МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2013
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2013
Proceedings of the Forty Second Spring Conference of the Union of Bulgarian Mathematicians Borovetz, April 2–6, 2013

KULLBACK-LEIBLER DIVERGENCE MONOTONICITY OF THE METROPOLIS-HASTINGS MARKOV CHAINS*

Dimiter P. Tsvetkov, Lyubomir Ya. Hristov, Ralitsa L. Angelova-Slavova

In this paper we consider classical stationary Markov Chains whose transition kernels satisfy the detailed balance condition. Such Markov Chains usually arise in the Metropolis-Hastings algorithm. It is shown that the sequence of the Kullback-Leibler divergences of the successive densities of the chain from the corresponding invariant distribution is decreasing.

Mertropolis-Hastings algorithm. The Metropolis-Hastings algorithm is one of the best recognized in the statistical applications algorithms (see e.g. [2], [5]). Consider a target distribution with density $\pi(\cdot)$ and proposal distribution with conditional density $q(\cdot|x)$, $x \in \mathbf{S}$, assumed that both have the same support $\mathbf{S}$ that is a measurable subset of some Euclidian space. It is assumed also that for density $q(\cdot|\cdot)$ we have some standard fast method for generating of random draws while $\pi(\cdot)$ is allowed to have quite complex relief. The Metropolis-Hastings scheme consists of the following steps. Choose first initial draw $x_{(0)}$. Suppose we have the current draw $x_{(t-1)}$. For the obtaining of the next draw $x_{(t)}$ one should

1) generate a "candidate" $x_* \sim q\left(x|x_{(t-1)}\right)$

2) compute

$$\alpha = \min\left[1, \frac{\pi(x_*)}{\pi\left(x_{(t-1)}\right)} \frac{q\left(x_{(t-1)}|x_*\right)}{q\left(x_*|x_{(t-1)}\right)}\right]$$

3) generate $u \sim U(0,1)$. If $0 < u < \alpha$ then accept the candidate and take $x_{(t)} = x_*$. If $u \geq \alpha$ then reject the candidate and take $x_{(t)} = x_{(t-1)}$.

This scheme defines transition kernel

$$\kappa(x \to x') = \min\left[1, \frac{\pi(x')}{\pi(x)} \frac{q(x|x')}{q(x'|x)}\right] q(x'|x) + \delta(x - x') \int \left(1 - \min\left[1, \frac{\pi(z)}{\pi(x)} \frac{q(x|z)}{q(z|x)}\right]\right) q(z|x)\, dz \tag{1}$$

where $\delta(x)$ is the Dirac function having the property

$$\int \varphi(x)\, \delta(x - x')\, dx = \varphi(x').$$

*****2000 Mathematics Subject Classification:** 94A17, 60J05, 97K80.
Key words: MCMC, Kullback-Leiber, Metropolis-Hastings.

Here the integral sign stands for an integration over $\mathbf{S}$ (which becomes a summation in the entirely discrete case). The notation $\kappa(x \to x')$ stands for the function of two variables x and x' associated with the conditional probability to move from state x to state x'.

The kernel (1) fulfills the normalizing condition

$$\int \kappa(x \to x')\, dx' = 1$$

and first of all this kernel satisfies the **detailed balance condition**

$$\pi(x)\, \kappa(x \to x') = \pi(x')\, \kappa(x' \to x) \tag{2}$$

which has to be verified only for $x \neq x'$. Actually in this case we have

$$\pi(x)\, \kappa(x \to x') = \pi(x')\, \kappa(x' \to x) = \min\left[\pi(x)\, q(x'|x), \pi(x')\, q(x|x')\right].$$

The transition kernel (1) defines a Markov Chain of random variables $\left(X_{(t)}\right)$ according to the following rule. Define the initial random variable $X_{(0)}$ with a proper initial distribution density $f_{(0)}(\cdot)$. For any next variable $X_{(t)}$ the distribution density is defined by the recurrent formula

$$f_{(t)}(x') = \int f_{(t-1)}(x)\, \kappa(x \to x')\, dx. \tag{3}$$

It follows immediately from the detailed balance condition that the target density $\pi(\cdot)$ is also **invariant** for the chain, i.e. it holds

$$\pi(x') = \int \pi(x)\, \kappa(x \to x')\, dx.$$

The Kullback-Leibler divergence. Given any two densities $f(\cdot)$ and $g(\cdot)$ over $\mathbf{S}$

$$D_{KL}(f||g) = \int f(x) \ln \frac{f(x)}{g(x)} dx$$

is called a Kullback-Leibler divergence (see [4]) of $g(\cdot)$ from $f(\cdot)$. It holds $D_{KL}(f||g) \geq 0$ with equality only in the case $f(\cdot) \equiv g(\cdot)$. Usually $D_{KL}(f||g)$ is used as a measure of distance between $g(\cdot)$ and $f(\cdot)$ nevertheless that it is not symmetric and triangle inequality do not hold either. One of the most useful statistical application of the Kullback-Leibler divergence is in the core of the Akaike Information Criteria (see e.g. [1]). In the proof hereafter we shall use the well known Jensen's inequality for the concave logarithmic function in the following form.

Proposition. *Suppose $h(\cdot)$ is a density over $\mathbf{S}$ and $F(\cdot)$ is a positive measureable function. Then*

$$\ln\left(\int F(x)\, h(x)\, dx\right) \geq \int (\ln F(x))\, h(x)\, dx.$$

We are ready to present the main result in a technical formulation.

Theorem 1. *Suppose that $\kappa(x \to x') : \mathbf{S} \times \mathbf{S} \to \mathbf{R}$ is a transition kernel with invariant density $\pi(\cdot)$ which satisfies the detailed balance condition* (2)*. Let also $h(\cdot)$ be some probability density over $\mathbf{S}$ with transformation*

$$K[h](x') = \int h(x)\, \kappa(x \to x')\, dx.$$

Then

$$D_{KL}(\pi||K[h]) \leq D_{KL}(\pi||h), \tag{4}$$

i.e. the Kullback-Leibler divergence do not increase.

Proof. Consider

$$\int \pi(x) \ln\left(\int h(z)\,\kappa(z \to x)\,dz\right) dx$$

and rewrite it in the following form by using of detailed balance condition (2)

$$\int \pi(x) \ln\left(\int \frac{h(z)}{\pi(z)}\pi(z)\,\kappa(z \to x)\,dz\right) dx = \int \pi(x) \ln\left(\int \frac{h(z)}{\pi(z)}\pi(x)\,\kappa(x \to z)\,dz\right) dx$$

therefore

$$\begin{aligned} &\int \pi(x) \ln\left(\int h(z)\,\kappa(z \to x)\,dz\right) dx \\ &\qquad = \int \pi(x) \ln \pi(x)\,dx + \int \pi(x) \ln\left(\int \frac{h(z)}{\pi(z)}\kappa(x \to z)\,dz\right) dx \end{aligned} \tag{5}$$

Here for the second addend in the right hand side of (5) we can apply the Jensen's inequality because $\kappa(x \to z)$ is a density for any x. Then we obtain

$$\begin{aligned} &\int \pi(x) \ln\left(\int h(z)\,\kappa(z \to x)\,dz\right) dx \\ &\qquad \geq \int \pi(x) \ln \pi(x)\,dx + \int \pi(x) \left(\int \ln\frac{h(z)}{\pi(z)}\kappa(x \to z)\,dz\right) dx \end{aligned} \tag{6}$$

Now changing the order of integration and taking into account that $\pi(\cdot)$ is invariant density for the kernel $\kappa(x \to x')$ we get

$$\begin{aligned} &\int \pi(x) \left(\int \ln\frac{h(z)}{\pi(z)}\kappa(x \to z)\,dz\right) dx \\ &\qquad = \int \ln\frac{h(z)}{\pi(z)} \left(\int \pi(x)\kappa(x \to z)\,dx\right) dz = \int \ln\frac{h(z)}{\pi(z)}\pi(z)\,dz \end{aligned}$$

Replacing the latter in (6) we receive

$$\int \pi(x) \ln\left(\int h(z)\,\kappa(z \to x)\,dz\right) dx \geq \int \pi(x) \ln \pi(x)\,dx + \int \ln\frac{h(z)}{\pi(z)} f(z)\,dz$$

which reduces immediately to

$$\int \pi(x) \ln\left(\int h(z)\,\kappa(z \to x)\,dz\right) dx \geq \int \pi(x) \ln h(x)\,dx.$$

Finally multiplying the latter with (-1) and adding to the both sides $\int \pi(x) \ln \pi(x)\,dx$

we obtain

$$\int \pi(x) \ln \pi(x)\, dx - \int \pi(x) \ln \left(\int h(z)\, \kappa(z \to x)\, dz \right) dx$$
$$\leq \int \pi(x) \ln \pi(x)\, dx - \int \pi(x) \ln h(x)\, dx.$$

Rewrite the latter in the form

$$\int \pi(x) \ln \frac{\pi(x)}{\int h(z)\, \kappa(z \to x)\, dz} dx \leq \int \pi(x) \ln \frac{\pi(x)}{h(x)} dx$$

which is exactly the inequality (4) what we want to prove. The proof is finished.

Here in the proof we use the specific properties of the kernel and the key role falls to the fact that the kernel satisfies detailed balance condition. From Theorem 1 immediately follows the validity of the following theorem.

Theorem 2. *Suppose we are given a Metropolis-Hastings Markov chain corresponding to the kernel* $\kappa(x \to x')$ *defined in* (1) *with some initial distribution density* $f_{(0)}(\cdot)$ *and density chain* $\left(f_{(t)}(\cdot)\right)$ *defined in* (3). *Then the sequence of Kullback-Leibler divergences*

$$D_{KL}\left(\pi || f_{(t)}\right) \tag{7}$$

does not increase.

Actually in the typical case the sequence (7) is strongly decreasing because in the typical common application case hereafter we have a strong Jensen's inequality

$$\ln \left(\int \frac{h(z)}{\pi(z)} \kappa(x \to z)\, dz \right) > \int \ln \frac{h(z)}{\pi(z)} \kappa(x \to z)\, dz.$$

Notes and conclusions. After this work was completed we found that the result of Theorem 1 (therefore also the result of Theorem 2) can be derived from the chain rule of the relative entropy (or from so called "data processing inequality") as it is announced in [3] for finite state space Markov chains. The proof we offer for Theorem 1 uses completely different approach than in [3] but uses essentially the detailed balance condition. Our results also are aimed to the general state space case.

The results of these theorems are not related to the question about the conditions under which the sequence $\left(f_{(t)}(\cdot)\right)$ converges in some usual metric to the target density $\pi(\cdot)$. Here one can refer for example to the common general settings and assumptions proposed in [5].

These theorems make in a certain sense clearer the well-known good behavior of the Metropolis-Hastings algorithm in the general practical case.

REFERENCES

[1] K. Burnham, D. Anderson. Model selection and multi-model inference. Springer-Verlag, 2002.

[2] N. Chen. Monte Carlo Methods in Bayesian Computation. Springer Verlag, 2000.

[3] T. Cover, J. Thomas. Elements of Information theory. Wiley-Interscience, Hoboken, 2006.

[4] S. Kullback. Information theory and statistics. Dover Publications, 1997.

[5] C. Robert, G. Casella. Monte Carlo Statistical Methods. Springer Verlag, 2004.

Dimiter Petkov Tsvetkov
Department of Mathematical Analysis and Applications
Faculty of Mathematics and Informatics
St. Cyril and St. Methodius University of Veliko Tarnovo
5000 Veliko Tarnovo, Bulgaria
e-mail: dimiter.tsvetkov@yahoo.com

Lyubomir Yanakiev Hristov
Department of Mathematical Analysis and Applications
Faculty of Mathematics and Informatics
St. Cyril and St. Methodius University of Veliko Tarnovo
5000 Veliko Tarnovo, Bulgaria
e-mail: lyubomir.hristov@gmail.com

Ralitsa Lubomirova Angelova-Slavova
Department of Communication Systems and Technologies
Vasil Levski National Military University at Veliko Tarnovo
5000 Veliko Tarnovo, Bulgaria
e-mail: rali_a@abv.bg

МОНОТОННОСТ НА КУЛБАК-ЛИБЛЕР ОТКЛОНЕНИЯТА ПРИ МЕТРОПОЛИС-ХЕСТИНГС МАРКОВСКИ ВЕРИГИ

Димитър П. Цветков, Любомир Я. Христов, Ралица Л. Ангелова-Славова

В тази статия се разглеждат класически стационарни Марковски вериги, които удовлетворяват условието за подробен баланс. Такива Марковски вериги обикновено възникват при Метрополис-Хестингс схеми за случаен избор. Доказано е, че редицата от Кулбак-Либлер отклоненията на последователните плътности на веригата от съответното стационарно разпределение е намаляваща.