Journal of the Operations Research
Society of Japan
Vol. 32, No. 1, March 1989

UPPER BOUNDS OF A MEASURE OF DEPENDENCE AND THE RELAXATION TIME FOR FINITE STATE MARKOV CHAINS

Masaaki Kijima
Tokyo Institute of Technology

(Received April 28, 1988; Revised September 22, 1988)

Abstract In this paper, we consider a measure of dependence between X_t and X_0 where X_t is an irreducible Markov chain on a finite state space. Namely, we define $d_t(X) = \sup_{f,g} Cor\,[f(\hat{X}_0), g(\hat{X}_t)]$. Here $\hat{X}_t$ is the stationary process associated with X_t and the supreme is taken over all real functions. An upper bound of $d_t(X)$, which is easy to calculate numerically, is derived. By showing a simple relation between $d_t(X)$ and the relaxation time $T_{REL}(X)$ of X_t, we also provide an upper bound of $T_{REL}(X)$. The bounds are shown to be tight when the Markov chain is reversible in time.

Key words. Markov chain, Finite time behavior, Relaxation time, Time reversibility

1 Introduction

It is well known that, under many circumstances, a Markov process X_t which converges to a stationary distribution π does so asymptotically at an exponential rate. That is,

$$\Pr[X_t \in A] - \pi(A) \sim O(\exp\{-rt\}) \quad \text{as} \quad t \to \infty. \tag{1.1}$$

Here the rate r has an eigenvalue interpretation and its reciprocal $T_{REL}(X) = r^{-1}$ is often called the relaxation time of the process. The relaxation time is useful in many applications since it indicates the order of magnitude of quantities such as the time for the effects of an external shock to a system to wear off (see e.g. Aldous [1] and references therein). However, the difficulty of finding the rate r has limited the use of the information in practice. In a recent paper by Aldous [1], it is pointed out that some measure of dependence between X_t and X_0 may be more informative for the finite time behavior of the process than the relaxation time. One possible variant provided there is

$$d_t(X) = \sup_{f,g} Cor[f(\hat{X}_0), g(\hat{X}_t)], \tag{1.2}$$

where $\hat{X}_t$ is the stationary process associated with X_t and the supreme is taken over all measurable functions. As for the relaxation time, the quantity $d_t(X)$ is hard to evaluate for general cases due to the complex form of its own. Only when the Markov process is finite and reversible in time, it is known that $T_{REL}(X) = r^{-1}$ and $d_t(X) = \exp\{-rt\}$ where $r = \min\{|\lambda_j|; \lambda_j \neq 0\}$ and λ_j are eigenvalues, which are all real, of the governing infinitesimal generator (see e.g. Keilson [2]). Other than these, no theoretical results have been found. The relation between $T_{REL}(X)$ and $d_t(X)$ is not clear. The purpose of this paper is to relate $T_{REL}(X)$ and $d_t(X)$ and to give upper bounds of them, which are easy to evaluate numerically, thereby providing useful information about the relaxation time as well as the finite time behavior of the Markov process under consideration.

In this paper, we shall study only finite state space Markov processes. Note that, in order to understand finite time properties, there is no loss of generality in restricting our attention to finite state spaces, see [1]. In the next section, we first consider a discrete time Markov chain. A measure of dependence similar to (1.2) for the discrete time case is defined and an upper bound of the measure is obtained for an ergodic Markov chain. The bound is related to the maximum eigenvalue of a symmetric matrix constructed from the governing transition probability matrix. The method employed is fully algebraic and seems to have no probabilistic interpretation. In Section 3, these results are applied for a continuous time irreducible Markov chain via a uniformization to obtain an upper bound of the measure $d_t(X)$. Remarks regarding the relation between the rate r in (1.1) and $d_t(X)$ are also stated. In particular, it will be shown that $d_t(X)$ converges to zero asymptotically at the rate r. Since the upper bound obtained is an exponential function, the exponential rate also provides an upper bound of the relaxation time.

2 Discrete Time Markov Chains and a Measure of Dependence

Let $\mathcal{N} = \{1, 2, \cdots, N\}$ $(N < \infty)$ be the state space and let X_n be a discrete time Markov chain on $\mathcal{N}$ governed by the transition probability matrix $\boldsymbol{A}$. It is assumed throughout this section that the Markov chain is ergodic or, equivalently, the matrix is primitive (see Seneta [5] for the definition). Then there exists a set of positive probabilities $e_k = \lim_{n\to\infty} \Pr[X_n = k]$ satisfying the equations $\boldsymbol{e}^T\boldsymbol{A} = \boldsymbol{e}^T$ and $\boldsymbol{e}^T\boldsymbol{1} = 1$ where $\boldsymbol{e} = (e_1, \cdots, e_N)^T$ and $\boldsymbol{1} = (1, \cdots, 1)^T$. Here T denotes the transpose. It should be noted that the corresponding stationary chain $\hat{X}_n$ is statistically determined only through the transition matrix $\boldsymbol{A}$. Hence it may be helpful to identify the matrix $\boldsymbol{A}$ rather than the stationary chain $\hat{X}_n$ in our notation. Accordingly, we define

$$d_n(\boldsymbol{A}) = \sup_{f,g} Cor[f(\hat{X}_0), g(\hat{X}_n)], \quad n \geq 1, \tag{2.1}$$

with $d_0(\boldsymbol{A}) = 1$.

Write the diagonal matrix having the diagonal elements e_k by $\boldsymbol{E}_D$, i.e., $\boldsymbol{E}_D = diag(e_k)$. It is not hard to see that

$$Cov[f(\hat{X}_0), g(\hat{X}_n)] = \boldsymbol{f}^T \boldsymbol{E}_D[\boldsymbol{A}^n - \boldsymbol{1}\boldsymbol{e}^T]\boldsymbol{g} \tag{2.2}$$

and

$$Var[f(\hat{X}_0)] = \boldsymbol{f}^T \boldsymbol{E}_D \boldsymbol{f} - (\boldsymbol{f}^T\boldsymbol{e})^2; \quad Var[g(\hat{X}_n)] = \boldsymbol{g}^T \boldsymbol{E}_D \boldsymbol{g} - (\boldsymbol{g}^T\boldsymbol{e})^2, \tag{2.3}$$

where $\boldsymbol{f} = (f(1), \cdots, f(N))^T$ and $\boldsymbol{g} = (g(1), \cdots, g(N))^T$ (cf. Keilson [2]). It follows from (2.1) through (2.3) that

$$d_n(\boldsymbol{A}) = \sup_{f,g} \frac{\boldsymbol{f}^T \boldsymbol{E}_D[\boldsymbol{A}^n - \boldsymbol{1}\boldsymbol{e}^T]\boldsymbol{g}}{\sqrt{\boldsymbol{f}^T \boldsymbol{E}_D \boldsymbol{f} - (\boldsymbol{f}^T\boldsymbol{e})^2}\sqrt{\boldsymbol{g}^T \boldsymbol{E}_D \boldsymbol{g} - (\boldsymbol{g}^T\boldsymbol{e})^2}}. \tag{2.4}$$

Let $\boldsymbol{R} = \boldsymbol{A} - \boldsymbol{1}\boldsymbol{e}^T$. Then $\boldsymbol{1}\boldsymbol{e}^T\boldsymbol{R} = \boldsymbol{R}\boldsymbol{1}\boldsymbol{e}^T = \boldsymbol{O}$, where $\boldsymbol{O}$ denotes the zero matrix, so that $\boldsymbol{A}^n = \boldsymbol{1}\boldsymbol{e}^T + \boldsymbol{R}^n$, $n \geq 1$. Write $\tilde{\boldsymbol{C}} = \boldsymbol{E}_D^{1/2}\boldsymbol{C}\boldsymbol{E}_D^{-1/2}$ for any square matrix $\boldsymbol{C}$. It is then readily seen that $\sqrt{\boldsymbol{e}}\sqrt{\boldsymbol{e}}^T\tilde{\boldsymbol{R}} = \tilde{\boldsymbol{R}}\sqrt{\boldsymbol{e}}\sqrt{\boldsymbol{e}}^T = \boldsymbol{O}$ and

$$\tilde{\boldsymbol{A}}^n = \sqrt{\boldsymbol{e}}\sqrt{\boldsymbol{e}}^T + \tilde{\boldsymbol{R}}^n, \quad n \geq 1, \tag{2.5}$$

where $\sqrt{\boldsymbol{e}} = (\sqrt{e_1}, \cdots, \sqrt{e_N})^T$. Let W be a subspace of R^N orthogonal to $\sqrt{\boldsymbol{e}}$, i.e. $W = \{\boldsymbol{y} : \boldsymbol{y}^T\sqrt{\boldsymbol{e}} = 0\}$. Then any $\boldsymbol{x} \in R^N$ can be decomposed as $\boldsymbol{x} = x_0\sqrt{\boldsymbol{e}} + \tilde{\boldsymbol{x}}$ with some $x_0 \in R$ and $\tilde{\boldsymbol{x}} \in W$. We note that the transformation $\tilde{\boldsymbol{R}}^n$ maps R^N into W. By letting $\boldsymbol{x} = \boldsymbol{E}_D^{1/2}\boldsymbol{f} = x_0\sqrt{\boldsymbol{e}} + \tilde{\boldsymbol{x}}$ and $\boldsymbol{y} = \boldsymbol{E}_D^{1/2}\boldsymbol{g} = y_0\sqrt{\boldsymbol{e}} + \tilde{\boldsymbol{y}}$, one has after a calculus that

$$d_n(\boldsymbol{A}) = \sup_{f,g} \frac{\tilde{\boldsymbol{x}}^T\tilde{\boldsymbol{R}}^n\tilde{\boldsymbol{y}}}{\|\tilde{\boldsymbol{x}}\|\,\|\tilde{\boldsymbol{y}}\|},$$

which is independent of x_0 and y_0. Here the vector norm $\|\cdot\|$ is defined as the Euclidian norm. Thus, one finally arrives at the expression

$$d_n(\boldsymbol{A}) = \sup_{\boldsymbol{x},\boldsymbol{y}\in W} \frac{\boldsymbol{x}^T\tilde{\boldsymbol{R}}^n\boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|} = \sup_{\boldsymbol{x},\boldsymbol{y}\in W} \frac{\boldsymbol{x}^T\tilde{\boldsymbol{A}}^n\boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}. \tag{2.6}$$

In what follows, we denote by $\lambda_j(\boldsymbol{C})$, $j = 1, \cdots, N$, the eigenvalues of square matrix $\boldsymbol{C}$. They are ordered so that $|\lambda_1(\boldsymbol{C})| \geq |\lambda_2(\boldsymbol{C})| \geq \cdots \geq |\lambda_N(\boldsymbol{C})|$ unless specified otherwise. If $\boldsymbol{C}$ is a stochastic matrix, then $\lambda_1(\boldsymbol{C}) = 1$. Also, since $\tilde{\boldsymbol{C}} = \boldsymbol{E}_D^{1/2}\boldsymbol{C}\boldsymbol{E}_D^{-1/2}$ is a similarity transform, one has $\lambda_j(\tilde{\boldsymbol{C}}) = \lambda_j(\boldsymbol{C})$. It should be noted from (2.5) that $\lambda_j(\tilde{\boldsymbol{R}}) = \lambda_{j+1}(\boldsymbol{A})$, $j = 1, \cdots, N-1$, and $\lambda_N(\tilde{\boldsymbol{R}}) = 0$. If stochastic matrix $\boldsymbol{A}$ is primitive, then $|\lambda_j(\tilde{\boldsymbol{R}})| < 1$ for all j. Also, the complex conjugate of complex number λ is denoted by $\bar{\lambda}$. For a complex vector $\boldsymbol{x} = (x_1, \cdots, x_N)^T$, its complex conjugate $\bar{\boldsymbol{x}}$ is defined by $\bar{\boldsymbol{x}} = (\bar{x}_1, \cdots, \bar{x}_N)^T$.

Some direct implications of (2.6) are given in the next theorem.

Theorem 1.
(1) $d_n(\boldsymbol{A}) = 0$ for some $n \geq 1$ if and only if $\boldsymbol{A} = \mathbf{1}\boldsymbol{e}^T$. If this is the case, $d_n(\boldsymbol{A}) = 0$ for all $n \geq 1$.
(2) If X_n is reversible in time (see below for the definition), i.e. $\tilde{\boldsymbol{A}}$ is symmetric, then $d_n(\boldsymbol{A}) = |\lambda_1(\tilde{\boldsymbol{R}})|^n$. Here $\lambda_j(\tilde{\boldsymbol{R}})$ are all real and strictly less than 1 in the magnitude.
(3) There exists a constant $0 < C \leq 1$ such that $d_n(\boldsymbol{A}) \geq C \max\{|Re(\lambda_j(\tilde{\boldsymbol{R}})^n)|\}$. As a consequence, if $\tilde{\boldsymbol{R}}$ has a real eigenvalue λ, then $d_n(\boldsymbol{A}) \geq C|\lambda|^n$.

Proof. We prove (3) only. Part (1) is trivial and (2) follows from a standard matrix diagonalization. To prove (3), let λ be such that $|Re(\lambda^n)| = \max\{|Re(\lambda_j(\tilde{\boldsymbol{R}})^n)|\}$. Suppose first that λ is complex. Then the complex conjugate $\bar{\lambda}$ is also an eigenvalue of $\tilde{\boldsymbol{R}}$ and $Re(\lambda^n) = Re(\bar{\lambda}^n)$. For the λ, let $\tilde{\boldsymbol{R}}\boldsymbol{u} = \lambda \boldsymbol{u}$ and $\boldsymbol{v}^T\tilde{\boldsymbol{R}} = \lambda \boldsymbol{v}^T$ where $\boldsymbol{u}$ and $\boldsymbol{v}$ are normalized so that $\boldsymbol{v}^T\boldsymbol{u} = 1$. Then $\tilde{\boldsymbol{R}}\bar{\boldsymbol{u}} = \bar{\lambda}\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{v}}^T\tilde{\boldsymbol{R}} = \bar{\lambda}\bar{\boldsymbol{v}}^T$ with $\bar{\boldsymbol{v}}^T\bar{\boldsymbol{u}} = 1$. Note that $\boldsymbol{v}^T\bar{\boldsymbol{u}} = 0$ since $\boldsymbol{v}^T\tilde{\boldsymbol{R}}\bar{\boldsymbol{u}} = \lambda\boldsymbol{v}^T\bar{\boldsymbol{u}} = \bar{\lambda}\boldsymbol{v}^T\bar{\boldsymbol{u}}$, but λ is complex. Define $\boldsymbol{\Delta} = \tilde{\boldsymbol{R}} - \lambda\boldsymbol{u}\boldsymbol{v}^T - \bar{\lambda}\bar{\boldsymbol{u}}\bar{\boldsymbol{v}}^T$. Then it is easy to see that $\boldsymbol{\Delta}\boldsymbol{u}\boldsymbol{v}^T = \boldsymbol{\Delta}\bar{\boldsymbol{u}}\bar{\boldsymbol{v}}^T = \boldsymbol{u}\boldsymbol{v}^T\boldsymbol{\Delta} = \bar{\boldsymbol{u}}\bar{\boldsymbol{v}}^T\boldsymbol{\Delta} = \boldsymbol{O}$. Hence, $\tilde{\boldsymbol{R}}^n = \lambda^n\boldsymbol{u}\boldsymbol{v}^T + \bar{\lambda}^n\bar{\boldsymbol{u}}\bar{\boldsymbol{v}}^T + \boldsymbol{\Delta}^n$. Choose $\boldsymbol{x} = \frac{1}{\sqrt{2}}(\boldsymbol{v} + \bar{\boldsymbol{v}})$ and $\boldsymbol{y} = \pm\frac{1}{\sqrt{2}}(\boldsymbol{u} + \bar{\boldsymbol{u}})$ where the sign is taken appropriately. If λ is real then $\tilde{\boldsymbol{R}}^n$ is decomposed as $\tilde{\boldsymbol{R}}^n = \lambda^n\boldsymbol{u}\boldsymbol{v}^T + \boldsymbol{\Delta}^n$. For this case, we choose $\boldsymbol{x} = \boldsymbol{v}$ and $\boldsymbol{y} = \pm\boldsymbol{u}$ appropriately. In the both cases, it follows that $\boldsymbol{x}^T\tilde{\boldsymbol{R}}^n\boldsymbol{y} = |Re(\lambda^n)|$. This proves Part (3). □

Remark 1. When $\tilde{\boldsymbol{R}}$ is symmetric, $|\lambda_1(\tilde{\boldsymbol{R}})|$ is readily calculated by a standard method without solving algebraic equations. Note that $\tilde{\boldsymbol{R}}^2$ is positive semi-definite so that $\lambda_j(\tilde{\boldsymbol{R}}^2) \geq 0$. The largest eigenvalue $\lambda_1(\tilde{\boldsymbol{R}}^2)$ can be efficiently found via e.g. the power method. Having $\lambda_1(\tilde{\boldsymbol{R}}^2)$, one then gets $|\lambda_1(\tilde{\boldsymbol{R}})| = \sqrt{\lambda_1(\tilde{\boldsymbol{R}}^2)}$.

To establish a desired bound of $d_n(\boldsymbol{A})$, we need two preliminaries. For any matrix $\boldsymbol{C}$, define a matrix norm by

$$||\boldsymbol{C}||_2 = \sup_{||\boldsymbol{x}||=1} ||\boldsymbol{C}\boldsymbol{x}|| = \sqrt{\sup_{||\boldsymbol{x}||=1} \boldsymbol{x}^T\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{x}}. \tag{2.7}$$

The matrix norm is deeply related to the singular value of the matrix. Since $\boldsymbol{C}\boldsymbol{C}^T$ is symmetric (in fact, positive semi-definite), the eigenvalues $\lambda_j(\boldsymbol{C}\boldsymbol{C}^T)$ are all real (non-negative). The singular values of $\boldsymbol{C}$, $\rho_j(\boldsymbol{C})$, are defined by $\rho_j(\boldsymbol{C}) = \sqrt{\lambda_j(\boldsymbol{C}\boldsymbol{C}^T)} = \sqrt{\lambda_j(\boldsymbol{C}^T\boldsymbol{C})}$. It is known that $||\boldsymbol{C}||_2 = \max\{\rho_j(\boldsymbol{C})\}$, see e.g. [4].

We next define the reversed process X_n^B of X_n. Let $\boldsymbol{A}_B = \boldsymbol{E}_D^{-1}\boldsymbol{A}^T\boldsymbol{E}_D$. It is easy to see that $\boldsymbol{A}_B$ is a stochastic matrix. The reversed process X_n^B is ergodic if and only if so is X_n. Also X_n^B has the same stationary distribution (e_k) as X_n. Let $\boldsymbol{R}_B = \boldsymbol{A}_B - \mathbf{1}\boldsymbol{e}^T$. One then has $\mathbf{1}\boldsymbol{e}^T\boldsymbol{R}_B = \boldsymbol{R}_B\mathbf{1}\boldsymbol{e}^T = \boldsymbol{O}$ so that $\boldsymbol{A}_B^n = \mathbf{1}\boldsymbol{e}^T + \boldsymbol{R}_B^n$, $n \geq 1$. It is easily seen that

$$\tilde{\boldsymbol{R}}_B = \boldsymbol{E}_D^{1/2}\boldsymbol{R}_B\boldsymbol{E}_D^{-1/2} = \tilde{\boldsymbol{R}}^T. \tag{2.8}$$

As a consequence, $d_n(\boldsymbol{A}) = d_n(\boldsymbol{A_B})$ for all $n \geq 0$. In particular, if $\boldsymbol{A_B} = \boldsymbol{A}$ or, equivalently, $\tilde{\boldsymbol{A}}$ is symmetric, then the stationary processes $\hat{X}_n$ and $\hat{X}_n^B$ are statistically undistinguishable and $\hat{X}_n$ (or X_n) is called reversible in time. The relation in (2.8) plays a key role in our analysis.

Theorem 2. Let X_n be an ergodic Markov chain governed by $\boldsymbol{A}$. Let $\tilde{\boldsymbol{R}}$ be as in (2.5) with $n = 1$. Then $\|\tilde{\boldsymbol{R}}\|_2 < 1$ and

$$d_n(\boldsymbol{A}) \leq \|\tilde{\boldsymbol{R}}^n\|_2 \leq \|\tilde{\boldsymbol{R}}\|_2^n, \quad n \geq 1.$$

Proof. To prove the first part of the theorem, consider the Markov process X_n^P governed by $\boldsymbol{A_P} = \boldsymbol{A}\boldsymbol{A_B}$. Note that the process is ergodic and has the same stationary distribution as X_n. Moreover, the associated stationary process is reversible in time since

$$\tilde{\boldsymbol{A}}_P = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{A}}_B = \sqrt{\boldsymbol{e}}\sqrt{\boldsymbol{e}}^T + \tilde{\boldsymbol{R}}\tilde{\boldsymbol{R}}^T \tag{2.9}$$

is symmetric. Hence, from Theorem 1(2) and (2.7), one has

$$\sqrt{d_n(\boldsymbol{A_P})} = \|\tilde{\boldsymbol{R}}\|_2^n, \quad n \geq 1. \tag{2.10}$$

Since the primitive matrix $\boldsymbol{A_P}$ has 1 as the Perron-Frobenius eigenvalue, $d_1(\boldsymbol{A_P}) < 1$ from (2.9) (see e.g. Seneta [5]). Thus $\|\tilde{\boldsymbol{R}}\|_2 < 1$. To derive the upper bound, let $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ be such that $d_n(\boldsymbol{A}) = \boldsymbol{x}_0^T\tilde{\boldsymbol{R}}^n\boldsymbol{y}_0$. Then

$$\begin{aligned} d_n(\boldsymbol{A}) &= \sqrt{\boldsymbol{x}_0^T(\tilde{\boldsymbol{R}}^n\boldsymbol{y}_0)(\tilde{\boldsymbol{R}}^n\boldsymbol{y}_0)^T\boldsymbol{x}_0} \\ &\leq \|\tilde{\boldsymbol{R}}^n\boldsymbol{y}_0\| \qquad \text{(by Schwart inequality)} \\ &\leq \sup_{\|\boldsymbol{x}\|=1} \|\tilde{\boldsymbol{R}}^n\boldsymbol{x}\| = \|\tilde{\boldsymbol{R}}^n\|_2. \end{aligned}$$

Since

$$\|\tilde{\boldsymbol{R}}^n\|_2 = \sup_{\boldsymbol{x}} \frac{\|\tilde{\boldsymbol{R}}\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\frac{\|\tilde{\boldsymbol{R}}^{n-1}\tilde{\boldsymbol{R}}\boldsymbol{x}\|}{\|\tilde{\boldsymbol{R}}\boldsymbol{x}\|} \leq \|\tilde{\boldsymbol{R}}\|_2\|\tilde{\boldsymbol{R}}^{n-1}\|_2,$$

one has the theorem. □

The usefullness of the upper bound in Theorem 2 is due to the geometric convergence of it to zero, by which one concludes that $d_n(\boldsymbol{A})$ converges to zero asymptotically at least at a geometric rate. As we shall see later, the matrix norm $\|\tilde{\boldsymbol{R}}\|_2$ also provides an upper bound of the relaxation time of X_n.

We next show that the upper bound in Theorem 2 is indeed attained. For this purpose, we will consider a particular case that includes the time reversible case. Suppose that $\tilde{\boldsymbol{R}}$ is normal, i.e. $\tilde{\boldsymbol{R}}\tilde{\boldsymbol{R}}^T = \tilde{\boldsymbol{R}}^T\tilde{\boldsymbol{R}}$. Then, standard spectral theory shows that

$$\tilde{\boldsymbol{R}}^n = \sum_{j=2}^{N} \lambda_j^n \boldsymbol{u}_j\bar{\boldsymbol{u}}_j^T = \sum_{j=2}^{N} \bar{\lambda}_j^n \bar{\boldsymbol{u}}_j\boldsymbol{u}_j^T, \quad n \geq 0, \tag{2.11}$$

where $\lambda_j = \lambda_j(\boldsymbol{A})$ which may be complex valued and $\boldsymbol{u}_j$ are the associated eigenvectors with $||\boldsymbol{u}_j|| = 1$. Since $d_n(\boldsymbol{A}) \geq \sup_{||\boldsymbol{x}||=1} \boldsymbol{x}^T \tilde{\boldsymbol{R}}^n \boldsymbol{x}$, one has

$$d_n(\boldsymbol{A}) \geq \sup_{||\boldsymbol{x}||=1} \sum_{j=2}^{N} \frac{1}{2}(\lambda_j^n + \bar{\lambda}_j^n)|\boldsymbol{x}^T \boldsymbol{u}_j|^2 = \max\{|Re(\lambda_j^n)|\}. \tag{2.12}$$

From (2.11), it is also easy to see that

$$\tilde{\boldsymbol{R}}^n \tilde{\boldsymbol{R}}^{n^T} = \sum_{j=2}^{N} |\lambda_j|^{2n} \boldsymbol{u}_j \bar{\boldsymbol{u}}_j^T, \quad n \geq 0.$$

It then follows that

$$d_n(\boldsymbol{A}) \leq \sqrt{d_n(\boldsymbol{A}_P)} = \sup_{||\boldsymbol{x}||=1} \sum_{j=2}^{N} |\lambda_j|^n |\boldsymbol{x}^T \boldsymbol{u}_j|^2 = \max\{|\lambda_j|^n\}. \tag{2.13}$$

However, when $\lambda_2(\boldsymbol{A})$ is real (recall that $\lambda_2(\boldsymbol{A})$ is maximum in the absolute value except $\lambda_1(\boldsymbol{A}) = 1$), $\max\{|Re(\lambda_j^n)|\} = \max\{|\lambda_j|^n\}$ so that $d_n(\boldsymbol{A}) = \sqrt{d_n(\boldsymbol{A}_P)}$. Note that the condition that $\tilde{\boldsymbol{R}}$ is normal has the following probabilistic interpretation. It is not hard to see that $\tilde{\boldsymbol{R}}$ is normal if and only if $\boldsymbol{A}\boldsymbol{A}_B = \boldsymbol{A}_B\boldsymbol{A}$. That is,

$$\frac{1}{1+n} \sum_{m=0}^{n} \boldsymbol{A}^m \boldsymbol{A}_B^{n-m} = \frac{1}{1+n} \sum_{m=0}^{n} \boldsymbol{A}_B^m \boldsymbol{A}^{n-m}. \tag{2.14}$$

for all $n \geq 1$. Equation (2.14) states that the process that proceeds m_1 steps by the original process and then $n - m_1$ steps by the reversed process is statistically undistinguishable from the process that proceeds m_2 steps first by the reversed process and then $n - m_2$ steps by the original process, where m_1 and m_2 are independently determined from a uniform distribution on $\{0, 1, \cdots, n\}$. The conditions are of course satisfied if $\tilde{\boldsymbol{R}}$ is symmetric, i.e., $\hat{X}_n$ is reversible in time.

Remark 2. The measure $d_n(\boldsymbol{A})$ in (2.1) can be defined for any positive and primitive matrix $\boldsymbol{A}$ in the following manner. Let $\boldsymbol{A}$ be such a matrix. Then there exists a unique positive eigenvalue which is larger in the absolute value than any other eigenvalues. Denote it by λ and let $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ with $||\boldsymbol{x}|| = 1$. Define $\boldsymbol{A}^* = \frac{1}{\lambda}\boldsymbol{E}_D^{-1/2}\boldsymbol{A}\boldsymbol{E}_D^{1/2}$ where $\sqrt{\boldsymbol{e}} = \boldsymbol{x}$ so that $\boldsymbol{e}^T\boldsymbol{1} = 1$. It is then evident that $\boldsymbol{A}^*$ is a stochastic matrix and the Markov process governed by $\boldsymbol{A}^*$ is ergodic. For this $\boldsymbol{A}^*$, one can define the measure $d_n(\boldsymbol{A}^*)$ in an obvious manner via (2.1). Thus the measure $d_n(\boldsymbol{A}) = \lambda^n d_n(\boldsymbol{A}^*)$ is endowed to any primitive matrix. An extension of this to an ML-matrix (see Seneta [5]) is also straightforward.

Before closing this section, we see a relation between the relaxation time $T_{REL}(X)$ and the measure $d_n(\boldsymbol{A})$ of a discrete time Markov chain. Suppose that X_n is ergodic. Then the transition probabilities $\Pr[X_n = j | X_0 = i]$ converge to stationary probabilities e_j asymptotically at geometric rates r_{ij}. That is,

$$\Pr[X_n = j | X_0 = i] - e_j \sim O(r_{ij}^n) \quad \text{as} \quad n \to \infty.$$

It is well known that $r_0 = r_{ij}$ for any i and j and $r_0 = |\lambda_2(\boldsymbol{A})|$. The relaxation time of X_n is defined as $T_{REL}(X) = (1 - r_0)^{-1}$. Consider in turn the speed of the convergence of $d_n(\boldsymbol{A})$ to zero. Since $d_n(\boldsymbol{A})$ converges to zero asymptotically at least at a geometric rate, we may let d be the convergence rate of $d_n(\boldsymbol{A})$. Of interest is the relation between r_0 and d. From (2.4), one easily sees that $d \leq r_0$ since the state space $\mathcal{N}$ is finite. On the other hand, taking $\boldsymbol{f} = \boldsymbol{u}_i$ and $\boldsymbol{g} = \pm\boldsymbol{u}_j$ appropriately where $\boldsymbol{u}_k$ is the kth unit vector, it follows that

$$d_n(\boldsymbol{A}) \geq C \sup_{i,j,\in\mathcal{N}} |\Pr[X_n = j | X_0 = i] - e_j| \tag{2.15}$$

for some C. Thus, $d \geq r_0$, from which one concludes that $d = r_0$. Therefore $d = r_0 \leq \|\tilde{\boldsymbol{R}}\|_2$ and $T_{REL}(X) \leq (1 - \|\tilde{\boldsymbol{R}}\|_2)^{-1}$, providing an upper bound of the relaxation time of X_n. Our emphasis is placed on the fact that it is in general extremely difficult to find $|\lambda_2(\boldsymbol{A})|$ numerically, while the matrix norm of a symmetric matrix is efficiently found by a standard method.

3 From Discrete to Continuous Time Chains

In this section, the results obtained so far are used to establish bounds of $d_t(X)$ in (1.2) via a uniformization (see e.g. Keilson [2], Kijima [3]). We note that the procedure below can be applied for a uniformizable semi-Markov process of Kijima [3]. But, to avoid inessential technicalities, we focus on a continuous time Markov chain.

Let X_t be a continuous time irreducible Markov chain on $\mathcal{N}$ governed by the infinitesimal generator $\boldsymbol{Q} = (q_{ij})$. Let ν be such that $\nu \geq \max\{|q_{ii}|\}$ and define $\boldsymbol{A}_\nu = \boldsymbol{I} + \frac{1}{\nu}\boldsymbol{Q}$, where $\boldsymbol{I}$ is the identity matrix. It is easy to see that $\boldsymbol{A}_\nu$ is stochastic. Moreover, by choosing ν sufficiently large, one can make $\boldsymbol{A}_\nu$ primitive. Fix ν so that $\boldsymbol{A}_\nu$ is so. Let $p_{ij}(t) = \Pr[X_t = j | X_0 = i]$ and let $\boldsymbol{P}(t) = (p_{ij}(t))$, $t \geq 0$. It is known [2], [3] that

$$\boldsymbol{P}(t) = \sum_{n=0}^{\infty} e^{-\nu t}\frac{(\nu t)^n}{n!}\boldsymbol{A}_\nu^n, \quad t \geq 0. \tag{3.1}$$

Let $e_k = \lim_{t\to\infty} p_{ik}(t)$ and $\boldsymbol{e} = (e_1, \cdots, e_N)^T$. Then, as in (2.4),

$$d_t(X) = \sup_{f,g} \frac{\boldsymbol{f}^T\boldsymbol{E}_D[\boldsymbol{P}(t) - \boldsymbol{1}\boldsymbol{e}^T]\boldsymbol{g}}{\sqrt{\boldsymbol{f}^T\boldsymbol{E}_D\boldsymbol{f} - (\boldsymbol{f}^T\boldsymbol{e})^2}\sqrt{\boldsymbol{g}^T\boldsymbol{E}_D\boldsymbol{g} - (\boldsymbol{g}^T\boldsymbol{e})^2}}. \tag{3.2}$$

Note that $\boldsymbol{e}^T\boldsymbol{Q} = (0, \cdots, 0)^T$ so that $\boldsymbol{e}^T\boldsymbol{A}_\nu = \boldsymbol{e}^T$. Thus, writing $\boldsymbol{A}_\nu = \boldsymbol{1}\boldsymbol{e}^T + \boldsymbol{R}_\nu$, one has from (3.1) that

$$\tilde{\boldsymbol{P}}(t) - \sqrt{\boldsymbol{e}}\sqrt{\boldsymbol{e}}^T = \sum_{n=0}^{\infty} e^{-\nu t}\frac{(\nu t)^n}{n!}\tilde{\boldsymbol{R}}_\nu^n, \quad t \geq 0.$$

By mimicking the arguments in Section 2, it then follows that

$$d_t(X) = \sup_{\boldsymbol{x},\boldsymbol{y}\in W} \sum_{n=0}^{\infty} e^{-\nu t}\frac{(\nu t)^n}{n!}\frac{\boldsymbol{x}^T\tilde{\boldsymbol{R}}_\nu^n\boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}. \tag{3.3}$$

Obvious implications of (3.3) are the following.

Theorem 3.

(1) $d_t(X) = 0$ for some $t > 0$ if and only if $\boldsymbol{Q} = \nu(\mathbf{1}\boldsymbol{e}^T - \boldsymbol{I})$ for some $\nu > 0$. If this is the case, $d_t(X) = 0$ for all $t > 0$.

(2) If X_t is reversible in time, then $d_t(X) = \exp\{-rt\}$ where $r = \min\{|\lambda_j(\boldsymbol{Q})|; \lambda_j(\boldsymbol{Q}) \neq 0\}$.

(3) If $\boldsymbol{Q}$ has a real eigenvalue λ apart from 0, then $d_t(X) \geq \frac{\boldsymbol{x}^T\boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}\exp\{-|\lambda|t\}$, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are right and left eigenvectors associated with λ respectively.

Proof. Statements (2) and (3) can be proved by noting the facts that $\lambda_j(\boldsymbol{Q}) = -\nu(1 - \lambda_j(\boldsymbol{A}_\nu))$, $j = 1, \cdots, N$, and $\boldsymbol{Q}$ and $\boldsymbol{A}_\nu$ share the same eigenvectors except the multiplicative factors. Part (1) is trivial. □

Let $\boldsymbol{Q}_B = \boldsymbol{E}_D^{-1}\boldsymbol{Q}^T\boldsymbol{E}_D$. Then $\boldsymbol{Q}_B$ is an infinitesimal generator governing the reversed process X_t^B of X_t. It is easy to see that X_t^B has the same stationary distribution as X_t. Recall that, for the discrete time case, the time reversible Markov chain governed by a transition probablity matrix of the form $\boldsymbol{A}\boldsymbol{A}_B$ provides an upper bound of the measure $d_n(\boldsymbol{A})$. For the continuous time case, this role is taken by the time reversible Markov chain governed by infinitesimal generator $\frac{1}{2}(\boldsymbol{Q} + \boldsymbol{Q}_B)$, as we shall see in the next theorem.

Theorem 4. Let X_t be an ergodic Markov chain governed by infinitesimal generator $\boldsymbol{Q}$ and let Y_t be a Markov chain governed by $\boldsymbol{Q}^* = \frac{1}{2}(\boldsymbol{Q} + \boldsymbol{Q}_B)$, where $\boldsymbol{Q}_B$ is defined as above. Then

$$d_t(X) \leq d_t(Y) = \exp\{\lambda_2(\boldsymbol{Q}^*)t\}, \quad t \geq 0, \tag{3.4}$$

and

$$T_{REL}(X) \leq T_{REL}(Y) = \frac{1}{|\lambda_2(\boldsymbol{Q}^*)|}, \tag{3.5}$$

where $\lambda_2(\boldsymbol{Q}^*)$ denotes the second largest negative eigenvalue of $\boldsymbol{Q}^*$.

Proof. From (3.3), one easily sees that

$$d_t(X) \leq \sum_{n=0}^{\infty} e^{-\nu t}\frac{(\nu t)^n}{n!}d_n(\boldsymbol{A}_\nu) \leq \inf_{\nu \geq \max\{|q_{ii}|\}} \exp\{-\nu(1 - \|\tilde{\boldsymbol{R}}_\nu\|_2)t\}, \tag{3.6}$$

where the second inequality follows from Theorem 2. Consider now the symmetric matrix

$$\tilde{\boldsymbol{A}}_\nu^T\tilde{\boldsymbol{A}}_\nu = \boldsymbol{I} + \frac{2}{\nu}\tilde{\boldsymbol{Q}}^* + \frac{1}{\nu^2}\tilde{\boldsymbol{Q}}^T\tilde{\boldsymbol{Q}}.$$

The definition of $\tilde{\boldsymbol{R}}_\nu$ yields that

$$\begin{aligned} -\nu(1 - \|\tilde{\boldsymbol{R}}_\nu\|_2) &= -\nu\left(1 - \sqrt{\lambda_2\left(\boldsymbol{I} + \frac{2}{\nu}\tilde{\boldsymbol{Q}}^* + \frac{1}{\nu^2}\tilde{\boldsymbol{Q}}^T\tilde{\boldsymbol{Q}}\right)}\right) \\ &= \frac{\lambda_2(2\tilde{\boldsymbol{Q}}^* + \frac{1}{\nu}\tilde{\boldsymbol{Q}}^T\tilde{\boldsymbol{Q}})}{1 + \sqrt{\lambda_2(\boldsymbol{I} + \frac{2}{\nu}\tilde{\boldsymbol{Q}}^* + \frac{1}{\nu^2}\tilde{\boldsymbol{Q}}^T\tilde{\boldsymbol{Q}})}} \end{aligned}$$

$$\leq \lambda_2(\tilde{Q}^* + \frac{1}{2\nu}\tilde{Q}^T\tilde{Q}).$$

Here $\lambda_2(C)$ is the second largest negative eigenvalue of symmetric ML-matrix C. Write $\lambda_2(\nu) = \lambda_2(2\tilde{Q}^* + \frac{1}{\nu}\tilde{Q}^T\tilde{Q})$ to derive its monotonicity property in terms of ν. Let $x(\nu)$ be the right eigenvector associated with $\lambda_2(\nu)$ and suppose $||x(\nu)|| = 1$. Then

$$\left(2\tilde{Q}^* + \frac{1}{\nu}\tilde{Q}^T\tilde{Q}\right) x(\nu) = \lambda_2(\nu)x(\nu). \tag{3.7}$$

Note that $x(\nu)$ is also the left eigenvector. By taking the derivative with respect to ν componentwise in (3.7) and then pre-multiplying $x^T(\nu)$ in the both sides, it follows that

$$\lambda_2'(\nu) = -\frac{1}{\nu^2}||\tilde{Q}x(\nu)||^2 \leq 0.$$

Thus, $\inf_{\nu\geq\max\{|q_{ii}|\}} \lambda_2(\tilde{Q}^* + \frac{1}{2\nu}\tilde{Q}^T\tilde{Q}) = \lambda_2(\tilde{Q}^*)$. This means that

$$\inf_{\nu\geq\max\{|q_{ii}|\}}\{-\nu(1 - ||\tilde{R}_\nu||_2)\} \leq \lambda_2(\tilde{Q}^*).$$

Thus, (3.4) follows from (3.6) and Theorem 3(2). To prove (3.5), we note that $d_t(X) \sim O(\exp\{-dt\})$ as $t \to \infty$. The existence of such d is guaranteed by (3.4). As for the discrete time case, it is not hard to see that $d = r$ where r is in turn the rate in (1.1). Therefore, (3.5) follows. □

If $Q = Q_B$, X_t is called reversible in time [2]. The extended notion of time reversibility described in Section 2 is also defined for the continuous time setting. Suppose $\tilde{Q}$ is normal, i.e. $QQ_B = Q_BQ$. Equivalently,

$$\frac{1}{t}\int_0^t P(t')P_B(t-t')dt' = \frac{1}{t}\int_0^t P_B(t')P(t-t')dt', \quad t > 0 \tag{3.8}$$

where $P_B(t) = (p_{ij}^B(t))$ with $p_{ij}^B(t) = \Pr[X_B(t) = j|X_B(0) = i]$. The probabilistic meaning of (3.8) is clear. Fix $\nu \geq \max\{|q_{ii}|\}$. If $\tilde{Q}$ is normal, so is $\tilde{A}_\nu = I + \frac{1}{\nu}\tilde{Q}$. Then, $\tilde{A}_\nu$ has the spectral representation as in (2.11). It follows that

$$d_t(X) \geq \sup_{||x||=1}\sum_{j=2}^{N}\exp\{-\nu(1-\lambda_j(A_\nu)t\}|x^Tu_j|^2 = \max\{\exp\{\xi_jt\}|\cos\eta_jt|;\xi_j \neq 0\} \tag{3.9}$$

where $\lambda_j(Q) = \xi_j + i\eta_j$ with $\xi_j \leq 0$. Note that the last term of (3.9) no longer depends on the choice of ν. On the other hand, when $\tilde{Q}$ is normal, it is not hard to show that $\lambda_j(Q^*) = \xi_j$, $j = 1,\cdots,N$. It follows that $d_t(X) \leq \max\{\exp\{\xi_jt\};\xi_j \neq 0\}$. Hence, if $QQ_B = Q_BQ$ and $\lambda_2(Q) \neq 0$ is real where $Re(\lambda_2(Q)) \geq Re(\lambda_j(Q))$, $d_t(X) = \exp\{\lambda_2(Q)t\}$, showing that the upper bound in (3.4) is in fact attained. When $\hat{X}_n$ is reversible in time, the conditions are trivially satisfied.

References

[1] Aldous, (1988), "Finite-time Implications of Relaxation Times for Stochastically Monotone Processes," *Probab. Th. Rel. Fields*, **77**, 137-145.

[2] Keilson, J. (1979), *Markov Chain Models – Rarity and Exponentiality*, Springer-Verlag, New York.

[3] Kijima, M. (1987), "Some Results for Uniformizable Semi-Markov Processes," *Austral. J. Statist.*, **29**, 193-207.

[4] Noble, B. and Daniel, J.W. (1977), *Applied Linear Algebra*, 2nd Ed. Prentice-Hall, Inc., New Jersey.

[5] Seneta, E. (1981), *Non-negative Matrices and Markov Chains*, 2nd Ed. Springer-Verlag, New York.

Masaaki KIJIMA
Department of Information Sciences
Tokyo Institute of Technology
Meguro-ku, Tokyo 152, Japan