

O PODSTATĚ A APLIKACÍCH MCMC METOD

Petr VOLF a Aleš LINKA¹

ÚTIA ČAV a TU Liberec, KAP

Abstract. We give an overview of MCMC algorithms, namely of the Gibbs sampler, Metropolis–Hastings algorithm and its variant using the simulated annealing. The conditions for convergence of resulting chain to its invariant distribution are presented. Then, the multivariate modifications of MH algorithm are discussed. A problem of MCMC synthesis, i. e. the search of the most probable state of a system is solved as an example.

Резюме: В статье изучаются МСМС методы. Предлагаются достаточные условия для сходимости распределения цепи Маркова – результата алгоритма Метрополиса–Астингса – к инвариантному распределению. Обсуждены некоторые многомерные варианты этого метода. В качестве примера решается задача синтеза – случайный поиск наиболее вероятного состояния системы.

1. ÚVOD

MCMC metody jsou postupy založené na intenzívních simulacích. V podstatě provádějí náhodné prohledávání určitého prostoru a pravidlo prohledávání (tj. pravidlo pro náhodné přechody od stavu k stavu) je voleno tak, aby výsledkem MCMC procedury byla posloupnost prvků (stavů) s rozdělením pravděpodobnosti konvergujícím k rozdělení námi požadovanému. Někdy také generujeme posloupnost prvků tak, aby konvergovala k prvku optimálnímu (kde optimalita je vyjádřena maximem pravděpodobnosti).

První případ se často používá k generování aposterioriního Bayesova rozdělení při odhadování parametrů složitých modelů. Příkladem je nastavení parametrů v modelech – neuronových sítích či v jiných matematických modelech, které odhadujeme pomocí trénovacích dat [24], [16]. Druhý případ (tj. například spojení MCMC metod se simulovaným žiháním) je určen k hledání nejpravděpodobnějších konfigurací stavů systému, například při rekonstrukci obrazové informace, [12], při řešení úloh rozpoznávání a shlukové analýzy. Jako příklad uvedeme model vztlínání tekutiny v textilním materiálu, kdy je cílem nalézt nejstabilnější (nejpravděpodobnější) stav systému.

Pokud v následujícím textu budeme mluvit o rozděleních pravděpodobnosti, budeme si pro jednoduchost představovat (pokud to situace dovolí), že pracujeme se spojitými distribucemi, tedy reprezentovanými hustotami (vzhledem k Lebesgueově míře).

¹Tato práce vznikla s podporou grantu GA ČR 201/97/0354 a grantu MŠMT VŠ/97084.

2. PŘEHLED NEJZNÁMĚJŠÍCH MCMC METOD

Je vyvinuto několik skupin metod jak vygenerovat výběr z určitého rozdělení a vyhnout se přitom výpočtu přesného tvaru tohoto rozdělení či jak generovat Markovovu posloupnost, jejíž rozdělení se blíží k onomu cílovému. Do první skupiny (vesměs jednodušších postupů) patří např. zamítací metoda, [1], nebo metoda „sampling – resampling“ (vážený bootstrap). MCMC procedury tvoří skupinu druhou:

Gibbsův algoritmus. (viz [3],[8],[18]) je MCMC metoda, která je vhodná pro případ mnohorozměrné veličiny $\mathbf{X} \in \mathcal{X} \subset R^p$. Nechť je naším cílem vygenerovat vzorek s (aspoň přibližně) rozdělením s hustotou $p(\mathbf{x})$. Označme $p_j(x_j|\mathbf{x}_{(-j)})$ hustoty podmíněných rozdělení, kde $\mathbf{x}_{(-j)} = (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_p)$.

Algoritmus začne z nějaké (zvolené) počáteční hodnoty $\mathbf{x}^{(0)}$ a postupně v každém kroku inovuje jednu složku \mathbf{x} , a to tak, že ji generuje právě z podmíněného rozdělení p_j . Takže například novou j -tou složku v $(m+1)$. cyklu dostaneme tak, že $x_j^{(m+1)}$ vygenerujeme pomocí hustoty

$$p_j(x_j|x_1^{(m+1)}, \dots, x_{j-1}^{(m+1)}, x_{j+1}^{(m)}, \dots, x_p^{(m)}).$$

Zjevně dostáváme náhodnou posloupnost $\mathbf{x}^{(m)}$, $m = 0, 1, 2, \dots$, která je Markovova. Není problém ukázat, že hustota invariantního rozdělení takového Markovovy posloupnosti je právě $p(\mathbf{x})$, a že rozdělení $\mathbf{x}^{(m)}$ při $m \rightarrow \infty$ k tomuto invariantnímu rozdělení konverguje. Podmínky jsou formulovány např. v [18], ale jak uvidíme, lze Gibbsův algoritmus považovat za případ algoritmu Metropolis–Hastingse a zkontrolovat podmínky pro konvergenci v rámci tohoto algoritmu.

Samozřejmě, zdaleka ne vždy známe tvar podmíněných pravděpodobností p_j . Pak je možné Gibbsův algoritmus při generování každé nové složky kombinovat se zamítací metodou, anebo použít proceduru Metropolis–Hastingse. Algoritmus Metropolis–Hastingse. (viz [3],[18],[23]) také vytváří Markovovu náhodnou posloupnost, a to následujícím způsobem. Nechť $\mathbf{x}^{(m)}$ je zatím poslední člen posloupnosti, nechť $q(\mathbf{x}|\mathbf{x}^{(m)})$ je nějaká (zatím zcela libovolná) hustota rozdělení pravděpodobnosti (tj. může být podmíněná hodnotou posledního členu posloupnosti, ale nemusí). Vygenerujeme s její pomocí “kandidáta” \mathbf{x}^* na další člen posloupnosti. Položme pak

$$(1) \quad \pi(\mathbf{x}^*, \mathbf{x}^{(m)}) = \frac{p(\mathbf{x}^*) \cdot q(\mathbf{x}^{(m)}|\mathbf{x}^*)}{p(\mathbf{x}^{(m)}) \cdot q(\mathbf{x}^*|\mathbf{x}^{(m)})},$$

($\pi(\mathbf{x}^*, \mathbf{x}^{(m)}) = 1$, pokud by jmenovatel byl nula), a přijmeme $\mathbf{x}^{(m+1)} = \mathbf{x}^*$ s pravděpodobností $\min\{1, \pi(\mathbf{x}^*, \mathbf{x}^{(m)})\}$. Pokud \mathbf{x}^* nepřijmeme, pokládáme $\mathbf{x}^{(m+1)} = \mathbf{x}^{(m)}$. Všimněme si, že ve jmenovateli je pravděpodobnost, podle které jsme vybrali nový člen, v čitateli pak pravděpodobnost přechodu od nového členu ke starému.

Takto tedy vzniká Markovova posloupnost, jejíž vlastnosti do značné míry závisí na vlastnostech rozdělení generujícího nové kandidáty, tj. rozdělení zde reprezentované hustotou q . I tu můžeme chápat jako hustotu rozdělení pravděpodobnosti přechodů pro nějakou (jinou) Markovovu posloupnost na \mathcal{X} (a to samozřejmě i v případě, kdy $q(\mathbf{x}'|\mathbf{x}) = q(\mathbf{x}')$ – tj. q by generovala i.i.d. posloupnost).

Z přijímacího pravidla je vidět, že případná neznalost normující konstanty u $p(\mathbf{x})$ zde nevádí. V případě mnohorozměrné náhodné veličiny \mathbf{X} můžeme MH algoritmus také použít (podobně jako Gibbsův) postupně pro inovaci jednotlivých složek.

Tento algoritmus má mnoho variant, jedna z nich, **algoritmus Metropolisův** (např. [3]) používá generující rozdělení **reversibilní**, tj. takové, že $q(\mathbf{x}'|\mathbf{x}) = q(\mathbf{x}|\mathbf{x}')$. Pak je přijímací pravidlo založeno jen na poměru $\pi = p(\mathbf{x}^*)/p(\mathbf{x}^{(m)})$. Vidíme zde jasně, že tyto algoritmy podporují přijímání prvků \mathbf{x} , které mají velkou pravděpodobnost $p(\mathbf{x})$. Pokud chceme získat jeden bod nejlépe reprezentující rozdělení $p(\mathbf{x})$, můžeme vzít průměr z hodnot $\mathbf{x}^{(m)}$ (případně po vynechání počáteční části řetězce). Můžeme také použít postup zvaný simulované žihání, které vede k nalezení prvku maximalizujícího pravděpodobnost (hustotu) $p(\mathbf{x})$.

2.1. Simulované žihání (simulated annealing). Až dosud jsme popisovali situaci, kdy na základě náhodného generování kandidátů a přijímacího pravidla generujeme (Markovovu) posloupnost postupných řešení. Představme si, že se chceme přiblížit k řešení, které maximalizuje nějakou funkci $g(\mathbf{x})$. Postupem popsaným dosud, s přijímacím pravidlem založeným na

$$\pi = \frac{\exp(g(\mathbf{x}^*))}{\exp(g(\mathbf{x}))} \cdot \frac{q(\mathbf{x}|\mathbf{x}^*)}{q(\mathbf{x}^*|\mathbf{x})},$$

bychom dostali řetězec $\mathbf{x}^{(s)}$, $s = 1, 2, \dots$, přičemž rozdělení veličiny $\mathbf{x}^{(s)}$ by se blížilo k $C \cdot \exp(g(\mathbf{x}))$. Pokud bychom ale použili přijímacího pravidla založeného na $\pi(s) = (\pi)^{1/T(s)}$ tak, že $T(s) > 0$, $T(s) \rightarrow 0$ pro $s \rightarrow \infty$ (kde s je číslo iterace), dostaneme (při vhodné volbě funkce T) přímo konvergenci $\mathbf{x}^{(s)}$ k $\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} g(\mathbf{x})$. Vhodná volba je například $T = K/\log(1 + s)$ (konvergence $T(s) \rightarrow 0$ by neměla být příliš rychlá, pak by totiž bylo nebezpečí, že přijímací pravidlo nás zavede do lokálního maxima a nedovolí nám z něj již vyskočit). V jednoduchém případě s rovnoměrnými či reversibilními pravděpodobnostmi q je tedy přijímací pravděpodobnost simulovaného žihání

$$\min \left\{ 1, \pi(s) = \exp \left([g(\mathbf{x}^*) - g(\mathbf{x})] \frac{1}{T(s)} \right) \right\}.$$

Více podrobností lze najít v článku [12] v Robustu'90.

3. O KONVERGENCI MH ALGORITMU

V klasické teorii Markovových řetězců, např. [6], se pracuje většinou s konečnými či spočetnými prostory stavů. S tím bychom ovšem například v případě stavů – parametrů regresního modelu nevystačili. Teorie tedy musí zahrnovat i nespočetné stavové prostory. Na druhé straně, MCMC většinou generuje řetězce homogenní, či nanejvýš s pomalu se měnícími přechodovými pravděpodobnostmi (například při užití simulovaného žlhaní), nebo se střídáním několika skupin přechodů (tzv. hybridní algoritmy). Podmínky pro konvergence Markovových řetězců s nespočetně stavy jsou formulovány např. v [18], podmínky pro detailní rovnováhu jsou zkoumány také v [7] a [23].

Uvažujme tedy Markovovu posloupnost se stavy z nespočetné množiny $E \subset \mathbf{R}^1$. Nechť $X^{(0)}, X^{(1)}, X^{(2)}, \dots$ je náš Markovův řetězec, $K(x, y)$ označme jeho jádro přechodů popisující pravděpodobnost přechodu z x do y . V případě MCMC je zpravidla substochastické, tj.

$$\int_E K(x, y) dy \leq 1, \quad R(x) = 1 - \int_E K(x, y) dy \geq 0,$$

kde $R(x)$ je pravděpodobnost, že řetězec zůstane ve stavu x . Dále označme

$$\begin{aligned} & K^{(t)}(x, y) \\ &= \int_E K^{(t-1)}(x, z) K(z, y) dz + K^{(t-1)}(x, y) \cdot R(y) + [R(x)]^{t-1} K(x, y) \end{aligned}$$

jádro popisující přechody za t kroků (opět může být substochastické). Označme ještě $p(x)$ hustotu invariantní distribuce řetězce a **předpokládejme, že je $p(x) > 0$ na E** . Tato hustota tedy splňuje

$$\int_A p(x) dx = \int_E P(X^{(1)} \in A | X^{(0)} = z) p(z) dz,$$

pro všechny měřitelné množiny $A \subset E$. Přitom je

$$P(X^{(1)} \in A | X^{(0)} = z) = \int_A K(z, y) dy + R(z) \cdot \mathbf{1}[z \in A].$$

Jádro K je **ireducibilní**, jestliže pro všechna $x \in E$ a pro všechny měřitelné $A \subset E$ je $P(X^{(t)} \in A | X^{(0)} = x) > 0$ pro nějaké $t > 0$. Dále, jádro je **aperiodické**, když pro žádné $n \geq 2$ neexistuje rozklad $\bigcup_{j=1}^n B_j = E$ takový, že řetězec pravidelně (periodicky) navštěvuje množiny B_j . Tyto definice jsou známy ze standardní teorie Markovových řetězců, a také zde jsou tyto vlastnosti základem ke konvergenci rozdělení řetězce k jeho invariantnímu rozdělení, viz [18]:

Tvrzení 1. Nechť p je hustota invariantního rozdělení řetězce $X^{(0)}, X^{(1)}, X^{(2)}, \dots$ a přechodové jádro K je ireducibilní a aperiodické, pak při $t \rightarrow \infty$, pro každé $x \in E$

$$\text{i) } \int_E |K^{(t)}(x, y) - p(y)| dy \rightarrow 0;$$

ii) pro reálnou a p -integrovatelnou funkci f ,

$$\frac{1}{t} \sum_{s=1}^t f(X^{(s)}) \rightarrow \int_E f(x) p(x) dx \quad \text{a. s.}$$

Tvrzení i) říká, že rozdělení $X^{(t)}$ konverguje k invariantnímu, tvrzení ii) - ergodičnost - pak mimo jiné, že průměr z členů řetězce je rozumným odhadem pro střední hodnotu invariantního rozdělení (a že to platí i pro transformaci řetězce).

3.1. Podmínky pro konvergenci. Nechť znovu $q(y|x)$ je hustota pravděpodobnosti pro navrhování nových prvků řetězce (na dosavadním prvku x záviset může, ale nemusí) a nechť $p(x)$ je hustota rozdělení, ke kterému chceme dospět. Pak přechodové jádro pro řetězec generovaný MH algoritmem je $K(x, y) = q(y|x) \cdot \alpha(x, y)$, kde

$$\alpha(x, y) = \begin{cases} \min \left\{ \frac{p(y) q(x|y)}{p(x) q(y|x)}, 1 \right\} & \text{když } p(x) q(y|x) > 0 \\ 1 & \text{když } p(x) q(y|x) = 0. \end{cases}$$

Není problém ukázat, že $p(x)$ skutečně je hustota invariantního rozdělení. Nejprve se ukáže, že platí podmínka **detailní rovnováhy**: $p(x) K(x, y) = p(y) K(y, x)$. Pak, pro každé $A \subset E$,

$$\begin{aligned} & \int_E P \left(X^{(1)} \in A | X^0 = z \right) p(z) dz = \\ &= \int_E \left\{ \int_A K(z, y) dy + R(z) \mathbf{1}[z \in A] \right\} p(z) dz = \\ &= \int_A \int_E K(y, z) p(y) dz dy + \int_E R(z) \mathbf{1}[z \in A] p(z) dz = \\ &= \int_A (1 - R(y)) p(y) dy + \int_A R(z) p(z) dz = \int_A p(y) dy. \end{aligned}$$

Ke konvergenci a ergodičnosti řetězce generovaného algoritmem typu MH sice stačí méně než podmínka detailní rovnováhy, platnost této podmínky je však snadné prokázat a dává i návod, jak algoritmy konstruovat. Samotná konvergence pak plyne z vlastností pravděpodobností (či jejich hustot) podle kterých navrhuji nové členy řetězce, viz [18]:

Tvrzení 2.

- i) Je-li q aperiodická (jako hustota pravděpodobností přechodu), pak K je také aperiodické,
- ii) Je-li q ireducibilní (na E), a $q(y|x) = 0$ právě když $q(x|y) = 0$, pak K je ireducibilní.

3.2. Modifikace mh algoritmu. Nejčastější modifikace spočívá v tom, že se pohybujeme ve vícerozměrném prostoru a inovujeme pomocí MH algoritmu jednu složku vektoru \mathbf{x} za druhou. Složky můžeme inovovat v ustáleném pořadí nebo je vybírat náhodně. Každý dílčí krok by měl splňovat podmínky kladené na MH algoritmus (tj. detailní rovnováhu) a stejné podmínky by měla splňovat i kombinace těchto kroků (důležitá je například podmínka ireducibility řetězce navrhovaných stavů). Znamená to pro každou složku j mít “navrhovací” pravděpodobnost $q_j(x_j^*|\mathbf{x})$. Pro takový krok inovující j -tou složku je pak přijímací pravděpodobnost (tj. toho, zda x_j^* vymění dosavadní x_j) $\min(1, \pi_j)$, s

$$\pi_j = \frac{p(\mathbf{x}^*)}{p(\mathbf{x})} \cdot \frac{q_j(x_j|\mathbf{x}^*)}{q_j(x_j^*|\mathbf{x})},$$

kde $\mathbf{x}^* = (x_1, \dots, x_{j-1}, x_j^*, x_{j+1}, \dots, x_p)$.

Všimněme si ještě, co se stane, když za navrhuující pravděpodobnosti vezmeme $p_j(x_j^*|\mathbf{x}_{(-j)})$, tj. podmíněné cílové pravděpodobnosti, tak jak je to u algoritmu Gibbse. Dostaneme pak $\pi_j = 1$, to znamená, že vždy novou komponentu x_j^* přijmeme. Vidíme, že z tohoto hlediska je Gibbsův algoritmus zvláštním případem MH algoritmu.

Obecnější situace lze popsat takto: mějme v každém kroku k dispozici několik možností, jak provést přechod k novému stavu. Mezi těmito možnostmi vybíráme (ať už náhodně nebo nenáhodně). Dokonce se může stát, že přechody vedou do různých prostorů. Například, v [24], když se konstruovala funkce z polynomiálních splinů, tak hledaným parametrem byly uzly těchto splinů, a hledal se i jejich optimální počet. Takže nutně docházelo k přechodům mezi prostory různých dimenzí. Pak je potřeba dát pozor na poměr $q_1(x_1|x_2)/q_2(x_2|x_1)$, kde např. q_1 označuje přechod z R^2 do R^1 a q_2 naopak. Může se stát, že zatímco q_2 je hustota spojitého rozdělení, je q_1 diskrétní pravděpodobnost. Hluběji se tímto případem zabývají [7] i [23]. Podstatné je (spíš než to, že obě pravděpodobnosti jsou definovány na jiných prostorech), aby navržené přechody měly vždy definovány i přechody opačné (realizovatelné s nenulovou pravděpodobností – či hustotou pravděpodobnosti). Vlastně to již vyjadřuje druhá část předpokladu ii) Tvrzení 2.

4. PŘÍKLAD – SIMULACE SMÁČENÍ VLÁKNA KAPALINOU

Představme si jednoduchou situaci: V nádobě máme kapalinu a kolmo do ní vnoříme vlákno z nějakého materiálu. Dojde k interakci mezi vláknem a kapalinou, která je dána vzájemnou adhezí materiálů. Situace se ustálí ve stavu, který by mohl být popsán jako stav s minimální potenciální energií. Přitom proti stoupání kapaliny po vlákně působí soudržnost kapaliny (koheze) a gravitace. Zkusme tuto situaci jednoduše simulovat. Půjde tedy o řešení úlohy **syntézy**, tj. budeme generovat Markovovu posloupnost stavů systému tak, abychom se blížili k stavům majícím určité rozdělení pravděpodobnosti,

případně k stavu toto rozdělení maximalizující. Využíváme zde modelu uvedeného v [15], program je vytvořen v MATLABu.

Představme si krychlovou síť $(x * y * z, x, y, z=1, 2, \dots, M)$, dohromady tedy $N = M^3$ buněk. Buňky budeme označovat indexem $i = 1, 2, \dots, N$, budeme ale mít i jejich x, y, z souřadnice x_i, y_i, z_i . Zavedme veličiny

$$\begin{aligned} k_i &= 1, & \text{je-li v místě } (i) \text{ kapalina,} \\ &= -1, & \text{není-li tam,} \\ s_i &= 1, & \text{je-li v onom místě vlákno (pak příslušné } k_i = -1), \\ &= 0, & \text{pokud ne.} \end{aligned}$$

Na začátku máme tedy fixovaná s_i a počáteční hladinu kapaliny K , tj.

$$\begin{aligned} k_i &= 1 & \text{pro } z_i \leq K, \text{ pokud } s_i = 0, \\ k_i &= -1 & \text{pro } z_i > K. \end{aligned}$$

Uvažujme tři koeficienty, a to gravitace, koheze, adheze, C_g, C_k, C_a . Dále uvažujme vzájemné působení, ať už kapaliny vzájemně, či kapaliny s vláknem, jen v sousedních k sobě přiléhajících buňkách naší sítě. Tedy každá buňka má 6 sousedních. Tyto okolní buňky k i -té označíme \mathcal{O}_i . Následující funkce H je Hamiltonián soustavy, který je zde úměrný (potenciální) energii systému:

$$(2) \quad H(\mathbf{k}) = \sum_{i=1}^N C_g k_i z_i - \sum_{i=1}^N \sum_{j \in \mathcal{O}_i} C_k k_i k_j - \sum_{i=1}^N \sum_{j \in \mathcal{O}_i} C_a k_i s_j.$$

Z jiného pohledu můžeme každému stavu, tj. konfiguraci $\mathbf{k} = \{k_i, i = 1, \dots, N\}$ přiřadit pravděpodobnost

$$(3) \quad P(\mathbf{k}) = C \cdot \exp(-H(\mathbf{k})),$$

kde C je normalizační konstanta. Tento typ modelu je Isingův model, viz i [12]. Je možné odvodit i podmíněné pravděpodobnosti svazující rozdělení stavu v dané buňce se stavy v buňkách sousedních.

Takto máme tedy zadánu úlohu a) generovat konfigurace odpovídající (aspoň přibližně) rozdělení pravděpodobnosti (3), b) najít konfiguraci $\mathbf{k} = \{k_i, i = 1, \dots, N\}$ maximalizující (3). Přitom, uvažujeme-li uzavřenou nádobu, zůstává $\sum_{i=1}^N k_i$ konstantní. Tato úloha je řešitelná metodami MCMC (a metodami podobnými).

Při generování markovské posloupnosti konfigurací, jejíž rozdělení konverguje k (3), můžeme, jak jsme viděli, postupovat zhruba dvěma způsoby:

- (1) Gibbsův algoritmus – vyžadoval by z (3) odvodit příslušná podmíněná rozdělení $P_i(k_i | \mathbf{k}_{(-i)})$ a z nich pak generovat postupně a

opakovaně hodnoty v jednotlivých buňkách. Protože jde o “nula–jedničkové” veličiny, není problém toto podmíněné rozdělení odvodit:

$$(4) \quad P_i(k_i = 0 | \mathbf{k}_{(-i)}) = \frac{1}{1 + \exp(-C_g z_i + C_a \sum_{\mathcal{O}_i} s_j + C_k \sum_{\mathcal{O}_i} k_j)},$$

$$P_i(k_i = 1 | \mathbf{k}_{(-i)}) = 1 - P_i(k_i = 0 | \mathbf{k}_{(-i)}).$$

Situace je zjednodušená tím, že (alespoň při normálních podmínkách, kdy C_g i $C_k > 0$) v úvahu připadá pouze změna na hladině kapaliny. Znamená to, že se v jednotlivých krocích budeme snažit v (náhodně) vybraném bodě (x, y) zvednout či snížit hladinu o 1 jednotku (tj. o 1 buňku v naší síti $x * y * z$ ve směru z). Problém ale je, jak při generování zachovat onu podmínku (alespoň vždy během několika kroků), že $\sum_{i=1}^N k_i = \text{const.}$ Z tohoto důvodu je v tomto případě Gibbsův algoritmus nevhodný. Proto použijeme raději

- (2) Algoritmus Metropolis–Hastings – např. následující variantu:

Náhodně (a rovnoměrně v síti $x*y$) zvolíme 2 body $(x_1, y_1), (x_2, y_2)$ a náhodně (opět rovnoměrně) rozhodneme, ve kterém navrheme uvažovat hladinu zvýšenou o 1, v druhém pak o 1 sníženou. Pak spočteme příslušné nové $H^* = H(\mathbf{k}^*)$ (kde v konfiguraci \mathbf{k}^* budou proti dosavadní \mathbf{k} změněny jen 2 hodnoty). Novou konfiguraci přijmeme s pravděpodobností

$$\min(1, \pi), \quad \text{kde} \quad \pi = \frac{\exp(-H^*)}{\exp(-H)} = \exp(H - H^*),$$

protože v MH algoritmu se vyskytující pravděpodobnosti $q(\mathbf{k}' | \mathbf{k})$ jsou zde reversibilní a vykrátí se (jde tedy vlastně o variantu Metropolis).

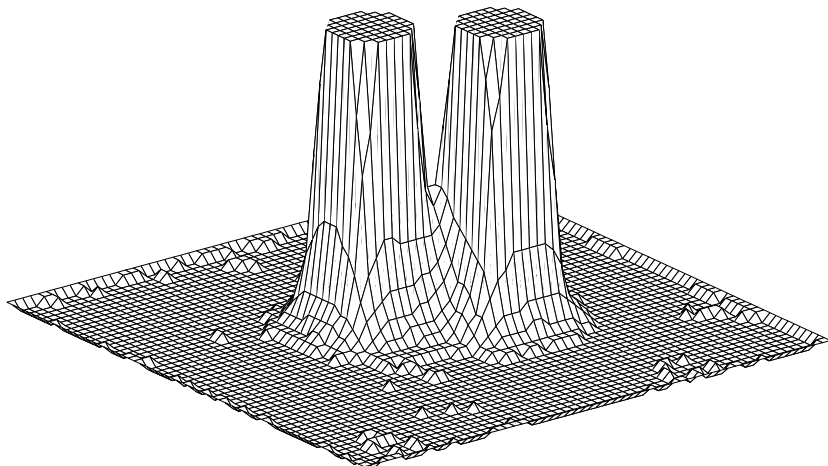
Jsou samozřejmě možné i jiné varianty postupu. Zajímavé je sledovat situaci s více vlákny, porovnávat situace s vlákny různě vzdálenými apod.

Jeden výsledek takového algoritmu (s dvěma vlákny) zde uvedeme. Zvolili jsme $C_g = 1$, $C_k = 4$, $C_a = 3$, počáteční hladinu 15. Bylo provedeno 500 ‘globálních’ iterací (tj. $M \times M \times 500$ pokusů inovovat dvojici hodnot). V každé buňce jsme pak za řešení vzali tu hodnotu (1 či -1), která se v buňce v posledních 200 iteracích vyskytla častěji (tj. prvních 300 iterací nebylo bráno v úvahu).

Řešili jsme i úlohu najít konfiguraci s maximální pravděpodobností. K tomu modifikujeme přijímací pravidlo, abychom dostali metodu simulovaného žihání. Tedy

$$\pi = \pi_m = \exp \left\{ (H - H^*) \cdot \frac{1}{T_m} \right\},$$

kde m je číslo ‘globální’ iterace, $T_m = 2 / \log(1 + m)$. Výsledek je na obr. 1.



Obr.1. Simulace smáčení dvou blízkých vláken

5. VYUŽITÍ MCMC V BAYESOVSKÉ ANALÝZE DAT

Představme si, že data \mathbf{y} , která měříme, jsou realizace náhodných veličin $\mathbf{Y} = Y_1, \dots, Y_n$, model “vzniku” dat nechť je popsán pomocí pravděpodobnosti (řekněme hustoty) $f(\mathbf{y}; \boldsymbol{\theta})$, kde $\boldsymbol{\theta} \in \Theta \subset R^p$ je neznámý parametr. Neurčitost hodnoty $\boldsymbol{\theta}$ se popisuje pomocí apriorního rozdělení – označme je (jeho hustotu) $g_0(\boldsymbol{\theta})$. Bayesovo pravidlo pak udává aposteriorní rozdělení jako podmíněné rozdělení hodnot parametru při napozorovaných datech, tj.

$$(5) \quad p(\boldsymbol{\theta}|\mathbf{y}) = C(\mathbf{y}) \cdot f(\mathbf{y}; \boldsymbol{\theta}) \cdot g_0(\boldsymbol{\theta}).$$

Zde $C(\mathbf{y})$ je normující člen, nezávisející na $\boldsymbol{\theta}$. Jako bodový odhad parametru $\boldsymbol{\theta}$ se nejčastěji používá modus aposteriorního rozdělení, $\hat{\boldsymbol{\theta}}(\mathbf{y}) = \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{y})$ nebo $\tilde{\boldsymbol{\theta}}(\mathbf{y}) = E(\boldsymbol{\theta}|\mathbf{y})$, tj. střední hodnota z aposteriorního rozdělení.

Problém samozřejmě bývá s výpočtem členu $C(\mathbf{y})$, ten ale k zjištění $\hat{\boldsymbol{\theta}}(\mathbf{y})$ vlastně nepotřebujeme. Jenže často (v případech mnohorozměrného $\boldsymbol{\theta}$) není snadné získat ani maximum, ani další charakteristiky aposteriorního rozdělení (mohly by nás zajímat především rozptyl a kvantily, ke zjištění “šíře” aposteriorního rozdělení a tím i spolehlivosti odhadu parametru). A tady nastupují metody, které jsou schopny aposteriorní rozdělení nasimulovat.

V dalším budeme s daty \mathbf{y} zacházet jako s “konstantou” – tj. jsme v situaci, kdy data jsou k dispozici a naším cílem je odhadnutí aposteriorního rozdělení parametru. Jde tedy jen o modifikaci metod, které jsme popsali v části 2.

5.1. Generování aposteriorního rozdělení. Cílová distribuce je nyní tedy dána hustotou $p(\boldsymbol{\theta}|\mathbf{y})$, kde \mathbf{y} jsou ‘pevná’ data. V **Gibbově algoritmu**

potřebujeme znát podmíněné aposteriorní hustoty

$$p_j(\theta_j | \boldsymbol{\theta}_{(-j)}, \mathbf{y}) = C(\boldsymbol{\theta}_{(-j)}, \mathbf{y}) \cdot f(\mathbf{y}; \boldsymbol{\theta}) \cdot g_{0j}(\theta_j | \boldsymbol{\theta}_{(-j)}),$$

kde g_{0j} je hustota podmíněné apriorní distribuce.

Algoritmus Metropolise–Hastingse. má nyní přijímací pravidlo založené na

$$\pi = \frac{p(\boldsymbol{\theta}^* | \mathbf{y})}{p(\boldsymbol{\theta} | \mathbf{y})} \cdot \frac{q(\boldsymbol{\theta} | \boldsymbol{\theta}^*, \mathbf{y})}{q(\boldsymbol{\theta}^* | \boldsymbol{\theta}, \mathbf{y})},$$

kde $\boldsymbol{\theta}$ je poslední člen řetězce, $\boldsymbol{\theta}^*$ je kandidát navržený na nový člen, $q(\boldsymbol{\theta}^* | \boldsymbol{\theta}, \mathbf{y})$ je hustota pravděpodobnosti (která, jak vidíme, může záviset nejen na stávajícím stavu řetězce $\boldsymbol{\theta}$, ale i na datech \mathbf{y}), ze které byl $\boldsymbol{\theta}^*$ navržen.

Všimněme si, jak se rozhodovací pravidlo zjednoduší, když budeme kandidáty na nové členy posloupnosti generovat přímo z apriorního rozdělení $g_0(\boldsymbol{\theta})$. Dostaneme (po dosazení za $p(\boldsymbol{\theta} | \mathbf{y})$ z Bayesova vzorce (5))

$$(6) \quad \pi(\boldsymbol{\theta}^*, \boldsymbol{\theta}) = \frac{f(\mathbf{y}; \boldsymbol{\theta}^*) \cdot g_0(\boldsymbol{\theta}^*)}{f(\mathbf{y}; \boldsymbol{\theta}) \cdot g_0(\boldsymbol{\theta})} \cdot \frac{g_0(\boldsymbol{\theta})}{g_0(\boldsymbol{\theta}^*)} = \frac{f(\mathbf{y}; \boldsymbol{\theta}^*)}{f(\mathbf{y}; \boldsymbol{\theta})},$$

neboli věrohodnostní poměr.

Jak jsme řekli, i v případě MH algoritmu můžeme postupovat po částech, tj. zde inovovat $\boldsymbol{\theta}$ po komponentách. To znamená v každém kroku vybrat j (náhodně nebo i ve stálém pořadí), navrhnout θ_j^* z nějaké podmíněné distribuce $q_j(\theta_j^* | \boldsymbol{\theta}, \mathbf{y})$, a přijmout θ_j^* (nové $\boldsymbol{\theta}^*$ je pak $(\theta_1, \dots, \theta_{j-1}, \theta_j^*, \theta_{j+1}, \dots, \theta_p)$) s pravděpodobností $\min(1, \pi)$, když

$$\pi = \frac{p_j(\theta_j^* | \boldsymbol{\theta}_{(-j)}, \mathbf{y})}{p_j(\theta_j | \boldsymbol{\theta}_{(-j)}, \mathbf{y})} \cdot \frac{q_j(\theta_j | \boldsymbol{\theta}^*, \mathbf{y})}{q_j(\theta_j^* | \boldsymbol{\theta}, \mathbf{y})}.$$

Opět, navrhovací distribuce nemusí záviset na dosavadním $\boldsymbol{\theta}$ a datech \mathbf{y} , můžeme použít i podmíněná apriorní rozdělení.

Protože příklady těchto řešení byly již na konferencích ROBUST uváděny, nebudeme je zde opakovat. Jen připomeneme, že jedním byl příklad rekonstrukce znehodnocené obrazové informace modelované pomocí Isingova modelu, v [12]. Aplikace MH algoritmu na optimalizaci výběru polynomiálních splinů v úloze nelineární regrese byla uvedena v [24]. Bohatým zdrojem informace o využití MCMC metod jsou také sborníky konference COMPSTAT'96 (např. [13], [14]) i sborník COMPSTAT'98, který se právě připravuje.

LITERATURA

- [1] Antoch J. a Vorlíčková D. (1992). *Vybrané metody statistické analýzy dat*. Academia, Praha.
- [2] Arjas E. a Liu L. (1995). *Assessing the losses caused by an industrial intervention – a hierarchical Bayesian approach*. JRSS C44, 357–368.
- [3] Bernardo J. M. a Smith A. F. M. (1994). *Bayesian Theory*. J. Wiley, New York.

- [4] Besag J. (1986). *On the statistical analysis of dirty pictures (with discussion)*. JRSS B **48**, 259–302.
- [5] Besag J., Green D., Higdon D. a Mengersen K. (1995). *Bayesian computation and stochastic systems*. Statist. Science **10**, 3–66.
- [6] Feller W. (1968). *An Introduction to Probability Theory and Its Applications*. J. Wiley, New York.
- [7] Gelman A. a Rubin D. B. (1992). *Inference from iterative simulation using multiple sequences (with discussion)*. Statist. Science **7**, 457–472.
- [8] Geman S. a Geman D. (1984). *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 724–741.
- [9] Geyer C. J. (1992). *Practical Markov chain Monte Carlo (with discussion)*. Statist. Science **7**, 473–511.
- [10] Green P. J. (1995). *Reversible jump MCMC computation and Bayesian model determination*. Biometrika **82**, 711–732.
- [11] Hastings W. K. (1970). *Monte Carlo sampling methods using Markov chains and their applications*. Biometrika **57**, 97–109.
- [12] Janžura M. (1990). *O jednom pravděpodobnostním algoritmu pro optimalizační úlohy*. Robust'90, JČMF, 84–88.
- [13] Janžura M. (1996). *Image processing, Markov chain approach*. Compstat'96, Physica Verlag, 89–100.
- [14] Linka A., Pícek J. a Volf P. (1996). *Monte Carlo method for likelihood regression analysis*. Compstat'96, Physica Verlag, 343–348.
- [15] Lukáš D. (1997). *Smáčení vláken vnořených kolmo do hladiny kapaliny*. Strutex 97, TU Liberec, 10–17.
- [16] Neal R. M. (1993). *Probabilistic Inference Using Markov Chain Monte Carlo*. Techn. Rep. CGR-TR-93-1, Univ. of Toronto.
<http://www.cs.utoronto.ca/~radford>.
- [17] Neal R. M. (1996). *Bayesian Learning for Neural Networks*. Springer Verlag, New York.
- [18] Roberts G. O. a Smith A. F. M. (1994). *Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms*. Stoch. Processes and Applic. **49**, 207–216.
- [19] Smid J., Volf P. a Rao G. (1997). *Monte Carlo approach to Bayesian regression modeling*. Computer Intensive Methods in Control and Signal Processing, Birkhauser, Boston, 169–180.
- [20] Smith A. F. M. a Gelfand A. E. (1992). *Bayesian statistics without tears: A sampling–resampling perspective*. The American Statistician, **46**, 84–88.
- [21] Tanner M. A. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer Verlag, New York.
- [22] Tierney L. (1994). *Markov chains for exploring posterior distributions*. Annals of Statist. **22**, 1701–1728.
- [23] Tierney L. (1995). *A Note on Metropolis–Hastings Kernels for General State Spaces*. Techn. Report No 606, School of Statistics, Univ. of Minnesota.
- [24] Volf P. (1996). *Bayesovský odhad parametrů modelu metodami MCMC*. Robust 96, JČMF, 273–284.