Bayesian Analysis (2007) **2**, Number 1, pp. 213–220

On the Multimodality of Random Probability Measures

George Kokolakis* and George Kouvaras†

Abstract. Nonparametric methods for density estimation are examined here. Within a Bayesian setting the construction of an absolutely continuous random probability measure is often required for nonparametric statistical analysis. To achieve this we propose a "partial convexification" procedure of a process, such as the Dirichlet, resulting in a multimodal distribution function with a finite expected number of modes. In agreement with convexity theory results, it is shown that the derived random probability measure admits a density with respect to Lebesgue measure.

Keywords: convexity, Dirichlet process, multimodal distribution functions, Polya trees, random probability measures.

1 Introduction

In nonparametric Bayesian analysis, the Dirichlet process (Ferguson 1973, 1974) is usually chosen to represent nonparametric prior information on a space of probability distributions. Among the most popular ways to derive random Dirichlet probability measures is the Urn representation scheme (Blackwell and MacQueen 1973), and the stick-breaking representation (Sethuraman 1994). The major drawback of a Dirichlet process is that it selects discrete distributions with probability one. To overcome this problem, alternative methods such as mixtures of Dirichlet processes (Antoniak 1974; Escobar and West 1995; Walker et al. 1999) and Polya trees (Lavine 1992, 1994) have been proposed.

In this paper, we present a Bayesian procedure with prior distributions on the space of probability measures on the real line that have finite expected number of modes. Section 2 describes the construction of random Dirichlet probability measures using a variant of Polya trees (Kokolakis 1983). In section 3, we review some concepts on unimodality in $\mathbb{R}$ and provide the theoretical background on "partial convexification" which is needed to implement our methodology. In section 4, we present some illustrative results and possible extensions are discussed.

*School of Applied Mathematics and Physical Sciences, National Technical University of Athens, Greece, mailto:kokolakis@math.ntua.gr

†School of Applied Mathematics and Physical Sciences, National Technical University of Athens, Greece, gkouv@math.ntua.gr

2 Construction of random Dirichlet probability measures

In this section, we construct random probability measures which are Dirichlet processes on some measurable space $(\mathcal{X}, \mathcal{F})$ with $\beta(\cdot)$ a σ–additive nonnull finite measure on $(\mathcal{X}, \mathcal{F})$. We assume that the σ–field $\mathcal{F}$ is generated by a sequence of nested partitions of $\mathcal{X}$, *i.e.*,

$$\mathcal{X} = B_0 \cup B_1, \qquad\qquad with \quad B_0 \cap B_1 = \emptyset$$

$$B_0 = B_{00} \cup B_{01}, \quad B_1 = B_{10} \cup B_{11}, \quad with \quad B_{00} \cap B_{01} = B_{10} \cap B_{11} = \emptyset$$

and in general

$$B_{\boldsymbol{e}} = B_{\boldsymbol{e}0} \cup B_{\boldsymbol{e}1}, \qquad\qquad with \quad B_{\boldsymbol{e}0} \cap B_{\boldsymbol{e}1} = \emptyset,$$

where $\boldsymbol{e}$ is a d-dimensional binary string, *i.e.*, $\boldsymbol{e} = (e_1, ..., e_d) \in \{0,1\}^d,\ d = 1, 2, ...$

Let $\theta_{\boldsymbol{e}} = Pr[B_{\boldsymbol{e}}]$, the random cell probabilities of the 2^d sets $B_{\boldsymbol{e}}, \boldsymbol{e} \in \{0,1\}^d,\ d = 1, 2, ...,$ and $u_{\boldsymbol{e}0} = Pr[B_{\boldsymbol{e}0} \,|\, B_{\boldsymbol{e}}]$ and $u_{\boldsymbol{e}1} = Pr[B_{\boldsymbol{e}1} \,|\, B_{\boldsymbol{e}}]$, the conditional (random) cell probabilities satisfying the relation $u_{\boldsymbol{e}0} + u_{\boldsymbol{e}1} = 1$ for all $\boldsymbol{e} \in \{0,1\}^{d-1}$, with $B_{\emptyset} = \mathcal{X}$. To generate a random cell probability $\theta_{\boldsymbol{e}}$ of a tree with d levels, we choose the path from the root of the tree to the leaf and form the corresponding product of conditional cell probabilities u's. A useful result here is the following theorem, special versions of which can be found in Ferguson (1974), in Kokolakis (1983) and in Kokolakis and Dellaportas (1996).

Theorem 1. *Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\boldsymbol{e}0}, \boldsymbol{\theta}_{\boldsymbol{e}1})$, $\boldsymbol{e} \in \{0,1\}^{d-1}$, $(d = 1, 2, ...)$, be the random cell probability vector at the d level and $u_{\boldsymbol{e}0}$ be the corresponding conditional cell probabilities up to the dth level. Suppose that $u_{\boldsymbol{e}0}$ are independent $Beta(\beta_{\boldsymbol{e}0}, \beta_{\boldsymbol{e}1})$ distributed with $\boldsymbol{e} \in \{0,1\}^{d-1}$.*
A necessary and sufficient condition that $\boldsymbol{\theta}$ be Dirichlet distributed with parameters $(\beta_{\boldsymbol{e}0}, \beta_{\boldsymbol{e}1})$, $\boldsymbol{e} \in \{0,1\}^{d-1}$ is that the following conditions be satisfied:

$$\beta_{\boldsymbol{e}0} + \beta_{\boldsymbol{e}1} = \beta_{\boldsymbol{e}}, \quad for \quad all \quad \boldsymbol{e} \in \{0,1\}^{d-1}, \quad d = 1, 2....$$

Remark 1. The above theorem allows us to construct a random Dirichlet process with parametric finite measure $\beta(\cdot)$ by taking $\beta_{\boldsymbol{e}} = \beta(B_{\boldsymbol{e}})$ for all $\boldsymbol{e} \in \{0,1\}^d, d = 1, 2, ...$

3 Partial convexification

Much of the work with unimodal distribution functions is based on a representation theorem due to Khincin and Shepp (see Feller 1971, p.158). In particular, when $\mathcal{X} = \mathbb{R}$ we have:

Theorem 2. *The c.d.f. F is unimodal, with mode at zero,* i.e., *F is convex on the negative real line and concave on the positive, if and only if it is of the form:*

$$F(t) = \int_0^1 G(t/u)\, du, \qquad t \in \mathbb{R}.$$

This means that F is the distribution of the product of two independent random variables U and Y, with U uniformly distributed on $(0,1)$ and Y having an arbitrary c.d.f. G.

It is easy to realize that Theorem 2 takes the following equivalent form:

Corollary 2.1. *The c.d.f. F is convex on the negative real line and concave on the positive, if and only if there exists a distribution function G on $\mathbb{R}$ such that F admits the representation:*

$$F(x) = G(0-) + \int_{\mathbb{R}} H_y(x) G(\,dy), \qquad x \in \mathbb{R}, \tag{1}$$

where $H_y(\cdot)$ is, for $y \neq 0$,

$$H_y(x) = \begin{cases} 0, & \dfrac{x}{y} \le 0, \\[2ex] \dfrac{x}{|y|}, & 0 < \dfrac{x}{y} < 1, \\[2ex] \dfrac{y}{|y|}, & \dfrac{x}{y} \ge 1, \end{cases}$$

and for $y = 0$, $H_0(\cdot)$ is degenerate at zero.

From the above representation of F we notice that when $y = 0$, we have $p \equiv Pr[X = 0] = Pr[Y = 0]$. When $y \neq 0$, $H_y(x)$ is a bounded function of x with bounded left and right derivatives. In addition, the derivative of $H_y(x)$ is bounded for all $x \in \mathbb{R} \setminus \{0, y\}$. Specifically, with fixed $y \neq 0$ and $x \neq 0, y$, the derivative of $H_y(x)$ is

$$h_y(x) \equiv DH_y(x) = \begin{cases} \dfrac{1}{|y|}, & 0 < \dfrac{x}{y} < 1, \\[2ex] 0, & \textit{otherwise}. \end{cases}$$

Applying the bounded convergence theorem we conclude that the c.d.f. F is differentiable a.e. in $\mathbb{R}$, and its derivative, wherever it exists, is

$$f(x) \equiv DF(x) = \int_{\mathbb{R}\setminus\{0\}} DH_y(x) G(\,dy) = \begin{cases} \displaystyle\int_{(-\infty, x)} \frac{1}{|y|} G(\,dy), & x < 0, \\[3ex] \displaystyle\int_{(x, +\infty)} \frac{1}{|y|} G(\,dy), & x > 0. \end{cases} \tag{2}$$

In this case $\int f(x)\,dx = 1-p$. When $p=0$, f is a density function.

Introducing the above result into (1) we get the following relation:

$$G(x-) + Pr[Y=x]\cdot I_{(x\geq 0)} = F(x) - xf(x), \qquad x \in \mathbb{R}. \tag{3}$$

When $\mathcal{X} = \mathbb{R}_+$, an immediate consequence of the above results is the following:

Corollary 2.2. *The c.d.f. F is concave on the positive real line if and only if there exists a distribution function G on $[0,+\infty)$ such that F admits the representation:*

$$F(x) = \int_{[0,+\infty)} H_y(x) G(\,dy), \qquad x \in \mathbb{R}_+, \tag{4}$$

and then

$$G(x) = F(x) - xf(x), \qquad x \in \mathbb{R}_+. \tag{5}$$

Results similar to (2), (4) and (5) can be found in Brunner (1992) and Hansen and Lauritzen (2002).

Remark 2. As a consequence of the Corollary 2.2, we always get a c.d.f. F on the positive real line with single mode at zero, no matter what the distribution G, we start with, is. Thus, the above procedure produces a very restrictive class of distributions.

"Partial convexification" of a c.d.f. G that produces a c.d.f. F with a finite number of modes, rather than a single mode, would be preferable. In our Bayesian model specification, we work with random c.d.f.'s produced by a Dirichlet process. Since a random Dirichlet probability distribution is with probability one discrete, we apply an alternative procedure. Instead of the Un(0,1) distribution we consider the Un(α, 1) distribution with $0<\alpha<1$. The parameter α can be fixed, or random with a prior distribution $p(\alpha)$, on the interval (0, 1).

Let us assume the following:

1. U is Uniformly distributed over the interval (α, 1) with α fixed on the interval (0, 1), *i.e.*,

$$U \sim \text{Un}(\alpha, 1), \qquad 0<\alpha<1.$$

2. Y is distributed according to G which is a Dirichlet process Di($\beta(\cdot)$) and $\beta(\cdot)$ a σ–additive nonnull finite measure on $(\mathbb{R}_+, \mathcal{B}_+)$, and

3. Y and U are independent.

Then

$$F(x) = \int_{(0,+\infty)} H_y(x) G(\,dy), \quad x \in \mathbb{R}_+, \tag{6}$$

where H_y stands for the Uniform distribution function over the interval $(\alpha y, y)$, *i.e.*,

$$H_y(x) = \begin{cases} 0, & x \leq \alpha y, \\ \dfrac{x - \alpha y}{(1-\alpha) y}, & \alpha y < x < y, \\ 1, & x \geq y, \end{cases} \tag{7}$$

and the c.d.f. F of the product $X = UY$ admits a.e. a derivative f. Specifically, the following equation is satisfied:

$$F(x) = G(x) + x f(x) - \frac{\alpha}{1-\alpha}\{G(\frac{x}{\alpha}) - G(x)\}, \qquad x \in \mathbb{R}_+. \tag{8}$$

It is interesting to notice that when Y has a discrete distribution $\{q_k = Pr[Y = k], \quad k = 1, 2, ...\}$, then

$$F(x) = \sum_{k=1}^{[x]} q_k + \frac{x}{1-\alpha} \sum_{k=[x]+1}^{[\frac{x}{\alpha}]} (\frac{1}{k} - \frac{\alpha}{x}) q_k, \qquad x \in \mathbb{R}_+, \tag{9}$$

where $[\cdot]$ stands for the integer part.

According to (6) we obtain a prior distribution on the subspace of multimodal c.d.f.'s. The expected number of modes of F increases from one, when $\alpha = 0$, to infinity, when $\alpha = 1$, having a finite number of modes when $0 < \alpha < 1$. This means that when $0 < \alpha < 1$, the c.d.f. $F(x)$ alternates between local concavities and local convexities, *i.e.*, a "partial convexification" of F is produced.

4 Application

For demonstration purposes we have simulated ten datasets from a mixture of two normal distributions, specifically $N(15, 3^2)$ and $N(30, 4^2)$, with weights $w_1/w_2 = 2/3$. The sample sizes have all been taken equal to 250. Dirichlet processes have been produced with parameter $50 \times$ Ga, where Ga stands for the Gamma distribution function with mean $\mu = 25$ and standard deviation $\sigma = 5$. The "partial convexification" procedure has been applied with the parameter α= 0.80.

In Figure 1, ten random draws from a Dirichlet process are presented (dotted lines) together with their locally convexified versions (continuous lines). In Figure 2, the posterior Dirichlet probability measures based on the datasets produced are presented with dotted lines. Their locally convexified versions are presented with continuous lines. We may notice the convexification procedure results in some overestimation of the posterior distributions. This could be avoided if, before updating the Dirichlet processes, we had solved the equation (8) with respect to G using a continuous version of the empirical distribution function in the place of F. This task, together with possible generalizations, such as dealing with higher dimensional sample spaces, representing correlation structures and using efficient algorithms, as in MacEachern (1994), is the object of future research.

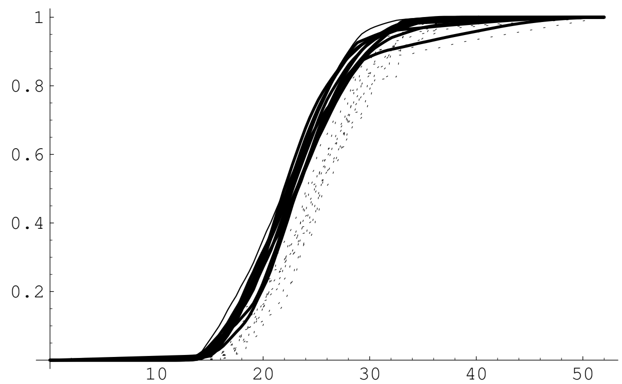

Figure 1: Locally convexified random Dirichlet probability distributions (continuous lines) and preconvexified random Dirichlet probability distributions (dotted lines).

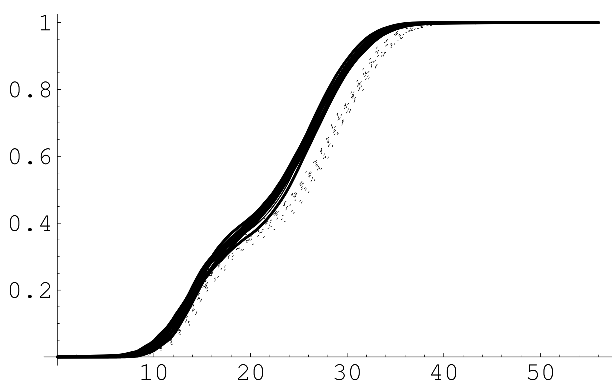

Figure 2: Locally convexified posterior Dirichlet probability distributions (continuous lines) and posterior Dirichlet probability distributions (dotted lines).

References

Antoniak, C. (1974). "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems." *Annals of Statistics*, 2: 1152–1174. 213

Blackwell, D. and MacQueen, J. (1973). "Ferguson distributions via Polya schemes." *Annals of Statistics*, 1: 353–355. 213

Brunner, L. (1992). "Bayesian nonparametric methods for data from a unimodal density." *Statistics and Probability Letters*, 14: 195–199. 216

Escobar, M. and West, M. (1995). "Bayesian density estimation and inference using mixtures." *JASA*, 90: 577–588. 213

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, volume 2. New York: Wiley, 2nd edition. 214

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *Annals of Statistics*, 1: 209–230. 213

— (1974). "Prior distributions on spaces of probability measures." *Annals of Statistics*, 2: 615–629. 213, 214

Hansen, M. and Lauritzen, S. (2002). "Nonparametric Bayes inference for concave distribution functions." *Statistica Neerlandica*, 56: 110–127. 216

Kokolakis, G. (1983). "A new look at the problem of classification with binary data." *Statistician*, 32: 144–152. 213, 214

Kokolakis, G. and Dellaportas, P. (1996). "Hierarchical modelling for classifying binary data." In *Bayesian Statistics 5*, 647–652. London: Oxford University Press. 214

Lavine, M. (1992). "Some aspects of Polya tree distributions for statistical modelling." *Annals of Statistics*, 20: 1222–1235. 213

— (1994). "More aspects of Polya tree distributions for statistical modelling." *Annals of Statistics*, 22: 1161–1176. 213

MacEachern, S. (1994). "Estimating normal means with a conjugate style Dirichlet process prior." *Communications in Statistics. Simulation and Computation*, 23: 727–741. 218

Sethuraman, J. (1994). "A constructive definition of Dirichlet prior." *Statistica Sinica*, 4: 639–650. 213

Walker, S., Damien, P., Laud, P., and Smith, A. (1999). "Bayesian nonparametric inference for random distributions and related functions (with discussion)." *Journal of the Royal Statistical Society - Series B*, 61: 485–528. 213