应用概率统计 第二十九卷
第五期 2013年10月

Chinese Journal of Applied Probability
and Statistics Vol.29 No.5 Oct. 2013

A Generalization of Dobrushin Coefficient *

MAO YONGHUA ZHANG MING⋆ ZHANG YUHUI

(*School of Mathematical Sciences, Beijing Normal University, Laboratory of Mathematics and Complex Systems, Ministry of Education, Beijing, 100875*)

Abstract

We generalize the well-known Dobrushin coefficient δ in total variation to weighted total variation δ_V, which gives a criterion for the geometric ergodicity of discrete-time Markov chains.

Keywords: V-norm, δ_V coefficient, geometric ergodicity.

AMS Subject Classification: 37A30, 47A35.

§1. Introduction and Main Results

In this paper, we generalize the classical Dobrushin coefficient, in order to give a criteria for geometric ergodicity of discrete-time Markov chains.

Let $X = (X_n)_{n\geq 0}$ be a discrete-time Markov chain taking values on a measurable space $(E, \mathscr{E})$. Denote by

$$P^n(x, A) := \mathsf{P}[X_n \in A|X_0 = x], \qquad x \in E,\ A \in \mathscr{E},$$

the n-step transition kernel. Set $P = P^1$ be the one-step kernel.

Throughout the paper, we assume that X is ψ-irreducible and aperiodic, c.f. [1].

We are interested in the geometrical ergodicity that is, there is an invariant probability measure π on $(E, \mathscr{E})$ and a constant $\rho \in [0, 1)$ and a function $C : E \to (0, \infty)$ such that

$$\|P^n(x, \cdot) - \pi\|_{\mathsf{Var}} \leq C(x)\rho^n, \qquad \text{for all } n > 0,\ \pi\text{-a.s. } x \in E, \tag{1.1}$$

where $\|\mu\|_{\mathsf{Var}} := 2\sup\limits_{A\in\mathscr{E}} |\mu(A)| = \sup\left\{\int_E f\mathrm{d}\mu : |f| \leq 1\right\}$ is the total variation for a signed measure μ.

*The research was supported in part by 985 Project (212011), 973 Project (2011CB808000), the National Natural Science Foundation of China (11131003), the Specialized Research Fund for the Doctoral Program of Higher Education (20100003110005) and the Fundamental Research Funds for the Central Universities.

⋆Corresponding author, E-mail: zhangming0408@mail.bnu.edu.cn.

Received October 23, 2012. Revised June 28, 2013.

From [1, 2], we know that (1.1) is equivalent to that there exist constants $\rho \in [0,1)$, $C < \infty$ and a π-a.s. finite function $V : E \to [1,\infty]$ such that

$$\|P^n(x,\cdot) - \pi\|_V \le CV(x)\rho^n, \tag{1.2}$$

where $\|\mu\|_V := \sup\Big\{\int_E f\mathrm{d}\mu : \|f\|_V \le 1\Big\}$ denotes the V-norm on signed measures. Recall that for a kernel $K(x,A)$ ($x \in E,\ A \in \mathscr{E}$),

$$\|K\|_V := \sup_{x\in E} \frac{\|K(x,\cdot)\|_V}{V(x)}, \tag{1.3}$$

and for a function f,

$$\|f\|_V := \sup_{x\in E} \frac{|f(x)|}{V(x)}, \tag{1.4}$$

c.f. [1, Chapter 14].

Specially, if in (1.1) $C(x) \le C < \infty$ for π-a.s. $x \in E$, (1.1) is strengthened to the so-called strong ergodicity. Equivalently in (1.2), we can choose $V : E \to [1,\infty)$ to be (upper) bounded. Recall that X is said to be strongly ergodic if there exist constants $\rho \in [0,1)$ and $C < \infty$ such that

$$\mathrm{ess}_\pi \sup_{x\in E} \|P^n(x,\cdot) - \pi\|_{\mathsf{Var}} \le C\rho^n, \qquad \text{for all } n > 0. \tag{1.5}$$

Dobrushin (1956) gave an elegant criterion for strong ergodicity in (1.5). Let

$$\delta(P) = \frac{1}{2}\mathrm{ess}_\pi \sup_{x,y\in E} \|P(x,\cdot) - P(y,\cdot)\|_{\mathsf{Var}}.$$

Then, X is strongly ergodic if and only if there exists n, such that $\delta(P^n) < 1$. $\delta(P)$ is now in term of Dobrushin coefficient.

In the context of E countable, an elegant description of $\delta(P)$ and strong ergodicity can be found in [4, § 6.1, § 6.3]. Although [4, § 6.3] only considered continuous-time Markov chains, these arguments remain valid for a general Markov process. Or this can be viewed as a special case where we will do in the following (by taking $V(x) \equiv 1$).

Now, we will generalize $\delta(P)$ to $\delta_V(P)$, which gives a criteria for geometric ergodicity in (1.1) or (1.2).

Definition 1.1 For a transition kernel P, and a π-a.s. finite function $V : E \to [1,\infty]$, we define the generalized Dobrushin coefficient $\delta_V(P)$:

$$\delta_V(P) := \mathrm{ess}_\pi \sup_{x,y\in E} \frac{1}{V(x)+V(y)} \|P(x,\cdot) - P(y,\cdot)\|_V. \tag{1.6}$$

Here is the main result.

Theorem 1.1 A ψ-irreducible and aperiodic Markov chain $X = (X_n)_{n\geq 0}$ is geometrically ergodic if and only if X is ergodic with stationary distribution π, and there is a π-a.s. finite function $V : E \to [1, \infty]$ such that $\pi(V) < \infty$ and $\delta_V(P^n) < 1$ for n large enough.

Note that when $V(x) \equiv 1$, $\delta_V(P) = \delta(P)$ and hence Theorem 1.1 generalizes the classical Dobrushin coefficient.

The key ingredient for the proof of Theorem 1.1 comes from an excellent observation. By defining a metric on E as in [5], we can transfer $\delta_V(P)$ to be a Wasserstein distance on space of probability measures.

§2. Proof of Theorem

For simplicity, in the following, sup is referred to $\mathrm{ess}_\pi \sup$.

Following [5], we introduce a metric on E. Let $V : E \to [1, \infty)$. For $x, y \in E$,

$$d_V(x,y) = \begin{cases} 0, & x = y; \\ V(x) + V(y), & x \neq y. \end{cases}$$

Then (E, d_V) is a complete metric space. Let φ be a function on (E, d_V), the Lipschitz norm of φ is

$$\|\varphi\|_{\mathrm{Lip}(d_V)} := \sup_{x\neq y} \frac{|\varphi(x) - \varphi(y)|}{d_V(x,y)} = \sup_{x\neq y} \frac{|\varphi(x) - \varphi(y)|}{V(x) + V(y)}.$$

For two (probability) measures μ_1, μ_2 on $(E, \mathcal{E})$, Wasserstein metric W_{d_V} is defined by

$$W_{d_V}(\mu_1, \mu_2) = \sup_{\varphi : \|\varphi\|_{\mathrm{Lip}(d_V)} \leq 1} \int \varphi \mathrm{d}(\mu_1 - \mu_2).$$

The following was proved in [5, Lemma 2.1].

Lemma 2.1 For two probability measures μ_1, μ_2 on $(E, \mathcal{E})$,

$$\|\mu_1 - \mu_2\|_V = W_{d_V}(\mu_1, \mu_2). \tag{2.1}$$

Remark 1 (1) For $V(x) \equiv 1$, both sides of (2.1) are reduced to the total variation of $\mu_1 - \mu_2$. For the left side, when $V(x) \equiv 1$,

$$\|\mu_1 - \mu_2\|_V = \sup\left\{ \int_E f \mathrm{d}(\mu_1 - \mu_2) : |f| \leq 1 \right\} = \|\mu_1 - \mu_2\|_{\mathsf{Var}}.$$

For the right side, d_V is nothing but discrete metric. Hence by [6, Theorem 5.7], $W_{d_V}(\mu_1, \mu_2) = \|\mu_1 - \mu_2\|_{\mathrm{Var}}$.

(2) As pointed out in [5], (2.1) is also true for measures μ_1, μ_2 with equal mass.

The following properties are essential to the proof of Theorem 1.1.

Proposition 2.1 Let P and $\widetilde{P}$ be probability transition kernels on $(E, \mathcal{E})$. Then

(i) $\delta_V(P\widetilde{P}) \le \delta_V(P)\delta_V(\widetilde{P})$,

(ii) $|\delta_V(P) - \delta_V(\widetilde{P})| \le \|P - \widetilde{P}\|_V$,

(iii) Let R be a transition kernel on $(E, \mathcal{E})$ such that $R(x, E) = 0$, for all $x \in E$ and $\|R\|_V < \infty$. Then $\|RP\|_V \le \|R\|_V \delta_V(P)$.

Proof (i) By the definition of $\delta_V(P)$ and Lemma 2.1, we have

$$W_{d_V}(P(x,\cdot), P(y,\cdot)) = \|P(x,\cdot) - P(y,\cdot)\|_V \le \delta_V(P)(V(x) + V(y)), \qquad \pi\text{-a.s. } x, y \in E.$$

Hence by the definition of W_{d_V}, we get

$$|P\varphi(x) - P\varphi(y)| \le \delta_V(P)\|\varphi\|_{\mathrm{Lip}(d_V)}(V(x) + V(y))$$

or

$$\frac{1}{V(x) + V(y)}|P\varphi(x) - P\varphi(y)| \le \delta_V(P)\|\varphi\|_{\mathrm{Lip}(d_V)},$$

and

$$\|P\varphi\|_{\mathrm{Lip}(d_V)} \le \delta_V(P)\|\varphi\|_{\mathrm{Lip}(d_V)}. \tag{2.2}$$

Since $\displaystyle\int (\widetilde{P}P)(x, \mathrm{d}z)\varphi(z) = \int (P\varphi)(z)\widetilde{P}(x, \mathrm{d}z)$, we get from (2.2) that

$$\begin{aligned} W_{d_V}((\widetilde{P}P)(x,\cdot), (\widetilde{P}P)(y,\cdot)) &= \sup_{\|\varphi\|_{\mathrm{Lip}(d_V)} \le 1} \int (P\varphi)(z)(\widetilde{P}(x, \mathrm{d}z) - \widetilde{P}(y, \mathrm{d}z)) \\ &\le \delta_V(P) \sup_{\|\psi\|_{\mathrm{Lip}(d_V)} \le 1} \int \psi(z)(\widetilde{P}(x, \mathrm{d}z) - \widetilde{P}(y, \mathrm{d}z)) \\ &= \delta_V(P) W_{d_V}(\widetilde{P}(x,\cdot), \widetilde{P}(y,\cdot)). \end{aligned}$$

Therefore by Lemma 2.1 again, we have

$$\begin{aligned} \delta_V(\widetilde{P}P) &= \sup_{x,y\in E} \frac{1}{V(x) + V(y)} \|(\widetilde{P}P)(x,\cdot) - (\widetilde{P}P)(y,\cdot)\|_V \\ &= \sup_{x,y\in E} \frac{1}{V(x) + V(y)} W_{d_V}((\widetilde{P}P)(x,\cdot), (\widetilde{P}P)(y,\cdot)) \\ &\le \delta_V(P) \sup_{x,y\in E} \frac{1}{V(x) + V(y)} W_{d_V}(\widetilde{P}(x,\cdot), \widetilde{P}(y,\cdot)) \\ &= \delta_V(P) \sup_{x,y\in E} \frac{1}{V(x) + V(y)} \|\widetilde{P}(x,\cdot) - \widetilde{P}(y,\cdot)\|_V \\ &= \delta_V(P)\delta_V(\widetilde{P}). \end{aligned}$$

(ii) Using the fact that $\left|\sup_x |u(x)| - \sup_x |v(x)|\right| \le \sup_x |u(x) - v(x)|$, we get

$$\begin{aligned}
&|\delta_V(P) - \delta_V(\widetilde{P})| \\
= & \left| \sup_{x,y\in E} \frac{1}{V(x)+V(y)} \|P(x,\cdot) - P(y,\cdot)\|_V - \sup_{x,y\in E} \frac{1}{V(x)+V(y)} \|\widetilde{P}(x,\cdot) - \widetilde{P}(y,\cdot)\|_V \right| \\
\le & \sup_{x,y\in E} \frac{1}{V(x)+V(y)} \sup_{|f|\le V} |(Pf(x) - Pf(y)) - (\widetilde{P}f(x) - \widetilde{P}f(y))| \\
\le & \sup_{x,y\in E} \frac{1}{V(x)+V(y)} \sup_{|f|\le V} [|Pf(x) - \widetilde{P}f(x)| + |Pf(y) - \widetilde{P}f(y)|] \\
\le & \sup_{x,y\in E} \frac{\|P(x,\cdot) - \widetilde{P}(x,\cdot)\|_V + \|P(y,\cdot) - \widetilde{P}(y,\cdot)\|_V}{V(x)+V(y)} \\
\le & \sup_{x\in E} \frac{\|P(x,\cdot) - \widetilde{P}(x,\cdot)\|_V}{V(x)} \\
= & \|P - \widetilde{P}\|_V,
\end{aligned}$$

where in the last inequality, we use the element that

$$\frac{u(x)+u(y)}{v(x)+v(y)} \le \max\left\{\frac{u(x)}{v(x)}, \frac{u(y)}{v(y)}\right\}.$$

(iii) Since $R(x,E) = 0$, for all $x \in E$, we have Jordan-Hahn decomposition $R = R^+ - R^-$, hence

$$\begin{aligned}
\|RP\|_V &= \sup_{x\in E} \frac{1}{V(x)} \|(R^+P)(x,\cdot) - (R^-P)(x,\cdot)\|_V \\
&= \sup_{x\in E} \frac{1}{V(x)} W_{d_V}((R^+P)(x,\cdot), (R^-P)(x,\cdot)).
\end{aligned}$$

It follows from (2.2) and Lemma 2.1 that,

$$\begin{aligned}
W_{d_V}((R^+P)(x,\cdot), (R^-P)(x,\cdot)) &= \sup_{\varphi:\|\varphi\|_{\mathrm{Lip}(d_V)}\le 1} \int \varphi(y)((R^+P)(x,\mathrm{d}y) - (R^-P)(x,\mathrm{d}y)) \\
&= \sup_{\varphi:\|\varphi\|_{\mathrm{Lip}(d_V)}\le 1} \int (P\varphi)(y)[R^+(x,\mathrm{d}y) - R^-(x,\mathrm{d}y)] \\
&\le \delta_V(P) \sup_{\phi:\|\phi\|_{\mathrm{Lip}(d_V)}\le 1} \int \phi(y)[R^+(x,\mathrm{d}y) - R^-(x,\mathrm{d}y)] \\
&= \delta_V(P) W_{d_V}(R^+(x,\cdot), R^-(x,\cdot)) \\
&= \delta_V(P) \|R^+(x,\cdot) - R^-(x,\cdot)\|_V.
\end{aligned}$$

Finally we obtain

$$\|RP\|_V \le \delta_V(P) \sup_{x\in E} \frac{1}{V(x)} \|R^+(x,\cdot) - R^-(x,\cdot)\|_V = \delta_V(P)\|R\|_V. \qquad \square$$

Proof of Theorem 1.1 Set $\Pi(x,\cdot)=\pi$, for all $x\in E$. As noted in [1], the function V in (1.6) can be chosen to be such that $\pi(V)<\infty$. Since $\delta_V(\Pi)=0$, we obtain

$$\delta_V(P^n)=|\delta_V(P^n)-\delta_V(\Pi)|\le\|P^n-\Pi\|_V=\sup_{x\in E}\frac{1}{V(x)}\|P^n(x,\cdot)-\pi\|_V\to 0.$$

Conversely, we have $\|I-\Pi\|_V\le\sup\limits_{x\in E}[V(x)+\pi(V)]/V(x)=1+\pi(V)<\infty$ and $(I-\Pi)(x,E)=0$ for all $x\in E$. Then

$$\|P^n-\Pi\|_V=\|P^n-\Pi P^n\|_V=\|(I-\Pi)P^n\|_V\le\|I-\Pi\|_V\delta_V(P^n),$$

and

$$\|P^n(x,\cdot)-\pi\|_V\le\|P^n-\Pi\|_V V(x)\le\|I-\Pi\|_V V(x)\delta_V(P^n).$$ □

References

[1] Meyn, S.P. and Tweedie, R.L., *Markov Chains and Stochastic Stability*, Springer-Verlag, 1996.

[2] Roberts, G.O. and Rosenthal, J.S., Geometric ergodicity and hybrid Markov chains, *Electronic Communications in Probability*, **2**(1997), 13–25.

[3] Dobrushin, R.L., Central limit theorem for nonstationary Markov chains I, II, *Theory of Probability & Its Applications*, **1(1)**(1956), 65–80, **1(4)**(1956), 329-383.

[4] Anderson, W.J., *Continuous-Time Markov Chains*, Springer-Verlag, 1991.

[5] Hairer, M. and Mattingly, J.C., Yet another look at Harris' ergodic theorem for Markov chains, *Seminar on Stochastic Analysis, Random Fields and Applications VI: Progress in Probability*, **63**(2011), 109–117.

[6] Chen, M.F., *From Markov Chains to Non-equilibrium Particle Systems*, World Scientufic, Second Edition, 2004.

[7] Kontoyiannis, I. and Meyn, S.P., Geometric ergodicity and the spectral gap of non-reversible Markov chains, *Probability Theory and Related Fields*, **154(1-2)**(2012), 327–339.

Dobrushin系数的推广

毛永华　张　铭　张余辉

(北京师范大学数学科学学院, 数学与复杂系统教育部重点实验室, 北京, 100875)

本文将一般的全变差距离下的Dobrushin系数δ推广到加权的全变差下的δ_V, 并利用δ_V系数得到了离散时间马氏链的几何遍历的判定准则.

关键词: V范数, δ_V系数, 几何遍历.

学科分类号: O211.62.