Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications ▮ (▮▮▮▮) ▮▮▮–▮▮▮

stochastic processes and their applications

www.elsevier.com/locate/spa

Cutoffs for product chains

Guan-Yu Chen[a], Takashi Kumagai[b,*]

[a] *Department of Appl. Math., National Chiao Tung University, Hsinchu 300, Taiwan*
[b] *RIMS, Kyoto University, Kyoto 606-8502, Japan*

Received 9 February 2017; received in revised form 18 September 2017; accepted 3 January 2018
Available online xxxx

Abstract

We consider products of ergodic Markov chains and discuss their cutoffs in total variation. Our framework is general in that rates to pick up coordinates are not necessary equal, and different coordinates may correspond to distinct chains. We give necessary and sufficient conditions for cutoffs of product chains in terms of those of coordinate chains under certain conditions. A comparison of mixing times between the product chain and its coordinate chains is made in detail as well. Examples are given to show that neither cutoffs for product chains nor for coordinate chains imply others in general.

MSC: 60J10; 60J27

Keywords: Product chains; Total variation and Hellinger distances; Cutoffs

1. Introduction

Let $\mathcal{X}$ be a countable set, K be an irreducible stochastic matrix indexed by $\mathcal{X}$ and π be a probability on $\mathcal{X}$. We write the triple $(\mathcal{X}, K, \pi)$ for a discrete time Markov chain on $\mathcal{X}$ with transition matrix K and stationary distribution π. It is well-known that if K is aperiodic, then $K^m(x, y)$ converges to $\pi(y)$ as m tends to infinity for all $x, y \in \mathcal{X}$. To quantify the convergence of K^m to π, we consider the (maximum) total variation and Hellinger distances, which are defined by

$$d_{\mathrm{TV}}(m) := \sup_{x\in\mathcal{X}, A\subset\mathcal{X}} \{K^m(x, A) - \pi(A)\}, \tag{1.1}$$

* Corresponding author.
E-mail addresses: gychen@math.nctu.edu.tw (G.-Y. Chen), kumagai@kurims.kyoto-u.ac.jp (T. Kumagai).

https://doi.org/10.1016/j.spa.2018.01.002
0304-4149/

Please cite this article in press as: G.-Y. Chen, T. Kumagai, Cutoffs for product chains, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.01.002.

and

$$d_H(m) := \sup_{x\in\mathcal{X}} \left(\frac{1}{2} \sum_{y\in\mathcal{X}} \left(\sqrt{K^m(x,y)} - \sqrt{\pi(y)} \right)^2 \right)^{1/2}. \tag{1.2}$$

As the above distances are non-increasing in m, it is natural to consider the mixing times of d_{TV} and d_H, which are respectively defined by

$$T_{\text{TV}}(\epsilon) := \inf\{m \ge 0 | d_{\text{TV}}(m) \le \epsilon\}, \quad T_H(\epsilon) := \inf\{m \ge 0 | d_H(m) \le \epsilon\}.$$

For weak convergence of distributions, the total variation distance arose naturally from the view point of probability, while the importance of the Hellinger distance is exemplified from the proof of Kakutani's dichotomy theorem in [11] for the study of infinite product measures. The following inequalities provide a comparison of the total variation and Hellinger distances, which are corollaries in [16] (see (25) on p. 365 for the details) and say

$$1 - \sqrt{1 - d_{\text{TV}}^2(m)} \le d_H^2(m) \le d_{\text{TV}}(m). \tag{1.3}$$

As a consequence, one obtains from (1.3) the following comparison of mixing times,

$$T_{\text{TV}}(\epsilon\sqrt{2-\epsilon^2}) \le T_H(\epsilon) \le T_{\text{TV}}(\epsilon^2), \quad \forall \epsilon \in (0,1). \tag{1.4}$$

We can further compare the cutoffs, introduced below, in total variation and Hellinger distance. Such a comparison will play a key role through this article.

In this article, we focus on the study of product chains and their cutoffs. To see a definition of product chains, let $(\mathcal{X}_i, K_i, \pi_i)_{i=1}^n$ be irreducible Markov chains and set

$$\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n, \quad \pi = \pi_1 \times \cdots \times \pi_n, \tag{1.5}$$

and

$$K = \sum_{i=1}^n p_i I_1 \otimes \cdots \otimes I_{i-1} \otimes K_i \otimes I_{i+1} \otimes \cdots \otimes I_n, \tag{1.6}$$

where I_j is the identity matrix indexed by $\mathcal{X}_j$, $A \otimes B$ denotes the tensor product of matrices A, B and $p_1, \dots, p_n$ are positive reals satisfying $p_1 + \cdots + p_n = 1$. It is obvious that K is a transition matrix on $\mathcal{X}$ with stationary distribution π. Thereafter, we call $(\mathcal{X}, K, \pi)$ the product chain of $(\mathcal{X}_i, K_i, \pi_i)_{i=1}^n$ according to the probability vector $(p_1, \dots, p_n)$.

Note that $p_1 = \cdots = p_n$ is one natural setting. In this case, Levin, Peres and Wilmer provide in [13, Theorem 20.7] tight bounds on the mixing time using spectral gaps of Markov chains in each coordinate. This leads to a sufficient condition on the cutoff of product chains. We are interested in obtaining necessary and sufficient conditions under a wider weighted product chain. While our motivation is purely mathematical, we note that it is a natural extension from the point of view of random media, in that the probability vector $(p_i)_{i=1}^n$ is considered as a random weight that gives preference to choose each coordinate. It is also natural from the perspective of Markov chain Monte Carlo (MCMC, for short) method, in that one cannot expect all the time that the MCMC algorithm is correctly implemented due to the restriction of the binary machinery. For instance, when $n = 7$, one has $p_i = 1/7$ but $1/7$ cannot be precisely stored with finitely many digits. Hence, an approximation to $1/7$ is made and this could result in various perturbations on transition matrices. In addition to the fluctuation from the noise or error within machines, it is natural to consider more generally that $(p_i)_{i=1}^n$ is any probability vector. By these reasons,