.html>>, Reading/Research Project/Visiting Students.

Príloha 1

SGED rozdelenie

SGED rozdelenie je zovšeobecnením GED rozdelenia (nazývaného aj EPD rozdelenie). GED rozdelenie sa používa na modelovanie ťažkých chvostov rozdelenia nejakej náhodnej premennej, pričom SGED rozdelenie umožňuje zároveň zachytiť aj asymetriu jej rozdelenia.

SGED rozdelenie je charakterizované dvojicou parametrov λ a ν . Parameter λ vyjadruje asymetriu tohto rozdelenia, zatiaľ čo ν meria, aké ťažké má chvosty. Ide o spojité rozdelenie, s hustotou

$$f(x; \lambda, \nu) = \left[1 + \operatorname{sgn}(\lambda x) \gamma \left(0.5 |\lambda x|^\nu, \frac{1}{\nu} \right) \right] \frac{\nu \exp(-0.5 |x|^\nu)}{2^{1+\frac{1}{\nu}} \Gamma(\frac{1}{\nu})}, \quad \nu, \lambda \in \mathbb{R}, \nu > 0,$$

kde $\Gamma(a) = \int_0^\infty e^{-t} t^{a-1} dt$ je gama funkcia, $\gamma(x, a) = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt$ je neúplná gama funkcia.

Párne (centrálne) momenty SGED rozdelenia sú dané

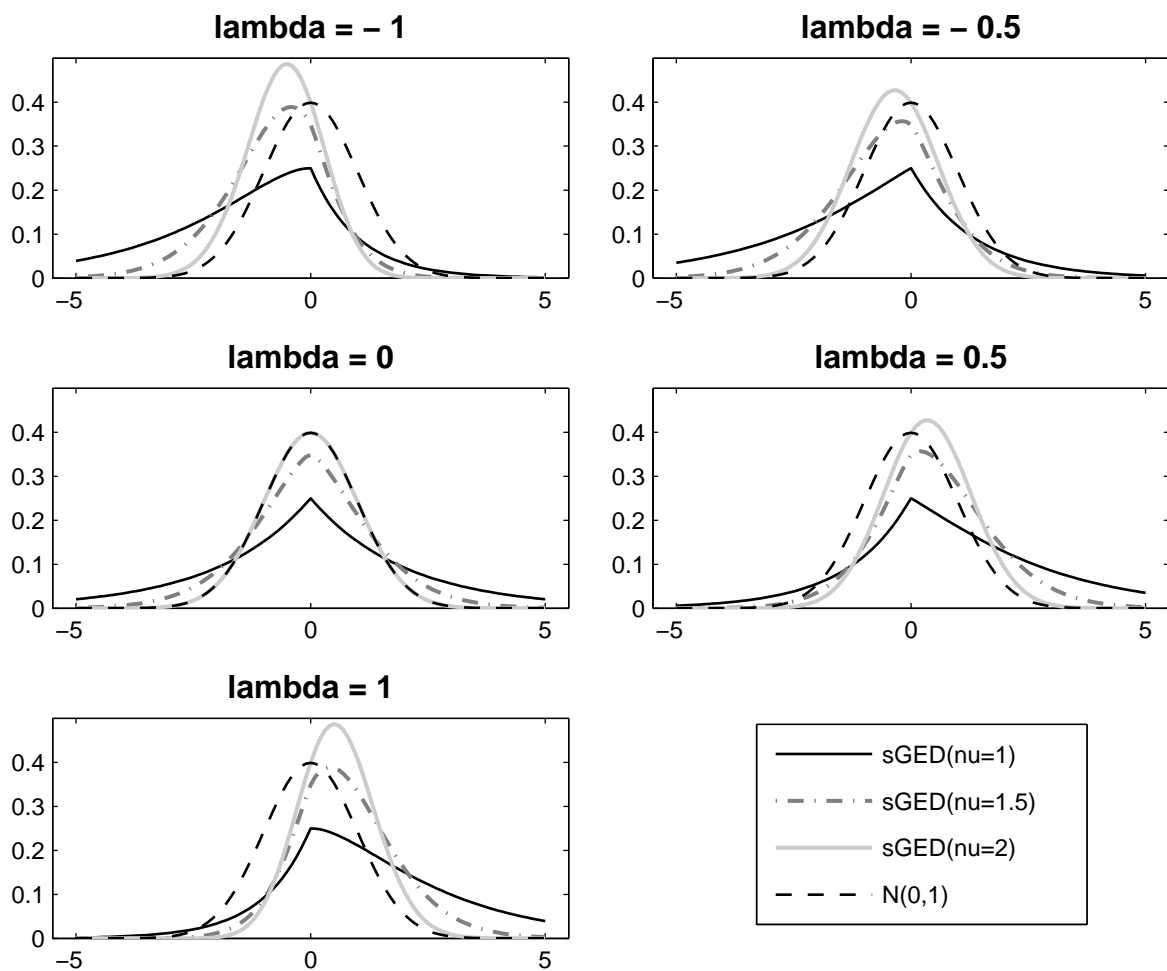
$$E[X^r] = 2^{\frac{r}{\nu}} \frac{\Gamma(\frac{r+1}{\nu})}{\Gamma(\frac{1}{\nu})}$$

a nepárne (centrálne) momenty sú

$$e[X^r] = 2^{\frac{r}{\nu}} \frac{\operatorname{sgn}(\lambda) \Gamma(\frac{r+1}{\nu})}{\Gamma(\frac{1}{\nu})} B\left(\frac{|\lambda|^\nu}{1+|\lambda|^\nu}, \frac{1}{\nu}, \frac{r+1}{\nu}\right),$$

kde $B(z, p, q)$ je distribučná funkcia beta rozdelenia.

Špeciálnymi prípadmi tohto rozdelenia sú normálne rozdelenie, ak $\lambda = 0$ a $\nu = 2$, a GED rozdelenie, ak $\lambda = 0$. Pre $\nu < 2$ má SGED rozdelenie ťažšie chvosty ako normálne rozdelenie, pre $\nu > 2$ zasa ľahšie chvosty. Zošíkmené bude, ak $\lambda \neq 0$, a to ľavostranne pri $\lambda < 0$ a pravostranne pri $\lambda > 0$. Tieto vlastnosti SGED rozdelenia sú zachytené na obrázku 7.



Obr. 7: SGED rozdelenie pre rôzne hodnoty parametrov

Príloha 2

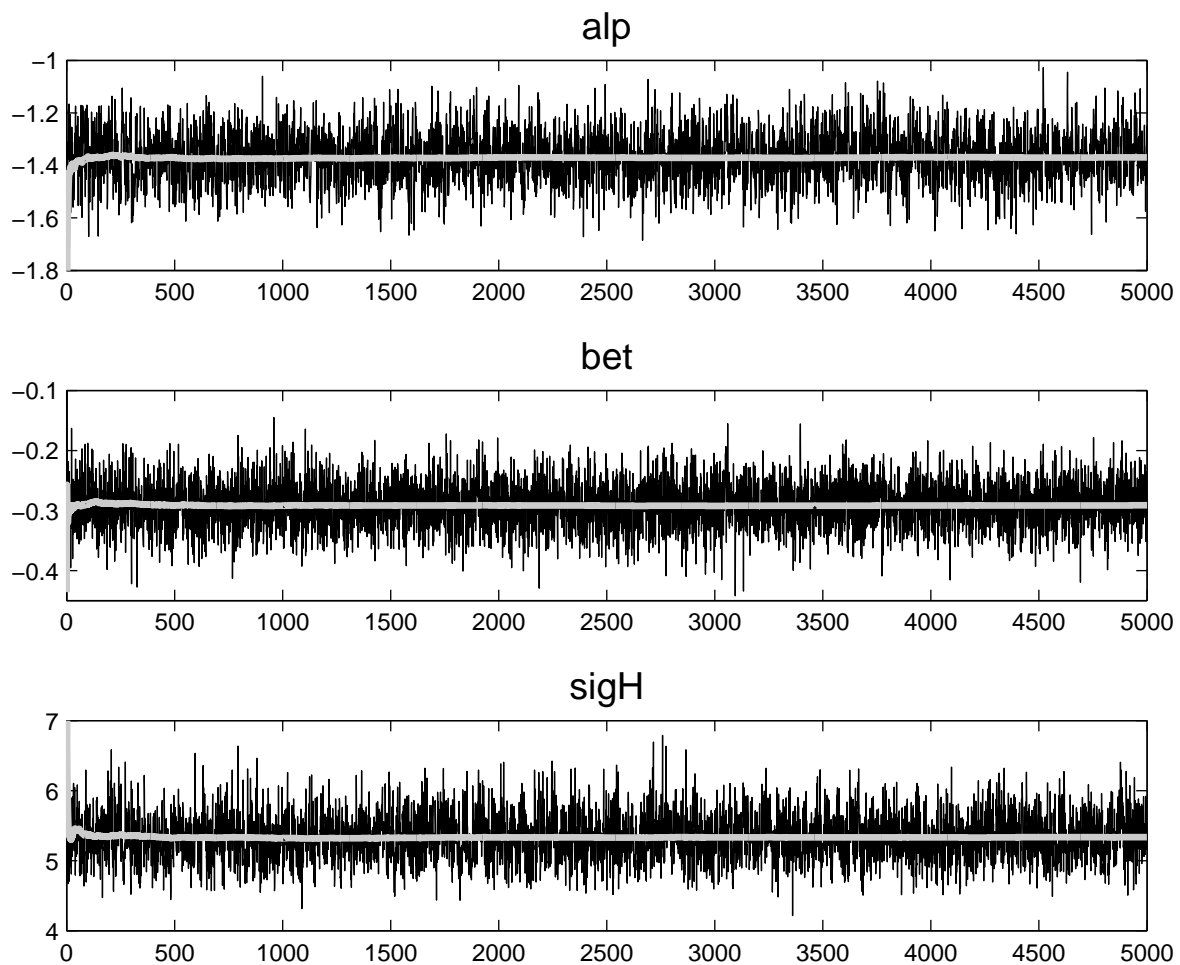
Výsledky praktickej časti

	SGED SV model		
	S&P500	DAX	SAX
μ	-1.1072	-0.3210	-0.5692
MCSE	0.0211	0.0065	0.0151
SD	0.6530	0.1085	0.1901
95%-ný int.spoľ.	[-1.1493,-1.0651]	[-0.3341,-0.3080]	[-0.5994,-0.5390]
ϕ	0.9357	0.8699	0.8665
MCSE	0.0043	0.0033	0.0033
SD	0.0418	0.0357	0.0389
95%-ný int.spoľ.	[0.9271,0.9442]	[0.8633,0.8764]	[0.8599,0.8731]
σ_H^2	0.0180	0.0639	0.1651
MCSE	0.0112	0.0049	0.0096
SD	0.0093	0.0417	0.0986
95%-ný int.spoľ.	[0.0158,0.0201]	[0.0542,0.0736]	[0.1459,0.1844]
λ	-0.2927	-1.3759	-0.0625
MCSE	0.0164	0.0623	0.0025
SD	0.2829	0.9488	0.0940
95%-ný int.spoľ.	[-0.3152,-0.2703]	[-1.5004,-1.2515]	[-0.0675,-0.0574]
ν	1.4885	2.0766	1.0332
MCSE	0.0210	0.0426	0.0033
SD	0.2754	0.5097	0.0851
95%-ný int.spoľ.	[1.4466,1.5304]	[1.9913,2.1618]	[1.0265,1.0398]

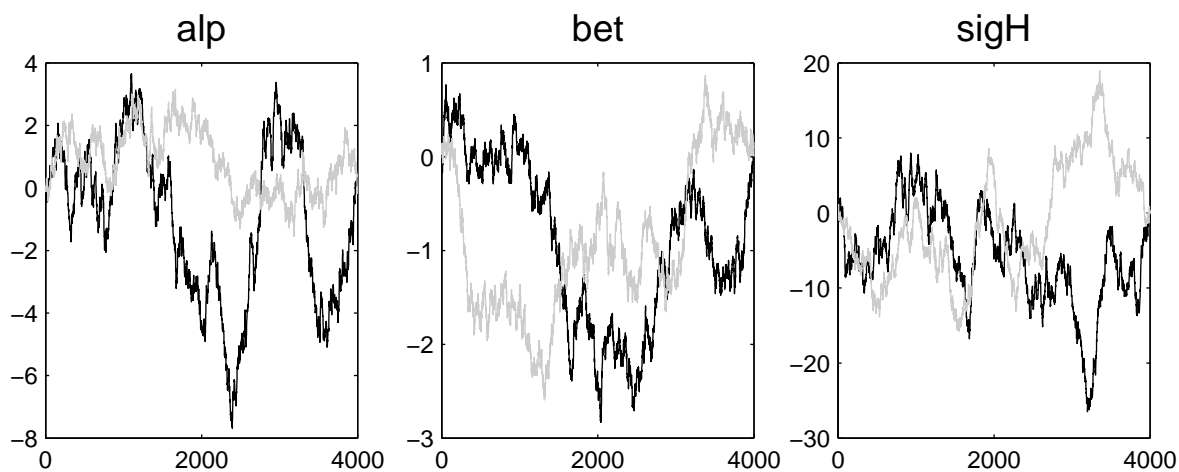
Tabuľka 4: Aposteriórne priemery a ich štatistiky v SGED SV modeli

		Taylorov SV model					
		RS + M-H			ARS		
		S&P500	DAX	SAX	S&P500	DAX	SAX
α		-1.77	-1.7821	-1.5879	-1.6214	-1.0304	-1.368
MCSE		0.0051	0.0056	0.0038	0.0025	0.002	0.0023
SD		0.0974	0.1066	0.0999	0.0913	0.0869	0.0957
95%-ný int.spol.		[-1.7801,-1.7599]	[-1.7931,-1.771]	[-1.5954,-1.5804]	[-1.6264,-1.6165]	[-1.0345,-1.0264]	[-1.3727,-1.3634]
β		-0.3388	-0.3164	-0.2859	-0.3421	-0.2941	-0.2913
MCSE		0.001	0.0011	0.0009	0.0006	0.0006	0.0006
SD		0.036	0.0378	0.0395	0.0351	0.0403	0.0388
95%-ný int.spol.		[-0.3408,-0.3368]	[-0.3186,-0.3141]	[-0.2877,-0.2840]	[-0.3433,-0.3409]	[-0.2953,-0.2929]	[-0.2925,-0.2900]
σ_H^2		4.6585	5.7628	6.4964	4.1722	4.2195	5.3398
MCSE		0.0164	0.0225	0.0165	0.0069	0.0057	0.0079
SD		4.3657	0.4553	4.9643	1.2599	1.38	1.7386
95%-ný int.spol.		[4.6258,4.6912]	[5.7178,5.8077]	[6.4633,6.5294]	[4.1585,4.1860]	[4.2081,4.2309]	[5.3241,5.3555]

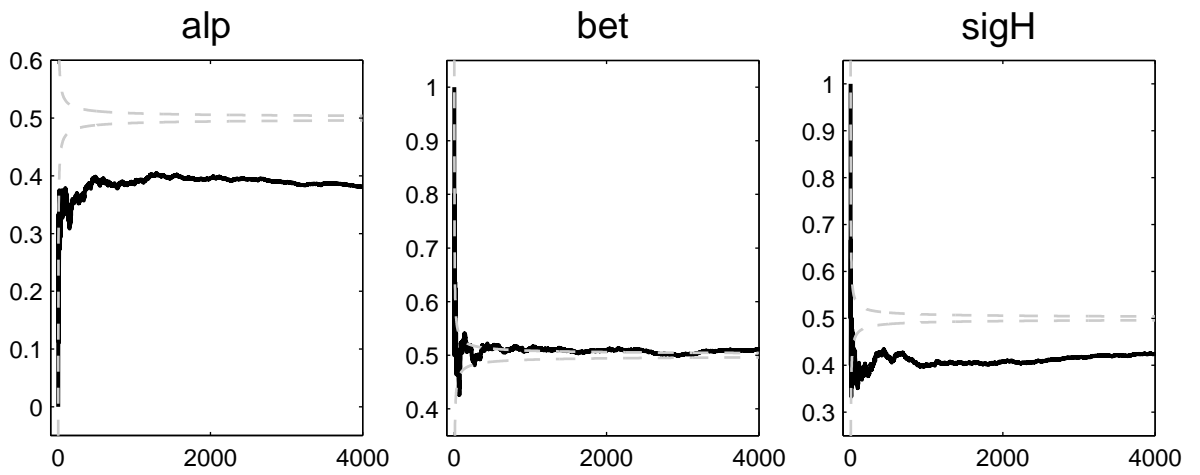
Tabuľka 5: Aposteriórne priemery a ich štatistiky v Taylorovom SV modeli



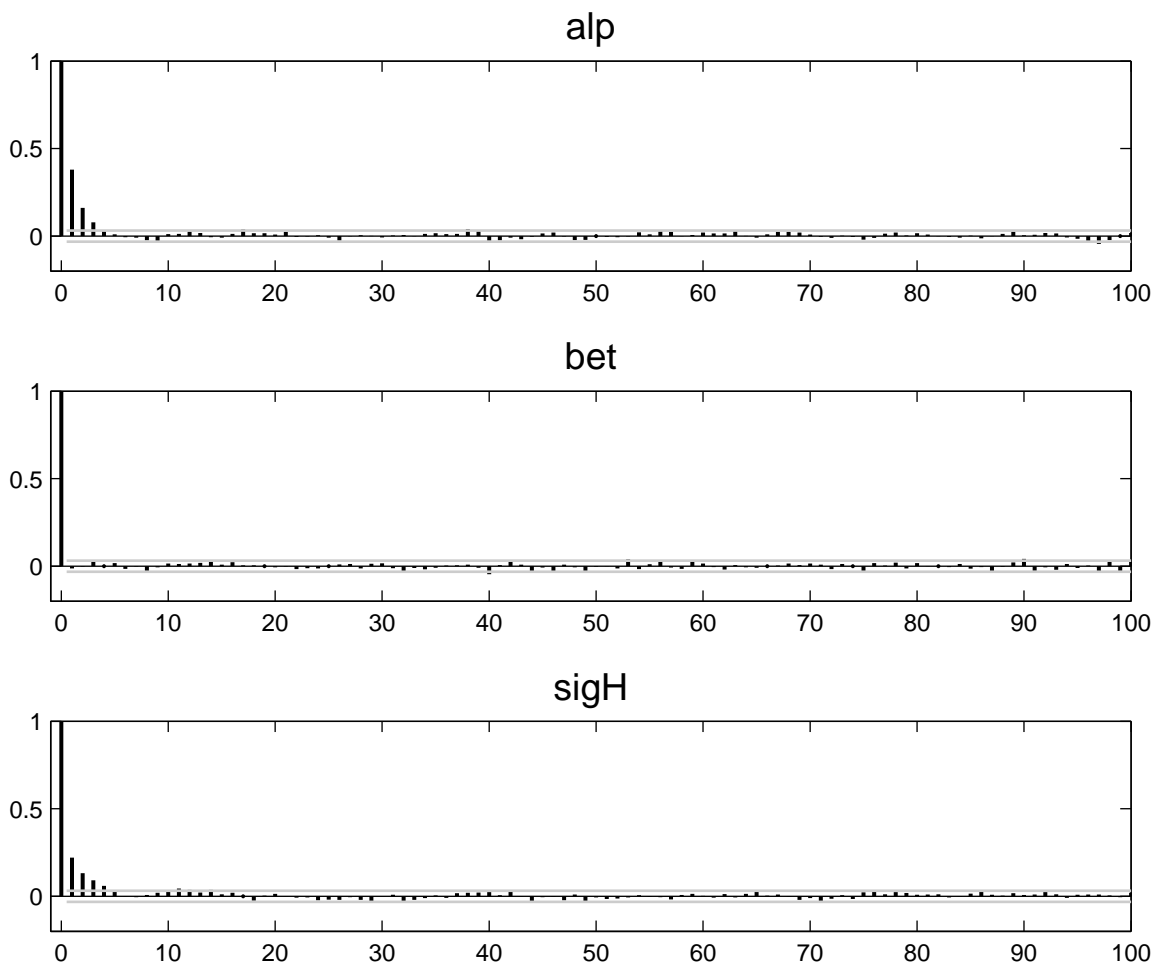
Obr. 8: Trajektórie parametrov s priebežnými priemerami, Taylorov SV model, SAX



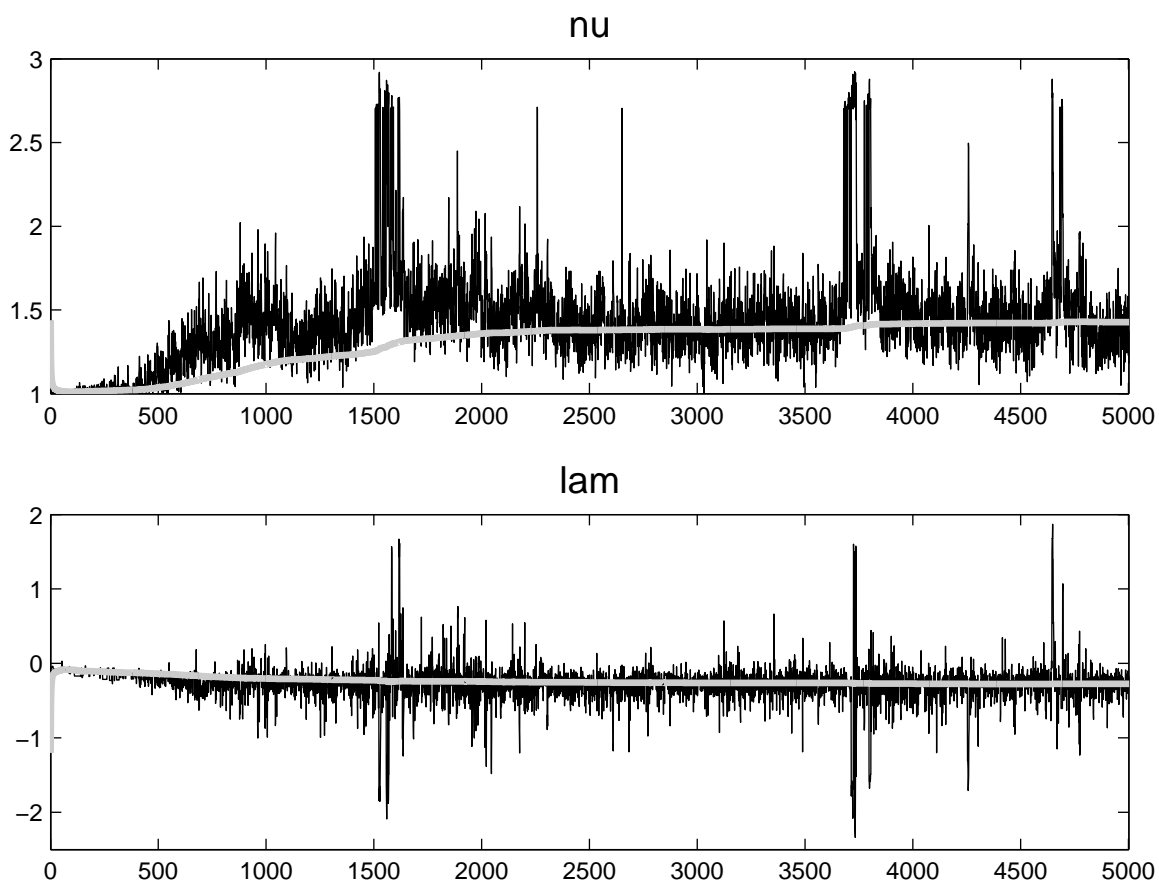
Obr. 9: Priebeh kumulatívnych súm parametrov, Taylorov SV model, SAX



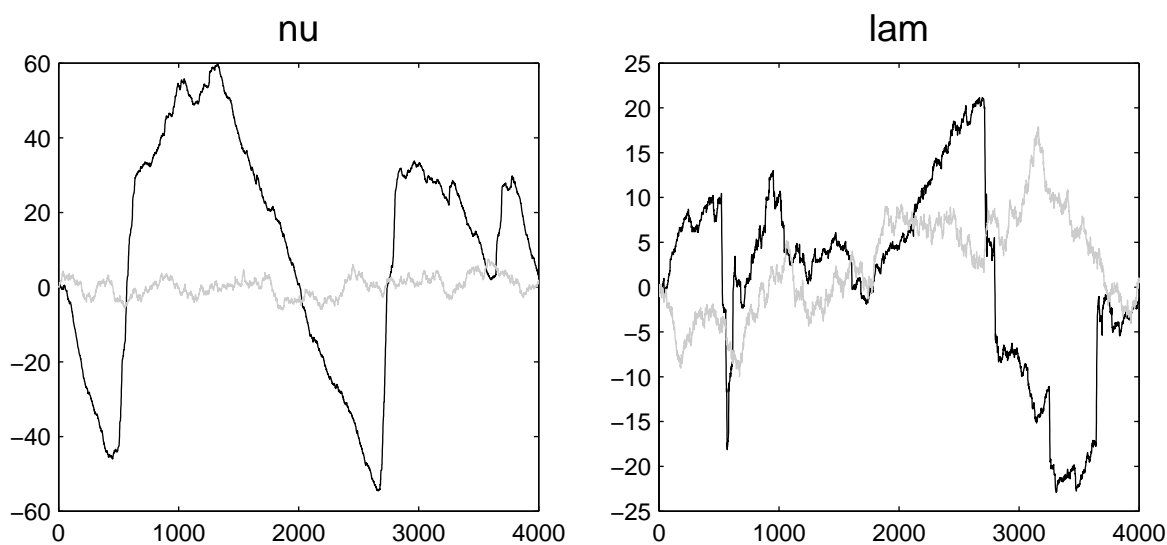
Obr. 10: Priebeh mier “strapatosti” parametrov, Taylorov SV model, SAX



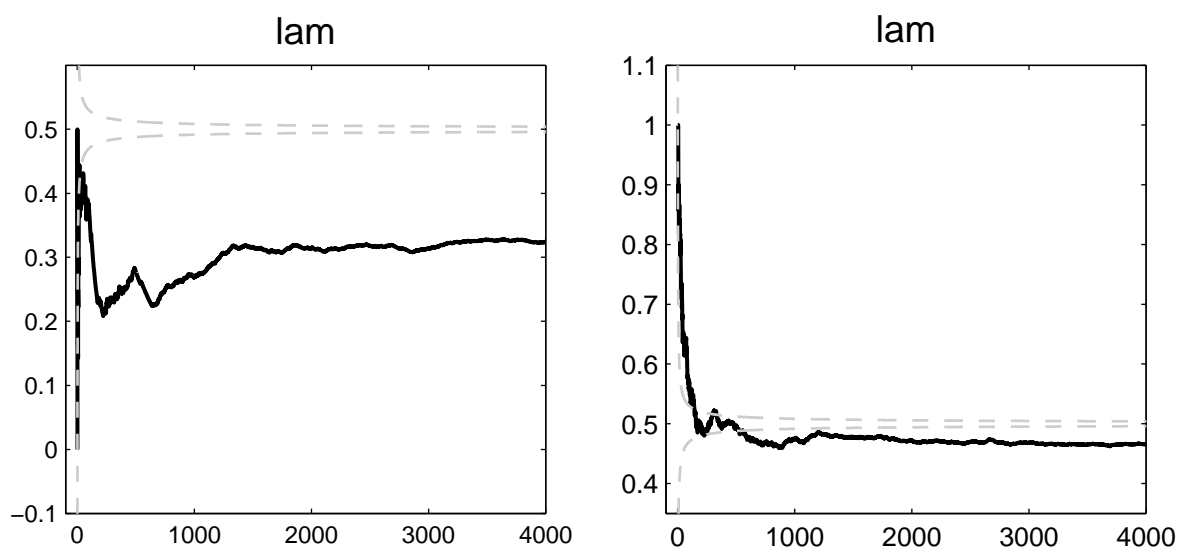
Obr. 11: Priebeh autokorelácií parametrov, Taylorov SV model, SAX



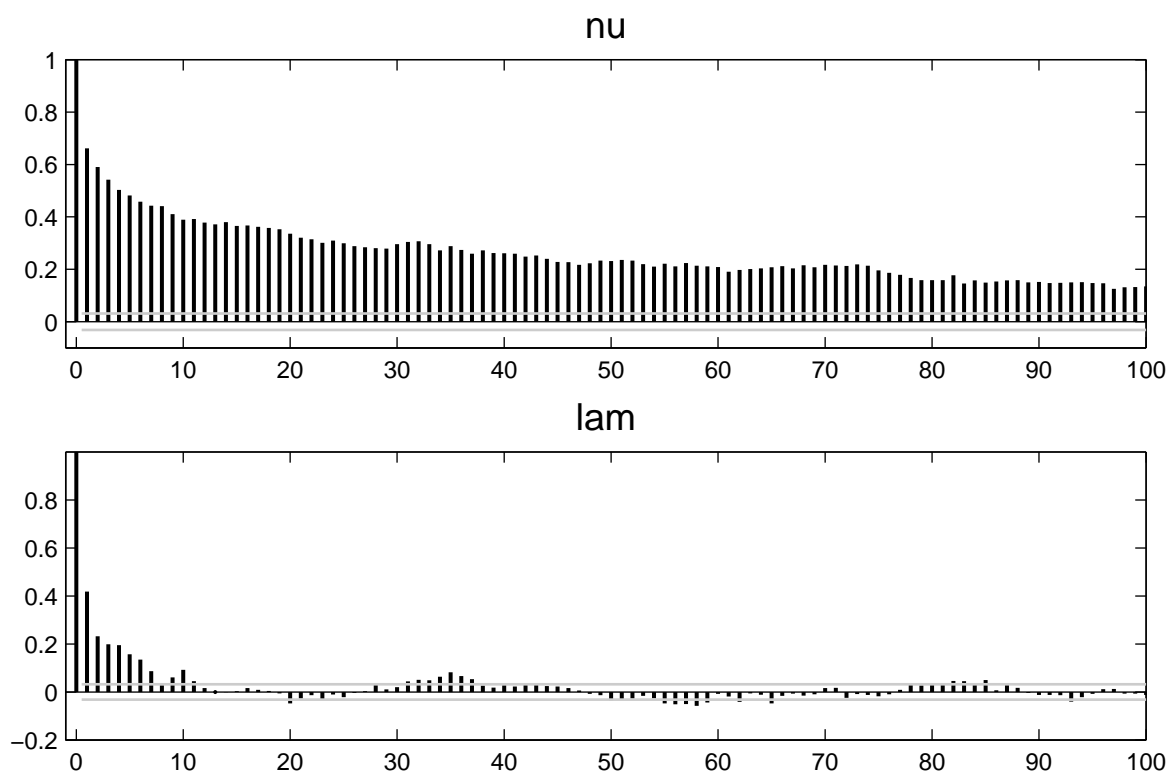
Obr. 12: Trajektórie parametrov s priebežnými priemermi, SGED SV model, S&P500



Obr. 13: Priebeh kumulatívnych súm parametrov, SGED SV model, S&P500



Obr. 14: Pribeh mier "strapatosti" parametrov, SGED SV model, S&P500



Obr. 15: Pribeh autokorelácií parametrov, SGED SV model, S&P500