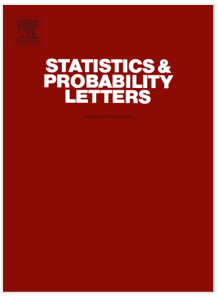
A generalized constructive definition for the Dirichlet process

S. Favaro, S.G. Walker



A generalized constructive definition for the Dirichlet process

S. Favaro, S.G. Walker

PII: S0167-7152(08)00209-5
DOI: 10.1016/j.spl.2008.04.001
Reference: STAPRO 5048

To appear in: *Statistics and Probability Letters*

Received date: 19 August 2007
Revised date: 31 March 2008
Accepted date: 1 April 2008

Please cite this article as: Favaro, S., Walker, S.G., A generalized constructive definition for the Dirichlet process. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.04.001

A generalized constructive definition for the Dirichlet process

S. Favaro [a,*], S.G. Walker [b]

[a] *Department of Decision Sciences, Univeristy "L. Bocconi", Viale Isonzo 25, Milano 20135, Italy*

[b] *Institute of Mathematics, Statistics and Actuarial Science, University of Kent, CT2 7NZ Canterbury, UK*

Abstract

In this paper we provide an alternative constructive definition for the Dirichlet process which generalizes the one given by Sethuraman

Key words: Random probability measures, Dirichlet process, Blackwell and MacQueen urn scheme, Bayesian Nonparametrics
MSC: Primary 60G57; secondary 62F15

1 Introduction

The Dirichlet process is a random probability measure (r.p.m.) whose characterization and properties were presented by Ferguson (1973) and Ferguson (1974) and further investigated by Blackwell (1973) and Blackwell and MacQueen (1973).

In order to define the Dirichlet process, let $(\mathcal{X}, \mathcal{T})$ be a Polish space endowed with the Borel σ-field $\mathscr{X}$ and consider the following associated spaces of measures $\mathcal{M}_{\mathcal{X}}$ and $\mathcal{P}_{\mathcal{X}}$. In particular, $\mathcal{M}_{\mathcal{X}}$ is the space of locally finite nonnegative measures on $(\mathcal{X}, \mathscr{X})$ endowed with the σ-field $\mathscr{M}_{\mathcal{X}}$ generated by the vague topology $\mathcal{V}$ which makes $(\mathcal{M}_{\mathcal{X}}, \mathcal{V})$ a Polish space, and $\mathcal{P}_{\mathcal{X}}$ is the space of probability measures on $(\mathcal{X}, \mathscr{X})$ endowed with its σ-field $\mathscr{P}_{\mathcal{X}}$ generated by the weak convergence topology $\mathcal{W}$ which makes $(\mathcal{P}_{\mathcal{X}}, \mathcal{W})$ a Polish space. Let $(\Omega, \mathscr{F}, p)$ be a probability space and let $\alpha \in \mathcal{M}_{\mathcal{X}}$, with total mass $a := \alpha(\mathcal{X})$.

* Corresponding author.
Email address: stefano.favaro@phd.unibocconi.it (S. Favaro).

A Dirichlet process on $(\mathcal{X}, \mathscr{X})$ with parameter α is a r.p.m. P such that, for any finite measurable partition $(B_1, \ldots, B_k)$ of $\mathcal{X}$, $\alpha(B_j) > 0$, for $j = 1, \ldots, k$, $(P(B_1; \omega), \ldots, P(B_k; \omega))$ is a random vector absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{k-1}$ with $\sum_{i=1}^{n} P(B_i, \omega) = 1$ and with Dirichlet distribution with parameter $(\alpha(B_1), \ldots, \alpha(B_k))$, $k \geq 2$.

Various authors have considered other characterizations and properties of the Dirichlet process. In particular, in this paper we are interested in the series representation of the Dirichlet process provided by Sethuraman (1994). Moving from the constructive definition of the Dirichlet process in Sethuraman (1994) we provide a more general construction. We will prove that the new construction is still a random measure on $(\mathcal{X}, \mathscr{X})$ giving probability one to the subset of discrete probability measures on $(\mathcal{X}, \mathscr{X})$ and its finite dimensional distributions are Dirichlet distributions. This establishes that the new construction is a Dirichlet process.

2 The generalized Sethuraman construction

Let $\alpha \in \mathcal{M}_{\mathcal{X}}$ with total mass a, let $\{n_i, i \geq 1\}$ be a fixed integer value sequence and let $(\Omega, \mathscr{F}, p)$ be a probability space supporting three independent sequences of r.v.s $\theta := \{\theta_i, i \geq 1\}$, $q := \{(q_{i,1}, \ldots, q_{i,n_i}), i \geq 1\}$ and $Y := \{(Y_{i,1}, \ldots, Y_{i,n_i}), i \geq 1\}$. The sequence θ is a sequence of independent r.v.s distributed according to a Beta d.f. with parameter (n_i, a), the sequence q is a sequence of independent random vectors identically distributed according to a Dirichlet d.f. with parameter $(1, \ldots, 1)$ and the sequence Y is a sequence of independent random vectors from a Pólya sequence with parameter α, i.e. if P_i are independent Dirichlet processes with parameter α, then for every $i \geq 1$, $Y_{i,1}, \ldots, Y_{i,n_i} | P_i$ are i.i.d. from P_i.

The condition of independence between the sequence of r.v.s θ, q and Y and the usual construction of a product measure implies the existence of the probability space $(\Omega, \mathcal{F}, p)$ supporting the sequence of r.v.s θ, q and Y and does not require any restrictions on $(\mathcal{X}, \mathscr{X})$, such as it being a Polish space. We now consider the sequence of r.v.s $\{p_i, i \geq 1\}$ obtained from the sequence of r.v.s θ, by the usual stick breaking construction $p_1 = \theta_1$ and $p_i = \theta_i \prod_{j=1}^{i-1}(1 - \theta_j)$ for $i \geq 2$. In particular, the stick breaking construction implies that $\sum_{i=1}^{n} p_i = 1 - \prod_{i=1}^{n}(1 - \theta_i) \to 1$ a.s. as $n \to \infty$. For any $\omega \in \Omega$ and $B \in \mathcal{X}$ define the map

$$P(\omega, B) := \sum_{i \geq 1} p_i(\omega) \sum_{j=1}^{n_i} q_{i,j}(\omega) \delta_{Y_{i,j}(\omega)}(B) \tag{1}$$

which is clearly a measurable map from $(\Omega, \mathcal{F})$ into $(\mathcal{P}_{\mathcal{X}}, \mathscr{P}_{\mathcal{X}})$ and takes values in the subset of discrete probability measures on $(\mathcal{X}, \mathscr{X})$.

In Theorem 2 we prove that the finite dimensional distributions of P are Dirichlet distributions. This establishes that P is a Dirichlet process with

parameter α. We first cosider the following lemma which recall a distributional equation having as unique solution the Dirichlet process with parameter α. As pointed out by a referee this type of distributional equations for the Dirichlet process appear already in James (2005) and for more general r.p.m., using the duality with the posterior distribution, are discussed explicitely in James (2002). We leave a short proof for clearity.

Lemma 1 *Let $\alpha \in \mathcal{M}_{\mathcal{X}}$ with total mass a, let $\{n_i, i \geq 1\}$ be a fixed integer value sequence and let θ, q and Y the three sequences of r.v.s above described. Then, for every $i \geq 1$, the distributional equation*

$$Q_i \stackrel{d}{=} \theta_i \sum_{j=1}^{n_i} q_{i,j} \delta_{Y_{i,j}} + (1-\theta_i) Q_i \tag{2}$$

has as its unique solution the Dirichlet process with parameter α.

Proof. For evey $i \geq 1$, let n independent r.v.s $\xi_{i,1}, \ldots, \xi_{i,n}$ such that $\xi_{i,j}$ is distributed according to a Beta d.f. with parameter $(1, n-j)$. From Theorem 1 in Jambunathan (1954) it follows that $q_{i,1} \stackrel{\mathrm{d}}{=} \xi_{i,1}$ and $q_{i,j} \stackrel{\mathrm{d}}{=} \xi_{i,j} \prod_{l=1}^{j-1}(1-\xi_{i,l})$, for $j = 2, \ldots, n$. By this stick breaking construction it follows by induction that $1 - \sum_{j=1}^{n-1} q_{i,j} = \prod_{j=1}^{n-1}(1-\xi_{i,j})$. Let $B_1, \ldots, B_k$ any measurable partition of $\mathcal{X}$, then it follows that given $Y_{i,n}$, $\xi_{i,n}(\delta_{Y_{i,n}}(B_1), \ldots, \delta_{Y_{i,n}}(B_k))$ has Dirichlet distribution with parameter $(\delta_{Y_{i,n}}(B_1), \ldots, \delta_{Y_{i,n}}(B_k))$. Using the stick breaking construction of $q_{i,j}$, for $j = 1, \ldots, n$

$$\sum_{j=1}^{n} q_{i,j}(\delta_{Y_{i,j}}(B_1), \ldots, \delta_{Y_{i,j}}(B_k)) = \sum_{j=1}^{n-1} q_{i,j}(\delta_{Y_{i,j}}(B_1), \ldots, \delta_{Y_{i,j}}(B_k))$$

$$+ \left(1 - \sum_{j=1}^{n-1} q_{i,j}\right) [\xi_{i,n}(\delta_{Y_{i,n}}(B_1), \ldots, \delta_{Y_{i,n}}(B_k))]$$

and, since by construction $\sum_{j=1}^{n} q_{i,j} = 1$, it easily follows that given $Y_{i,1}, \ldots, Y_{i,n}$, $\sum_{j=1}^{n} q_{i,j}(\delta_{Y_{i,j}}(B_1), \ldots, \delta_{Y_{i,j}}(B_k))$ has Dirichlet distribution with updated parameter $(\alpha(B_1) + \sum_{j=1}^{n} \delta_{Y_{i,j}}(B_1), \ldots, \alpha(B_k) + \sum_{j=1}^{n} \delta_{Y_{i,j}}(B_k))$. This shows that the Dirichlet process with parameter α satisfies the distributional equation (2). The uniqueness of the solution follows by Lemma 3.3 in Sethuraman (1994). □

Theorem 2 *Let $\alpha \in \mathcal{M}_{\mathcal{X}}$ with total mass a, let $\{n_i, i \geq 1\}$ be a fixed integer value sequence and let P defined by (1). Then, for any measurable partition $B_1, \ldots, B_k$ of $\mathcal{X}$, $(P(B_1), \ldots, P(B_k))$ is distributed according to a Dirichlet distribution with parameter $(\alpha(B_1), \ldots, \alpha(B_k))$.*

Proof. From Lemma 1, the distributional equation

$$P = \theta_1 \sum_{j=1}^{n_1} q_{1,j}(\delta_{Y_{1,j}}(B_1), \ldots, \delta_{Y_{1,j}}(B_k)) + (1-\theta_1)P.$$

has unique solution the Dirichlet process with parameter α. Then, to prove the theorem, we can use arguments similar to those used in Lemma 1. In particular, if we define $\tilde{P} := (P(B_1), \ldots, P(B_k))$, then

$$\sum_{i=1}^{m} p_i \sum_{j=1}^{n_i} q_{i,j}(\delta_{Y_{i,j}}(B_1), \ldots, \delta_{Y_{i,j}}(B_k)) + \left(1 - \sum_{i=1}^{m} p_i\right) \tilde{P}$$

$$= \sum_{i=1}^{m-1} p_i \sum_{j=1}^{n_i} q_{i,j}(\delta_{Y_{i,j}}(B_1), \ldots, \delta_{Y_{i,j}}(B_k))$$

$$+ \left(1 - \sum_{i=1}^{m-1} p_i\right) \left(\theta_m \sum_{j=1}^{n_m} q_{m,j}(\delta_{Y_{m,j}}(B_1), \ldots, \delta_{Y_{m,j}}(B_k)) + (1-\theta_m)\tilde{P}\right)$$

and

$$\theta_m \sum_{j=1}^{n} q_{m,j}(\delta_{Y_{m,j}}(B_1), \ldots, \delta_{Y_{m,j}}(B_k)) + (1-\theta_m)\tilde{P}$$

has Dirichlet distribution with parameter $(\alpha(B_1), \ldots, \alpha(B_k))$. Then, it follows that

$$\sum_{i=1}^{m} p_i \sum_{j=1}^{n_i} q_{i,j}(\delta_{Y_{i,j}}(B_1), \ldots, \delta_{Y_{i,j}}(B_k)) + \left(1 - \sum_{j=1}^{m} p_i\right) \tilde{P}$$

has Dirichlet distribution with parameter $(\alpha(B_1), \ldots, \alpha(B_k))$, and the result follows by taking the limit for $m \to \infty$. □

The following remark underlines the connection between the constructive definition of the Dirichlet process that we have proposed and the one given in Sethuraman (1994).

Remark 3 *The measurable map (1) generalizes the measurable map used in Sethuraman (1994). In particular, if in (1) we set $n_i = 1$ for all $i \geq 1$, then*

$$P(\omega, B) = \sum_{i \geq 1} p_i(\omega) \sum_{j=1}^{1} q_{i,j}(\omega)\delta_{Y_{i,j}(\omega)}(B) = \sum_{i \geq 1} p_i(\omega)\delta_{Y_{i,1}(\omega)}(B) \qquad (3)$$

which corresponds to the measurable map used in Sethuraman (1994). Moreover, if we fix $n_i = n$ for all $i \geq 1$ then, as a consequence of Theorem 2 and of convergence results in Sethuraman and Tiwari (1982) if follows $P \to P^*$ weakly as $n \to \infty$, where P^* is a Dirichlet process with parameter α.

3 Discussion

The series representation provided by Sethuraman (1994) has been widely used in several areas of Bayesian nonparametrics. In particular, it has been used to simulate the Dirichlet process via deterministic and random approximations (see Ishwaran and Zarepour (2002), Muliere and Tardella (1998) and the references therein) and to define new approaches for computing functionals of the Dirichlet process (see Diaconis and Kemperman (1995), Guglielmi (2001)).

In this paper, we have introduced a more general constructive definition of the Dirichlet process. From the original constructive definition in Sethuraman (1994), for every $i \geq 1$, we replaced the random measures δ_{Y_i} involved in the series representation, by an appropriate random convex linear combination of dependent random measures $\delta_{Y_{i,1}}, \ldots, \delta_{Y_{i,n_i}}$, where $Y_{i,1}, \ldots, Y_{i,n_i}$ are from a Pólya sequence.

The new series representation has more flexibility due to the introduction of the integer value sequence $\{n_i, i \geq 1\}$. In particular, on the basis of the large number of applications of the series representation it seems natural and interesting to investigate the consequences of the new series representation in Bayesian nonparametrics.

Acknowledgements

The authors are grateful to the Associate Editor and to the referees for their valuable comments and suggestions.

References

Blackwell, D., 1973. Discretness of Ferguson selections. Ann. Stat. 1, 356–358.

Blackwell, D., MacQueen, J. B., 1973. Ferguson distributions via Pólya urn schemes. Ann. Stat. 1, 353–355.

Diaconis, P. and Kemperman, J.B.K., 1995. Some new tools for Dirichlet priors, in: Bernardo, J., Berger, J. O., Dawid, A. and Smith A. F. M. (Eds.), Bayesian Statistics, Oxford University Press, pp. 95-104.

Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. Ann. Stat. 1, 209–230.

Ferguson, T.S., 1974. Prior distributions on spaces of probability measures. Ann. Stat. 2, 615–629.

Guglielmi, A. and Tweedie, R., 2001. MCMC estimation of the law of the mean of a Dirichlet process. Bernoulli. 4, 573–592.

Ishwaran, H. and Zarepour, M., 2002. Exact and approximate sum-representation for the Dirichlet process. Canad. J. Stat. 30, 269–283.

James, L.F., 2002. Poisson Process Partition Calculus with applications to Exchangeable models and Bayesian Nonparametrics. MatharXive, arXiv:math/0205093v1.

James, L.F., 2005. Functionals of Dirichlet processes, the Cifarelli-Regazzini identity and beta-gamma processes. Ann. Stat. 33, 647–660.

Jambunathan, M.V., 1954. Some properties of beta and gamma distributions. Ann. Math. Stat. 25, 401–405.

Muliere, P. and Tardella, L., 1998. Approximating distributions of random functional of Ferguson-Dirichlet priors. Canad. J. Stat. 26, 283–297.

Sethuraman, J. and Tiwari, T.C., 1982. Convergence of Dirichlet measures and the interpretation of their parameter, in: Gupta, S. S. and Berger, J. O. (Eds.), Statistical Decision Theory and Related Topics III, pp. 305-315.

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