Random Multivariate Multimodal Distributions

George Kouvaras\textsuperscript{1} and George Kokolakis\textsuperscript{2}

\textsuperscript{1} National Technical University of Athens
Department of Mathematics
Zografou Campus
15780 Athens, Greece
(e-mail: gkouv@math.ntua.gr)

\textsuperscript{2} National Technical University of Athens
Department of Mathematics
Zografou Campus
15780 Athens, Greece
(e-mail: Kokolakis@math.ntua.gr)

Abstract. Bayesian nonparametric inference for unimodal and multimodal random probability measures on a finite dimensional Euclidean space is examined. After a short discussion on several concepts of multivariate unimodality, we introduce and study a new class of nonparametric prior distributions on the subspace of random multivariate multimodal distributions. This class in a way generalizes the very restrictive class of random unimodal distributions. A flexible constructional approach is developed using a variant of Khinchin’s representation theorem for unimodal distributions.

Keywords: Convexity, Dirichlet process, Unimodality-Multimodality, Polya trees, Random probability measures.

1 Introduction

Much of nonparametric Bayesian inference has proceeded by modelling the unknown cumulative distribution function (c.d.f.) as a stochastic process. In a fundamental paper, [Ferguson, 1973], a random process, called the Dirichlet process, was defined as a distribution on \((\mathcal{P}, \mathcal{S})\), where \(\mathcal{P}\) is the collection of all probability measures on a measurable space \((\mathcal{X}, \mathcal{F})\), endowed with a \(\sigma\)-algebra \(\mathcal{S}\). The major drawback of a Dirichlet process is that it selects discrete distributions with probability one [Ferguson, 1973] and [Blackwell, 1973].

Several different classes of nonparametric priors, which all contain the Dirichlet process as a particular case, have been proposed. It seems worth mentioning, among others, the mixture of Dirichlet processes [Antoniak, 1974], which is a Dirichlet process where the base measure is itself random and the mixture of Dirichlet process prior [Lo, 1984], which is a convolution of a Dirichlet process with an appropriate kernel. After the work of previous authors, the study of absolutely continuous random probability measures has

In this paper, we present a Bayesian nonparametric inference for unimodal and multimodal random probability measures on a finite dimensional Euclidean space that have finite expected number of modes. As a consequence, we get a random probability measure that admits a derivative almost everywhere in $\mathbb{R}^d$. The paper is organized as follows. Section 2 has the essential theoretical background on multivariate unimodality to implement our methodology. In section 3, a detailed description of partial convexification procedure is provided. Random bivariate multimodal probability measures are constructed and possible modifications and extensions are discussed in section 4.

2 Univariate and multivariate unimodality

An important property of a distribution is unimodality. A univariate c.d.f. $F$ is said to be unimodal with mode (or vertex) at $m$, if $F$ is convex on $(-\infty, m)$ and concave on $(m, \infty)$. We make use of unimodality to get absolutely continuous distribution functions.

2.1 Unimodality on $\mathbb{R}$

For univariate distributions there is a well known representation theorem due to Khinchin (see [Feller, 1971], p.158) that refers to the classical univariate unimodality.

**Theorem 1.** A real valued random variable $X$ has a unimodal density at 0 if and only if it is a product of two independent random variables $U$ and $Y$, with $U$ uniformly distributed on $(0, 1)$ and $Y$ having an arbitrary distribution.

This can be expressed in the following equivalent form cf. [Shepp, 1962], [Brunner, 1992] and [Kokolakis and Kouvaras, 2007].

**Theorem 2.** 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:

$$F(x) = G(x) + xf(x),$$  \hspace{1cm} (1)

for all $x$ points of continuity of $G$. 

2.2 Unimodality on \( \mathbb{R}^d \)

For multivariate distributions, however, there are several different ways that unimodality is defined. Among the main types of multivariate unimodality there are the following: the beta unimodality, which is generated by the Beta distribution, \( \text{Beta}(\kappa, \nu) \), instead of Uniform distribution on the interval \((0, 1)\), and contains the classical univariate unimodality and some of the existing multivariate notions of unimodality as special cases (star unimodality, \( \nu \)-unimodality), the linear unimodality, which is characterized by the unimodality of the distribution of any linear combination of the components of a random vector and the strong unimodality, which is defined as a convolution of unimodal distributions. An extended study of different types of unimodality and their useful consequences can be found in [Dharmadhikari and Kumar, 1988] and [Bertin et al., 1997].

In what follows we focus our attention on the Khinchin’s classical unimodality extended to the multivariate case. For the sake of simplicity we restrict ourselves to the bivariate case. Results for higher dimensions can be easily derived. According to the classical unimodality [Shepp, 1962] we have:

**Theorem 3.** The c.d.f. \( F \) is unimodal at 0 if and only if there is a random vector \((X_1, X_2)\) with c.d.f. \( F \), such that

\[
(X_1, X_2) = (Y_1 U_1, Y_2 U_2),
\]

where \((Y_1, Y_2)\) and \((U_1, U_2)\) are independent random vectors, \((U_1, U_2)\) is uniformly distributed on the unit square and \((Y_1, Y_2)\) having an arbitrary c.d.f. \( G \).

Equivalent to Theorem 3 is the following.

**Theorem 4.** The c.d.f. \( F \) is unimodal at 0 if and only if for all \((x_1, x_2)\) points of continuity of \( G \),

\[
F(x_1, x_2) = G(x_1, x_2) + x_1 \frac{\partial F(x_1, x_2)}{\partial x_1} + x_2 \frac{\partial F(x_1, x_2)}{\partial x_2} - x_1 x_2 \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2},
\]

where \( G \) is an arbitrary c.d.f.

According to the above procedure, i.e. by the component wise multiplication of two independent random vectors \((Y_1, Y_2)\) and \((U_1, U_2)\), where the latter is uniformly distributed on the unit square, we always get a c.d.f. \( F \) with a single mode at zero, no matter what the distribution \( G \), we start with, is. To overcome the limitation of getting always a single mode at zero we propose the following “partial convexification” procedure.

3 Univariate and multivariate partial convexification

Partial convexification procedure of a c.d.f. \( G \) relies on using \( U(\alpha, 1) \) distributions, with \( 0 < \alpha < 1 \), instead of \( U(0, 1) \) [Kokolakis and Kouvaras, 2007].
The parameter \( \alpha \) can be fixed, or random with a prior distribution \( p(\alpha) \), on the interval \((0, 1)\). According to this, 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.

**Definition 1.** The \( d \)-variate c.d.f. \( F \) is called partially convexified if there exists a random vector \( X = (X_1, \ldots, X_d) \) with c.d.f. \( F \), such that

\[
(X_1, \ldots, X_d) = (Y_1 U_1, \ldots Y_d U_d),
\]

where \( Y = (Y_1, \ldots, Y_d) \) and \( U = (U_1, \ldots, U_d) \) are independent vectors, \( U \) is uniformly distributed on the rectangle \((\alpha_1, 1) \times \ldots \times (\alpha_d, 1)\) and \( Y \) having an arbitrary \( d \)-variate c.d.f. \( G \).

### 3.1 Partial convexification on \( \mathbb{R} \)

Using the partial convexification procedure, Theorem 2 can be expressed [Kokolakis and Kouvaras, 2007] in the following form.

**Theorem 5.** The c.d.f. \( F \) is partially convexified if there exists a distribution function \( G \) on \( \mathbb{R} \) such that \( F \) admits the representation:

\[
F(x) = x f(x) + \frac{1}{1 - \alpha} [G(x) - \alpha G(\frac{x}{\alpha})],
\]

for all \( x, x/\alpha \), points of continuity of \( G \).

### 3.2 Partial convexification on \( \mathbb{R}^2 \)

A generalization to the bivariate case using Uniform distributions \( U(\alpha_i, 1) \), \( i = 1, 2 \), with parameters \( \alpha_i \) fixed in the interval \((0, 1)\), is as follows.

**Theorem 6.** The c.d.f. \( F \) is partially convexified if there exists a distribution function \( G \) on \( \mathbb{R}^2 \) such that \( F \) admits the representation:

\[
F(x_1, x_2) = x_1 F_{x_1}(x_1, x_2) + x_2 F_{x_2}(x_1, x_2) - x_1 x_2 f(x_1, x_2) + P(x_1, x_2)
\]

where

\[
P(x_1, x_2) = \frac{1}{(1 - \alpha_1)(1 - \alpha_2)} \times \{G(x_1, x_2) - \alpha_1 G(\frac{x_1}{\alpha_1}, x_2) - \alpha_2 G(x_1, \frac{x_2}{\alpha_2})
\]

\[
+ \alpha_1 \alpha_2 G(\frac{x_1}{\alpha_1}, \frac{x_2}{\alpha_2})]\},
\]

for all \((x_1, x_2), (x_1/\alpha_1, x_2), (x_1, x_2/\alpha_2)\) and \((x_1/\alpha_1, x_2/\alpha_2)\) points of continuity of \( G \), with subscripts in (6) denoting partial derivatives.
Proof. We have:
\[
 F(x_1, x_2) = P[X_1 \leq x_1, X_2 \leq x_2] = P[U_1 Y_1 \leq x_1, U_2 Y_2 \leq x_2] \\
 = \int_{\mathbb{R}^2} P[U_1 Y_1 \leq x_1, U_2 Y_2 \leq x_2 \mid Y_1 = y_1, Y_2 = y_2] G(dy_1, dy_2) \\
 = \int_{\mathbb{R}^2} H_{y_1}(x_1) H_{y_2}(x_2) G(dy_1, dy_2), \tag{8}
\]
where the functions \( H_{y_i}(\cdot) \) \((i = 1, 2)\) for \( y_i \neq 0 \) are as follows:

For \( x_i \geq 0 \),
\[
 H_{y_i}(x_i) = \begin{cases} 
 0, & x_i \leq \alpha_i y_i, \\
 \frac{x_i - \alpha_i y_i}{(1 - \alpha_i)y_i}, & \alpha_i y_i < x_i < y_i, \\
 1, & x_i \geq y_i, 
\end{cases} \quad (i = 1, 2)
\]

for \( x_i < 0 \)
\[
 H_{y_i}(x_i) = \begin{cases} 
 0, & x_i \leq y_i, \\
 \frac{y_i - x_i}{(1 - \alpha_i)y_i}, & y_i < x_i < \alpha_i y_i, \\
 1, & x_i \geq \alpha_i y_i, 
\end{cases} \quad (i = 1, 2)
\]
and \( H_0(\cdot) \) is degenerate at zero.

In the sequel we will only refer to the first quadrant \( I_1 = \{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0\} \). Thus in \( I_1 \), the expression (8) can be written as follows:

\[
 F(x_1, x_2) = \int_{(-\infty, x_1] \times (-\infty, x_2]} G(dy_1, dy_2) \\
 + \int_{(-\infty, x_1] \times (x_2, \infty]} \frac{x_2 - \alpha_2 y_2}{(1 - \alpha_2)y_2} G(dy_1, dy_2) \\
 + \int_{(x_1, \infty) \times (-\infty, x_2]} \frac{x_1 - \alpha_1 y_1}{(1 - \alpha_1)y_1} G(dy_1, dy_2) \\
 + \int_{(x_1, \infty) \times (x_2, \infty]} \frac{x_1 - \alpha_1 y_1}{(1 - \alpha_1)y_1} \frac{x_2 - \alpha_2 y_2}{(1 - \alpha_2)y_2} G(dy_1, dy_2). \tag{9}
\]

With fixed \( y_i \neq 0 \), the functions \( H_{y_i}(\cdot) \) \((i = 1, 2)\) are bounded with bounded left and right derivatives a.e. in \( I_1 \). Applying the bounded convergence
theorem we conclude that the c.d.f. $F$ is differentiable a.e. in $I_1$, with respect to the Lebesgue measure, and their 1st and 2nd order mixed derivatives, wherever they exist, are as follows:

$$F_{x_1}(x_1, x_2) = \frac{\partial F(x_1, x_2)}{\partial x_1}$$

$$= \int_{(x_1, \frac{x_1}{\alpha_1})} \frac{1}{(1-\alpha_1)y_1} G(dy_1, dy_2)$$

$$+ \int_{(x_1, \frac{x_1}{\alpha_1})} \frac{1}{(1-\alpha_1)y_1} \frac{x_2 - \alpha_2 y_2}{(1-\alpha_2)y_2} G(dy_1, dy_2),$$

$$F_{x_2}(x_1, x_2) = \frac{\partial F(x_1, x_2)}{\partial x_2}$$

$$= \int_{(0, x_1) \times (x_2, \frac{x_2}{\alpha_2})} \frac{1}{(1-\alpha_2)y_2} G(dy_1, dy_2)$$

$$+ \int_{(x_1, \frac{x_1}{\alpha_1}) \times (x_2, \frac{x_2}{\alpha_2})} \frac{x_1 - \alpha_1 y_1}{(1-\alpha_1)y_1} \frac{1}{(1-\alpha_2)y_2} G(dy_1, dy_2),$$

and

$$f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2}$$

$$= \int_{(x_1, \frac{x_1}{\alpha_1}) \times (x_2, \frac{x_2}{\alpha_2})} \frac{1}{(1-\alpha_1)y_1(1-\alpha_2)y_2} G(dy_1, dy_2).$$

Introducing the above results in (9) we get finally:

$$F(x_1, x_2) = x_1 F_{x_1}(x_1, x_2) + x_2 F_{x_2}(x_1, x_2) - x_1 x_2 f(x_1, x_2) + P(x_1, x_2), \quad (10)$$

where

$$P(x_1, x_2) = \frac{1}{(1-\alpha_1)(1-\alpha_2)} G(x_1, x_2) - \frac{\alpha_1}{(1-\alpha_1)(1-\alpha_2)} G\left(\frac{x_1}{\alpha_1}, -x_2\right)$$

$$- \frac{\alpha_2}{(1-\alpha_1)(1-\alpha_2)} G\left(x_1, \frac{x_2}{\alpha_2}\right) + \frac{\alpha_1 \alpha_2}{(1-\alpha_1)(1-\alpha_2)} G\left(\frac{x_1}{\alpha_1}, -\frac{x_2}{\alpha_2}\right).$$

We may notice that when the points $(x_1, x_2)$, $(x_1/\alpha_1, x_2)$, $(x_1, x_2/\alpha_2)$ and $(x_1/\alpha_1, x_2/\alpha_2)$ are all continuity points of $G$, then $P(x_1, x_2)$ takes the form:

$$P(x_1, x_2) = \frac{1}{(1-\alpha_1)(1-\alpha_2)} \times \left\{G(x_1, x_2) - \alpha_1 G\left(\frac{x_1}{\alpha_1}, x_2\right) - \alpha_2 G\left(x_1, \frac{x_2}{\alpha_2}\right)

+ \alpha_1 \alpha_2 G\left(\frac{x_1}{\alpha_1}, \frac{x_2}{\alpha_2}\right)\right\}. \quad (11)$$

Expressions (10) and (11) hold to the others quadrants as well. When the origin $(0,0)$ is a continuity point of $G$ then $f$ is a density.
4 Random bivariate multimodal probability measures

In our Bayesian model specification we assume the following:

- \((Y_1, Y_2) \sim G\), where \(G\) is a random c.d.f. produced by a Dirichlet process, using the tree structure presented in [Kokolakis and Kouvaras, 2007].
- \((U_1, U_2)\) is uniformly distributed on the rectangle \((\alpha_1, 1) \times (\alpha_2, 1)\) with \(\alpha_i\) fixed on the interval \((0, 1)\).
- \((Y_1, Y_2)\) and \((U_1, U_2)\) are independent.

By definition, the c.d.f. \(F\) of the random vector \((X_1, X_2) = (Y_1 U_1, Y_2 U_2)\) will be random partially convexified c.d.f. and thus we take a prior distribution on the subspace of bivariate multimodal c.d.f.’s. Various generalizations of the present result for higher dimensions and correlation structures are possible. Such models can be useful alternatives to the standard Bayesian mixtures models for multivariate multimodal distributions.

References


