Markov chain Monte Carlo estimation of the law of the mean of a Dirichlet process

ALESSANDRA GUGLIELMI[1] and RICHARD L. TWEEDIE[2†]

[1]*CNR-IAMI, via Ampère 56, 20131 Milano, Italy. E-mail: alessan@iami.mi.cnr.it*
[2]*Division of Biostatistics, A460 Mayo Building, 420 Delaware Street SE, University of Minnesota, Minneapolis MN 55455-0378, USA.*

The distribution $\mathcal{M}_\alpha$ of the mean Γ_α of a Dirichlet process on the real line, with parameter α, can be characterized as the invariant distribution of a real Markov chain Γ_n. In this paper we prove that, if α has finite expectation, the rate of convergence (in total variation) of Γ_n to Γ_α is geometric. Upper bounds on the rate of convergence are found which seem effective, especially in the case where α has a support which is not doubly infinite. We use this to study an approximation procedure for $\mathcal{M}_\alpha$, and evaluate the approximation error in simulating $\mathcal{M}_\alpha$ using this chain. We include examples for a comparison with some of the existing procedures for approximating $\mathcal{M}_\alpha$, and show that the Markov chain approximation compares well with other methods.

Keywords: Dirichlet process; Markov chain Monte Carlo; Markov chains with general state space; mean functional; rate of convergence

1. Introduction

In this paper we investigate the rate of convergence (in total variation) of a particular Markov chain Γ_n to the mean functional Γ_α of a Dirichlet process P_α on $\mathbb{R}$ with parameter α, and then show that by drawing from the distributions of Γ_n for large n, we can obtain an effective method of simulating from the distribution $\mathcal{M}_\alpha$ of Γ_α itself.

To define this process, we let α be a finite measure on $\mathbb{R}$ (endowed with the Borel σ-field $\mathscr{B}(\mathbb{R})$), and write $a := \alpha(\mathbb{R})$ for the total mass of α. We write $\alpha_0(\cdot) = \alpha(\cdot)/a$ for the probability measure found by normalizing α. The Dirichlet process on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with parameter α is then (Ferguson 1973) a random probability measure P_α such that, for any finite measurable partition $(A_1, \ldots, A_k)$ of $\mathbb{R}$, $\alpha(A_j) > 0$, $j = 1, \ldots, k$, the distribution of $(P_\alpha(A_1; \omega), \ldots, P_\alpha(A_k; \omega))$ is Dirichlet with parameter $(\alpha(A_1), \ldots, \alpha(A_k))$, $k \geq 2$.

We write $\Gamma_\alpha = \int x P_\alpha(\mathrm{d}x)$ and denote its distribution, and also its distribution function (df), by $\mathcal{M}_\alpha$. Denote the smallest interval containing the support of α by $[L, U]$, $-\infty \leq L < U \leq \infty$: that is,

$$\alpha((-\infty, L) \cup (U, \infty)) = 0, \tag{1}$$

and no smaller closed interval has this property. Of course, if $L = -\infty$, the interval is

denoted by $(-\infty, U)$, and analogously if $U = +\infty$. Although some integral transforms and the df itself have been analytically determined (see Cifarelli and Regazzini 1990; Regazzini *et al.* 2000), the distribution $\mathcal{M}_\alpha$ is not easy to evaluate numerically.

In order to simulate effectively from $\mathcal{M}_\alpha$, we will demonstrate that we can use a characterization of the law of a Dirichlet process given by Feigin and Tweedie (1989) to develop a Markov chain Monte Carlo algorithm and that, using this, we can find analytic bounds on the approximation error. We restrict our attention to Γ_α. Note that our results can be applied to more general linear functionals $\int g(x) P_\alpha(\mathrm{d}x)$ of a Dirichlet process on an arbitrary Polish space $(X, \mathscr{B}(X))$, $g: X \to \mathbb{R}$, since this random variable has the same distribution as $\Gamma \alpha_g$, where $\alpha_g(A) = \alpha(g^{-1}(A))$, A in $\mathscr{B}(\mathbb{R})$.

The Markov chain characterization is as follows. Let $(X_n, Y_n)_{n \geqslant 1}$ be a sequence of independent and identically distributed vectors, defined on some probability space $(\Omega, \mathscr{F}, P)$, such that X_n and Y_n are independent, with $X_n \sim \alpha_0$, and $Y_n \sim Beta(a, 1)$, for all $n = 1, 2 \ldots$. If δ_x denotes the measure degenerate at x, P_0 is any probability on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and

$$P_n = (1 - Y_n)\delta X_n + Y_n P_{n-1}, \qquad n \geqslant 1, \tag{2}$$

then Feigin and Tweedie (1989) show that $\{P_n, n \geqslant 0\}$ is a Markov chain whose invariant distribution is the law of a Dirichlet process on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with parameter α. Moreover, if $\Gamma_n := \int x P_n(\mathrm{d}x)$, so that

$$\Gamma_n = (1 - Y_n)X_n + Y_n\Gamma_{n-1}, \qquad n \geqslant 1, \tag{3}$$

then $\{\Gamma_n, n \geqslant 0\}$ is a Markov chain with real states, and it is shown in Theorem 2 of Feigin and Tweedie (1989) that under the condition

$$\int_{\mathbb{R}} \log(1 + |x|)\alpha_0(\mathrm{d}x) < \infty, \tag{4}$$

the chain Γ_n has invariant distribution $\mathcal{M}_\alpha$, the law of Γ_α.

One of our goals is to show that, if $P^n(x, A) = P(\Gamma_n \in A | \Gamma_0 = x)$, then, provided $\int |x|\alpha_0(\mathrm{d}x) < \infty$, we have, for some $\rho < 1$,

$$\|P^n(x, \cdot) - \mathcal{M}_\alpha\| < R(x)\rho^n, \qquad \forall x \in \mathbb{R}. \tag{5}$$

Hence $\mathcal{M}_\alpha$ can be estimated by generating from Γ_n, with n fixed, and the approximation error between the empirical df of a sample from Γ_n and the df $\mathcal{M}_\alpha$ decreases geometrically. A second goal is then to find estimates of ρ that can be used to bound that approximation error.

We wish to approximate the df $\mathcal{M}_\alpha$ of Γ_α, using the empirical distribution of a sample of size k from the nth-step distribution of the approximating Markov chain Γ_n, starting from an appropriate point x. Denote by $\mathcal{M}^{n,x}(\cdot)$ and $\mathcal{M}_k^{n,x}(\cdot)$ the df of Γ_n and the empirical df of a sample of size k from the distribution of Γ_n given $\Gamma_0 = x$, respectively. Then we have, for any fixed n,

$$\sup_{t \in \mathbb{R}}|\mathcal{M}_k^{n,x}(t) - \mathcal{M}_\alpha(t)| \leqslant \sup_{t \in \mathbb{R}}|\mathcal{M}_k^{n,x}(t) - \mathcal{M}^{n,x}(t)| + \sup_{t \in \mathbb{R}}|\mathcal{M}^{n,x}(t) - \mathcal{M}_\alpha(t)|.$$

To handle the first term, we could use (for example) the Dvoretzy–Kiefer–Wolfowitz

inequality. Indeed, if F is any df on $\mathbb{R}$, and F_k is the empirical distribution of a sample of size k from F, then this inequality states that

$$P\left(\sup_x |F_k(x) - F(x)| < \frac{s}{\sqrt{k}}\right) \geqslant 1 - 58\mathrm{e}^{-2s^2}, \qquad s > 0 \tag{6}$$

(see, for instance, Serfling 1980). Choosing, for example, $s = 2.1$, we have a probability greater than 0.99 of the first term being less than $2.1/\sqrt{k}$, which is within 0.01 for $k = 50\,000$, and within 0.03 for $k = 5000$.

The second term on the right is obviously less than $\|P^n(x, \cdot) - \mathcal{M}_\alpha\|$, which can be bounded as in (5) under appropriate circumstances: the goal of the next several sections is to develop this bound, which will then dictate the distribution from which we will draw the $50\,000$ or 5000 samples.

A number of other approximation procedures have recently been proposed in the literature. Muliere and Secchi (1996) suggest a bootstrap procedure, which they call a 'proper Bayesian bootstrap'. Guglielmi (1998) proposes a more 'classical' approximation technique: the df $\mathcal{M}_\alpha$ of Γ_α is approximated by the df of the mean of a different Dirichlet process whose parameter is supported on a finite number of points. This new measure is obtained from α by a 'truncation and discretization' procedure, and the approximation error in the supremum metric is determined in Guglielmi (1998). In this case, the approximating df is simple enough to be numerically computed, thanks to the finiteness of the support. See Regazzini *et al.* (2000, Section 6) for an updated version of this approximation technique. Finally, a method that is somewhat similar to that which we propose is given by Muliere and Tardella (1998). Rather than using the Markov chain P_n in (2), their technique is based on a different constructive definition of a Dirichlet process by Sethuraman (1994). He shows that P_α can be viewed, for all ω, as the random probability measure

$$\sum_{i=1}^{\infty} p_i(\omega)\delta_X i_{(\omega)}(\cdot), \tag{7}$$

where $p_1 = 1 - Y_1$, $p_j = (1 - Y_j)Y_{j-1} \cdots Y_1$, $j \geqslant 2$, X_j and Y_j are independent random variables, with $X_j \sim \alpha_0$ and $Y_j \sim Beta(a, 1)$, for all j. Based on this, Muliere and Tardella (1998) define a random probability measure $P_\varepsilon(\cdot, \omega)$ that is the sum of $n_\varepsilon(\omega)$ terms in the series (7) (plus a normalizing random factor), where n_ε is a random variable. They generate independent random draws from $P_\varepsilon(\cdot, \omega)$ and approximate the distribution of a functional of P_α via the empirical df obtained by taking functionals of the observed $P_\varepsilon(\cdot, \omega)$. The same idea of approximating the infinite series representing a Dirichlet process by a finite random sum appears in Florens and Rolin (1994), where the authors suggest stopping the sum at some level, fixed in advance, as well as in Gelfand and Kottas (1999). However, the use of the random stopping time n_ε introduces more variability into the algorithm. We note that Muliere and Tardella (1998) consider z-quantile functionals, as well as linear functionals of P_α.

This paper is organized as follows. In Section 2 we present some basic definitions of the theory of Markov chains with general state space and the characterization of the law of a Dirichlet process as the invariant distribution of the Markov chain P_n. Sections 3–4 contain

our main result: the rate of convergence of Γ_n and the determination of the approximation error. Numerical applications are illustrated in Section 5.

2. The approximating Markov chain

We will use results recently developed for Markov chains taking real (and indeed general state-space) values: for definitions not given below, see Nummelin (1984) and Meyn and Tweedie (1993).

We consider a Markov chain $\{\Phi_n, n \geq 0\}$ with state space $\mathscr{X}$, $\mathscr{X}$ being a Borel subset of $\mathbb{R}$, not necessarily countable, endowed with σ-algebra $\mathscr{B}(\mathbb{R}) \cap \mathscr{X}$. Such a Markov chain Φ_n, with transition probabilities $P^n(x, A) = P(\Phi_n \in A | \Phi_0 = x)$, $n \geq 0$, is called *Harris ergodic* if and only if there exists a probability measure π on $(\mathscr{X}, \mathscr{B}(\mathbb{R}) \cap \mathscr{X})$, called an *invariant* or *limiting* distribution, such that

$$\lim_{n \to \infty} \|P^n(x, \cdot) - \pi\| = 0, \qquad \forall x \in \mathscr{X};$$

here $\|P_1(\cdot) - P_2(\cdot)\| = \sup_{A \in \mathscr{B}(\mathbb{R})} |P_1(A) - P_2(A)|$ denotes the total variation distance between two probability measures.

A Harris ergodic chain Φ_n with invariant distribution π is said to be *geometrically ergodic*, or to *converge geometrically*, if there exist $0 < \rho < 1$ and a non-negative function R on $\mathscr{X}$ with $\int R(x)\pi(\mathrm{d}x) < \infty$, such that

$$\|P^n(x, \cdot) - \pi\| \leq R(x)\rho^n, \qquad \forall x \in \mathscr{X}.$$

The chain is said to be *uniformly ergodic*, or to *converge uniformly*, if

$$\sup_x \|P^n(x, \cdot) - \pi\| \to 0, \qquad n \to \infty,$$

or, equivalently, if there exist $0 < \rho < 1$ and a positive constant R independent of x, such that

$$\|P^n(x, \cdot) - \pi\| \leq R\rho^n, \qquad \forall x \in \mathscr{X}.$$

We shall see below that the Feigin–Tweedie chain is, under suitable conditions, either geometrically or even uniformly ergodic. To prove this we need the idea of *small sets* and *drift functions*.

A set C is called small if there exist $n_0 \geq 1$, $\varepsilon > 0$ and a probability measure φ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that

$$P^{n_0}(x, A) \geq \varepsilon \varphi(A), \qquad \forall A \in \mathscr{B}(\mathbb{R}), \forall x \in C.$$

The chain Φ_n satisfies the *geometric Foster–Lyapunov drift condition* if there exist a function $V : \mathbb{R} \mapsto [1, \infty)$, a small set C and constants $b < \infty$, $0 < \lambda < 1$, such that, for all x in $\mathscr{X}$,

$$PV(x) := \int P(x, \mathrm{d}y)V(y) \leq \lambda V(x) + bI_C(x). \tag{8}$$

When (8) holds then the chain is known to be geometrically ergodic (see Meyn and Tweedie, 1993, Chapter 15).

Finally, we say that a real-valued chain Φ_n is *stochastically monotone* if $x \mapsto P(\Phi_n < s|\Phi_0 = x)$ is a non-increasing function for any $n \geqslant 1$ and any $s \in \mathbb{R}$.

The key results for the chain Γ_n are the following.

Theorem 1. (*i*) *The chain* Γ_n *is a stochastically monotone Markov chain.*

(*ii*) *If* (4) *holds, then the chain is Harris ergodic with limiting distribution* $\mathscr{M}_\alpha$.

(*iii*) *If, further,*

$$\mathrm{E}|X_1| = \int_{\mathbb{R}} |x|\alpha_0(\mathrm{d}x) < \infty, \tag{9}$$

then Γ_n *is geometrically ergodic.*

(*iv*) *If the support of* α *is bounded then the chain is uniformly ergodic.*

Proof. (i) We first prove that Γ_n is stochastically monotone. We will let $P(z, A) := P(\Gamma_n \in A|\Gamma_{n-1} = z)$, $z \in \mathbb{R}$, $A \in \mathscr{B}(\mathbb{R})$ denote the transition kernel of Γ_n, and choose any $z_1 < z_2$, $s \in \mathbb{R}$. Then

$$P(z_1, (-\infty, s)) = P(\Gamma_1 < s|\Gamma_0 = z_1) = P((1 - Y_1)X_1 + Y_1 z_1 < s)$$

$$\geqslant P((1 - Y_1)X_1 + Y_1 z_2 < s) = P(z_2, (-\infty, s)).$$

(ii) Next we consider the Harris ergodicity of the chain. This was proved by Feigin and Tweedie (1989) under (4). It is not obvious from their proof, however, that the limiting measure is $\mathscr{M}_\alpha$, and we now clarify this. As in (1) of Feigin and Tweedie (1989), P_α satisfies the distributional equality

$$P_\alpha = (1 - Y)\delta_X + YP_\alpha,$$

where $X \sim \alpha_0$, $Y \sim Beta(a, 1)$. Taking expectations, we have

$$\Gamma_\alpha = (1 - Y)X + Y\Gamma_\alpha \tag{10}$$

also. From (10) it is immediate that the law of Γ_α is invariant for the chain Γ_n, and the uniqueness of this invariant measure for a Harris ergodic chain then shows that the limit is $\mathscr{M}_\alpha$ as required.

(iii) Next, we turn to the question of geometric ergodicity. Under (9), we shall show that Γ_n satisfies the Foster–Lyapunov condition for $V(x) = 1 + |x|$. Note first that

$$PV(x) = \mathrm{E}[1 + |(1 - Y_1)X_1 + Y_1 x|]$$

$$\leqslant 1 + \mathrm{E}(1 - Y_1)\mathrm{E}|X_1| + |x|\mathrm{E}Y_1$$

$$= 1 + \frac{\mathrm{E}|X_1|}{a + 1} + \frac{a}{a + 1}|x|.$$

Therefore, we must find a set C such that

$$1 + \frac{\mathrm{E}|X_1|}{a+1} + \frac{a}{a+1}|x| \leq \lambda(1 + |x|) + bI_C(x). \tag{11}$$

Now, if $C = [-K(\lambda), K(\lambda)]$, where

$$K(\lambda) := \frac{1 - \lambda + \mathrm{E}|X_1|/(a+1)}{\lambda - a/(a+1)}, \tag{12}$$

then (11) holds for all

$$\lambda \in \left(\frac{a}{a+1}, 1\right), \qquad b \geq 1 - \lambda + \frac{\mathrm{E}|X_1|}{a+1}. \tag{13}$$

To show that the chain is geometrically ergodic we need to establish that the set $C = [-K(\lambda), K(\lambda)]$ is small. This is essentially used in proving Harris ergodicity in Feigin and Tweedie (1989): they rely on the fact that Γ_n is weak Feller with Lebesgue measure as its irreducibility measure, so all compact sets are test sets for the drift conditions from Tweedie (1975).

However, to establish analytic upper bounds on the rate of convergence, we need a constructive proof of the fact that compact sets are small, and we need the actual values in the minorization. We carry out this step in the next section.

(iv) Finally, we show that the chain is essentially uniformly ergodic when the support of α is bounded. Since $K(\lambda) \to \infty$ for $\lambda \to a/(a+1)$, in this case it is possible to choose λ such that $C = [-K(\lambda), K(\lambda)] \supseteq [L, U]$, where $[L, U]$ is the support of α as in (1).

Moreover, by construction, if the chain Γ_n starts from $x \in C \supseteq [L, U]$, it can never leave C; in fact any superset of $[L, U]$ is absorbing, from (3). Since we have that such bounded intervals are small, if we restrict the chain to this absorbing set the whole space is small. We therefore have from Chapter 15 of Meyn and Tweedie (1993) that on this absorbing set the chain is uniformly ergodic.   □

Remark. To conclude this section, we observe that the choice of the function V is linked to condition (9). Indeed, if we consider $V(x) = 1 + |x|$ as test function in the Foster–Lyapunov condition, (9) must be assumed; conversely, if we assume (9), then the most 'natural' test function is exactly $1 + |x|$.

It is not known if $(\Gamma_n)_n$ is still geometrically ergodic if, instead of (9), one assumes only (4). We conjecture that in this case the chain is no longer geometrically ergodic.

To support this conjecture, we note that when (9) does not hold we can no longer use $V(x) = 1 + |x|$ as the test function in the geometric Foster–Lyapunov condition. The natural test function under (4) is $V(x) = 1 + \log(1 + |x|)$: but for this function, (8) typically does not hold. Indeed, a necessary condition for (8) to hold is

$$PV(x) \leq \lambda V(x) + B, \qquad \text{for some } 0 < \lambda < 1 \text{ and } B < \infty \tag{14}$$

(see Meyn and Tweedie, 1993, Lemma 15.2.8). But when $V(x) = 1 + \log(1 + |x|)$, (14) is equivalent to assuming that

$$\mathrm{E}\left[\log \frac{1 + |(1 - Y_1)X_1 + Y_1 x|}{(1 + |x|)^\lambda}\right]$$

is uniformly bounded in x. If, say, the support of α_0 is $\mathbb{R}^+$, $a = 1$ and $x > 0$, then

$$
\begin{aligned}
\mathrm{E}\left[\log \frac{1 + |(1 - Y_1)X_1 + Y_1 x|}{(1 + |x|)^\lambda}\right] &= \int_0^\infty \mathrm{d}A_0(x_1) \int_0^1 \log(1 + x_1 + y_1(x - x_1))\,\mathrm{d}y_1 - \lambda \log(1 + x) \\
&\geqslant \int_0^\infty \left(\frac{(1 + x)\log(1 + x)}{x} - 1\right)\mathrm{d}A_0(x_1) - \lambda \log(1 + x) \\
&= \left(\frac{1}{x} + 1 - \lambda\right)\log(1 + x) - 1 \to +\infty, \qquad \text{as } x \to +\infty,
\end{aligned}
$$

where A_0 denotes the distribution function corresponding to α_0. Since, for any given V, (8) is a sufficient but not necessary condition for geometric ergodicity, this does not show that the chain is not geometrically ergodic, but it does indicate that proving such ergodicity is not straightforward.

3. The structure of small sets

The goal in this section is to prove directly that bounded intervals are small, and to actually estimate the constants in the minorization for these sets.

We will write K for $K(\lambda)$ given by (12), and $C = [-K, K]$. Since $\Gamma_1 = (1 - Y_1)X_1 + Y_1\Gamma_0 = Y_1(\Gamma_0 - X_1) + X_1$ and Y_1 has density with respect to the Lebesgue measure on $\mathbb{R}$, then the conditional distribution of Γ_1, given $\Gamma_0 = x$, also has a density. More precisely, for any x, $A \in \mathcal{B}(\mathbb{R})$,

$$
\begin{aligned}
P(x, A) &= \alpha_0(\{x\})\delta_x(A) + (1 - \alpha_0(\{x\}))P(Y_1(x - X_1) + X_1 \in A | X_1 \neq x) \\
&\geqslant \delta P(Y_1(x - X_1) + X_1 \in A | X_1 \neq x),
\end{aligned}
$$

where

$$
0 < \delta = 1 - \sup \alpha_0(\{x\}) \leqslant 1, \tag{15}
$$

the supremum being taken on the set of all discontinuity points of A_0. Now the density of $(1 - Y_1)X_1 + Y_1 x$, given $X_1 = x_1$ and $X_1 \neq x$, is

$$
f_{Y_1}\left(\frac{z - x_1}{x - x_1}\right)\frac{1}{|x - x_1|},
$$

where f_{Y_1} denotes the density of Y_1. Thus the conditional density $p(x, z)$ for Γ_1, given $\Gamma_0 = x$ and $X_1 \neq x$, is

$$p(x, z) = \int_{\mathbb{R}-\{x\}} a\left(\frac{z-x_1}{x-x_1}\right)^{a-1} \frac{1}{|x-x_1|} I_{(0,1)}\left(\frac{z-x_1}{x-x_1}\right) \mathrm{d}A_0(x_1)$$

$$= \int_{(-\infty,x)} a\left(\frac{z-x_1}{x-x_1}\right)^{a-1} \frac{1}{x-x_1} I_{(0,1)}\left(\frac{z-x_1}{x-x_1}\right) \mathrm{d}A_0(x_1)$$

$$+ \int_{(x,\infty)} a\left(\frac{x_1-z}{x_1-x}\right)^{a-1} \frac{1}{x_1-x} I_{(0,1)}\left(\frac{x_1-z}{x_1-x}\right) \mathrm{d}A_0(x_1) \tag{16}$$

$$= \begin{cases} \int_{(-\infty,z)} a\dfrac{(z-x_1)^{a-1}}{(x-x_1)^a} \, \mathrm{d}A_0(x_1), & \text{if } z < x, \\[2em] \int_{(z,\infty)} a\dfrac{(x_1-z)^{a-1}}{(x_1-x)^a} \, \mathrm{d}A_0(x_1), & \text{if } z > x. \end{cases}$$

We now find the density of a minorizing measure by computing $\inf_{x \in C} p(x, z)$.

By the mean value theorem and Lebesgue dominated convergence theorem, if $x > z$,

$$\frac{\partial}{\partial x} p(x, z) = -a \int_{(-\infty,z)} a\frac{(z-x_1)^{a-1}}{(x-x_1)^a} \frac{1}{x-x_1} \, \mathrm{d}A_0(x_1) > -\infty$$

and, for $x < z$,

$$\frac{\partial}{\partial x} p(x, z) = a \int_{(z,\infty)} a\frac{(x_1-z)^{a-1}}{(x_1-x)^a} \frac{1}{x_1-x} \, \mathrm{d}A_0(x_1) < \infty,$$

so that $\partial p(x, z)/\partial x$ is positive for $x < z$ and negative for $x > z$. Thus

$$p(x, z) \geqslant \min(p(-K, z), \, p(K, z)) =: p_0(z).$$

Therefore, for all $x \in C$ and all A, we have

$$P(x, A) \geqslant \delta \int_A p(x, z) \, \mathrm{d}z \geqslant \delta \varepsilon_0 \int_A p_0(z) \, \mathrm{d}z / \varepsilon_0 \tag{17}$$

provided

$$\varepsilon_0 := \int \min(p(-K, z), \, p(K, z)) \, \mathrm{d}z$$

$$= \int_D p(-K, z) \, \mathrm{d}z + \int_{D^c} p(K, z) \, \mathrm{d}z > 0$$

where $D := \{z : p(-K, z) \leqslant p(K, z)\}$. (This construction is similar to that of Rosenthal 1995, Lemma 6(b).)

To show that $\varepsilon_0 > 0$, choose any $z \in (L, U)$. Since $z < U$, by definition $\alpha_0(z, \infty) > 0$, and since $z > L$, by definition $\alpha_0(-\infty, z) > 0$; and so $p(-K, z) > 0$ and $p(K, z) > 0$. Hence $p_0(z) > 0$ on (L, U), and so $\varepsilon_0 > 0$ as required.

This proves directly that the set C is small, satisfying the minorizing equation with $n = 1$ and

$$\varepsilon = \delta\varepsilon_0 = [1 - \sup \alpha_0(\{x\})]\left[\int \min(p(-K, z), \, p(K, z)) \, \mathrm{d}z\right].\tag{18}$$

4. Rates of convergence

We use the minorization above to establish rates of convergence of Γ_n. We find rates in three separate cases, when we can apply uniform ergodicity, stochastic monotonicity and general rates. In this section we specifically use $P^n(x, \cdot) = P(\Gamma_n \in \cdot | \Gamma_0 = x)$, $x \in \mathbb{R}$, to denote the transition law of Γ_n.

4.1. Uniform ergodicity for α with finite support

If the support of α_0 is bounded (so that $-\infty < L < U < \infty$), then when the chain Γ_n starts from $x \in [L, U]$, we can take the state space to be $[L, U]$ itself, which is small with ε defined as in (18) by

$$\varepsilon = [1 - \sup \alpha_0(\{x\})]\left[\int \min(p(L, z), \, p(U, z)) \, \mathrm{d}z\right].\tag{19}$$

It follows from Theorem 16.2.4 in Meyn and Tweedie (1993) that

$$\|P^n(x, \cdot) - \mathscr{M}_\alpha\| \leqslant (1 - \varepsilon)^n, \qquad x \in [L, U].\tag{20}$$

Note that there is an error in their statement, since they use a different definition of total variation distance, and hence their result should have a factor of 2 multiplying $(1 - \varepsilon)^n$.

The bound above may be poor if the set $[L, U]$ is large, leading to small ε. In this case the next case will also apply, and may give better bounds than the uniform ones above, especially for small initial values.

4.2. Geometric ergodicity: 'one-sided' α and stochastic montonicity

Suppose that the interval $[L, U]$ is finite on one side or the other: for convenience, assume that $-\infty < L$. Then the set $[L, \infty)$ is again absorbing and we can restrict attention to this set. We will use the drift condition (11) and also the stochastic monotonicity of Theorem 1(i).

We take $C = [L, K]$, where K is defined by (12) and λ, b are defined by (13), and apply results of Roberts and Tweedie (2000) to the chain Γ_n in this case. Take $d = 1 + K(\lambda)$, and ε as defined in (18). We define

$$J = d + \lambda^{-1}(b - \varepsilon), \qquad \eta = \frac{\log(J/(1 - \varepsilon))}{\log \lambda^{-1}},$$

$$\zeta(x) = V(x)\mathscr{M}_\alpha((-\infty, x]) + \int_{(x,\infty)} V(z)\mathscr{M}_\alpha(dz), \qquad \xi(x) = \frac{\log \zeta(x)}{\log \lambda^{-1}}. \qquad (21)$$

With these values we have the following result.

Theorem 2. *Suppose* $\mathrm{E}|X_1| < \infty$ *and* $\Gamma_0 = x \in [L, \infty)$.

(i) *If* $J < 1$, *then for* $n > \xi(x) + \eta(1 - \varepsilon)/(\lambda^\eta - (1 - \varepsilon))$,

$$\|P^n(x, \cdot) - \mathscr{M}_\alpha\| \le \zeta(x)\frac{\varepsilon}{1 - J}\lambda^n;$$

(ii) *If* $J \ge 1$, *and* $\rho = (1 - \varepsilon)^{\eta^{-1}}$, *for* $n > \xi(x) + \eta(1 - \varepsilon)/\varepsilon$,

$$\|P^n(x, \cdot) - \mathscr{M}_\alpha\| \le \zeta(x)^{\log \rho/\log \lambda}\frac{\mathrm{e}\varepsilon(n - \xi(x) + \eta)}{\eta}\rho^n.$$

(iii) *If* $J \ge 1$, *then for any* $1 > \omega > \rho = (1 - \varepsilon)^{\eta^{-1}}$,

$$\|P^n(x, \cdot) - \mathscr{M}_\alpha\| \le \frac{[1 - (1 - \varepsilon)/\omega]\omega^{-\xi(x)}}{1 - (1 - \varepsilon)\omega^{-\eta}}\omega^n.$$

Proof. (i) and (ii) are a direct corollary of Theorem 2.2 in Roberts and Tweedie (2000). For (iii), we can use the same method of proof as in that theorem, but use (33) of Roberts and Tweedie (1999) rather than Theorem 5.1 of Roberts and Tweedie (1999). □

A number of remarks can be made about this result. Firstly, $\zeta(x)$ is generally unknown, even if it can bounded by

$$\zeta(x) \le V(x) + \int_{\mathbb{R}} V(z)\mathscr{M}_\alpha(dz) \le V(x) + \frac{b}{1 - \lambda} < \infty,$$

as noted by Roberts and Tweedie (2000). This bound is often very rough as the examples below show: in many cases it might be better to estimate $\zeta(x)$ by

$$Z_V^{(m)} = V(x)\frac{1}{m}\sum_{i=1}^m I_{(-\infty,x]}(\Gamma_i) + \frac{1}{m}\sum_{i=1}^m I_{(x,+\infty)}(\Gamma_i)(1 + |\Gamma_i|)$$

for m big enough, using the law of large numbers for Markov chains.

Secondly, we observe that the convergence of $P^n(x, \cdot)$ to $\mathscr{M}_\alpha$ is uniform for those initial x such that $V(x) \le R$, R being any positive constant, since

(i′) if $J < 1$, then for $n > \tilde{\xi} + \eta(1 - \varepsilon)/(\lambda^\eta - (1 - \varepsilon))$, where $\tilde{\xi} = \log(R + b/(1 - \lambda))/\log \lambda^{-1}$,

$$\|P^n(x, \cdot) - \mathscr{M}_\alpha(\cdot)\| \le \left(R + \frac{b}{1 - \lambda}\right)\frac{\varepsilon}{1 - J}\lambda^n;$$

(ii′) if $J \geqslant 1$, $\rho = (1 - \varepsilon)^{\eta^{-1}}$, then, for $n > \tilde{\xi} + \eta(1 - \varepsilon)/\varepsilon$,

$$\|P^n(x, \cdot) - \mathscr{M}_\alpha(\cdot)\| \leqslant \left(R + \frac{b}{1 - \lambda}\right)^{\log \rho/\log \lambda} \frac{e\varepsilon(n + \eta)}{\eta} \rho^n; \tag{22}$$

both (i′) and (ii′) imply $\sup_{x \leqslant R'} \|P^n(x, \cdot) - \mathscr{M}_\alpha(\cdot)\| \to 0$, $n \to \infty$, for any R'. A similar result holds for the uniform version of (iii).

The result (ii′) implies that the chain Γ_n is 'almost' geometric at rate ρ, and of course is geometric at any rate slower than ρ.

Thirdly, note that the bound (ii) or (22) is very tight if ε is very small; but the values of n for which the bound holds are correspondingly large, of the order of ε^{-1}.

4.3. Geometric and other rates: α with doubly infinite support

When α has support which is unbounded on both sides, we cannot use the stochastic monotonicity results and we turn to the results of Roberts and Tweedie (1999). There are two different approaches given in that paper.

First, we can use the results of Theorems 5.1 and 5.2 of Roberts and Tweedie (1999). As they show, we can construct a bivariate test function, $h(x, y) = [V(x) + V(y)]/2$, which satisfies the drift condition (for the bivariate kernel denoted P_2 and the test set $C_2 = C \times C$)

$$P_2 h \leqslant \lambda_2 h + b_2 I_{C_2},$$

where

$$\lambda_2 = \lambda + \frac{b}{2(1 + d)}, \qquad b_2 = \frac{b(1 + 2d)}{2 + 2d}. \tag{23}$$

For $\lambda_2 < 1$, we require $d = 1 + K(\lambda) > b/(2(1 - \lambda)) - 1$. The definition of $K(\lambda)$ shows that if we choose λ close enough to the minimal value $a/(a + 1)$ then this can always be achieved, although the resulting set C may get to be somewhat large and this might result in ε becoming unacceptably small.

With these values we now construct

$$J_2 = d + \lambda_2^{-1}(b_2 - \varepsilon), \qquad \eta_2 = \frac{\log(J_2) - \log(1 - \varepsilon)}{\log \lambda_2^{-1}}$$

$$\zeta_2(x) = V(x) + \int V(z) \mathscr{M}_\alpha(\mathrm{d}z)/2, \qquad \xi_2(x) = \frac{\log \zeta_2(x)}{\log \lambda_2^{-1}}. \tag{24}$$

Following Roberts and Tweedie (1999), let $n' = n - \zeta_2(x)$ and define $\hat{\beta}_n$ as

$$\hat{\beta}_n = \begin{cases} \lambda_2^{-1}, & J_2 < 1; \\ (1 - \varepsilon)^{-1/\eta_2}(1 + \eta_2/n')^{-1/\eta_2}, & J_2 \geqslant 1. \end{cases}$$

As a corollary of Theorem 5.1 and of (33) in Roberts and Tweedie (1999) we then have the following general bounds.

Theorem 3. *Suppose* $\mathrm{E}|X_1| < \infty$.

    (*i*) *If $J_2 < 1$, then for $n > \xi_2(x) + \eta_2(1 - \varepsilon)/[\lambda_2^{\eta_2} - (1 - \varepsilon)]$,*

$$\|P^n(x, \cdot) - \mathscr{M}_\alpha\| \leq \zeta_2(x)\frac{1 - \lambda_2(1 - \varepsilon)}{1 - J_2}\lambda_2^n.$$

    (*ii*) *If $J_2 \geq 1$, then for $n > \xi_2(x) + \eta_2(1 - \varepsilon)/\varepsilon$,*

$$\|P^n(x, \cdot) - \mathscr{M}_\alpha\| \leq (1 - \hat{\beta}_n(1 - \varepsilon))(1 + n'/\eta_2)(\hat{\beta}_n)^{-n'}.$$

    (*iii*) *If $J_2 \geq 1$, then, for any $1 > \omega > \rho = (1 - \varepsilon)^{\eta_2^{-1}}$,*

$$\|P^n(x, \cdot) - \mathscr{M}_\alpha\| \leq \frac{[1 - (1 - \varepsilon)/\omega]\omega^{-\xi_2(x)}}{1 - (1 - \varepsilon)\omega^{-\eta_2}}\omega^n.$$

Note again that in these bounds we need the value $\zeta_2(x)$ (which is notationally somewhat hidden by the use of n'). This can be bounded, again inefficiently in many contexts, by $(V(x) + b/(1 - \lambda))/2$; alternatively, it could be estimated by the law of large numbers using the sum $(V(x) + \sum_1^n V(\Gamma_i)/n)$.

    As a second approach, from Theorem 4.2 of Roberts and Tweedie (1999), we have, using the constants in (21), the following result.

Theorem 4. *The partial sums of the transition laws P^n converge to $\mathscr{M}_\alpha$ with a bound given by*

$$\left\|n^{-1}\sum_1^n P^m(x, \cdot) - \mathscr{M}_\alpha\right\| \leq n^{-1}\left\{2 + \left[\log V(x) + \log\left(\frac{b}{1 - \lambda}\right) + \frac{2}{\varepsilon}\log\left(\frac{J}{1 - \varepsilon}\right)\right]\Big/\log\lambda^{-1}\right\}. \tag{25}$$

Although this is not a bound on geometric convergence, it still serves the purpose from the point of simulations. If we set *n* large enough that the right-hand side of (25) is small enough to meet our approximation criteria, then we can draw from the distribution $n^{-1}\sum_1^n P^m(x, \cdot)$ by first drawing a value of *m* uniformly in the range $\{1, 2, \ldots, n\}$ and then drawing from the resultant distribution $P^m(x, \cdot)$. This two-stage sampling procedure then has the error of approximation in (25). Often, this shows that we need a smaller *n* than is indicated by the geometric bounds.

5. Applications

Here we give some examples of the approximation of the df $\mathscr{M}_\alpha(t)$ by $\mathscr{M}_k^{n,x}(t)$, selecting a suitable *n*, once the approximation error has been fixed. We consider examples mainly from Muliere and Tardella (1998) and Guglielmi (1998), in order to compare the bounds obtained with their results.

Example 1. Let $\alpha_0 = U(0, 1)$, so that the support of α_0 is the interval $[0, 1]$, and suppose that $a = 1$. In this case $\mathrm{E}|X_1| = \mathrm{E}X_1 = \frac{1}{2}$, $K(\lambda) = (1 - \lambda + \frac{1}{4})/(\lambda - \frac{1}{2})$.

If we choose $\lambda \leqslant \frac{7}{8}$, then $K(\lambda) \geqslant 1$, so that the whole interval $[0, 1]$ is small. Thus Γ_n, starting from any x in $[0, 1]$, is uniformly ergodic with rate of convergence bounded by $\rho = 1 - \varepsilon$, where

$$\varepsilon = \int_D p(0, z) \, \mathrm{d}z + \int_{D^c} p(1, z) \, \mathrm{d}z = \int_{1/2}^1 (-\log z) \, \mathrm{d}z + \int_0^{1/2} (-\log(1 - z)) \, \mathrm{d}z = 1 - \log 2.$$

For instance, if $n \geqslant 13$, $\sup_x \|P^n(x, \cdot) - \mathcal{M}_\alpha\| \leqslant 0.009$, while if $n \geqslant 38$, $\sup_x \|P^n(x, \cdot) - \mathcal{M}_\alpha\| \leqslant 10^{-6}$. If we simulate a sample of size $k = 10\,000$ from Γ_n, then $\sup_{t \in \mathbb{R}} |\mathcal{M}_k^{n,x}(t) - \mathcal{M}^{n,x}(t)| < 0.021$ with probability almost one for any n, as in Section 1. Hence, with high probability,

$$\sup_{t \in \mathbb{R}} |\mathcal{M}_k^{n,x}(t) - \mathcal{M}_\alpha(t)| < 0.03$$

for any $n \geqslant 13$. The same approximation error was achieved in Guglielmi (1998), who considered, as an approximating function, the df $\mathcal{M}_\alpha r$ of the mean of a Dirichlet process with parameter equal to a discrete measure with $r = 1001$ points of mass. In Figure 1 we plot the empirical df for $n = 13$, $x = 0$, and both $k = 500$ and $k = 5000$, respectively, together with the exact df. The convergence, even for this low value of n, is clear. If we plot the empirical df for $k = 10\,000$, then this is visually indistinguishable from the exact df.

Figure 2 displays the trace plot of Γ_n for three randomly generated initial values; the plot shows that for different initial values the convergence is essentially immediate.

Suppose now, on the other hand, that $a = 50$. In this case $[0, 1]$ is still small, but now $\varepsilon = 0.33 \times 10^{-16}$, which would imply choosing n greater than about 0.14×10^{18} in order to guarantee that $\sup_x \|P^n(x, \cdot) - \mathcal{M}_\alpha\| \leqslant 0.01$ using (20).

However, we can choose λ such that $K(\lambda)$ is less than 1 and $P(x, A) \geqslant \varepsilon\mu(A \cap (K, 1])$, for all $x \in [0, K]$, μ being the Lebesgue measure on the real line, and $\varepsilon = \int \min (p(0, z), p(K, z)) \, \mathrm{d}z$. Moreover, estimating $\zeta(0)$ by Z_V^m, with, say, $m = 10^6$, we found that the best bound in Theorem 2(ii) was obtained for $\lambda = 0.999$, $b = 0.011$, $\varepsilon = 0.574\,135 \times 10^{-11}$, giving $\sup_{x \in [0,1]} \|P^n(x, \cdot) - \mathcal{M}_\alpha\| \leqslant 0.01$ for n greater than about 0.62×10^{15}. In this case we obtain an improvement using Theorem 2 over (20), even though neither set of values is of practical use.

In Figure 3(a) we give the trace plot of Γ_n for five randomly generated intial values; observe that in this case the convergence seems to happen at about $n = 150$. Thus, in spite of the fact that n above is very large, simulating the corresponding Markov chain is very easy, and the actual convergence seems to be very much faster than any of these theoretical upper bounds. Figure 3(b) shows the histogram of 5000 values from Γ_{150}.

For $a = 100$, Theorem 2 gives n much greater than before in the bounds, but a trace plot shows that the convergence actually seems to occur at about $n = 250$.

Example 2. Next we consider an example from Muliere and Tardella (1998) when the parameter is $\alpha + \sum_{i=1}^2 \delta_{x_i}$, $x_1 = 0.05$, $x_2 = 0.1$, α being the uniform distribution on $(0, 1)$, so the the parameter is not a continuous distribution. This measure is the 'posterior' parameter of a Dirichlet process, after two observations, since, as is well know, the posterior distribution
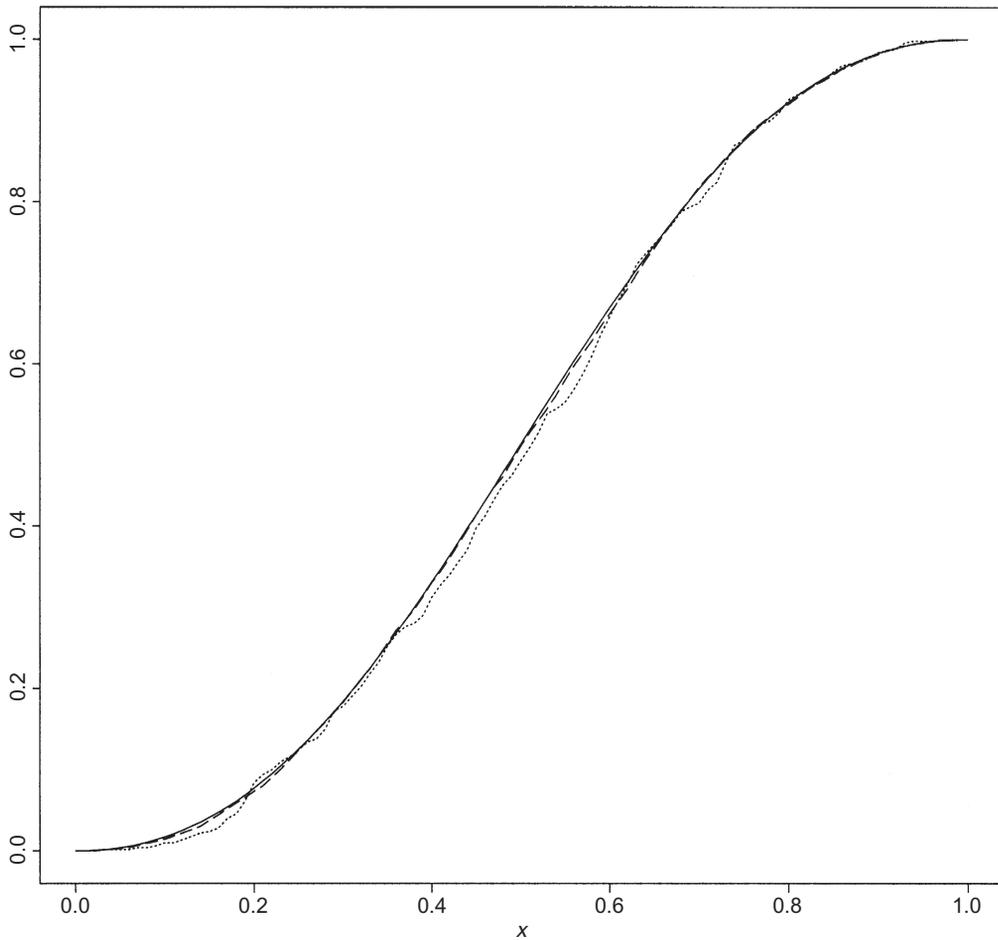
Figure 1. Empirical dfs of Γ_{13} from $\Gamma_0 = 0$ for $k = 500$ ($\cdots$), $k = 5000$ ($- - -$) and exact df (———) of Γ_α, when $\alpha = U(0, 1)$.
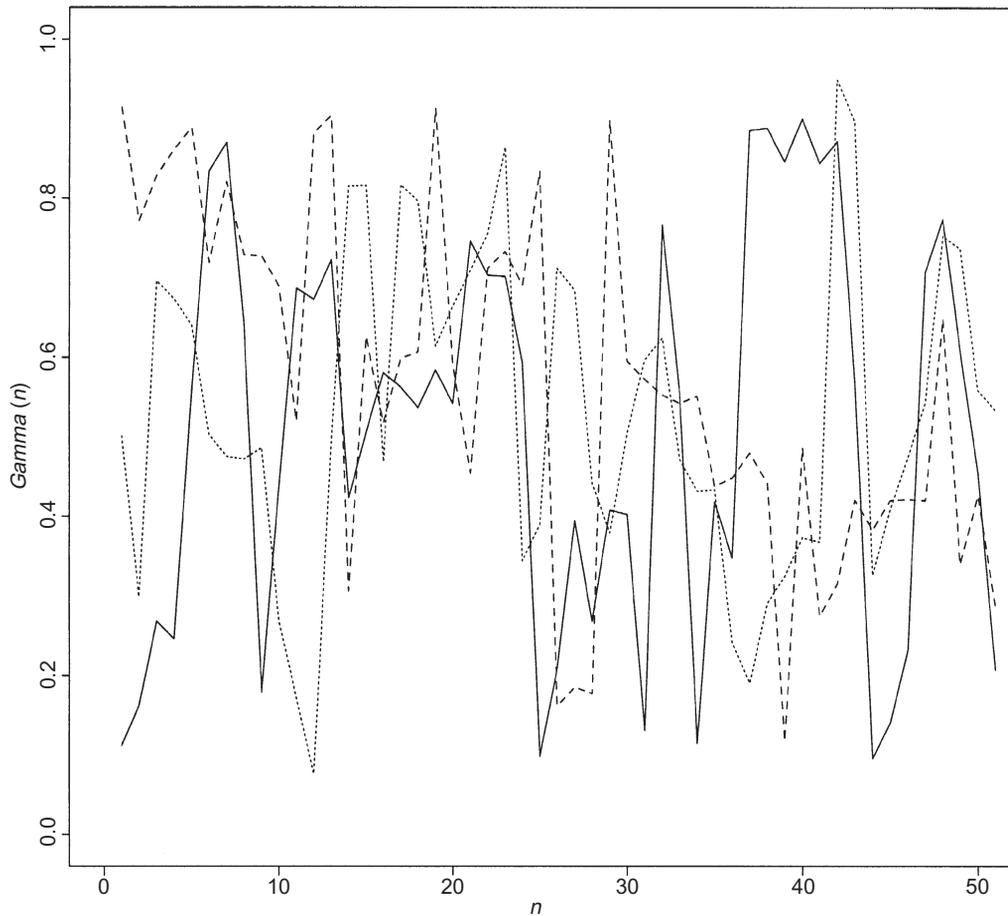
of P_α given X_1, X_2 is still the distribution of a Dirichlet process, now with parameter $\alpha + \sum_{i=1}^{2} \delta X_i$.

We have computed the bound in Theorem 2(ii) with $\varepsilon = \delta \varepsilon_0'$, $\delta = 1 - 1/a$, $a = 3$, and $\varepsilon_0 \geqslant \varepsilon_0'$, where

$$\varepsilon_0' := \int_0^{x_1} p(K, z)\,\mathrm{d}z + \int_{K/(K+1)}^{1} p(0, z)\,\mathrm{d}z, \qquad K < 1,$$

$$\varepsilon_0' := \int_0^{x_1} p(1, z)\,\mathrm{d}z + \int_{1/2}^{1} p(0, z)\,\mathrm{d}z, \qquad K \geqslant 1,$$

since ε_0' is easier to compute than ε_0.

Figure 2. Trace plot of Γ_n for three randomly generated initial values, when $\alpha = U(0, 1)$.

If λ is less than 0.903, then K is greater than 1, and the whole space $[0, 1]$ is small; in this case $\varepsilon = 0.005\,062$, so that we obtain a uniform error less than 0.01 when $n \geqslant 908$.

On the other hand, for $\lambda = 0.95$ and $b = 0.104\,17$, ε is equal to $0.018\,001$; now, Theorem 2(ii) gives the same error for $n \geqslant 4061$ (estimating $\zeta(0)$ by Z_V^m, $m = 10^6$). In this case the bounds in Theorem 2 (and 3 and 4) are not as good as the uniform ergodicity bounds in (20), even though we have improved the value of ε.

Example 3. Finally, let α_0 be a standard normal distribution, and let $a = 10$. If $\lambda = 0.98$, then d is greater than $b/(2(1 - \lambda)) - 1$ and, defining λ_2 and b_2 as in (23), and $b = 0.0926$, we obtain
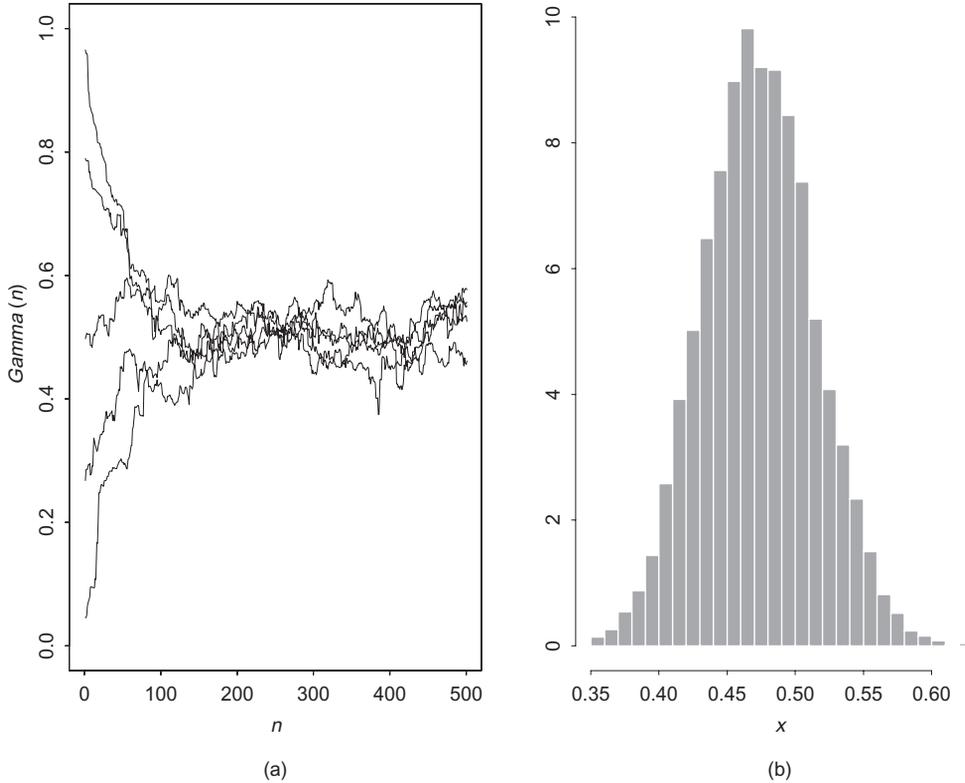
Figure 3. (a) Trace plot of Γ_n for five randomly generated initial values and (b) histogram of 5000 values from Γ_{150}, when $\alpha = 50U(0, 1)$.

$$\varepsilon = \varepsilon_0 = \int_{-\infty}^{0} p(K, z)\, \mathrm{d}z + \int_{0}^{+\infty} p(-K, z)\, \mathrm{d}z$$

$$= \int_{-\infty}^{0} \mathrm{d}z \int_{-\infty}^{z} a\frac{(z - x_1)^{a-1}}{(K - x_1)^a}\, \mathrm{d}A_0(x_1) + \int_{0}^{+\infty} \mathrm{d}z \int_{z}^{+\infty} a\frac{(x_1 - z)^{a-1}}{(x_1 + K)^a}\, \mathrm{d}A_0(x_1) = 1.089 \times 10^{-3}.$$

We apply Theorem 3(ii), estimating $\zeta(0)$ by Z_V^m, to obtain an error less than 0.01 for $n \geqslant 1\,012\,555$.

In this case the guaranteed convergence is after a relatively reasonable number of iterations, but again the trace plot of Γ_n for five randomly generated initial values (Figure 4) shows that convergence seems to occur after a very much smaller number of iterations.
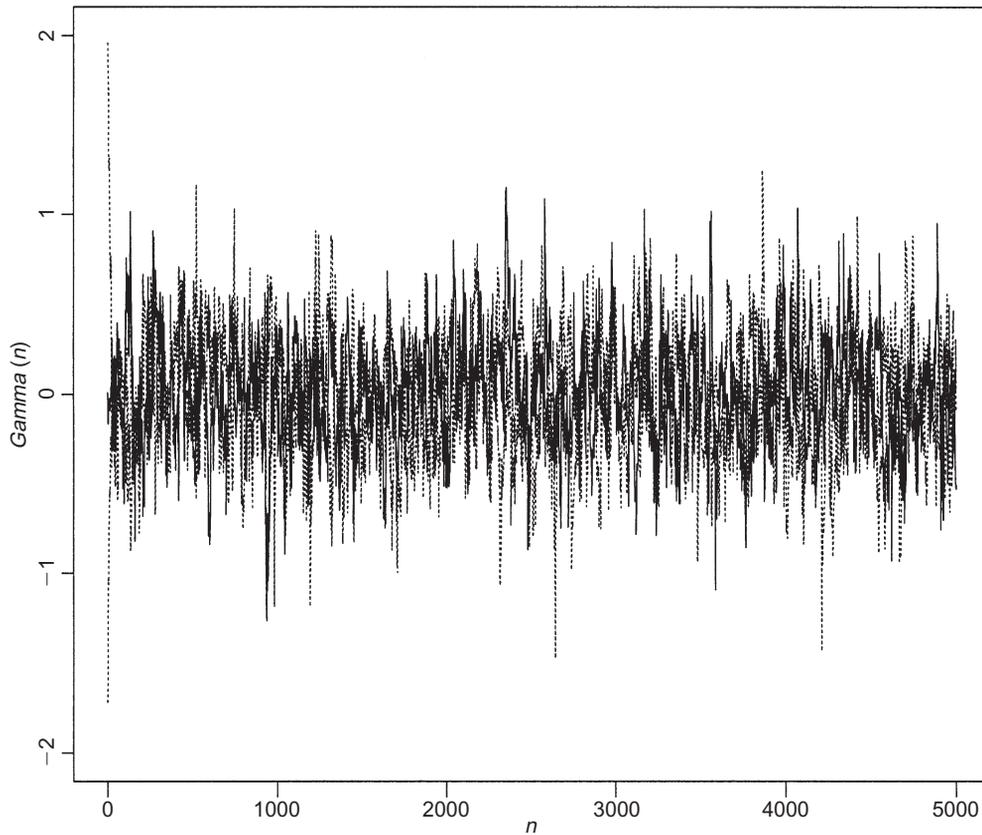
Figure 4. Trace plot of Γ_n for five randomly generated initial values when $\alpha = 10N(0, 1)$.

6. Conclusions

In this paper we suggest a new approximation procedure for the distribution of the mean functional of a Dirichlet process, since this distribution is difficult to handle even though an analytic expression is known. This Markov chain Monte Carlo procedure is essentially based on simulating draws from random variables that belong to a Markov chain converging (in total variation distance) to the mean functional, and seems to be very easy to implement. We need only draw n independent couples of random variables (X_i, Y_i) and to compute Γ_n recursively by (3); this step has to be repeated k times to obtain the approximating function (i.e. the empirical distribution from Γ_n), and so we can either use many short runs or subsample one long run at steps of size n, where n is chosen to meet the preset approximation criteria.

Because of the Markovian nature of the algorithm, we can find analytic upper bounds on the number of iterations needed in setting these approximation criteria. In all the cases that

we examined the upper bounds seemed rather conservative, compared with an assessment of trace plots to consider convergence, but in several of the examples they were still very effective, although in other cases they were rather larger than can be used in practice.

The examples considered show that the larger a becomes, the bigger n must be in order to obtain the same approximation error. This remark could be anticipated from (3), since, if a is large, Y_n is very close to 1 (its mean is about 1 and its variance is very small), so that Γ_n is very close to Γ_{n-1} and the chain does not move enough to enable rapid convergence.

A different approximation scheme, proposed in Muliere and Tardella (1998), relies also on simulation using draws from the distribution of the random vector (X_n, Y_n) as does ours, but, contrary to our method, the number of draws is determined by a random stopping rule (therefore introducing into the simulation one more variable compared to our procedure). Observe that Muliere and Tardella (1998) provide bounds in terms of the Prohorov distance between the mean functionals, and the corresponding computational effort grows (on average) linearly with $a = \alpha(\mathbb{R})$. Here a much stronger distance, the total variation, can be guaranteed, thus justifying the need for more complex theoretical results to obtain finer upper bounds, although, not unreasonably given the strength of the metric, the computational effort appears to grow faster than linearly with a, as is also attested in Guglielmi (1998) and Regazzini *et al.* (2000).

The two approximations, together with that proposed by Florens and Rolin (1994), are strongly connected, as pointed out by one of the referees. Indeed, when truncating the series representation (7) to a finite index n,

$$\sum_{i=1}^{n} p_i(\omega)\delta_{X_i(\omega)}(\cdot), \qquad (26)$$

one obtains a random measure that does not sum to one, whose remaining mass is $R_n(\omega) = 1 - \sum_{l=1}^{n} p_l(\omega) = \prod_{i=1}^{n} Y_i(\omega)$. Here the random probability measure P_n, as defined in (2), assigns this mass to a distribution δ_{x_0} degenerate at some fixed point x_0, while Florens and Rolin (1994) use the random measure (26) scaled by $1 - R_n(\omega)$, that is,

$$\frac{\sum_{i=1}^{n} p_i(\omega)\delta_{X_i(\omega)}(\cdot)}{1 - R_n(\omega),}$$

as an approximating random probability measure. On the other hand, Muliere and Tardella (1998) define n by a random rule and assign the remaining mass R_n to $\delta_{X_0(\omega)}$, where $X_0(\omega)$ is a random point. Hence, all these approximations will provide similar random quantities if the same n is used, although the features of the techniques are different.

Our approximation procedure also seems numerically more efficient than that in Guglielmi (1998) and Regazzini *et al.* (2000), where a 'classical' approximation method is proposed, involving no simulation and considering, as the approximation error, the distance in the Lévy metric between the exact df and the df corresponding to a discrete parameter measure. For the examples considered here, we found that the execution times do not differ too much when the support of α is bounded and a is relatively small (for instance, when a has the same order of magnitude as 1). When a increases, the simulation procedure

proposed here results in a larger index n of the random variable Γ_n from which we generate the empirical distribution, but the increase in execution time is not striking. Conversely, if a increases, the 'classical' procedure brings an addition in the number of points in the support of the approximating parameter, yielding a much longer execution time.

The stochastic monotonicity of the chain should enable use of the Propp–Wilson 'perfect sampling' algorithm, and this has been sketched out in the case of α having bounded support in Guglielmi *et al*. (2000). In further work we shall explore whether the approach of Corcoran and Tweedie (2001) can be exploited when α has unbounded support.

Overall the Markov chain Monte Carlo procedure appears to provide a simply implemented procedure that has the highly desirable property that in general we can guarantee the order of approximation using Markov chain theory.

Acknowledgement

References

Cifarelli, D.M. and Regazzini, E. (1990) Distribution functions of means of a Dirichlet process. *Ann. Statist.*, **18**, 429–442.

Corcoran, J. and Tweedie, R.L. (2001) Perfect sampling of Harris recurrent Markov chains. *Ann. Appl. Probab.* To appear.

Feigin, P.D. and Tweedie, R.L. (1989) Linear functionals and Markov chains associated with Dirichlet processes. *Math. Proc. Cambridge Philos. Soc.*, **105**, 579–585.

Ferguson, T.S. (1973) A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 209–230.

Florens, J.-P. and Rolin, J.M. (1994) Bayes, bootstrap, moments. Discussion Paper 94.13, Institut de Statistique, Université Catholique de Louvain.

Gelfand, A.E. and Kottas, A. (1999) Full nonparametric Bayesian inference for single and multiple sample problems. Technical report, University of Connecticut.

Gugliemi, A. (1998) Numerical analysis for the distribution function of the mean of a Dirichlet process. Quaderno IAMI 98.01, CNR-IAMI.

Guglielmi, A., Holmes, C.C. and Walker, S.G. (2000) Perfect simulation involving a continuous and unbounded space. Technical Report TR-00-08, Department of Mathematics, Imperial College, London University.

Meyn, S.P. and Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability*. Berlin: Springer-Verlag.

Muliere, P. and Secchi, P. (1996) Bayesian nonparametric predictive inference and bootstrap techniques. *Ann. Inst. Statist. Math.*, **48**, 663–673.

Muliere, P. and Tardella, L. (1998) Approximating distributions of random functionals of Ferguson–Dirichlet priors. *Canad. J. Statist.*, **26**, 283–297.

Nummelin, E. (1984) *General Irreducible Markov Chains and Non-negative Operators*. Cambridge: Cambridge University Press.

Regazzini, E., Guglielmi, A. and Di Nunno, G. (2000) Theory and numerical analysis for exact distributions of functionals of a Dirichlet process. Quaderno IAMI 00.12, CNR-IAMI.

Roberts, G.O. and Tweedie, R.L. (1999) Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process. Appl.*, **80**, 211–229. Correction (2001): *Stochastic Process. Appl.*, **91**, 337–338.

Roberts, G.O. and Tweedie, R.L. (2000) Rates of convergence of stochastically monotone stochastic processes. *J. Appl. Probab.*, **37**, 359–373.

Rosenthal, J.S. (1995) Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.*, **90**, 558–566.

Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Sethuraman, J. (1994) A constructive definition of Dirichlet priors. *Statist. Sinica*, **4**, 639–650.

Tweedie, R.L. (1975) Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Process. Appl.*, **3**, 385–403.