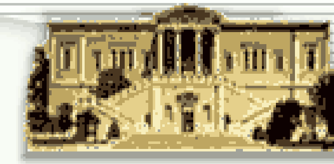


Convexification and Multimodality of Random Probability Measures

George Kokolakis† & George Kouvaras*

Department of Applied Mathematics
National Technical University of Athens, Greece

Isaac Newton Institute, Cambridge, UK , August 2007

† Email: kokolakis@math.ntua.gr   Web site: http://www.math.ntua.gr/~kokolakis/
∗ Email: gkouv@math.ntua.gr

Outline of the talk

- Our Goal
- Random Probability Measures
- Univariate Unimodality
- Multivariate Unimodality
- Multimodality - Partial Convexification
- Application - Bivariate Case
- References

Our Goal

Construction of unimodal and multimodal random probability measures on a finite dimensional Euclidean space by developing a flexible approach based on Bayesian nonparametric models and Convexity Theory. As a result, we get random probability measures that admit derivatives almost everywhere in $\mathbb{R}^d$.

Random Probability Measures

- The mostly used random probability measures are those derived from the **Dirichlet Process** (DP) [Ferguson 1973, 1974]. The parameter measure α of a DP, when normalized, is a prior guess about the unknown sampling distribution function. Thus, the DP's represent nonparametric prior information on the space of probability measures corresponding to the unknown sampling distribution functions.

- The major drawback of the DP's is that they select discrete distributions with probability one. Since the posterior distribution of a DP is also a DP with parameter $\alpha + \sum_{i=1}^{n} \delta_{x_i}$, we can never get an absolutely continuous one, no matter what the sampling distribution and the sample size are.

Random Probability Measures

- Among the basic classes of nonparametric priors, containing the DP as a particular case, are the following:
  - **Mixtures of DP's** [Antoniak, 1974], by taking mixtures of the base measure.
  - **Mixtures of DP's** [Lo, 1984], by taking convolutions of a DP with an appropriate kernel.
  - **Polya Trees** [Lavine, 1992], [Lavine, 1994] and [Kokolakis and Dellaportas, 1996].
  - **Mixtures for convex sets** [Hoff, 2003].
- Another approach, based on the convexification of a DP, produces multimodal random probability measures admitting densities:
  - **Partial Convexification** [Kokolakis and Kouvaras, 2007].

Univariate Unimodality

Definition

A univariate c.d.f. F is said to be **unimodal** about zero, if F is convex on the negative real line and concave on the positive.

Some consequences of the above definition are:

- If F is unimodal about zero, then apart from a possible mass at zero, F is absolutely continuous.
- If F is unimodal about zero, then the left and right derivatives of F exist everywhere except possibly at zero and are equal to each other almost everywhere.

Univariate Unimodality

There is a well known representation theorem due to Khinchin (see [Feller, 1971, p.158]) useful to achieve univariate unimodality.

Theorem 1 A real valued random variable X has a unimodal distribution with mode at zero 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] and [Brunner, 1992].

Theorem 2 The c.d.f. F is unimodal if and only if there exists a distribution function G such that F satisfies the equation:
$F(x) = G(x) + xf(x)$, for all $x \in \mathbb{R}$, points of continuity of G.

Univariate Unimodality

Theorems 1 and 2 can be considered as particular cases of the Choquet-type representation theorem according which every unimodal about 0 c.d.f. F is a generalized mixture of the uniform distributions U_y $(0 \le |y| < \infty)$ and vice versa. Specifically we have the following:

Theorem 3 The c.d.f. F is unimodal about 0 if and only if there exists a probability measure G on the Borel σ-field $\mathcal{B} = \mathcal{B}(\mathbb{R})$ such that

$$F(A) = \int_{\mathbb{R}} U_y(A) G(dy), \quad \forall\, A \in \mathcal{B}.$$

This means that every unimodal F belongs to the closed convex hull of the set of $\{U_y, 0 \le |y| < \infty\}$.

Multivariate Unimodality

For multivariate distributions, there are several different types of unimodality. The notion of convexity plays an important role in all these definitions. An extended study of different types of unimodality and their useful consequences can be found in [Dharmadhikari and Joag-dev, 1988] and in [Bertin et al., 1997]. Among the main definitions of unimodality in $\mathbb{R}^d$ we have the following ones:

- **Star Unimodality** about **0**. The c.d.f. F is called *star unimodal* about **0** if it belongs to the closed convex hull of the set of all uniform distributions on sets S in $\mathbb{R}^d$ which are star-shaped about **0**. (A set S in $\mathbb{R}^d$ is called *star-shaped* about **0**, if for any $\mathbf{x} \in S$ the line segment $[\mathbf{0}, \mathbf{x}]$ is contained in S).

Multivariate Unimodality

- **Block Unimodality** about **0**. The c.d.f. F is called *block unimodal* about **0** if it belongs to the closed convex hull of the set of all uniform distributions on rectangles having opposite vertices the points **0** and **x** for all **x** $\epsilon\ \mathbb{R}^d$.
- **Central Convex Unimodal** about **0**. The c.d.f. F is called *central convex unimodal* about **0** if it belongs to the closed convex hull of the set of all uniform distributions on centrally symmetric convex bodies in Rd.
- **Logconcave distributions.** A probability measure P on **B**d = **B** (Rd) is called *logconcave* if for all nonempty sets A, B in **B**d and for all θ ϵ (0,1), we have: P[θA +(1-θ)B] ≥ {P[A]}θ {P[B]}1-θ. (This implies that both the c.d.f. F and the p.d.f. f are logconcave).

Multivariate Unimodality

Other types of multivariate unimodality are:

- **Linear Unimodality** about **0**, when the c.d.f. of any linear combination of the components of $\mathbf{X}$ is univariate unimodal about 0.
- **Convex Unimodality** about **0**, when for every c > 0, the set {**x**: f(**x**) ≥ c} is a centrally symmetric convex set.
- **α-Unimodality** about **0**, when $\mathbf{X} = U^{1/\alpha}Y$, where U is uniform on (0, 1) and **Y** is independent of U.
- **Beta Unimodality** about **0**, when $\mathbf{X} = U\mathbf{Y}$, where U has a Beta distribution with parameter α, β and **Y** is independent of U. (When β = 1 then F is α-unimodal).

Multivariate Unimodality

In what follows we refer only to *block unimodality*.

According to [Shepp, 1962] we have:

Theorem 4 The c.d.f. F is (block) unimodal about **0** if and only if there exists a random vector $\mathbf{X}=(X_1,\ldots,X_d)^T$ with c.d.f. F such that $\mathbf{X}=(Y_1U_1,\ldots,Y_dU_d)^T$, where $\mathbf{U}=(U_1,\ldots,U_d)^T$ is uniformly distributed on the unit cube $(0,1)^d$ and $\mathbf{Y}$ is independent of $\mathbf{U}$ having an arbitrary c.d.f. G.

This means that the c.d.f. F is the component-wise log-convolution of the Uniform on the unit cube $(0,1)^d$ c.d.f. U with G. We write $F=U\otimes G$.

Multivariate Unimodality

When d=2 Shepp shows the following:

<u>Theorem 5</u> The c.d.f. F is unimodal about **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.

Multivariate Unimodality

Theorems 4 and 5 can be also considered as particular cases of the Choquet-type representation theorem in $\mathbb{R}^d$. Specifically we have the following:

Theorem 6 The c.d.f. F is unimodal about **0** if and only if there exists a probability measure G on the Borel σ-field $\mathcal{B}_d=\mathcal{B}(\mathbb{R}^d)$ such that

$$F(A) = \int U_{\mathbf{y}}(A)G(d\mathbf{y}), \quad \forall\, A \in \mathcal{B}_d.$$

This again means that the c.d.f. F belongs to the closed convex hull of the set of $\{U_{\mathbf{y}}, 0 \le \|\mathbf{y}\| < \infty\}$.

Multimodality - Partial Convexification

According to the above procedure, i.e. by the component wise multiplication of two independent random vectors **Y** and **U**, where the latter is uniformly distributed on the unit d-cube, we always get a c.d.f. F with a single mode at **0**, no matter what the distribution G, we start with, is. To overcome the limitation of getting always a c.d.f. F with a single mode at **0**, we propose the following "**partial convexification**" procedure that results to multimodal multivariate distributions.

Partial Convexification in $\mathbb{R}$

In Kokolakis and Kouvaras (2007), partial convexification of a univariate c.d.f. G was based on using a $U(\alpha,1)$ distribution, instead of $U(0, 1)$, with $0 < \alpha < 1$.

The parameter α can be fixed or random with a prior distribution $p(\alpha)$ on the interval (0, 1).

The number of the modes of the log-convolution $F=U\otimes G$ is one, when $\alpha = 0$, and increases up to the number of the modes of G as α approaches 1. This means that when $G \sim DP$ and $0 < \alpha < 1$, the c.d.f. F alternates between local concavities and local convexities, i.e. a “partial convexification” of F is produced.

Partial Convexification in $\mathbb{R}$

According to above definition, Theorem 2 can be expressed in the following form [Kokolakis and Kouvaras, 2007].

Theorem 7 If the c.d.f. F is partially convexified then there exists a distribution function G on $\mathbb{R}$ such that F admits the representation:

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

for all x and x/α points of continuity of G, with $0 < \alpha < 1$.

Partial Convexification in $\mathbb{R}^d$

Definition

The d−variate c.d.f. F is called *partially block convexified*, or simply *partially convexified*, if there exists a random vector $\mathbf{X}=(X_1,\ldots,X_d)^T$ with c.d.f. F, such that $\mathbf{X}=(Y_1U_1,\ldots,Y_dU_d)^T$, where $\mathbf{U}=(U_1,\ldots,U_d)^T$ is uniformly distributed on the d-cube $(\alpha_1, 1)\times\ldots\times(\alpha_d, 1)$ and $\mathbf{Y}$ is independent of $\mathbf{U}$ having an arbitrary c.d.f. G.

Partial Convexification in $\mathbb{R}^2$

A generalization of Theorem 7 to the bivariate case using Uniform distributions $U(\alpha_i, 1)$, $i = 1, 2$, with parameters α_i fixed in the interval (0, 1), is as follows. [Kouvaras and Kokolakis, 2008].

Theorem 8 If the c.d.f. F is partially convexified then there exists a distribution function G on $\mathbb{R}^2$ such that F admits the representation:

$$F(x_1, x_2) = x_1 F_{x_1}(x_1, x_2) + x_2 F_{x_2}(x_1, x_2) - x_1 x_2 f(x_1, x_2) + Q(x_1, x_2)$$

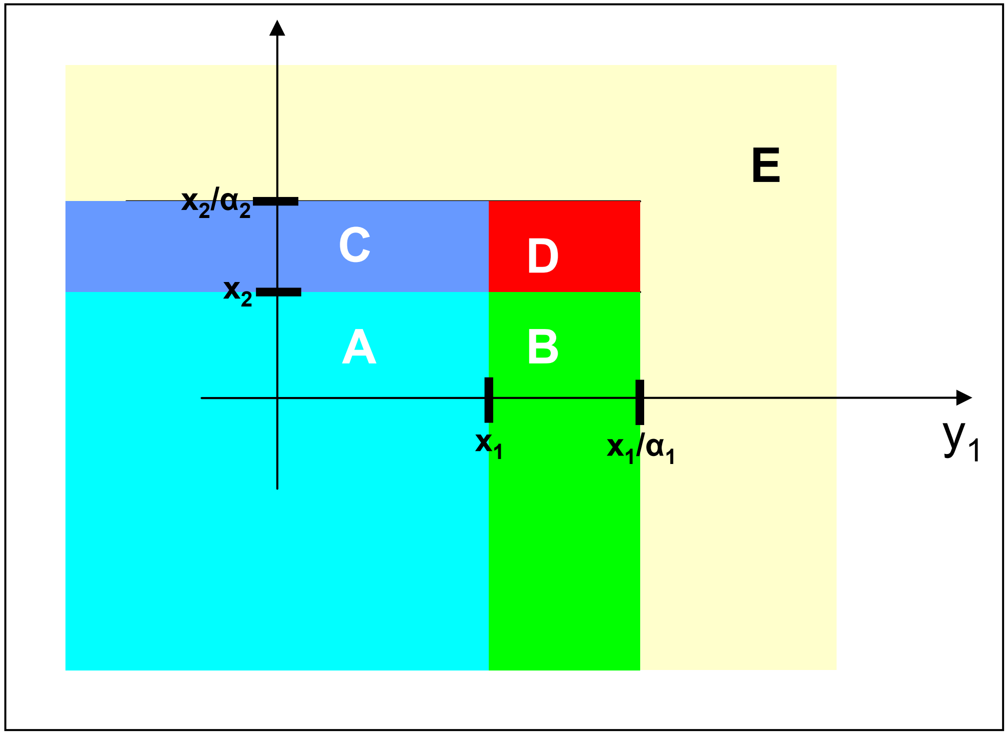
where

$$Q(x_1, x_2) = \frac{1}{(1-\alpha_1)(1-\alpha_2)} \times \{G(x_1, x_2) - \alpha_1 G(\frac{x_1}{\alpha_1}, x_2) - \alpha_2 G(x_1, \frac{x_2}{\alpha_2}) + \alpha_1 \alpha_2 G(\frac{x_1}{\alpha_1}, \frac{x_2}{\alpha_2})\}$$

Partial Convexification in $\mathbb{R}^2$

<u>Sketch of the proof</u>

Let $\mathbf{x}$ in $I=\{(x_1,x_2): x_1, x_2 > 0\}$. We consider the following partition of $\mathbb{R}^2$.



Partial Convexification in $\mathbb{R}^2$

We have for any **x** in the first quadrant:

$$F(\mathbf{x}) = P[X_1 \le x_1, X_2 \le x_2] = P[U_1 Y_1 \le x_1, U_2 Y_2 \le x_2]$$

$$= \int_{\mathbb{R}^2} H_{\mathbf{y}}(\mathbf{x}) G(d\mathbf{y}),$$

where

$$H_{\mathbf{y}}(\mathbf{x}) \equiv \mathrm{H}(\mathbf{x}, \mathbf{y}) = \begin{cases} 1, & (\mathbf{y} \in A) \\ \dfrac{\frac{x_1}{y_1} - \alpha_1}{1 - \alpha_1}, & (\mathbf{y} \in B) \\ \dfrac{\frac{x_2}{y_2} - \alpha_2}{1 - \alpha_2}, & (\mathbf{y} \in C) \\ \displaystyle\prod_{i=1}^{2} \dfrac{\frac{x_i}{y_i} - \alpha_i}{1 - \alpha_i}, & (\mathbf{y} \in D) \\ 0, & (\mathbf{y} \in E) \end{cases}$$

Partial Convexification in $\mathbb{R}^2$

Thus, we have the following decomposition :

$$
\begin{aligned}
F(\mathrm{x}) &= \int_{\mathbb{R}^2} H_{\mathrm{y}}(\mathrm{x}) G(d\mathrm{y}) \\
&= \int_A G(d\mathrm{y}) \\
&+ \int_B \frac{x_1 - \alpha_1 y_1}{(1-\alpha_1) y_1} G(d\mathrm{y}) \\
&+ \int_C \frac{x_2 - \alpha_2 y_2}{(1-\alpha_2) y_2} G(d\mathrm{y}) \\
&+ \int_D \frac{x_1 - \alpha_1 y_1}{(1-\alpha_1) y_1} \cdot \frac{x_2 - \alpha_2 y_2}{(1-\alpha_2) y_2} G(d\mathrm{y}).
\end{aligned}
$$

Partial Convexification in $\mathbb{R}^2$

With fixed y_1, $y_2 \neq 0$, the function $H_{\mathbf{y}}(\mathbf{x})$ is bounded with bounded above and below first and second order derivatives with respect to $\mathbf{x}$ a.e. in I.

Applying the bounded convergence theorem we conclude that the c.d.f. F is differentiable a.e. in I, and thus it finally takes the required expression, provided that (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 points of continuity of G.

The same procedure applied to the other quadrants yields the same result.

Partial Convexification in $\mathbb{R}^d$

A generalization of Theorem 8 to the d-dimensional case is as follows.

Theorem 9 If the c.d.f. F is partially convexified then there exists a distribution function G on $\mathbb{R}^d$ such that F admits the representation:

$$F(\boldsymbol{x}) = \sum_{i=1}^{d} x_i F_{x_i}(\boldsymbol{x}) - \sum\sum_{1\le i<j\le d} x_i x_j F_{x_i x_j}(\boldsymbol{x}) + \sum\sum\sum_{1\le i<j<k\le d} x_i x_j x_k F_{x_i x_j x_k}(\boldsymbol{x})$$
$$- \cdots + (-1)^{d-1} f(\boldsymbol{x}) \prod_{i=1}^{d} x_i + Q(\boldsymbol{x}),$$

where $$Q(\boldsymbol{x}) = \frac{1}{\prod_{i=1}^{d}(1-\alpha_i)} \times \{G(\boldsymbol{x}) - \sum_{i=1}^{d} \alpha_i G_i(\boldsymbol{x}) + \sum\sum_{1\le i<j\le d} \alpha_i \alpha_j G_{i,j}(\boldsymbol{x})$$
$$- \sum\sum\sum_{1\le i<j<k\le d} \alpha_i \alpha_j \alpha_k G_{i,j,k}(\boldsymbol{x}) + \cdots + (-1)^d G_{1,2,\cdots,d}(\boldsymbol{x}) \prod_{i=1}^{d} \alpha_i\}$$

and $$G_{i,j,k}(\boldsymbol{x}) = G(x_1, \cdots, \frac{x_i}{\alpha_i}, \cdots, \frac{x_j}{\alpha_j}, \cdots, \frac{x_k}{\alpha_k}, \cdots, x_d).$$

Our Model

In our Bayesian nonparametric model specification we assume the following:

- $\mathbf{Y}=(Y_1,\ldots,Y_d)^T \sim G$, where G is a random c.d.f. produced by a DP in $\mathbb{R}^d$.
- $\mathbf{U}=(U_1,\ldots,U_d)^T$ is uniformly distributed on the d-cube $(\alpha_1, 1)\times\ldots\times(\alpha_d, 1)$ with α_i fixed on the interval (0, 1), (i=1,…,d).
- **Y** and **U** are independent.

Then $\mathbf{X}=\mathbf{U}\bullet\mathbf{Y}\equiv(U_1Y_1,\ldots,U_dY_d) \sim F$ partially convexified.

Thus $F=U\otimes G$ is the component-wise log-convolution of U and G.

Application

We consider the generation of a random c.d.f G from a DP[c·β(·)], where c=2, the normalized parameter measure β(·) corresponds to the mixture $b(\mathbf{y})=w_1N(\mathbf{y} \mid \mu_1, \Sigma_1) + w_2N(\mathbf{y} \mid \mu_2, \Sigma_2)$ with $w_1=0.4=1-w_2$ and

$$\mu_1 = \begin{pmatrix} 8 \\ 8 \end{pmatrix}, \quad \mu_2 = \begin{pmatrix} 20 \\ 20 \end{pmatrix} \text{ and } \Sigma_1 = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 16 & 4 \\ 4 & 25 \end{pmatrix}.$$

Application

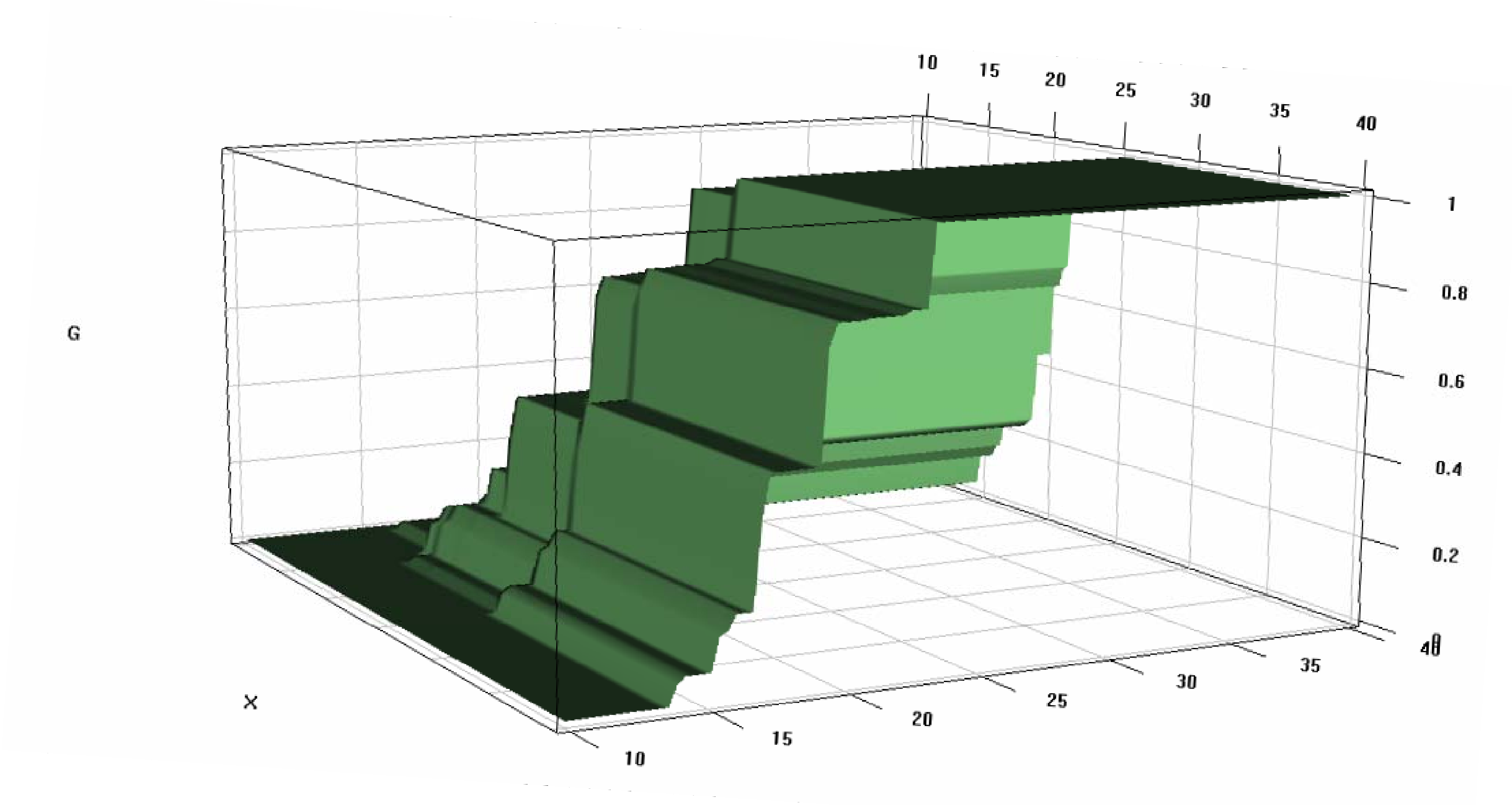


Figure 1. Random c.d.f. G from DP[c·β(·)].

Application

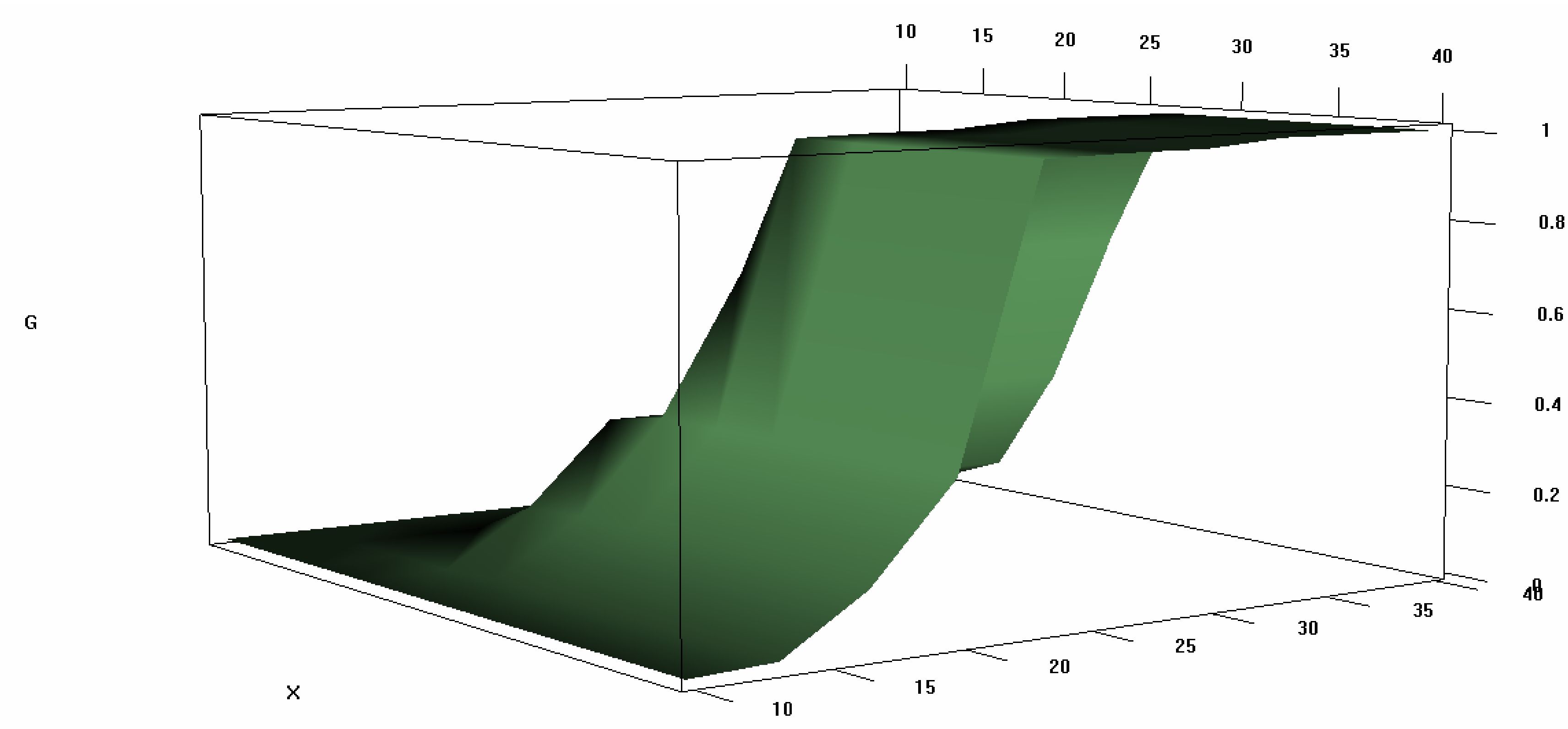


Figure 2. Partially Convexified c.d.f. F=U⊗G, with $\alpha_1=\alpha_2=0.2$.

Future Research

- Consider randomly distributed lower vertex **α**.
- Investigate hidden properties.
- Combine with the likelihood – Deconvolute the empirical distribution function.
- Work with other types of multivariate unimodality. (Star, α-unimodality, Beta-Unimodality,…).

Conclusion

This presentation addresses the following interesting question.

Q: How can we create a useful Bayesian nonparametric model that places a prior over multivariate distributions with a varying number of modes.

A: The answer provided here is based on generalized mixtures of uniform distributions with a DP. We developed the notion of partially convexified distributions based on the component-wise log-convolution of a multivariate Uniform distribution over the d-cube $(\alpha_1, 1)\times\ldots\times(\alpha_d, 1)$ with $0 < \alpha_i < 1$ $(i=1,\ldots,d)$ and a $G \sim DP$. Our model pins down the number of modes of the random probability measure produced between the extremes *one* and *infinity*.

References

- **Antoniak, C.E**. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *Annals of Statistics*, 2, 1152-1174.
- **Bertin, E.M.J., Cuculescu, I. and Theodorescu, R**. (1997). *Unimodality of Probability Measures*. Kluwer Academic Publishers.
- **Brunner, L.J**. (1992). Bayesian nonparametric methods for data from a unimodal density. *Statistics and Probability Letters*, 14, 195-199.
- **Dharmadhikari, S.W. and Joag-dev**. (1988). *Unimodality, Convexity, and Applications*. Academic Press, New York.
- **Ferguson, Th. S**. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1, 209-230.
- **Ferguson, Th. S**. (1974). Prior distributions on spaces of probability measures. *Annals of Statistics*, 2, 615-629.
- **Feller, W**. (1971). *An Introduction to Probability Theory and its Applications*, (2nd edition), John Wiley and Sons, New York.
- **Hansen, M. and Lauritzen, S**. (2002). Nonparametric Bayes inference for concave distribution functions. *Statistica Neerlandica*, 56, 110-27.

References

- **Hoff, P.D**. (2003). Nonparametric estimation of convex models via mixtures. *Annals of Statistics*, 31, 174-200.
- **Kokolakis, G. and Dellaportas, P**. (1996). Hierarchical modelling for classifying binary data. *Bayesian Statistics* 5, 647-652.
- **Kokolakis, G. and Kouvaras, G**. (2007). On the multimodality of random probability measures. *Bayesian Analysis*, 2, 213-220.
- **Kouvaras, G. and Kokolakis, G**. (2008). Random multivariate multimodal distributions. In *Recent Advances in Stochastic Modelling and Data Analysis*. World Scientific. (to appear).
- **Lavine, M**. (1992). Some aspects of Polya tree distributions for statistical modelling. *Annals of Statistics*, 20, 1222-1235.
- **Lavine, M**. (1994). More aspects of Polya tree distributions for statistical modelling. *Annals of Statistics*, 22, 1161-1176.
- **Shepp, L.A**. (1962). Symmetric random walk. *Transactions of the American Mathematical Society* 104, 144-153.