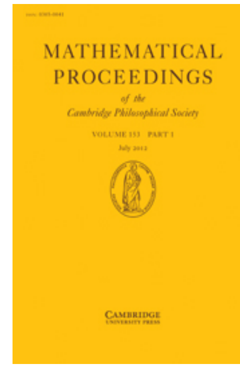

Linear functionals and Markov chains associated with Dirichlet processes

Paul D. Feigin and Richard L. Tweedie

Mathematical Proceedings of the Cambridge Philosophical Society / Volume 105 / Issue 03 / May 1989, pp 579 - 585
DOI: 10.1017/S0305004100077951, Published online: 04 October 2011

Math. Proc. Camb. Phil. Soc. (1989), **105**, 579
Printed in Great Britain

Linear functionals and Markov chains associated with Dirichlet processes

BY PAUL D. FEIGIN
Technion, Haifa and Division of Mathematics and Statistics, CSIRO

AND RICHARD L. TWEEDIE
Siromath Pty. Ltd., Sydney and Bond University, Gold Coast, Australia

(*Received* 20 *June* 1988)

Abstract

By investigating a Markov chain whose limiting distribution corresponds to that of the Dirichlet process we are able directly to ascertain conditions for the existence of linear functionals of that process. Together with earlier analyses we are able to characterize those functionals which are a.s. finite in terms of the parameter measure of the process. We also show that the appropriate Markov chain in the space of measures is only weakly convergent and not Harris ergodic.

1. *Introduction*

The family of Dirichlet processes provides a way of producing a mathematically tractable family of prior distributions on the set $\Pi(\Xi)$ of all probability measures on the Borel sets of a complete separable metric space Ξ. Their definition, characterization and essential properties were extensively presented by Ferguson [**4**, **5**] and further elaborated on by Blackwell and MacQueen [**2**] and Blackwell [**1**]. Following these basic works various authors have considered other characterizations and properties of the Dirichlet process – see, for example, Hannum *et al.* [**6**], Doss and Sellke [**3**] and Yamato [**13**].

Here we consider the Dirichlet process as the limiting or invariant measure for a particular Markov chain with state space $\Pi(\Xi)$. We show that all the (finite) linear functionals of this measure-valued Markov chain are themselves Markov chains. This fact gives the process a very special structure which we exploit in our proofs. The approach allows one to obtain readily conditions for the existence of linear functionals of the Dirichlet process itself (see Theorem 2). A further analysis based on the approach of Hannum *et al.* [**6**] then shows that these conditions are necessary. The results are summarized in our main new result – Theorem 4. We also consider the implications of this result to the question of tail-behaviour of Dirichlet processes and show that we have generalized the results of Doss and Sellke [**3**] to general spaces Ξ.

Our starting point is a known characterization of the Dirichlet process P with parameter α whose distribution on $\Pi(\Xi)$ we denote by $\Delta(\alpha)$. The parameter is often called the prior measure and is a finite measure on Ξ. Let $M = \alpha(\Xi)$, $Q = M^{-1}\alpha$ and

P, X, Y be independently distributed with distributions $\Delta(\alpha)$ on $\Pi(\Xi)$, Q on Ξ, and Beta$(M, 1)$ on $[0, 1]$, respectively. Then, according to this characterization $\Delta(\alpha)$ satisfies

$$\Lambda[P] = \Lambda[(1-Y)\,\delta_X + YP], \tag{1}$$

where $\Lambda[\cdot]$ denotes 'law of' and δ_x is a distribution function degenerate at $x \in \Xi$ (see the proof of Theorem 1 below).

Equation (1) can be recognised as the equation for a stationary measure for the Markov chain defined via the recursion

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

where $P_0 \in \Pi(\Xi)$ is arbitrary and $\{(X_n, Y_n)\}$ is an i.i.d. sequence with the same joint distribution as (X, Y) above. We show below how the limit theory for Markov chains on general state spaces (see Tweedie [**12**]) together with the general theory of convergence of random measures may now be applied to prove that (1) has a unique solution. That this solution is the Dirichlet process $\Delta(\alpha)$ then follows from the defining property of the latter. The details will be given below.

The Markov chain approach is particularly convenient for analysing the existence and properties of linear functionals $\int g\,dP$ because they also turn out to be derivable as strong limits of Markov chains on $\mathbb{R}$. These functionals were considered by Hannum *et al.* [**6**] and their existence was also investigated by Doss and Sellke [**3**] who referred to them as moments.

We note that (1) is also the basis of a related construction of the Dirichlet process discussed in Sethuranam and Tiwari [**9**].

2. *Existence of an invariant measure*

We prove that the Markov chain defined in (2) has a unique invariant measure which we may then identify as the Dirichlet process measure $\Delta(\alpha)$. We first use a result of Kallenberg [**7**].

LEMMA 1. *For a complete separable metric space Ξ, and random measures P_n on $\Pi(\Xi)$,*

$$\int g\,dP_n \xrightarrow{d} \text{some } P_g \quad \text{for all bounded continuous } g \text{ on } \Xi \tag{3}$$

implies that there exists a random measure P such that

$$P_n \xrightarrow{d} P \quad \text{with} \quad \int g\,dP \overset{d}{=} P_g \tag{4}$$

for all bounded continuous g. (Here $\xrightarrow{d}$ denotes convergence in law.)

Proof. The result follows directly from lemma 5·1 of Kallenberg [**7**]. ∎

The above lemma reduces the problem of convergence of the P_n in (2) to one of showing weak convergence for bounded continuous linear functionals $\int g\,dP_n$. In fact a stronger result follows directly from the theory of Harris ergodicity for Markov chains.

LEMMA 2. *Suppose that g is a bounded (Borel) measurable function on Ξ and $\{P_n\}$ is defined as in equation (2). Then $\{G_n = \int g\,dP_n\}$ is a Markov chain on $\mathbb{R}$ with a unique* *stationary measure* Π_g. *In particular,* G_n *converges in law for* Π_g*-almost all starting points* G_0.

Proof. From the definition of G_n and the defining equation (2) we obtain

$$G_n = (1-Y_n)\,g(X_n) + Y_n\,G_{n-1};\, n \geqslant 1. \tag{5}$$

We may therefore conclude that G_n is a Markov chain on $\mathbb{R}$.

In fact, from (5) it follows that G_n is restricted to the compact set $[-\|g\|, \|g\|]$ where $\|g\| = \sup_\Xi |g(x)|$. Now we know that a Markov chain on a compact space has at least one finite invariant measure provided that it is weak Feller (see Rosenblatt [8]); that is, if the measures $\mathbb{P}(G_n \in .\,|\,G_{n-1} = x)$ are weakly continuous in x. To prove the weak Feller property it suffices to observe that, since

$$\mathbb{P}(G_n \leqslant y\,|\,G_{n-1} = x) = \mathbb{P}((1-Y_n)\,g(X_n) \leqslant y - xY_n) = \int_0^1 \mathbb{P}\left(g(X_n) \leqslant \left[\frac{y-zx}{1-z}\right]\right)\mathbb{P}(Y_n \in dz),$$

we have

$$\mathbb{P}(G_n \leqslant y|\,G_{n-1} = x) - \mathbb{P}(G_n \leqslant y|\,G_{n-1} = x+\delta)$$
$$= \int_0^1 \mathbb{P}\left(g(X_n) \in \left[\frac{y-zx-z\delta}{1-z}, \frac{y-zx}{1-z}\right]\right)\mathbb{P}(Y_n \in dz).$$

Since, for any Q and g, the distribution of $g(X_n)$ has at most a countable number of atoms, and since Y_n has a continuous distribution, $\{G_n\}$ is weak Feller.

If we can now show that $\{G_n\}$ is ϕ-irreducible for some finite measure ϕ, then (see Tweedie [**10**]) the chain is positive recurrent and the invariant measure is unique.

Set $\underline{g} = \text{ess–inf}_\Xi\, g(x), \bar{g} = \text{ess–sup}_\Xi\, g(x)$ (where Q is the reference measure on Ξ) and set $\phi(A) = \lambda(A \cap [\underline{g}, \bar{g}])$ with λ denoting Lebesgue measure. Then, since Y_1 has a density on $[0, 1]$, the conditional distribution of $G_1 = Y_1 G_0 + (1-Y_1)\,g(X_1)$, given G_0 and X_1, has a density with respect to λ on $[G_0, g(X_1)]$ for $G_0 < g(X_1)$ (and similarly for $G_0 > g(X_1)$). Since $g(X_1)$ can take values equal or arbitrarily close to $\underline{g}$ and $\bar{g}$ we obtain the result that $\phi(A) > 0$ implies $\mathbb{P}(G_1 \in A\,|\,G_0) > 0$, the required ϕ-irreducibility. ∎

We note that the above convergence results do not depend on P_0 and so we obtain

THEOREM 1. *There exists a unique invariant measure for* $\{P_n\}$ *on* $\Pi(\Xi)$ *which satisfies equation* (1) *and it is the Dirichlet process measure* $\Delta(\alpha)$.

Proof. Lemmas 1 and 2 guarantee the existence of a unique invariant measure for the process described in equation (2). The uniqueness itself follows since the distribution of P is determined by the distributions of all the linear functionals $\int g\,dP$ (see Kallenberg [7], theorem 3·1). That this distribution corresponds to that of the Dirichlet process can be shown as follows. Consider the defining property (D): for any finite measurable partition $\{B_1, \ldots, B_m\}$ of Ξ,

$$(P(B_1), \ldots, P(B_m)) \sim \text{Dirichlet}\,(\alpha(B_1), \ldots, \alpha(B_m)). \tag{6}$$

It is not difficult to show that if P satisfies (D) then so does $(1-Y)\,\delta_X + YP$ (see Ferguson [**4**]). Hence the unique solution to equation (1) has property (D) and so must be the Dirichlet process with parameter α. ∎

In the last section we make a technical, although interesting observation concerning the nature of the convergence of P_n. Although the functionals G_n give rise to Harris ergodic chains and the convergence of $\Lambda[G_n \mid G_0]$ is in total variation on $\Pi(\mathbb{R})$, whenever α has a diffuse part on Ξ we only have *weak* convergence of $\Lambda(P_n \mid P_0)$ on $\Pi(\Xi)$.

3. *Finiteness of linear functionals*

We seek minimal conditions for the existence of linear functionals $\int g\,dP$, where P is a Dirichlet process with parameter α. For the case $\Xi = \mathbb{R}$, Doss and Sellke [**3**] analyse the related problem of tail-behaviour of P whereas Hannum *et al.* [**6**] give the distribution of $\int g\,dP$ when $\int |g|\,d\alpha < \infty$.

Although the Markov chain analysis quite readily gives the sufficiency of the logarithmic moment (with respect to α – see the statement of Theorem 2) the necessity of this condition will be proved separately by refining the technique of Hannum *et al.* [**6**]: see Theorem 3. We note that this refinement could also provide an alternate proof of the sufficiency.

THEOREM 2. *If* $\int \log(1+|g(\xi)|)\,d\alpha(\xi) < \infty$ *then* $|\int g\,dP| < \infty$ *a.s.* $[\Delta(\alpha)]$.

Proof. Choose P_0 so that $\int |g|\,dP_0 < \infty$ and define $\{G_n\}$ via (5). Our result will follow if we show that G_n has an invariant or limiting distribution. We now prove that under the stated condition G_n is Harris ergodic.

The proof of ϕ-irreducibility is virtually identical to that of Lemma 2; the fact that $\underline{g}$ and $\bar{g}$ may be infinite does not change the result.

We now turn to the conditions for ergodicity and define $f(u) = \log(1+|u|)$. We will show that

$$E[f(G_1) \mid G_0] \leqslant f(G_0) - \epsilon \quad (G_0 \in K^c) \tag{7}$$

and

$$E[f(G_1) \mid G_0] \text{ is bounded} \quad (G_0 \in K),$$

for some compact $K \subset \mathbb{R}$ and $\epsilon > 0$. This condition is sufficient for Harris ergodicity (see, for example, Tweedie [**12**]).

Indeed

$$f(G_1) \leqslant f((1-Y_1)|g(X_1)| + Y_1|G_0|) \tag{8}$$

$$= f(G_0) + \log\left[\frac{1+Y_1|G_0| + (1-Y_1)|g(X_1)|}{(1+|G_0|)\,Y_1}\right] + \log(Y_1). \tag{9}$$

Note that as $|G_0| \uparrow \infty$ the argument of the first logarithm converges monotonically to 1. Since $E|\log(Y_1)| < \infty$, and since

$$\log(1+Y_1|G_0| + (1-Y_1)|g(X_1)|) \leqslant \log(1+Y_1|G_0|) + \log(1+|g(X_1)|) \tag{10}$$

we may conclude that under the condition of the theorem $E[f(G_1) \mid G_0]$ is finite, and is in fact bounded for $G_0 \in K$. Now choose $\epsilon = -\frac{1}{2}E\log(Y_1)$ and k large enough to ensure that

$$E\log\left[\frac{1+Y_1 k + (1-Y_1)|g(X_1)|}{(1+k)\,Y_1}\right] < \epsilon. \tag{11}$$

Then it follows immediately that

$$E[f(G_1) \mid G_0] \leqslant f(G_0) - \epsilon \quad (G_0 \in K^c) \tag{12}$$

where $K = \{y \in \mathbb{R} : |y| < k\}$, and the theorem is proved. ∎

We now turn to the converse, using a quite different approach.

THEOREM 3. *Suppose* $g > 0$ *and* $\int \log(1+g)\,d\alpha = \infty$. *Then*

$$\int g\,dP = \infty \ a.s.\,[\Delta(\alpha)]. \tag{13}$$

Proof. We consider the proofs of theorems 2·3 and 2·5 from Hannum *et al.* [6]. Take a sequence of bounded measurable positive functions $g_n \uparrow g$ (pointwise on Ξ). Firstly we note that there is no need to restrict Ξ to $\mathbb{R}$ in their proofs. Secondly, for bounded g_n we may replace the characteristic function by Laplace transforms and conclude that, for each n,

$$\mathbb{P}\left(\int g_n\,dP \leqslant x\right) = \mathbb{P}(T_n^x \leqslant 0) \tag{14}$$

where T_n^x has Laplace transform given by

$$\log E(\exp[-sT_n^x]) = -\int_\Xi \log[1+s(g_n-x)]\,d\alpha. \tag{15}$$

Now fixing x and $s < 1/(x+1)$, we note that $1+s(g_n-x) > s(1+g_n)$. Hence

$$\int_\Xi \log[1+s(g_n-x)]\,d\alpha \geqslant \int_\Xi \log(1+g_n)\,d\alpha + M\log(s), \tag{16}$$

which converges to ∞ as $n\to\infty$ by the condition of the theorem. Thus since $E[\exp(-sT_n^x)]$ tends to 0, so do $\mathbb{P}(T_n^x \leqslant 0)$ and $\mathbb{P}(\int g_n\,dP \leqslant x)$, and hence

$$\mathbb{P}\left(\int g\,dP \leqslant x\right) = 0 \quad \text{for} \quad \text{all } x > 0 \tag{17}$$

which is the desired result. ∎

We summarize these last two results as follows.

THEOREM 4. *For a Dirichlet process* P *with parameter* α, *the integral* $\int |g|\,dP$ *is finite or infinite a.s.* $[\Delta(\alpha)]$ *according as* $\int \log(1+|g|)\,d\alpha$ *is finite or infinite.*

Proof. The result follows from Theorems 2 and 3. Note that, in the infinite case, $\int g\,dP$ may be undefined, or may be infinite, a.s. ∎

We conclude this section by showing that this approach can reproduce the results of Doss and Sellke [3] concerning the tail-behaviour of P for the special case that $\Xi = \mathbb{R}$. Assume that $\alpha(\mathbb{R}) = 1$ and that g is a positive and monotonically increasing function on $\mathbb{R}$. For such g,

$$\int \log(1+g)\,d\alpha < \infty \Leftrightarrow \int \log(g)\,d\alpha < \infty \tag{18}$$

and so the preceding result gives us that

$$\int \log(g)\,d\alpha < \Rightarrow \int g\,dP < \infty \quad \text{a.s.} \tag{19}$$

$$\Rightarrow g(x)(1-P(x)) \to 0 \quad \text{a.s.} \quad \text{as} \quad x\to\infty \tag{20}$$

where we use P to denote the distribution function corresponding to the measure P.

From equations (18–20) it follows that

$$\frac{1-P(x)}{h(1-\alpha(x))} \to 0 \quad \text{a.s.} \quad \text{as} \quad x \to \infty \tag{21}$$

if

$$\int_a^\infty -\log\left[h(1-\alpha(x))\right] d\alpha(x) < \infty \quad \text{for some } a > 0. \tag{22}$$

If α (as a distribution function) is differentiable then (22) is a direct consequence of

$$\int_0^1 -\log h(u)\, du < \infty, \tag{23}$$

which is exactly the condition of Doss and Sellke [3]. Of course, our results work for other spaces Ξ, whereas their approach depends on properties of Gamma processes on $\mathbb{R}$. Moreover, we note that if α is differentiable then we do not need a convexity condition on h. The differentiability of α can be removed but we do not pursue these details here.

4. *The nature of the convergence*

We prove that on $\Pi(\Xi)$ the convergence of P_n cannot be strengthened to complete set-wise convergence.

THEOREM 5. *Suppose that α has a diffuse part (i.e. it is not purely atomic). Then there exists no subset $H \subset \Pi(\Xi)$ such that $\{P_n\}$ restricted to H is Harris ergodic.*

Proof. We refer to Tweedie [11] for notation. From p. 303 of Tweedie [11], for a chain which is Harris ergodic on a set H there exists a continuous component for the chain which is everywhere non-trivial on H (note that Harris ergodicity implies that the improperly essential set E in the Harris representation ($\mathscr{H}$) is not present). Hence ($\mathscr{T}'$) holds for the process on H, and from p. 301 of Tweedie [11] ($\mathscr{G}$) holds: that is, there exists no uncountable collection of points $\{P_\gamma\} \subset \Pi(\Xi)$ such that for some $0 < \beta < 1$ the measures

$$\{V_\beta(\,.\,|P_\gamma) = \Sigma\, \mathbb{P}(P_n \in .\,|P_0 = P_\gamma)\, \beta^n\} \tag{24}$$

are mutually singular.

Let $J(P)$ denote the set of atoms of P and $S_D = \text{support}\,(\alpha_D)$ denote the support of the diffuse part of α. Define

$$H = \{P \in \Pi(\Xi) : P \text{ atomic}, J(P) \subset S_D \cup J(\alpha)\}. \tag{25}$$

Let $\{C_\gamma\}$ denote the (uncountable) collection of countable non-finite subsets of $S_D \backslash J(\alpha)$ and for each γ let

$$A_\gamma = \{P \in H : J(P) = J(\alpha) \cup C_\gamma\}. \tag{26}$$

Then we may choose $P_\gamma \in A_\gamma, P_{\gamma'} \in A_{\gamma'}$ for $C_\gamma \neq C_{\gamma'}$. Now, given $P_0 = P_\gamma$, with probability 1 we have

$$P_n \in B_\gamma = \{P \in H : J(P) \supset J(\alpha) \cup C_\gamma\} \tag{27}$$

whereas with probability 0 we have $P_n \in B_{\gamma'}$. The latter holds since $X_n \in C_{\gamma'} \backslash C_\gamma$ with probability 0 (see (2)). Thus the measures $V_\beta(\,.\,|P_\gamma)$ are mutually singular.

Now the sets A_γ cover H. Condition ($\mathscr{G}$) implies that this cover must be countable. However, since $\{C_\gamma\}$ is uncountable, so must $\{A_\gamma\}$ be. This contradicts ($\mathscr{G}$), and hence the chain P_n on $\Pi(\Xi)$ cannot be Harris ergodic on H.

We note, moreover, that $\Delta(\alpha)$ gives probability 1 to H and that any *structural* subset of H of the form

$$\{P \in H : J(P) \subset S'; \alpha(S') < \alpha(\Xi)\} \tag{28}$$

is not closed. Hence P_n cannot be Harris ergodic on any structural subset of H. For any other subset of H the original argument may be invoked to show that P_n cannot be Harris ergodic on it either. ▮

REFERENCES

[1] D. Blackwell. Discreteness of Ferguson selections. *Ann. Statist.* **1** (1973), 356–358.

[2] D. Blackwell and J. B. MacQueen. Ferguson distributions via Polya urn schemes. *Ann. Statist.* **1** (1973), 353–355.

[3] H. Doss and T. Sellke. The tails of probabilities chosen from a Dirichlet prior. *Ann. Statist.* **10** (1982), 1302–1305.

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

[5] T. S. Ferguson. Prior distributions on spaces of probability measures. *Ann. Statist.* **2** (1974), 615–629.

[6] R. C. Hannum, M. Hollander and N. A. Langberg. Distributional results for random functionals of a Dirichlet process. *Ann. Probab.* **9** (1981), 665–670.

[7] O. Kallenberg. *Random Measures* (Academic Press, 1976).

[8] M. Rosenblatt. *Markov Processes: Structure and Asymptotic Behavior* (Springer-Verlag, 1971).

[9] J. Sethuranam and R. Tiwari. Convergence of Dirichlet measures and the interpretation of their parameter. In *Proceeding of the Third Purdue Symposium on Statistical Decision Theory Related Topics* (Academic Press, 1982).

[10] R. L. Tweedie. Criteria for classifying general Markov chains. *Adv. in Appl. Probab.* **8** (1976), 737–771.

[11] R. L. Tweedie. Topological aspects of Doeblin decompositions for Markov chains. *Z. Warsch. Verw. Gebiete* **46** (1979), 299–305.

[12] R. L. Tweedie. Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In *Probability, Statistics and Analysis* (Cambridge University Press, 1983).

[13] H. Yamato. Characteristic functions of means of distributions chosen from a Dirichlet process. *Ann. Probab.* **12** (1984), 262–267.