Applicable Analysis and Discrete Mathematics
available online at http://pefmath.etf.bg.ac.yu

Appl. Anal. Discrete Math. **3** (2009), 52–63. doi:10.2298/AADM0901052R

TRANSIENT LIMITS

Fred Richman, Katarzyna Winkowska-Nowak

Let $\mathbf{A}$ be a Markov matrix depending on a small parameter σ, and $\mathbf{C}_n$ the average of the first n powers of $\mathbf{A}$. The stationary distributions of $\mathbf{A}$ are the rows of $\mathbf{S} = \lim_{n\to+\infty} \mathbf{C}_n$. The *limiting stationary distributions* are the rows of $\lim_{\sigma\to 0} \mathbf{S}$. We investigate *transient limits* of the sequence $\mathbf{C}_n$. These idempotent Markov matrices come up implicitly in an algorithm to compute limiting stationary distributions. They represent the intermediate-term behavior of the Markov chain at different time scales.

1. THE MOTIVATING EXAMPLE

Consider a network of m sites, each of which can be either in state -1 or 1. The state of the network is described by a vector $\mathbf{v}$ of length m whose entries are ± 1. In addition, there is an m-by-m influence matrix $\mathbf{C}$ that describes the interaction between the sites, and a noise vector $\mathbf{X}$ of length m consisting of independent random variables X_i having a common density function of the form $\dfrac{1}{\sigma} f\left(\dfrac{x}{\sigma}\right)$, where f is a symmetric nonnegative function such that $\int_r^{+\infty} f(x)\,dx > 0$ for all r, so the noise is not bounded. The positive parameter σ represents the (average) strength of the noise.

The stepping equation

$$\mathbf{v}' = \operatorname{sgn}(\mathbf{X} + \mathbf{C}\mathbf{v})$$

gives the state $\mathbf{v}'$ of the system in terms of the state $\mathbf{v}$ at the previous time. This defines a Markov chain with 2^m states and a strictly positive transition matrix $\mathbf{A}(\sigma)$. The matrices $\mathbf{A}^n(\sigma)$ converge to a matrix $\mathbf{A}^\infty(\sigma)$ each row of which is equal to the unique stationary distribution $\mathbf{p}(\sigma)$ for $\mathbf{A}(\sigma)$. In [**2**] an algorithm was given for computing the limiting stationary distribution $\lim_{\sigma\to 0} \mathbf{p}(\sigma)$, which exists for

2000 Mathematics Subject Classification. Primary 60J10, Secondary 40A05.
Keywords and Phrases. Markov chain, transient limit.

a large class of families of Markov chains, including those coming from networks with many standard noise density functions (see Section 3).

In the computation of the limiting stationary distribution in [**2**], other distributions on the states appear. These distributions, although transient, persist over long periods of time at small noise levels. They may be thought of as "transient stationary distributions" and they reflect some of the dynamics of the chain, not just its long term behavior. Although the limiting stationary distribution is independent of starting state, the transient stationary distributions need not be. These distributions are instances of the general notion of a *transient limit*.

A simple example is a network with 5 sites arranged in a line, where each site influences its neighbor and itself, that is, $c_{ij} = 1$ if $|i - j| \leq 1$, and $c_{ij} = 0$ otherwise. It turns out that the limiting stationary distribution assigns a probability of 1/2 to each of the two homogeneous states, which we can picture as

● ● ● ● ● and ○ ○ ○ ○ ○

These states require a noise of at least 2 to change. If we choose a noise level and time period so that we are likely to see many occurrences of noise 1 but no occurrences of noise 2, then the system will end up in one of the two homogeneous states with probability depending on the initial state. If we choose a noise level and time period so that we are unlikely to see any occurrences of noise 1, and we start in a stable state like

● ● ○ ○ ○ or ○ ○ ● ● ●

which require a noise of at least 1 to change, then the network will typically stay in that state during that time period. Note that the unstable state

○ ○ ● ● ○

will quickly switch to

○ ○ ● ● ●

The possibility of choosing such noise levels and time periods depends on the density function f. For the normal density $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ we can do it, but for the Cauchy density $\pi / \left(1 + x^2\right)$ we cannot.

2. TRANSIENT LIMITS

Let $A_n(\sigma)$ be a family of sequences in a metric space, (S, d), indexed by $\sigma \in \mathbf{R}^+$. A **window** is a pair of real-valued functions $1 \leq u(\sigma) \leq v(\sigma)$ such that $\lim_{\sigma \to 0} u/v = 0$. We say that L is a **transient limit** of $A_n(\sigma)$ if there is a window $u \leq v$ such that

$$\lim_{\sigma \to 0} \sup_{u \leq n \leq v} d\big(A_n(\sigma), L\big) = 0.$$

Note that we need only consider small σ in $\mathbb{R}^+$ because our interest is in what happens as σ goes to zero.

The idea here is that, for small values of σ, the sequence $A_n(\sigma)$ is very close to L for a long time, and this approximation gets better, and the time gets longer, as σ gets smaller. That's the sense in which L is a limit of the sequence. The limit is *transient* in that, for any given σ, the sequence $A_n(\sigma)$ might eventually get far away from L.

As a simple example, consider $A_n(\sigma) = (1-\sigma)^n$. The transient limits of $A_n(\sigma)$ are 0 and 1. For 0 we can take the window $u = 1/\sigma^2$ and $v = 1/\sigma^3$. Zero, of course, is the actual limit of $A_n(\sigma)$ for each $\sigma < 1$. For the transient limit 1, which is truly transient, we can take the window $u = 1$ and $v = 1/\sqrt{\sigma}$.

In the example of Section 1, $A_n(\sigma)$ is the Markov matrix $\mathbf{A}^n(\sigma)$. This sequence converges for small σ to a matrix each of whose rows is a distribution which is approximately 1/2 on each of the two states 11111 and 00000. However, for a long period of time, each row of $\mathbf{A}^n(\sigma)$ contains an entry which is approximately equal to 1 for one of the states 11111, 00000, 11000, 11100, 00111, or 00011.

Theorem 1. *Suppose the limit* $A_\infty(\sigma) = \lim_{n\to+\infty} A_n(\sigma)$ *exists for each sufficiently small* σ *in* $\mathbb{R}^+$. *If* $\lim_{\sigma\to 0} A_\infty(\sigma)$ *exists, then it is a transient limit of* $A_n(\sigma)$.

Proof. Choose $u \geq 1$ so that $\sup_{u\leq n} d\left(A_n(\sigma), \lim_{n\to+\infty} A_n(\sigma)\right) < \sigma$ and let $v = u/\sigma$. □

The limit $\lim_{\sigma\to 0} A_\infty(\sigma)$ is called the **ultimate limit** of the sequence $A_n(\sigma)$. The term "transient limit" is a bit of a misnomer in this case, but it is convenient to include the ultimate limit under this heading.

Consider the indexed Markov matrix

$$\mathbf{A}(\sigma) = \begin{pmatrix} 1-\sigma & \sigma & 0 \\ 0 & 1-\sigma^2 & \sigma^2 \\ 0 & 0 & 1 \end{pmatrix}.$$

We will show that the powers of $\mathbf{A}(\sigma)$ form a sequence with exactly three transient limits (including the ultimate limit), namely,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

To this end, we will look at a more general chain with states $1, 2, \ldots, m$, and positive transition probabilities only from i to i and from i to $i+1$, such that the probability of going from i to $i+1$ decreases rapidly with i.

Lemma 2. *Let* $\mathbf{A}$ *be an* m*-by-*m *Markov matrix,* $p, q \in [0,1]$, *and* $k \leq m$. *Suppose*

- $a_{jj} + a_{j,j+1} = 1$ *for each* $j < m$, *and* $a_{mm} = 1$,

- *$a_{k,k+1} = q$ if $k < m$,*
- *$a_{i,i+1} \geq p$ for each $i < k$.*

Then for each $\varepsilon > 0$ there exists $\delta > 0$, depending only on ε and k, such that if $np > 1/\delta$ and $nq < \delta$, then $\left|a_{ik}^{(n)} - 1\right| < \varepsilon$ for each $i \leq k$.

If, in addition, $a_{i,i+1} \leq q$ for each $i > k$, then $\left|a_{ii}^{(n)} - 1\right| < \varepsilon$ for each $i > k$.

Proof. Let X_n be the state that the MARKOV chain is in after n steps. We will show that $P_i(X_n \geq k) = \sum_{j \geq k} a_{ij}^{(n)}$ is within $\varepsilon/2$ of 1 if $np > 1/\delta$. (Note that $P_i(X_n \geq k) = 1$ for $i \geq k$.) We will also show that if $i \leq k$, then $P_i(X_n \leq k)$ is within $\varepsilon/2$ of 1 if $nq < \delta$. Hence if $i \leq k$, then $P_i(X_n = k) = a_{ik}^{(n)}$ is within ε of 1 if $np > 1/\delta$ and $nq < \delta$. We will also indicate how to choose δ.

First note that $P_i(X_n \geq k) = \sum_{j \geq k} a_{ij}^{(n)}$ is an increasing function of the probabilities $a_{t,t+1}$ for $t < k$, so by taking all those probabilities equal to p, we get a lower bound on $\sum_{j \geq k} a_{ij}^{(n)}$, namely

$$\sum_{u=k-i}^{n} \binom{n}{u} p^u (1-p)^{n-u}.$$

To show that this is within $\varepsilon/2$ of 1 if $np > 1/\delta$, it suffices to show that

$$\binom{n}{u} p^u (1-p)^{n-u} < \frac{\varepsilon}{2k}$$

for each $u < k - i$. But

$$\binom{n}{u} p^u (1-p)^{n-u} \leq \frac{(np)^u}{u!\,(1-p)^u} (1-p)^n$$

and

$$\log(np)^u (1-p)^n = u \log(np) + n \log(1-p) \leq k \log(np) - np$$

and this will be less than $\log \varepsilon - \log(2k)$ if $np > 1/\delta$ for a suitable choice of δ, depending only on ε and k.

On the other hand, if $i \leq k$, then $P_i(X_n \leq k) \geq P_k(X_n = k) = (1-q)^n$. But

$$(1-q)^n = (1-q)^{(1/q)nq}$$

and we can make $(1-q)^{1/q}$ arbitrarily close to $1/\varepsilon$ by requiring nq, hence q, to be sufficiently small. So $(1-q)^n$ can be made arbitrarily close to $(1/e)^0 = 1$ by requiring nq to be sufficiently small. Pick δ to meet this condition as well as the condition of the preceding paragraph.

Finally, if $a_{i,i+1} \leq q$ for $i > k$, then $(1-q)^n$ is also a lower bound for $a_{ii}^{(n)}$ for any $i > k$. □

Theorem 3. *Consider the indexed m-by-m Markov matrix, whose entries depend on σ,*

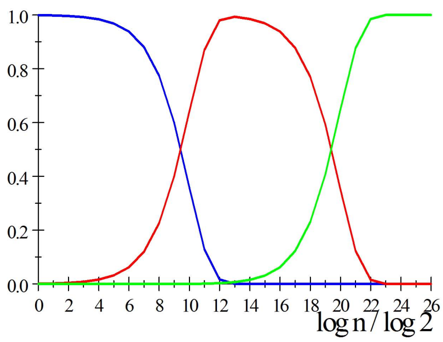
$$\mathbf{A}(\sigma) = \begin{pmatrix} 1-s_1 & s_1 & 0 & \cdots & 0 & 0 \\ 0 & 1-s_2 & s_2 & \cdots & 0 & 0 \\ 0 & 0 & 1-s_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-s_{m-1} & s_{m-1} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

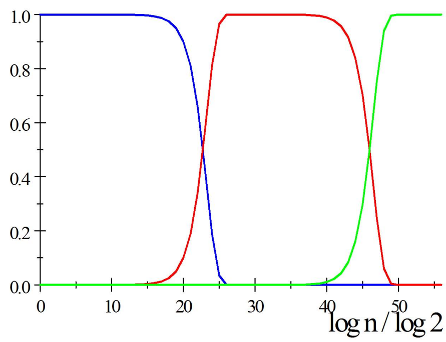
where $s_0 = 1$, $s_k > 0$, and $s_k/s_{k-1} \to 0$, as $\sigma \to 0$, for $k = 1, \ldots, m-1$. For $k = 1, \ldots, m$, let L_k be the Markov matrix such that $(L_k)_{ik} = 1$ for $i \leq k$ and $(L_k)_{ii} = 1$ for $i > k$. Then L_k is a transient limit of $\mathbf{A}^n(\sigma)$. Indeed, if we set $\tau_k = 1/s_k$ for $k = 0, \ldots, m$, and $\tau_{m+1} = \infty$, then for each $\theta \in (0, 1/2)$ and $k = 0, \ldots, m$ we have

$$\lim_{\sigma\to 0} \sup \left\{ d\left(A_n(\sigma), L_k\right) : \left(\tau_{k-1}\right)^{1-\theta} \left(\tau_k\right)^{\theta} < n < \left(\tau_{k-1}\right)^{\theta} \left(\tau_k\right)^{1-\theta} \right\} = 0.$$

Proof. Setting $u = \left(\tau_{k-1}\right)^{1-\theta} \left(\tau_k\right)^{\theta}$ and $v = \left(\tau_{k-1}\right)^{\theta} \left(\tau_k\right)^{1-\theta}$ gives a window because $u/v = \left(s_k/s_{k-1}\right)^{1-2\theta} \to 0$. Given $\varepsilon > 0$ choose $\delta > 0$ as in Lemma 2. Set $p = s_{k-1}$ and $q = s_k$ so $u = p^{\theta-1}q^{-\theta}$ and $v = p^{-\theta}q^{\theta-1}$. If $u \leq n \leq v$, then for sufficiently small values of σ we have $nq \leq vq = (q/p)^{\theta} \leq \delta$ and $np \geq up = (p/q)^{\theta} \geq 1/\delta$. The result follows from Lemma 2. □

The displayed limit in Theorem 3 continues to hold if the τ_k are replaced by τ'_k such that τ_k/τ'_k is bounded and bounded away from zero. The τ_k are *cutoffs*, analogous to the cutoff phenomenon described in [**1**]. The parameter corresponding to σ in [**1**] is a discrete parameter m, the size of the square matrix A. In our situation we get multiple cutoffs rather than just a single cutoff.





The cutoffs for the example are $\tau_1 = 1/\sigma$ and $\tau_2 = 1/\sigma^2$. To get a picture of what these transient limits are like, we give above the graphs of $a_{11}^{(n)}, a_{12}^{(n)}, a_{13}^{(n)}$, the distribution on the states at time n if you start in state 1, versus $\log_2 n$ for $\sigma = 10^{-3}$ and $\sigma = 10^{-7}$.

For $\sigma = 10^{-3}$, the cutoffs are at $\log_2 10^3 = 9.97$ and $\log_2 10^6 = 19.93$. For $\sigma = 10^{-7}$, the cutoffs are at $\log_2 10^7 = 23.25$ and $\log_2 10^{14} = 46.51$.

3. TIME SCALES AND SUITABLE DENSITY FUNCTIONS

An **indexed probability** is a function $\delta(\sigma)$, defined on some nonempty interval $(0, r]$, that is either identically 0 or strictly positive, and has a limit as σ goes to 0. In the motivating example, the entries of the transition matrix $\mathbf{A}(\sigma)$ are indexed probabilities. Normally we are interested in functions mapping into $[0, 1]$, hence the term *probability*, but it is convenient to allow the broader definition because it is closed under addition. We will not need to distinguish indexed probabilities that agree on some nonempty interval of the form $(0, r]$ in the intersection of their domains, so we will not keep track of what those domains are, for example, when the functions are added or multiplied. An indexed probability $\delta(\sigma)$ can be thought of as a *perturbed probability*: the base probability is $\lim_{\sigma\to 0} \delta(\sigma)$ and the small quantity σ causes a perturbation.

Preorder the indexed probabilities by setting $\delta \preceq \gamma$ if $\delta = \gamma = 0$, or if $\gamma \neq 0$ and $\lim_{\sigma\to 0} \delta(\sigma)/\gamma(\sigma)$ exists. This is the **time-scale preorder**. Write $\delta \prec \gamma$ if $\gamma \neq 0$ and $\lim_{\sigma\to 0} \delta(\sigma)/\gamma(\sigma) = 0$. Note that if $\delta \preceq \gamma$, then either $\gamma \preceq \delta$ or $\delta \prec \gamma$. The preorder gives rise to an equivalence relation $\delta \approx \gamma$ defined by requiring both $\delta \preceq \gamma$ and $\gamma \preceq \delta$. An equivalence class of indexed probabilities determines a **time scale**: we think of the *reciprocal* of an indexed probability as a time scale because $1/p$ is roughly how long you have to wait to see an event that has probability p. Longer time scales thus correspond to smaller probabilities. Occasionally we are interested in the finer order given by setting $\delta \leq \gamma$ if $\delta = \gamma = 0$, or if $\gamma \neq 0$ and $\lim_{\sigma\to 0} \delta(\sigma)/\gamma(\sigma)$ exists and is at most 1. We write the corresponding equivalence relation as $\delta \equiv \gamma$ if we wish to emphasize that it is not equality of functions but rather asymptotic equality. Note that $\delta \approx \gamma$ if and only $\delta \equiv a\gamma$ for some positive number a.

An **indexed Markov chain** M is a family M_σ of Markov chains indexed by the interval $(0, 1]$. The entries in the transition matrix $\mathbf{A}$ of M are indexed probabilities. We consider indexed Markov chains for which the set Π of all products of entries of $\mathbf{A}$, together with 0 and 1, is linearly preordered under $\delta \preceq \gamma$, hence also preordered under $\delta \leq \gamma$. Call such an indexed Markov chain **suitable**. Note that if $\delta \preceq \gamma$, then $\delta + \gamma \approx \gamma$, so Π is closed under addition (up to time-scale equivalence). In computing with a suitable indexed Markov chain, we need to deal with conditional probabilities of the form δ/γ where δ, γ are in Π and $\delta \preceq \gamma$. These are also indexed probabilities and the set Π^* of such quotients is still linearly preordered. Note that $\Pi \subset \Pi^*$.

The probabilities that occur in the networks of [**2**] are products of probabilities of the form

$$[r](\sigma) = \int_{r}^{+\infty} \frac{1}{\sigma} f\Big(\frac{x}{\sigma}\Big)\, dx = \int_{r/\sigma}^{+\infty} f\,(y)\, dy,$$

where f is a fixed symmetric density function (typically everywhere positive). As we are interested only in the values of $[r]$ for small σ, we care only about the behavior of f near $+\infty$. Indeed, the asymptotic behavior of f determines $[r]$ up to time-scale equivalence. More precisely, if $\lim\limits_{x\to+\infty} f(x)/g(x) = a$, then $\lim\limits_{\sigma\to 0}[r]_f(\sigma)/[r]_g(\sigma) = a$. To see this, choose σ so that $|f(x)/g(x) - a| \le \varepsilon$ for $x \ge r/\sigma$. Then

$$\left|\frac{[r]_f(\sigma)}{[r]_g(\sigma)} - a\right| = \left|\frac{\int_{r/\sigma}^{+\infty} \left(f\,(y) - ag(y)\right)\, dy}{\int_{r/\sigma}^{+\infty} g\,(y)\ dy}\right| = \left|\frac{\int_{r/\sigma}^{+\infty} g(y)\Big(\frac{f(y)}{g(y)} - a\Big)dy}{\int_{r/\sigma}^{+\infty} g\,(y)\ dy}\right| \le \varepsilon.$$

So if $\lim\limits_{x\to+\infty} f(x)/g(x) = a$, then f and g give rise to the same *bracket space*, that is, the space of products of indexed probabilities of the form $[r]$, up to time scale equivalence. Moreover, scaling does not affect the bracket space: if $\lim\limits_{x\to+\infty} g(x) = f(\lambda x)$, then

$$[r]_g \equiv \frac{1}{\lambda}\,[\lambda r]_f$$

so g and f give rise to the same bracket space.

Most common density functions f give rise to a bracket space in which all elements are comparable, hence to a suitable Markov chain. For example, any integrable rational function with positive coefficients, like the Cauchy density function, any exponential power density function, like the Laplace and normal density functions, the logistic density function, and so on. Basically, this is because density functions that are asymptotic to x^{-p} for $p > 1$, or to $e^{x^{-b}}$ for $b \ge 1$, give rise to bracket spaces in which all elements are comparable. Notice if we combine all the brackets from these two given families of density functions, then *all* the elements in the resulting space are comparable.

In [**2**] it is shown that any suitable Markov chain has a limiting stationary distribution. To see that some such suitability condition is needed, consider the transition matrix

$$A\,(\sigma) = \begin{pmatrix} 1 - s_1 & s_1 \\ s_2 & 1 - s_2 \end{pmatrix},$$

where, for σ small, $s_1 = e^{-1/\sigma}$ and $s_2 = \left(1 + \sin^2(1/\sigma)\right) e^{-1/\sigma}$. The stationary distribution of $A\,(\sigma)$ is

$$\left(\frac{s_2}{s_1 + s_2}, \frac{s_1}{s_1 + s_2}\right) = \left(\frac{1 + \sin^2(1/\sigma)}{2 + \sin^2(1/\sigma)}, \frac{1}{2 + \sin^2(1/\sigma)}\right)$$

which does not approach a limit as σ goes to 0. The problem is that s_1 and s_2 are not comparable in the time-scale order.

Only a finite number of time scales are critical for the behavior of a suitable indexed Markov chain. These are the time scales that appear in the running of the algorithm of [**2**] as the largest off-diagonal terms in the reduced matrix (after some invisible states are eliminated, and some sets of states are combined into mixed states). The windows between these critical time scales (cutoffs) give rise to distinct transient limits.

We can cook up a density function with incomparable brackets. For this purpose, it is convenient to to prove the following theorem.

Theorem 3. *Let $a_1, a_2, \ldots$ be a sequence of positive real numbers. Then there exists a continuous, integrable, monotone decreasing, positive function f such that, for each i,*

$$\int_i^{i+1} f(x)\, dx = a_i$$

if and only if

1. *$a_i \geq a_{i+1}$ for all i,* 2. *$\sum_{i=1}^{+\infty} a_i$ converges,*
3. *for no i does it hold that $a_i = a_{i+1} > a_{i+2} = a_{i+3}$.*

Proof. Let f be such a function. As f is monotone decreasing, (1) holds, and as f is integrable, (2) holds. If $a_i = a_{i+1} > a_{i+2} = a_{i+3}$, then, because f is monotone decreasing and continuous, $f(x) = a_i$ for $i \leq x \leq i+2$ and $f(x) = a_{i+2}$ for $i+2 \leq x \leq i+4$. But then f would not be continuous at a_{i+2}, so (3) must hold.

Conversely, suppose the sequence of positive numbers a_i satisfies the three conditions. Define f as follows. If $a_i = a_{i+1}$, then set $f(i) = f(i+1) = f(i+2) = a_i$. Note that we need (3) for this to be possible. For those $i > 1$ for which $f(i)$ is not yet defined, set $f(i) = (a_{i-1} + a_i)/2$. Setting $f(1) = 2a_1 - f(2)$ we have defined f on each positive integer. Note that in each case $f(i) \geq a_i \geq f(i+1)$. Now, for each $i > 1$ such that $a_{i-1} > a_i > a_{i+1}$, choose x_i between i and $i+1$ so that

$$\frac{f(i) + a_i}{2}(x_i - i) + \frac{a_i + f(i+1)}{2}(i + 1 - x_i) = a_i$$

and define $f(x_i) = a_i$. Finally, let f be the piecewise linear function determined by these function values. □

Note that we can arrange for the f of the theorem to be infinitely differentiable by replacing each linear piece by an infinitely differentiable function g which is symmetric with respect to the midpoint of the segment and all of whose derivatives are equal to zero at the endpoints. For example, we can choose g to be a function of the form

$$g(x) = c \int_a^x \exp\left(\frac{-1}{(t-a)(b-t)}\right) dt + d$$

on the interval (a, b).

To construct an example of a density function with incomparable brackets, let s_i be a sequence of positive numbers with $s_i \leq s_{i+1}$ for $i \geq 2$ and

$$\lim_{j\to+\infty} s_1 s_2 \cdots s_j = 0.$$

Set

$$a_i = s_1 s_2 \cdots s_i \left(1 - s_{i+1}\right).$$

Note that $\sum\limits_{j=i}^{+\infty} a_i$ forms a collapsing sum adding up to $s_1 s_2 \cdots s_i$. It is not difficult to show that the sequence a_i satisfies the conditions of the theorem. Consider the particular sequence s_i given by

$$1, 2^{-1}, 3^{-1/2}, 3^{-1/2}, 2^{-1/4}, 2^{-1/4}, 2^{-1/4}, 2^{-1/4},$$
$$3^{-1/8}, 3^{-1/8}, 3^{-1/8}, 3^{-1/8}, 3^{-1/8}, 3^{-1/8}, 3^{-1/8}, 3^{-1/8}, \ldots$$

Let f be the function constructed by the theorem. We want to show that the brackets [1] and [2] associated with this density function are incomparable. So consider the ratio

$$\frac{[2]\,(1/i)}{[1]\,(1/i)} = \frac{\int_{2i}^{+\infty} f\left(x\right)\,dx}{\int_{i}^{+\infty} f\left(x\right)\,dx} = s_{i+1} s_{i+2} \cdots s_{2i}$$

For $i = 1$ the ratio is $s_2 = 1/2$, for $i = 2$ it is $s_3 s_4 = 1/3$, for $i = 4$ it is $s_5 s_6 s_7 s_8 = 1/2$, for $i = 8$ it is $s_9 s_{10} \cdots s_{16} = 1/3$ and so on. So the ratio does not converge as i goes to infinity.

4. CESARO SUMS

The powers of an arbitrary MARKOV matrix $\mathbf{A}$ need not converge because $\mathbf{A}$ need not be aperiodic. However, the CESARO sums

$$C_n = \frac{1 + \mathbf{A} + \cdots + \mathbf{A}^{n-1}}{n}$$

smooth out any periodicity and always converge to a matrix whose rows are the stationary distributions of $\mathbf{A}$. The simplest example of this phenomenon occurs when

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

To see this phenomenon play out with transient limits, consider the two indexed MARKOV matrices

$$\begin{pmatrix} 1-\sigma & \sigma \\ \sigma & 1-\sigma \end{pmatrix} \text{ and } \begin{pmatrix} \sigma & 1-\sigma \\ 1-\sigma & \sigma \end{pmatrix}.$$

Each sequence of powers has the ultimate limit $\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ but the first also has a transient limit $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ in the window between 1 and $1/\sigma$. The second does not have a transient limit in that window. However, if we take the sequence of CESARO sums of the powers, then the second matrix has the ultimate limit as the transient limit in the window between 1 and $1/\sigma$. This example illustrates why you want to use CESARO sums for computing the transient limits. Indeed, the algorithm of [**2**] computes the cutoffs for the transient limits of the CESARO sums.

It is instructive to see exactly how the algorithm treats the second matrix. It looks for the largest off-diagonal element, which is $1 - \sigma \equiv 1$, then ignores everything at larger time scales and considers the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. From each ergodic component of this matrix (only one in this case), it forms a mixed state with internal probabilities the stationary distribution on that ergodic component. When the mixed states are unpacked at the end, the original states are resurrected using these probabilities. In the present case, the matrix immediately reduces to one mixed state with internal probabilities equal to that of the ultimate limit, and the time scale $1/\sigma$ never comes into play.

5. IDEMPOTENTS

We want to show that if $\mathbf{A}(\sigma)$ is an indexed family of m-by-m MARKOV matrices, then the transient limits of $\mathbf{A}^n(\sigma)$ are commuting idempotent MARKOV matrices and there are at most m of them. More generally, we want to show this for the transient limits of the sequence of CESARO sums

$$C_n(\sigma) = \frac{1 + A(\sigma) + \cdots + A^{n-1}(\sigma)}{n}.$$

That this is indeed a more general result is a consequence of the following theorem.

Theorem 5. *Let $A(\sigma)$ be an indexed family of elements of a normed ring such that the set of powers $A^n(\sigma)$ is bounded for each σ. If L is a transient limit of $A^n(\sigma)$, then L is a transient limit of the Cesaro sums*

$$C_n(\sigma) = \frac{1 + A(\sigma) + \cdots + A^{n-1}(\sigma)}{n}.$$

Proof. There is a window $u \le v$ such that

$$\lim_{\sigma \to 0} \sup_{u \le n \le v} |A^n(\sigma) - L| = 0.$$

Let b be a bound on $|A^n(\sigma) - L|$ and let

$$\varepsilon = \sup_{u \le n \le v} |A^n(\sigma) - L|.$$

Then, for $u \leq n \leq v$,

$$|C_n(\sigma) - L| \leq \frac{ub + (n-u)\varepsilon}{n} \leq \frac{u}{n}b + \varepsilon$$

so for $\sqrt{uv} \leq n \leq v$ we have $|C_n(\sigma) - L| \leq b\sqrt{u/v} + \varepsilon$ which goes to zero with σ. □

The set of powers $A^n(\sigma)$ is clearly bounded for $A(\sigma)$ a Markov matrix.

The set of Markov matrices is metrically closed, and is closed under taking powers and averages, so if $\mathbf{A}(\sigma)$ is an indexed family of Markov matrices, then any transient limit of the Cesaro sums of powers of $\mathbf{A}(\sigma)$ is a Markov matrix. The proof that it is idempotent may be carried out in the context of a normed ring.

Theorem 6. *Let L be a transient limit of the Cesaro sums*

$$C_n(\sigma) = \frac{1}{n}\sum_{i=0}^{n-1} A^i(\sigma).$$

in a normed ring. If the set $\{A^n(\sigma) : n \in \mathbb{N}, \sigma \in \mathbb{R}^+\}$ is bounded, then $L^2 = L$.

Proof. Let b be a bound on the powers $A^n(\sigma)$. We will show that we can approximate L arbitrarily closely by elements U, V and UV that are bounded by b. Then

$$L^2 - L = L(L-U) + (L-V)U - (L-UV)$$

shows that L is idempotent.

Because L is a transient limit of the matrices $C_n(\sigma)$, for each $\varepsilon > 0$ we can choose σ and n so that $C_n, \ldots, C_{3n}$ are within ε of L. Note that for all i

$$A^i C_n = \frac{n+i}{n} C_{n+i} - \frac{i}{n} C_i$$

so

$$A^i C_n - L = \frac{n+i}{n}(C_{n+i} - L) - \frac{i}{n}(C_i - L).$$

Thus, for $n \leq i < 2n$, each $A^i C_n$ is within 4ε of L. So their average

$$\frac{1}{n}\sum_{i=n}^{2n-1} A^i C_n = A^n C_n C_n$$

is within 4ε of L. Set $U = A^n C_n$ and $V = C_n$. Then U is within 4ε of L, V is within ε of L, and UV is within 4ε of L. □

Again, if $A(\sigma)$ is a Markov matrix, the boundedness condition is automatic. Finally, we show that the transient limits of the Cesaro sums commute and are comparable.

Theorem 7. *Let $A(\sigma)$ be an indexed family in a normed ring, and L and M be transient limits of the Cesaro sums $C_n(\sigma)$. If the powers of $A(\sigma)$ are uniformly* *bounded in n and σ, then L and M are commuting idempotent matrices, and either $LM = L$ or $LM = M$.*

Proof. We know that L and M are idempotent by Theorem 6. There exist windows $u_L \le v_L$ and $u_M \le v_M$ such that

$$\lim_{\sigma\to 0} \sup_{u_L\le n\le v_L} |C_n(\sigma) - L| = 0,$$
$$\lim_{\sigma\to 0} \sup_{u_M\le n\le v_M} |C_n(\sigma) - M| = 0.$$

If these two windows overlap for arbitrarily small values of σ, then $L = M$. Otherwise $v_L \le u_M + 1$, say, for arbitrarily small σ. Choose such σ so that also $v_M \ge 2u_M$ and $v_L - u_L \ge 1$. Choose $j = \lceil u_L \rceil$ and $n = \lceil u_M \rceil$. Then $u_L \le j \le v_L$ and $u_M \le n \le n + j \le v_M$ and $j/n = \eta$ can be made as small as we please.

Let b be a bound on the powers of A, hence on the Cesaro sums C_k and on L and M. We need $b\eta$ to be small. The basic identity for Cesaro sums is

$$A^i C_n = \frac{n+i}{n} C_{n+i} - \frac{i}{n} C_i$$

so, for $i \le j$,

$$\left|A^i C_n - M\right| \le \frac{n+i}{n} |C_{n+i} - M| + \frac{2bi}{n} \le (1+\eta)\,\varepsilon + 2b\eta.$$

From this we get, by averaging, that $|C_j C_n - M| \le (1+\eta)\,\varepsilon + 2b\eta$. Thus we can approximate L by C_j and M by $C_j C_n$ and C_n. □

There are at most m distinct m-by-m mutually commuting idempotent Markov matrices with the property that for any two matrices L and M, either $LM = L$ or $LM = M$. Indeed, such matrices represent projections onto a chain of subspaces of an m-dimensional vector space, and because we are dealing with Markov matrices, the projections are nonzero.

REFERENCES

1. P. Diaconis: *The cutoff phenomenon in finite Markov chains.* Proc. Natl. Acad. Sci. USA, **93** (1996), 1659–1664.
2. F. Richman, K. Winkowska-Nowak: *Computing limiting stationary distributions of small noisy networks.* Journal of Applied Probability, **39** (2002), 1–18.

Florida Atlantic University,
Boca Raton FL 33431,
USA
E–mail: richman@fau.edu

(Received February 23, 2008)
(Revised September 18, 2008)

Warsaw University,
Iss Ul. Stawki 5/7, Warsaw 02-183,
Poland
E–mail: kasia@math.fau.edu