A Marginal Ergodic Theorem

Michael Lavine

Duke University, Durham, NC, USA.

Summary. In recent years there have been several papers giving examples of Markov Chain Monte Carlo (MCMC) algorithms whose invariant measures are improper (have infinite mass) and which therefore are not positive recurrent, yet which have subchains from which valid inference can be derived. These are nonergodic (not having a limiting distribution) Markov chains (MC's) that can be written, possibly after transformation, as $Z = \{Z(n) = Z(z, n); n \geq 0\} = \{(X(n), Y(n); n \geq 0\}$ for which the subchain $X(n)$ is ergodic (has a limiting distribution). This paper gives a *marginal ergodic theorem* which *(a)* gives conditions on Z guaranteeing that the subchain X is ergodic, *(b)* gives a formula for computing the limiting distribution in case it exists, and *(c)* gives a formula for bounding the $\liminf$ and $\limsup$ as $n \to \infty$ of the distribution of $X(n)$ in case the limit does not exist.

1. Background

In recent years several papers including Hobert [2001], Liu and Wu [1999], Meng and van Dyk [1999] and van Dyk and Meng [2001] have appeared which use MCMC samplers based on improper posteriors to make valid inference about a proper target posterior. The following example is in Hobert [2001]. It or a close relative has appeared in Liu and Wu [1999], Meng and van Dyk [1999], van Dyk and Meng [2001] and Meng and van Dyk [1997].

Example 1 Let $U \sim t(\nu, \mu, 1)$, the t distribution with ν degrees of freedom and unknown location parameter μ. (ν is known.) We can express U as

$$U|\alpha, q \sim \mathrm{N}\left(\mu, \frac{\alpha}{q}\right), \quad q|\alpha \sim \mathrm{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2\alpha}\right), \quad p(\alpha) \propto \alpha^p$$

for some $p < -.5$. Then the posterior for (μ, α, q) is improper but an MCMC chain based on alternate draws from $[q|\mu, \alpha, u]$ and $[\mu, \alpha|q, u]$ yields a subchain for μ that not only has a proper limiting distribution but yields i.i.d. draws.

In this and similar examples there is a vector of parameters X in a space X having a proper "target" distribution, often but not necessarily a posterior distribution from a Bayesian analysis. The vector X is augmented with Y to form $Z = (X, Y)$, where $Y \in \mathcal{Y}$ and $Z \in \mathcal{Z}$. In Example 1 $X = \mu$ and $Y = (\alpha, q)$. A Markov chain (MC) $Z(n) = (X(n), Y(n))$ is constructed such that $\lim_{n\to\infty} X(n)$ has the desired target distribution even though $\lim_{n\to\infty} Z(n)$ is irrelevant or might not even exist. The cited papers give various sufficient conditions on $Z(n)$ to ensure that $X(n)$ has the desired limiting distribution.

We use the notation $K(z|z')$ or $K(x, y|x', y')$ to denote both transition probabilities and their densities. We assume throughout that K is irreducible and aperiodic. See Borovkov [1998], Robert and Casella [1999] or any book on Markov chains for definitions. A measure μ on $\mathcal{Z}$ is said to be an *invariant* measure for K if, for every measurable set B, $\mu(B) =$

$\int K(B|z')\mu(dz')$. Invariant σ-finite measures are unique up to multiplication by constants (Borovkov [1998], pg. 5 or Robert and Casella [1999], pg. 155).

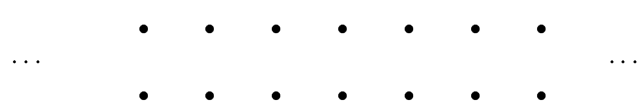
We will deal generally with the situation in which there is a proper target distribution for X, say $\pi_X(x)$. We wish to construct an MC with transition kernel $K(x, y|x', y')$ for which the subchain X converges in distribution to the target. Usually Y is introduced either to simplify or accelerate convergence; but the purpose of Y is irrelevant for this paper. We seek conditions on K such that the subchain X converges in distribution to the target. Hobert [2001] shows that if $X(n)$ is a Markov chain and, in addition, μ has a density π that can be written as $\pi(z) = \pi(x, y) = \pi_X(x)\pi_Y(y)$ and the integral

$$\int\int k(x, y|x', y')\pi_Y(y')\, dx\, dy'$$

is free of x', then the subchain X is positive recurrent with invariant density π_X. Liu and Wu [1999] and Meng and van Dyk [1999] show that if K is the limit of transition kernels K_i, where K_i has μ_i as its stationary distribution and μ_i has π_X as its X marginal, then K also has π_X as its limiting X marginal. Liu and Wu [1999] also show that if Y is a locally compact transformation group with a prior corresponding to a Haar measure then a particular MC has the desired limiting distribution. In this paper we impose less severe conditions, similar to those needed for the usual ergodic theorem.

To illustrate the delicacy of the situation, consider the following example.

Example 2 Let $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{\ldots, -1, 0, 1, \ldots\}$ and $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$. The following picture shows the points $z \in \mathcal{Z}$.



Let the measure μ assign mass 1 to every point; that is, for every $z \in \mathcal{Z}$, $\mu(z) = 1$. Consider the following two MCs.

MC 2A. be determined by *(a)* $y = y' - 1, y'$, or $y' + 1$ with probabilities .25, .5 and .25 and *(b)* $x \sim$ Bern(.5) independent of x', y' and y. MC 2a has μ as its invariant measure because for each (x, y), $\mu(x, y) = \sum_{x', y'} K(x, y|x', y')\mu(x', y')$. (There are only six nonzero terms in the summation corresponding to the six possible (x', y') pairs from which (x, y) might have come.

MC 2B. Change variables from (x, y) to (u, v) according to

$$u(x, y) = \begin{cases} 0 & \text{if } x = 0, \\ 1 & \text{if } x = 1,\ y \text{ is odd}, \\ 2 & \text{if } x = 1,\ y \text{ is even}. \end{cases}$$

$$v(x, y) = \begin{cases} y & \text{if } x = 0, \\ \frac{y-1}{2} & \text{if } x = 1,\ y \text{ is odd}, \\ \frac{y-2}{2} & \text{if } x = 1,\ y \text{ is even}. \end{cases}$$

The following picture shows the points $z \in \mathcal{Z}$ labelled by their value of v.

...	-2	-2	-1	-1	0	0	1	1	2	2	...
	-4	-3	-2	-1	0	1	2	3	4	5	

Within any set of three points where v is constant, u assumes the values 0, 1 and 2, with $u = 0$ occurring on the bottom row and $u = 1, 2$ on the top row. Let $K_{2b}(u, v|u', v')$ be determined by *(a)* $v = v' - 1, v'$, or $v' + 1$ with probabilities .25, .5 and .25 and *(b)* $u = 0, 1$ or 2 with probabilities 1/3, 1/3 and 1/3, independent of u', v' and v. MC 2b has μ as its invariant measure because for each (u, v), $\mu(u, v) = \sum_{u',v'} K(u, v|u', v')\mu(u', v')$.

MC 2a and MC 2b both have μ as their invariant measure. However, in MC 2a the long run frequency of $x = 0$ is .5 and hence the limiting distribution of X is Bern(.5), but in MC 2b the long run frequency of $u = 0$, and hence $x = x(u, v) = 0$, is 1/3 and the limiting distribution of X is Bern(1/3).

The point of the example is that a given invariant measure, if it's improper, does not have a unique limiting distribution, even for a marginal that appears to have an embedded proper distribution. A given transition kernel may have a unique limiting distribution; but two transition kernels that share the same invariant measure may have different limiting distributions. This phenomenon occurs only for improper invariant measures. Two transition kernels that share a proper invariant measure necessarily have the same limiting distribution.

In Example 2 we can write μ as a product: $\mu(x, y) = p(y) \times p(x|y) = 2 \times .5$ where $p(x|y)$ does not depend on y. One might hope, then, that $p(x|y)$ would have a natural interpretation as the marginal distribution of x and that any MC for which μ is invariant would have Bern(.5) as its limiting distribution for $X(n)$. The example shows that this hope is unrealized. In fact, the μ-marginal measure of $\{(x, y) : x = 0\}$ is ∞.

Example 2 can be analyzed by existing theorems. In that example, $X(n)$ is a Markov chain. And, as we saw earlier, μ does factor so Hobert's theorem may apply. For MC 2a, Hobert's condition is

$$\sum_x \sum_{y'} K_{2a}(x, y|x', y') = \sum_{y'} K_{2a}(y|x', y') = 2$$

which is free of x' and therefore the limiting distribution of $X(n)$ under MC 2a is Bern(.5). On the other hand, for MC 2b Hobert's condition is

$$\sum_x \sum_{y'} K_{2b}(x, y|x', y') = \sum_{y'} K_{2b}(y|x', y')$$

which does depend on x' and therefore Hobert does not guarantee that MC 2b has the right limiting distribution. On yet another hand,

$$\sum_u \sum_{v'} K_{2b}(u, v|u', v') = \sum_{v'} K_{2b}(v|u', v')$$

is free of v' so the limiting distribution of $U(n)$ under MC 2b is Multinomial $(1/3, 1/3, 1/3)$ and therefore the limiting distribution of $X(n)$ is Bern(2/3).

Also, μ is the limit of μ_i where μ_i is defined to be μ restricted to $\{(x, y) : |y| \leq i\}$ and K_{2a} is the limit of $K_{2a,i}$ where $K_{2a,i}$ is K_{2a} restricted to $\{(x, y) : |y| \leq i\}$ and suitably modified at the boundary so that μ_i is its stationary distribution. Therefore Liu and Wu

[1999] and Meng and van Dyk [1999] guarantee that K_{2a} has a limiting X marginal of Bern(.5). On the other hand, μ is also the limit of ν_i where ν_i is simply μ restricted to $\{(x,y) : |y| \leq i \text{ if } x = 0 \text{ and } |y| \leq 2i \text{ if } x = 1\}$. But K_{2a} is not the limit of transition kernels having ν_i as their stationary distributions. However, K_{2b} is the limit of transition kernels $K_{2b,i}$ having ν_i as their stationary distributions; so Liu and Wu [1999] and Meng and van Dyk [1999] guarantee that K_{2b} has a limiting X marginal of Bern(2/3).

In Examples 1 and 2 $X(n)$ is an MC. Examples 3 and 4 show that $X(n)$ may converge even when it's not Markov and in situations not covered by existing theorems.

Example 3 Let $\mathcal{Z}$ be as in Example 2. Let transition kernel $K_3(x,y|x',y')$ be defined by *(a)* $y = y' - 1, y'$, or $y' + 1$ with probabilities .4, .2 and .4 and *(b)* $x \sim$ Bern(1/3) if y is even; $x \sim$ Bern(2/3) if y is odd. It is easily verified that $X(n)$ is not Markov under K_3 and that $X(n)$ has limiting distribution Bern(.5).

Example 3 can be handled by existing theorems after a transformation of variables. Let

$$u(x,y) = \lfloor y/2 \rfloor$$
$$v(x,y) = x$$
$$w(x,y) = \begin{cases} 0 & \text{if } y \text{ is even,} \\ 1 & \text{if } y \text{ is odd} \end{cases}$$

where $\lfloor y \rfloor$ is the *floor* of y, the largest integer less than or equal to y. The kernel K_3 is Markov in (v,w), satisfies Hobert's criterion and is the limit of kernels having proper invariant measures. Therefore any of the existing theorems show that the limiting distribution of (v,w) is the measure $\pi_{3,V,W}$ defined by

$$\pi_{3,V,W}(v,w) = \begin{cases} 1/3 & \text{if } w = 0,\ v = 1, \\ 1/3 & \text{if } w = 1,\ v = 0, \\ 2/3 & \text{if } w = 0,\ v = 0, \\ 2/3 & \text{if } w = 1,\ v = 1. \end{cases}$$

which has marginal $\pi_{3,V} =$ Bern(.5).

Example 4 Let $\mathcal{Z}$ be as in Example 3. Let transition kernel $K_4(x,y|x',y')$ be defined by *(a)* $y = y' - 1, y'$, or $y' + 1$ with probabilities .4, .2 and .4 and *(b)* $x \sim$ Bern$(.5 + 1/(|y| + 2))$.

Example 4 cannot be handled by the existing theorems. Intuition says that K_4 is arbitrarily close to K_{2a} outside of bounded sets so we expect the limiting distribution of $X(n)$ to be Bern(.5). This expectation is correct, as the Marginal Ergodic Theorem will show.

2. A Marginal Ergodic Theorem

We adopt and adapt the proof of the Ergodic Theorm in Borovkov [1998] (Thm. 2.2, pp. 17–25) which is based on a renewal argument. See also the numerous references therein. Let $Z(n)$ be a Markov Chain (MC) in $(\mathcal{Z}, \mathcal{B}_{\mathcal{Z}})$. When we want to emphasize the initial state we write $Z(n) = Z(z,n)$, the state at time n, where z is the state at time $n = 0$. $Z(n)$ is divided, possibly after transformation, into components $Z(n) = (X(n), Y(n)) = (X(z,n), Y(z,n))$.

The state space and σ-field are $(\mathcal{X} \times \mathcal{Y}, \mathcal{B}_\mathcal{X} \times \mathcal{B}_\mathcal{Y})$. We assume throughout that $Z(n)$ is aperiodic and irreducible. For n-step transition probabilities from initial state z to a set $B_\mathcal{Z} \in \mathcal{B}_\mathcal{Z}$ we write

$$P(z, n, B_\mathcal{Z}) = \Pr[Z(z,n) \in B_\mathcal{Z}]$$

The usual ergodic theorem requires the following two conditions. (See Borovkov [1998], pg. 11, 12.) Let there exist a set $V \in \mathcal{B}_\mathcal{Z}$, a probability measure φ on $(\mathcal{Z}, \mathcal{B}_\mathcal{Z})$, a number $p \in (0,1)$, and an integer $n_0 \geq 1$ such that

$$\sup_{z \in V} \mathbf{E}\tau(z) < \infty, \qquad \text{Condition I}$$

where $\tau(z) = \min\{k \geq 1 : Z(z,k) \in V\}$ is the time at which the Markov chain $Z(n)$, having started in state z, returns to the set V and

$$P(z, n_0, B_\mathcal{Z}) > p\varphi(B_\mathcal{Z}) \qquad \text{Condition II}'$$

for every $B_\mathcal{Z} \in \mathcal{B}_\mathcal{Z}$ and every $z \in V$. Condition I is *positive recurrence*; Condition II$'$ is a *mixing* condition. Conditions I and II$'$ (along with irreducibility and aperiodicity) are sufficient to imply ergodicity of $Z(n)$. They are implied by the assumption of *Harris recurrence*, the assumption usually seen in the statistics literature on MCMC.

Because we are concerned with limiting distributions on $\mathcal{X}$ and not on all of $\mathcal{Z}$, the marginal ergodic theorem requires the following modification of Condition II$'$. Let there exist a set $V \in \mathcal{B}_\mathcal{Z}$, a probability measure φ on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$, a number $p \in (0,1)$, and an integer $n_0 \geq 1$ such that

$$P(z, n_0, B_\mathcal{X} \times \mathcal{Y}) \equiv \Pr[X(z, n_0) \in B_\mathcal{X}] > p\varphi(B_\mathcal{X}) \qquad \text{Condition II}$$

for every $B_\mathcal{X} \in \mathcal{B}_\mathcal{X}$ and every $z \in V$.

Before stating the main theorem we first describe the extended Markov chain Z^* on which the renewal argument is based. For each $z \in \mathcal{Z}$, let ψ_z be any probability measure on $(\mathcal{Z}, \mathcal{B}_\mathcal{Z})$ satisfying $P(z, 1, B_\mathcal{Z}) > p\psi_z(B_\mathcal{Z})$ for all $B_\mathcal{Z} \in \mathcal{B}_\mathcal{Z}$ and having $\mathcal{X}$-marginal φ; i.e. $\psi_z(B_\mathcal{X} \times \mathcal{Y}) = \varphi(B_\mathcal{X})$. (The measure having $\mathcal{X}$-marginal φ and conditional $P(Y(n)|X(n-1))$ just as in the Z chain will do.) For each $z \in \mathcal{Z}$, define the probability measure

$$Q_z(\cdot) \equiv \frac{1}{1-p}\left(P(z,1,\cdot) - p\psi_z(\cdot)\right).$$

Define the extended chain Z^* by $Z^*(n) = (\tilde{Z}(n), W(n))$ where *(a)* $\tilde{Z}(n) \in \mathcal{Z}$ and *(b)* $W(n)$ is an i.i.d. sequence of Bern(p) random variables, where p is from Condition II. The distribution of $Z^*(n)$ is given by

$$\textit{(a)} \quad W(n) \sim \text{i.i.d. Bern}(p)$$

and

$$\textit{(b1)} \quad \tilde{Z}((z,w),1,B_\mathcal{Z}) = Z(z,1,B_\mathcal{Z})$$

if $z \notin V$ and

$$\textit{(b2a)} \quad \tilde{Z}((z,1),1,B_\mathcal{Z}) = \psi_z(B_\mathcal{Z})$$

$$\textit{(b2b)} \quad \tilde{Z}((z,0),1,B_\mathcal{Z}) = Q_z(B_\mathcal{Z})$$

if $z \in V$. That is, if $z \notin V$ then $\tilde{Z}(n)$ has the same transition probabilities as $Z(n)$, independent of $W(n)$; but if $z \notin V$ then $\tilde{Z}(n)$ has transition probabilities either ψ_z or Q_z according to whether $W(n)$ is 1 or 0, respectively. Because $W(n) \sim \text{Bern}(p)$, the marginal transition probabilities of $\tilde{Z}(n)$ are

$$p\psi_z(\cdot) + (1-p)Q_z(\cdot) = P(z, 1, \cdot),$$

the same as $Z(n)$, so the distributions of $Z(n)$ and $\tilde{Z}(n)$ coincide.

Let $V^* = \{z^* : \tilde{z} \in V, w = 1\}$. The process Z^* has the following property: $\Pr[X^*(n+1) \in B_{\mathcal{X}} | z^*(n)]$ is independent of z^* provided that $z^* \in V^*$. So the trajectory of Z^* is divided into cycles by the times of arrival in V^*. In the usual ergodic theorem, the cycles are actually i.i.d., and hence renewals, because the first step in each cycle is taken according to distribution φ which does not depend on z^*. In the marginal ergodic theorem the cycles are not i.i.d. because the first step in a cycle is taken according to distribution ψ_z which does depend on z. However, every ψ_z has the same $\mathcal{X}$-marginal, and hence so do the first steps of all the cycles.

For each $z' \in V$, put

$$\tau^*(z') = \min\{k \geq 1 : Z^*((z',1),k) \in V^*\}$$
$$a(z') = \mathbf{E}\tau^*(z') = \sum_{k=1}^{\infty} P_k(z')$$

where $P_k(z') = \Pr[\tau(z') \geq k]$. Let

$$a^\ell = \inf_{z'} a(z') \qquad a^u = \sup_{z'} a(z').$$

Theorem 1 (Marginal Ergodic Theorem) *Let Condition I and Condition II hold. In addition suppose that*

the $\tau^*(z)$ are uniformly summable Condition III

That is, for every $\epsilon > 0$ there exists N such that $\sum_N^\infty P_k(z^*) < \epsilon$ for every z.

Then the set functions π^ℓ and π^u on X defined by:

$$\pi^u(B_{\mathcal{X}}) = \frac{1}{a^\ell} \sum_{k=1}^{\infty} \sup_{z' \in V} P_k(z') P(z', k, B_{\mathcal{X}} \times \mathcal{Y})$$

and

$$\pi^\ell(B_{\mathcal{X}}) = \frac{1}{a^u} \sum_{k=1}^{\infty} \inf_{z' \in V} P_k(z') P(z', k, B_{\mathcal{X}} \times \mathcal{Y})$$

satisfy

$$\liminf_{n\to\infty} P(z, n, B_{\mathcal{X}} \times \mathcal{Y}) \geq \pi^\ell(B_{\mathcal{X}})$$

and

$$\limsup_{n\to\infty} P(z, n, B_{\mathcal{X}} \times \mathcal{Y}) \leq \pi^u(B_{\mathcal{X}})$$

Proof of Theorem 1. For convenience, assume that $n_0 = 1$ in Condition II. Suppose that $Z^*(0) = (z, 1) \in V^*$. Then at time $n = 0$ the chain Z^* is at the beginning of a cycle. Denote by $\tau_1^*, \tau_2^*, \ldots$ the lengths of successive cycles (returns to V^*). Let $T_0 = 0$ and, for $k \geq 1$, $T_k = \sum_{j=1}^{k} \tau_j^*$; T_k is the k-th time when Z^* arrives back in V^*. We call the T_k's "quasi-renewal" times.

Let $\nu(n) = \min\{k : T_k \geq n\}$ and $\gamma(n) = n - T_{\nu(n)-1}$. Then $\nu(n) - 1$ is the number of completed cycles by time $n - 1$ and $\gamma(n)$ is the elapsed time since the last completed cycle. Let

$$h(z, n) = \sum_{k=0}^{\infty} \Pr[T_k = n | Z^*(0) = (z, 1)]$$

be the local quasi-renewal function. Then $\lim_{n\to\infty} h(z, n) \in [1/a^u, 1/a^\ell]$. From the perspective of time n, there exists some $k \in \{1, \ldots, n\}$ such that $\gamma(n) = k$, $n - k$ was the time of the last quasi-renewal prior to n and $Z^*(n - k) = (z', 1)$ for some $z' \in V$. Therefore,

$$P^*((z, 1), n, B_{\mathcal{X}} \times \mathcal{Y} \times \{0, 1\}) = \sum_{k=1}^{n} h(z^*, n - k) \int_{z' \in V} P_k(z') P(z', k, B_{\mathcal{X}} \times \mathcal{Y})\, dF_\bullet(z')$$

where $F_\bullet$ is the distribution of $Z(n - k)$ given that $Z^*(0) = z^*$ and that $n - k$ is a quasi-renewal time. Let

$$R^\ell(z', n) = \frac{1}{a^\ell} - h(z', n), \qquad R^u(z', n) = \frac{1}{a^u} - h(z', n).$$

Because the $\tau^*(z)$'s are uniformly summable (Condition III), for every $\epsilon > 0$ there exists N_1 such that

$$\sup_{z', n} R^\ell(z', n) \sum_{n > N_1} \sup_{z' \in V^*} P_k(z') > -\frac{\epsilon}{2}$$

and there exists N_2 such that $n > N_2$ implies

$$N_1 \sup_{z' \in V} R^\ell(z', n) \sup_{k, z'} P_k(z') P(z', k, B_{\mathcal{X}} \times \mathcal{Y}) > -\frac{\epsilon}{2}$$

Therefore

$$\begin{aligned}
&\lim_{n\to\infty} \pi^u(B_{\mathcal{X}}) - P^*((z,1), n, B_{\mathcal{X}} \times \mathcal{Y} \times \{0,1\}) \\
&= \lim_{n\to\infty} \frac{1}{a^\ell} \sum_{k=1}^{\infty} \sup_{z' \in V} P_k(z') P(z', k, B_{\mathcal{X}} \times \mathcal{Y}) \\
&\qquad - \sum_{k=1}^{n} h(z, n-k) \int_{z'} P_k(z') P(z', k, B_{\mathcal{X}}) \, dF_\bullet(z') \\
&\geq \lim_{n\to\infty} \sum_{k=1}^{n} R^\ell(z, n-k) \sup_{z' \in V} P_k(z') P(z', k, B_{\mathcal{X}} \times \mathcal{Y}) \\
&\quad + \frac{1}{a^\ell} \sum_{k=n+1}^{\infty} \sup_{z' \in V} P_k(z') P(z', k, B_{\mathcal{X}} \times \mathcal{Y}) \\
&= \lim_{n\to\infty} \sum_{k=1}^{N_1} R^\ell(z, n-k) \sup_{z' \in V} P_k(z') P(z', k, B_{\mathcal{X}} \times \mathcal{Y}) \\
&\quad + \sum_{k=N_1+1}^{n} R^\ell(z, n-k) \sup_{z' \in V} P_k(z') P(z', k, B_{\mathcal{X}} \times \mathcal{Y}) \\
&\quad + \frac{1}{a^\ell} \sum_{k=n+1}^{n} \sup_{z' \in V} P_k(z') P(z', k, B_{\mathcal{X}} \times \mathcal{Y}) \\
&\geq -\frac{\epsilon}{2} - \frac{\epsilon}{2} + 0 = -\epsilon.
\end{aligned}$$

Therefore,

$$\limsup P(z, n, B_{\mathcal{X}} \times \mathcal{Y}) = \limsup P^*(z^*, n, B_{\mathcal{X}} \times \mathcal{Y} \times \{0,1\}) \leq \pi^u(B_{\mathcal{X}}) \tag{1}$$

and similarly

$$\liminf P(z, n, B_{\mathcal{X}} \times \mathcal{Y}) \geq \pi^\ell(B_{\mathcal{X}}). \tag{2}$$

Finally, in case $Z^*(0) \notin V^*$, by irreducibility there exists with probability 1 a finite n such that $Z^*(n) \in V^*$. Because the $\limsup$ and $\liminf$ do not depend on the first n steps, Equations 1 and 2 still hold. (See Borovkov [1998] pg. 23 for formalities.)

3. Discussion

It is worth examining in more detail the structure of π^ℓ and π^u. In the ergodic theorem there is but a single π defined as

$$\begin{aligned}
\pi(B_{\mathcal{Z}}) &= \frac{1}{a} \sum_{k=1}^{\infty} P_k P(k, B_{\mathcal{Z}}) \\
&= \frac{1}{\mathbf{E}\tau^*} \sum_{k=1}^{\infty} \Pr\left[Z^*(1) \notin V^*, \ldots, Z^*(k-1) \notin V^*, Z^*(k) \in B_{\mathcal{Z}}\right]
\end{aligned}$$

The chain $Z(\cdot)$ is divided into i.i.d. cycles. The distribution of $Z(n)$, $\mathcal{L}[Z(n)]$, is written as $\sum_k \mathcal{L}[Z(n)|\gamma(n) = n-k]\Pr[\gamma(n) = n-k]$. The quantities a, τ^*, P_k and $P(k, B_{\mathcal{Z}})$ are free of z because the first step in each cycle is made according to distribution φ which is free of z. For the marginal ergodic theorem, the first step of a cycle is made according to ψ_z which is not free of z, although its X-marginal is. Because the distribution of each cycle depends on the state z' in which the cycle starts, we end up with bounds that are the supremum and infimum with respect to the different possible values of z'.

In any given example there may be several valid choices of n_0, V, p and φ. For the ergodic theorem these choices make no difference to the value of π. For the marginal ergodic theorem it sometimes happens that π^u and π^ℓ do depend on these choices. Though the theorem holds for any valid choice, there may be particular choices that yield tight bounds while others yield loose bounds. In the examples that follow we use this idea to advantage by taking the infimum of π^u and the supremum of π^ℓ over a class of valid n_0's or V's.

Example 5 (Example 2 continued) For K_{2a} $X(n) \sim$ Bern(.5) independent of $Z(n-1)$ and therefore, as we saw earlier, the limiting distribution of $X(n)$ is Bern(.5). What does the Marginal Ergodic Theorem say? We can take $V = \mathcal{Z}$ and $n_0 = 1$ because *(a)* Condition I is satisfied trivially with $\tau(z) = 1$ for all z and *(b)* Condition II is satisfied with $n_0 = 1$, $\varphi =$ Bern(.5) and any $p \in (0, 1)$. Then

$$\begin{aligned} a^u = a^\ell &= 1/p \\ P_k(z') \equiv \Pr[\tau^* \geq k] &= (1-p)^{k-1} \quad \text{for all } z' \\ P(z', k, 1) = P(z', k, 0) &= 0.5 \quad \text{for all } z',\ k \end{aligned}$$

and thus $\pi^\ell = \pi^u \equiv \pi$ where $\pi(0) = \pi(1) = 0.5$. Because π^ℓ and π^u coincide the marginal ergodic theorem says that $X(n)$ has a limiting distribution equal to π.

Similar reasoning applies to any MC for which $X(n)$ is independent of $Z(n-1)$. For instance, Example 1 yields i.i.d. draws for μ. Therefore π^ℓ and π^u coincide and the marginal ergodic theorem says that there is a limiting distribution π for μ and tells us what it is. Of course π is the distribution of the i.i.d. draws, a fact previously noted by Hobert [2001].

Example 6 (Example 4 continued) In Example 4 we can also take $V = \mathcal{Z}$ and $n_0 = 1$. Then, for $k = 1$ and $B_{\mathcal{X}} = \{1\}$,

$$\begin{aligned} &\sup_{z' \in V} P_k(z')P(z', k, B_{\mathcal{X}}) = P_1((0,0))P((0,0), 1, 1) = .2 + (.8)(\frac{5}{6}) = \frac{13}{15} \\ &> \inf_{z' \in V} P_k(z')P(z', k, B_{\mathcal{X}}) = P_1((\infty,0))P((\infty,0), 1, 1) = 1/2. \end{aligned}$$

Similar inequalities hold so that for all $k = 1, 2, \ldots$ and $B_{\mathcal{X}} = \{0\}, \{1\}$,

$$\sup_{z' \in V} P_k(z')P(z', k, B_{\mathcal{X}}) > \inf_{z' \in V} P_k(z')P(z', k, B_{\mathcal{X}}).$$

Therefore π^ℓ and π^u differ, and it seems that the Marginal Ergodic Theorem, while true, fails to give tight bounds in a case where a limiting distribution actually exists. However, for any m, we may also choose $V = \{(x, y) : |y| > m\}$. As m increases,

$$\sup_{z' \in V} P_1(z')P(z', 1, 1) = P_1((m, 0))P((m, 0), 1, 1) \to 1/2$$
$$= \inf_{z' \in V} P_1(z')P(z', 1, 1)$$

and similarly for other values of k and $B_{\mathcal{X}}$. In fact, for any $B_{\mathcal{X}}$,

$$\lim_{m \to \infty} \pi^u(B_{\mathcal{X}}) - \pi^\ell(B_{\mathcal{X}}) = 0$$

Because the Theorem is true for any V satisfying the conditions we see that Example 4 does have a limiting distribution π, that π is determined solely by the set of z's outside any bounded set, and that $\pi = \text{Bern}(.5)$.

Example 7 (Example 3 continued) In Example 3 we can also take $V = \mathcal{Z}$ and $n_0 = 1$. Then, for $k = 1$ and $B_{\mathcal{X}} = \{1\}$, $\sup_{z' \in V} P_1(z')P(z', 1, 1)$ occurs when y' is even; the sup is equal to $.2(1/3) + .8(2/3) = .6$. The inf occurs when y' is odd and is equal to $.2(2/3) + .8(1/3) = .4$. So again π^ℓ and π^u do not coincide even though there is a limiting distribution. However, we may, if we choose, use a larger value of n_0. We would declare n to be a quasi-renewal time only if $Z^*(n) \in V^*$ and there had been an interval of at least n_0 steps since the previous quasi-renewal. Therefore, $P_k(z') = 0$ for $k < n_0$ and $P(z', k, B_{\mathcal{X}})$ matters only for values of $k \geq n_0$. Since $\lim_{n \to \infty} P((z'), n, B_{\mathcal{X}}) = .5$ regardless of z', π^ℓ and π^u converge as n_0 goes to ∞ and the marginal ergodic theorem correctly identifies the limiting distribution.

In all the examples so far, $X(n)$ has had a limiting distribution. In the next example $X(n)$ does not have a limiting distribution. Instead, the long run frequency of $X(n) = 0$ and $X(n) = 1$ varies between two bounds which are correctly identified by π^ℓ and π^u.

Example 8 Let $\mathcal{Z}$ be as in Example 2. Let transition kernel $K_8(x, y|x', y')$ be defined by *(a)* $y = y' - 1, y'$, or $y' + 1$ with probabilities .4, .2 and .4 and *(b)* $x \sim \text{Bern}(1/3)$ if $y < 0$; $x \sim \text{Bern}(2/3)$ if $y \geq 0$. In this example we can take $V = \mathcal{Z}$, $n_0 = 1$, $\varphi = \text{Bern}(.5)$ and $p = 2/3$. Then

$$P_k(z') \equiv \Pr[\tau^* \geq k] = (1 - p)^{k-1} \quad \text{for all } z'$$
$$a^\ell = a^u = 1/p = 3/2$$
$$\inf_{z'} P(k, z', 1 \times \mathcal{Y}) = \inf_{z'} P(k, z', 0 \times \mathcal{Y}) = 1/3 \quad \text{for all } k$$
$$\sup_{z'} P(k, z', 1 \times \mathcal{Y}) = \sup_{z'} P(k, z', 0 \times \mathcal{Y}) = 2/3 \quad \text{for all } k$$

so

$$\pi^\ell(0) = \pi^\ell(1) = \frac{2}{3} \sum (1/3)^{k-1} 1/3 = 1/3$$
$$\pi^u(0) = \pi^u(1) = \frac{2}{3} \sum (1/3)^{k-1} 2/3 = 2/3$$

and these quantities do not change if we take a smaller V or larger n_0.

In fact, the MC defined by K_8 alternates between spending an increasingly large amount of time in the positive Y-axis and an increasingly large amount of time in the negative Y-axis. In the first case, the sampling frequency of X approaches $\text{Bern}(2/3)$; in the second it approaches Bern(1/3). The marginal ergodic theorem correctly identifies the fact that the long run frequency of X alternates between the two values.

Finally, it should be noted that when $Y(n)$ is defined independently of $X(n)$, then $X(n)$ can be viewed as a *recursive chain* or a *Markov chain in a random environment*. Following Borovkov [1998] (pg. 166), a sequence $X(n)$ is called a recursive chain governed by the sequence $Y(n)$ if for all $n \geq 0$ and $B \in \mathcal{B}_{\mathcal{X}}$,

$$\Pr[X(n+1) \in B|\mathcal{F}_n] = \Pr[X(n+1) \in B|X(n), Y(n)] \quad \text{a.s.}$$

where $\mathcal{F}_n$ is the σ-field generated by $X(0), \ldots, X(n), Y(0), \ldots, Y(n)$. There are ergodic theorems for recursive chains. See, for example, Theorem 14.1 (pg. 178) and other theorems in Borovkov's Chapter 14. They all require some version of stationarity on the sequence $Y(n)$ and hence are less general, for our purposes, than the Marginal Ergodic Theorem.

References

A. A. Borovkov. *Ergodicity and Stability of Stochastic Processes.* John Wiley, Chichester, England, 1998.

James P. Hobert. Stability relationships among the Gibbs sampler and its subchains. *Journal of Computational and Graphical Statistics*, page to appear, 2001.

Jun S. Liu and Ying Nian Wu. Parameter expansion scheme for data augmentation. *Journal of the American Statistical Association*, 94:1264–1274, 1999.

Xiao-Li Meng and David van Dyk. The EM algorithm — an old folk song sung to a fast new tune (with Discussion. *Journal of the Royal Statistical Society (Series B)*, 60:511–567, 1997.

Xiao-Li Meng and David van Dyk. Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika*, 86:301–320, 1999.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods.* Springer-Verlag, New York, 1999.

David A. van Dyk and Xiao-Li Meng. The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10:1–50, 2001.