

MONOTONE MEASURES OF ERGODICITY FOR MARKOV CHAINS¹

J. KEILSON

University of Rochester

William E. Simon Graduate School of Business Administration

Rochester, NY 14627 USA

O.A. VASICEK

KMV Corporation

San Francisco, CA ZIP? USA

(Received October, 1997; Revised February, 1998)

The following paper, first written in 1974, was never published other than as part of an internal research series. Its lack of publication is unrelated to the merits of the paper and the paper is of current importance by virtue of its relation to the relaxation time. A systematic discussion is provided of the approach of a finite Markov chain to ergodicity by proving the monotonicity of an important set of norms, each measures of egodicity, *whether or not time reversibility is present*. The paper is of particular interest because the discussion of the relaxation time of a finite Markov chain [2] has only been clean for time reversible chains, a small subset of the chains of interest. This restriction is not present here. Indeed, a new relaxation time quoted quantifies the relaxation time for all finite ergodic chains (cf. the discussion of $Q_1(t)$ below Equation (1.7)]. This relaxation time was developed by Keilson with A. Roy in his thesis [6], yet to be published.

Key words: Markov Chains, Ergodic, Relaxation Time.

AMS subject classifications: 60J10, 60J99.

1. Introduction

Let $N(t)$ be a finite homogeneous Markov chain in continuous time on the state space $\mathbf{N} = \{1, 2, \dots, K\}$ which is irreducible and hence ergodic.

Let $\underline{p}^T(t) = (p_n(t))$ be the state probability vector at time t with $p_n(t) = P[N(t) = n]$. Let $\underline{e}^T = (e_n)$ be the ergodic vector $\underline{e}^T = \lim_{t \rightarrow \infty} \underline{p}^T(t)$.

Consider the norm function with probability weights $q_n > 0$,

¹This paper is dedicated to R. Syski on the occasion of his special issue.

$$h_\alpha(\underline{p}, \underline{q}) = \left\{ \sum_1^K q_n \left(\frac{p_n}{q_n} \right)^\alpha \right\}^{\frac{1}{\alpha}}, \quad -\infty < \alpha < \infty \tag{1.1}$$

$$h_0(\underline{p}, \underline{q}) = \exp \left[\sum_1^K q_n \log \frac{p_n}{q_n} \right]$$

defined for arbitrary probability vectors $\underline{p}, \underline{q}$ supported on \mathbf{N} . Of interest are the related functions of time

$$R_\alpha(t) = h_\alpha(\underline{p}_t, \underline{e}) = \left\{ \sum_1^K e_n \left(\frac{p_n(t)}{e_n} \right)^\alpha \right\}^{\frac{1}{\alpha}} \tag{1.2}$$

$$R_0(t) = \exp \left[\sum_1^K e_n \log \frac{p_n}{q_n} \right]$$

and

$$Q_\alpha(t) = \left\{ \sum_1^K p_n(t) \left(\frac{p_n(t)}{e_n} \right)^\alpha \right\}^{\frac{1}{\alpha}}; Q_0(t) = \exp \left\{ \sum_1^K p_n(t) \log \left(\frac{p_n(t)}{e_n} \right) \right\}. \tag{1.3}$$

The functions $h_\alpha(\underline{w}, \underline{e})$ are vector norms on R_K . $R_\alpha(t)$ then describes a time dependent norm function. The function $Q_\alpha(t)$ has related properties.

It will be shown (cf. Theorem 1) that $Q_\alpha(t)$ is strictly decreasing in t for $\alpha > -1$, and strictly increasing in t for $\alpha < -1$ when $\underline{p}_0 \neq \underline{e}$, i.e., when the chain is not stationary. When $\alpha = -1$, $Q_{-1}(t) = 1$ for all t . A similar monotonicity of $R_\alpha(t)$ with t is demonstrated in Theorem 2. We note that $\lim_{t \rightarrow \infty} Q_\alpha(t) = 1$ for all α , $t \rightarrow \infty$. In particular,

$$Q_0(t) = \exp\{H(\underline{p}(t), \underline{e})\} \tag{1.4}$$

and

$$R_0(t) = \exp\{-H(\underline{e}, \underline{p}(t))\} \tag{1.5}$$

where [4, 5],

$$\{H(\underline{p}, \underline{q})\} = \sum_n p_n \log \left(\frac{p_n}{q_n} \right) \tag{1.6}$$

is the divergence of the distribution \underline{p} from the distribution \underline{q} , an entity closely related to entropy and other concepts in information theorem. The monotonicity of $H(\underline{e}, \underline{p}(t))$ and $H(\underline{p}(t), \underline{e})$ for arbitrary chains has been known [5]. A value of α of special interest corresponding to weighted quadratic distance is $\alpha = 1$ for which

$$Q_1(t) = \sum_n \frac{p_n^2(t)}{e_n} \tag{1.7}$$

and this is strictly decreasing in t if the chain is not stationary.

In the time reversible case, the monotonicity of (1.7) is well known [2, 3]. Indeed, for this case, the quadratic distance to ergodicity

$$D(t) = \sqrt{\sum_n \frac{[p_n(t) - e_n]^2}{e_n}} = \sqrt{Q_1(t) - 1} \tag{1.8}$$

is strictly decreasing by virtue of the symmetry of $\underline{e}_D^{1/2} [\underline{p}(t) - \underline{1} \underline{e}^T] \underline{e}_D^{-1/2}$ and its associated spectral representation [2]. That $Q_1(t)$ and $D(t)$ are monotone decreasing for all ergodic chains is striking. The function $D^2(t)$ decreases strictly to zero for every finite non-stationary homogeneous ergodic chain.

The monotonicity of the distance to ergodicity (1.8) does not appear to extend easily to the full family $R_\alpha(t)$ and $Q_\alpha(t)$.

It is shown in [6] that [cf. Remark 3.2.2], for any ergodic chain and initial distribution,

$$D(t) \leq D(0)e^{-|\theta_1|t} = \sqrt{\sum_n \frac{[p_n(0) - e_n]^2}{e_n}} e^{-|\theta_1|t} \tag{1.9}$$

where $-\theta_1$ is the smallest of the positive real singular values of $\underline{e}_D^{1/2} Q \underline{e}_D^{-1/2}$ (cf. [1]). When the chain is reversible in time, this agrees with the known result [2]. Consequently, $T_{rel} = \frac{1}{|\theta_1|}$ is a natural extension of the relaxation time for time reversible ergodic chains to all ergodic chains.

The distance $D(t)$ and the relaxation time also play a role in the covariance function $R_f(\tau) = \text{cov}[f_{J(t)}, f_{J(t+\tau)}]$ of any stationary ergodic chain $J(t)$. Here (cf. [2]),

$$R_f(\tau) = \underline{f}^T \underline{e}_D [\underline{p}(\tau) - \underline{1} \underline{e}^T] \underline{f} = \underline{f}^T \underline{e}_D [\underline{p}(\tau) - \underline{1} \underline{e}^T] \underline{e}_D^{-1/2} \underline{e}_D^{1/2} \underline{f} = \underline{g}^T(\tau) \underline{b}$$

where

$$\underline{g}^T(\tau) = \frac{1}{\underline{f}^T \underline{e}} \underline{f}^T \underline{e}_D [\underline{p}(\tau) - \underline{1} \underline{e}^T] \underline{e}_D^{-1/2}; \underline{b} = (\underline{f}^T \underline{e}) \underline{e}_D^{1/2} \underline{f}.$$

Since f may be made positive by adding a constant without altering $R_f(\tau)$, it follows that $\frac{1}{\underline{f}^T \underline{e}} \underline{f}^T \underline{e}_D [\underline{p}(\tau) - \underline{1} \underline{e}^T]$ is of the form $\underline{p}^T(t) - \underline{e}^T$ needed for Equation (1.8) and

that Equation (1.9) is then relevant. One then has from the Schwartz inequality

$$|R_f(\tau)| \leq \sqrt{\underline{g}^T(\tau) \underline{g}(\tau)} \sqrt{\underline{b}^T \underline{b}} \leq \sqrt{\underline{g}^T(0) \underline{g}(0)} \sqrt{\underline{b}^T \underline{b}} e^{-|\theta_1|t}. \tag{1.10}$$

Note that for the Frobenius norm of $\underline{r}(t) = \underline{e}_D^{1/2} [\underline{p}(t) - \underline{1} \underline{e}^T] \underline{e}_D^{-1/2}$,

$$\|r(t)\|_{FROB}^2 = \text{Trace}[\underline{r}(t) \underline{r}^T(t)] = \sum_{m,n} e_m \frac{[p_{mn}(t) - e_n]^2}{e_n}$$

and this is strictly decreasing in time. This allows from Equation (1.7) with $p_n(0) = \delta_{mn}$, weighting by e_m , and summation over m . Indeed

$$\begin{aligned} \sum_{m,n} e_m \frac{[p_{mn}(t) - e_n]^2}{e_n} &= \sum_m e_m D_m^2(t) \leq \sum_m e_m D_m^2(0) e^{-2|\theta_1|t} \\ &= e^{-2|\theta_1|t} \text{Trace} \left[\underline{e}_D [\underline{I} - \underline{1} \underline{e}^T] \underline{e}_D^{-1} [\underline{I} - \underline{e} \underline{1}^T] \right] \end{aligned}$$

and

$$\|r(t)\|_{FROB} \leq e^{-|\theta_1|t} \sqrt{N-1}.$$

When $\alpha = \pm\infty$,

$$Q_\infty(t) = \max_n \frac{p_n(t)}{e_n} \tag{1.10}$$

is a decreasing function of t ; and

$$Q_{-\infty}(t) = \min_n \frac{p_n(t)}{e_n} \tag{1.11}$$

is an increasing function of t .

2. Some Basic Lemmas

Lemma 1: For $y > 0$, let

$$g_\alpha(y) = \frac{1}{\alpha+1}y^{\alpha+1} - \frac{1}{\alpha}y^\alpha + \frac{1}{\alpha(\alpha+1)} \tag{2.1}$$

$$g_0(y) = y - 1 - \log y; \quad g_{-1}(y) = y^{-1} - 1 + \log y.$$

Then for all real α and for all $y > 0$

$$g_\alpha(y) \geq 0 \text{ with equality if and only if } y = 1. \tag{2.2}$$

Proof:

$$g_\alpha(y) = \int_1^y z^\alpha(1 - z^{-1})dz, \quad y > 1$$

$$g_\alpha(y) = \int_y^1 z^\alpha(z^{-1} - 1)dz, \quad 0 < y < 1.$$

Consequently, $g_\alpha(y) \geq 0$. Moreover, strict inequality holds for $y \neq 1$.

Lemma 2: Let $\underline{R} = (r_{mn})$ be a doubly conservative matrix, i.e.,

$$r_{mn} \geq 0, \quad m \neq n, \quad \sum_m r_{mn} = \sum_n r_{mn} = 0.$$

Then, for all real $\alpha \neq 0$ and all $x_n > 0, n = 1, 2, \dots, K$

$$\frac{1}{\alpha} \sum_m \sum_n x_m r_{mn} x_n^\alpha \leq 0, \tag{2.3}$$

and

$$\sum_m \sum_n x_m r_{mn} (1 + \log x_n) \leq 0. \tag{2.4}$$

Moreover, equality holds if and only if $x_m = x_n$ whenever $r_{mn} > 0, m, n = 1, 2, \dots, K$.

Proof: Let $\alpha \neq 0, -1$. Then from simple algebra with $y_{mn} = \frac{x_n}{x_m}$

$$\frac{1}{\alpha} \sum_m \sum_n x_m r_{mn} x_n^\alpha = - \sum_m \sum_{n \neq m} r_{mn} x_m^{\alpha+1} g_\alpha(y_{mn}). \tag{2.5}$$

By taking $\alpha \rightarrow 0$, (2.5) yields

$$\sum_m \sum_n x_m r_{mn} (1 + \log x_n) = \lim_{\alpha \rightarrow 0} \sum_m \sum_n x_m r_{mn} (1 + \frac{x_n^\alpha - 1}{\alpha})$$

$$= \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} \sum_m \sum_n x_m r_{mn} x_n^\alpha = - \sum_m \sum_{n \neq m} r_{mn} x_m g_0(y_{mn}).$$

Lemma 2 then follows from Lemma 1.

Remark 1: Note that for special value $\alpha = 1$

$$\begin{aligned} \sum_m \sum_n x_m r_{mn} x_n &\leq 0 \text{ for all real } \underline{x}, \text{ because } \sum_m \sum_n x_m r_{mn} x_n \\ &= -\frac{1}{2} \sum_m \sum_n r_{mn} (x_m - x_n)^2 + \frac{1}{2} \sum_m r_{mn} x_m^2 + \frac{1}{2} \sum_m \sum_n r_{mn} x_n^2 \end{aligned}$$

and the last two terms are zero.

3. The Main Result

Theorem 1: For a finite ergodic chain in continuous time, let

$$Q_\alpha(t) = \left\{ \sum_1^K p_n(t) \left(\frac{p_n(t)}{e_n} \right)^\alpha \right\}^{\frac{1}{\alpha}}; \quad Q_0(t) = \sum_1^K p_n(t) \log \left(\frac{p_n(t)}{e_n} \right).$$

If the chain is not stationary, $Q_\alpha(t)$ is a strictly decreasing function of t on $[0, \infty)$ when $\alpha > -1$ and a strictly increasing function of t on $[0, \infty)$ when $\alpha < -1$.

Proof: Since $p_n(t)$ is differentiable, $Q_\alpha(t)$ is also differentiable. Then for $\alpha \neq 0$

$$\frac{d}{dt} Q_\alpha(t) = \frac{\alpha + 1}{\alpha} (Q_\alpha(t))^{1-\alpha} \left\{ \sum_1^K p'_n(t) \left(\frac{p_n(t)}{e_n} \right)^\alpha \right\}. \tag{3.1}$$

Let $\underline{p}(t) = [p_{mn}(t)]$ be the transition matrix for the chain $N(t)$ so that $\underline{p}(t) = \exp[\underline{S}t]$ where $\underline{S} = [s_{mn}]$ is the infinitesimal generator of the chain. Let $r_{mn} = e_m s_{mn}$. Then $\underline{r} = [r_{mn}]$ is doubly conservative in the sense of Lemma 2. Let $x_m(t) = p_m(t)/e_m$. Then

$$\begin{aligned} \frac{1}{\alpha} \left\{ \sum_1^K p'_n(t) \left(\frac{p_n(t)}{e_n} \right)^\alpha \right\} &= \frac{1}{\alpha} \sum_n \sum_m p_m(t) s_{mn} \left(\frac{p_n(t)}{e_n} \right)^\alpha \\ &= \frac{1}{\alpha} \sum_m \sum_n x_m(t) r_{mn} x_n^\alpha(t). \end{aligned} \tag{3.2}$$

The expression $x_m(t) > 0$ for $t > 0$, a property of ergodic chains in continuous time. By Lemma 2, the expression (3.2) is always non-positive for $t > 0$. From (3.1) it follows that $\frac{d}{dt} Q_\alpha(t)$ is non-positive for $\alpha + 1 > 0$, $\alpha \neq 0$, and non-negative for $\alpha + 1 < 0$. It attains the value 0 iff $s_{mn} > 0$ implies $x_m(t) = p_m(t)/e_m = x_n(t) = p_n(t)/e_n$. For chains with positive transition rates between all pairs of states, this is possible only if $p_n(t) = e_n$ for all n , i.e., if the chain is stationary. The result then follows for all ergodic chains. When $\alpha = 0$

$$\frac{d}{dt} Q_0(t) = Q_0(t) \frac{1}{\alpha} \sum_m \sum_n x_m(t) r_{mn} (1 + \log x_n(t)) \leq 0$$

with equality iff the chain is stationary.

Similar monotonicity properties hold for the family $R_\alpha(t)$. From (1.2) and (1.3) we have

$$R_\alpha(t) = [Q_{\alpha-1}(t)]^{(\alpha-1)/\alpha}. \quad (3.3)$$

When $\alpha > 1$, or $\alpha < 0$, $R_\alpha(t)$ and $Q_{\alpha-1}(t)$ have the same monotonicity properties with t . When $0 < \alpha < 1$, $R_\alpha(t)$ increases when $Q_{\alpha-1}(t)$ decreases. When $\alpha = 0$, we have $\frac{d}{dt}R_0(t) = -R_0(t) \sum_n p'_n(t) \frac{e_n}{p_n(t)}$ which for non-stationary chains is strictly positive for any $t > 0$ by virtue of (3.2) and Lemma 1. This yields the following theorem.

Theorem 2: *Under the conditions of Theorem 1, $R_\alpha(t)$ is strictly decreasing in t for $\alpha > 1$, and strictly increasing in t for $\alpha < 1$. $R_1(t) = 1$.*

4. Ergodic Chains in Discrete Times

The results of the previous section apply also to discrete time finite Markov chains with the strict monotonicity replaced by weak monotonicity. Since the state probabilities $p_n(t)$ can be zero, we will restrict ourselves to the case $\alpha \geq 0$ only. In keeping with convention, $p \log p$ will be defined to be zero whenever $p = 0$.

Theorem 3a: *Let $\alpha \geq 0$. For a finite ergodic Markov chain, the sequence $Q_\alpha(t)$, $t = 0, 1, 2, \dots$ is a non-increasing sequence.*

Theorem 3b: *If the elements of the one step transition matrix are positive and the chain is not stationary, then $Q_\alpha(t)$, is strictly monotone in t .*

The details of the proofs are similar to that for the continuous time case and are somewhat tedious. They will not be given here.

References

- [1] Amir-Moez, A.R. and Fass, A.I., *Elements of Linear Spaces*, Edward Bros., Ann Arbor, MI 1961.
- [2] Keilson, J., *Markov Chain Models- Rarity and Exponentiality*, Springer-Verlag, Applied Mathematical Sciences Series 28 1979.
- [3] Kendall, D.G., Unitary dilations of Markov transition operators and the corresponding integral representations for transition probability matrices, (ed. by U. Grenander), *Probability and Statistics*, Almqvist and Wiksell, Stockholm 1959.
- [4] Kullback, S., *Information Theory and Statistics*, Dover Publications, Inc., New York 1968.
- [5] Renyi, A., *Foundations of Probability*, Holden-Day, Inc., San Francisco 1970.
- [6] Roy, A., *Transient Analysis of Queueing Systems*, Ph.D. Thesis, The University of Rochester 1996. This is available from UMI, Ann Arbor, MI.