SEQUENTIAL BARYCENTER ARRAYS AND RANDOM PROBABILITY MEASURES

Theodore Hill†

Georgia Institute of Technology

Michael Monticino

University of North Texas

Abstract. This article introduces and develops a constructive method for generating random probability measures with a prescribed mean or distribution of the means. The method involves sequentially generating an array of barycenters which uniquely defines a probability measure. Basic properties of the generated measures are presented, including conditions under which almost all the generated measures are continuous, or almost all are purely discrete, or almost all have finite support. Among the applications considered are models for distribution of mass, average-optimal control problems, and experimental approximation of universal constants.

1. Introduction.

The main purpose of this article is to introduce a general and natural method for constructing random probability measures with any prescribed mean or distribution of the means. This method complements classical and recent constructions (e.g., Dubins and Freedman (1967), Ferguson (1973, 1974), Graf, Mauldin and Williams (1986), Mauldin, Sudderth and Williams (1992) and Monticino (1996)), none of which generate random measures with *a priori* specified means. In fact, even the calculation of the distribution of the means for those constructions is difficult (cf., Cifarelli and Regazzani (1990) and Monticino (1995)).

1991 *Mathematics Subject Classification*. Primary 60A10, 62A15, 60G57. Secondary 60G57, 60G30.

Key words and phrases. sequential baricenter arrays, random probability measures, random distributions, random homeomorphisms, distribution of mass.

† Research partially supported by NSF grant DMS-95-03375.

Typeset by $\mathcal{AMS}$-TEX

The new method presented here, which is based on sequential barycenters, satisfies Ferguson's (1974) two basic requirements that such constructions have large support and are analytically manageable. The construction is easy to implement and is robust, allowing generation of random measures which are either (almost surely) discrete or continuous, as desired. Since many problems in probability and analysis involve distributions with given means, the new construction will perhaps prove a useful tool in a variety of applications, several which are outlined in section 4 below.

Section 2 contains basic definitions, properties and examples of sequential barycenter arrays of (non-random) distributions. Section 3 presents the new construction of random distributions based on generation of their successive barycenters, rather than direct construction of the distribution function itself as in previous methods.

As noted in Dubins and Freedman (1967), any method for generating random probability distributions may also be viewed as a method for generating random convex functions (since the indefinite integral of a distribution function is convex), stochastic processes (whose sample paths are distribution functions), or random homeomorphisms (if the generated distributions are strictly increasing and continuous), but such interpretations will be left to the interested reader.

2. Sequential Barycenter Arrays.

This section introduces the notion of a sequential barycenter array (SBA) and develops some basic properties of the probability measures defined by the arrays. These SBA's, although not named as such, are used in standard proofs of Skorohod's embedding theorems (e.g., Billingsley (1986, Section 37)), and it is the reversal of this standard procedure which is the foundation for the construction of the random measures given in the next section.

Throughout this section, let X be a real-valued random variable with distribution function F, such that $E[|X|] < \infty$.

Definition 2.1. The **F-barycenter of (a,c]**, $b_F(a,c]$, is given by

$$b_F(a,c] = \begin{cases} E[X|X \in (a,c]] = \frac{\int_{(a,c]} x dF(x)}{F(c)-F(a)}, & \text{if } F(c) > F(a) \\ a, & \text{if } F(c) = F(a). \end{cases}$$

Example 2.2. Let X be uniformly distributed over $[0,1]$. Then $b_F(a,c] = \frac{c+a}{2}$, for any $0 \le a \le c \le 1$.

Example 2.3. Let X be binomially distributed with $n = 2$ and $p = \frac{1}{2}$. Then $b_F(0,1] = 1$, $b_F(1,2] = 2$, $b_F(0,2] = \frac{4}{3}$, and $b_F(a,0] = 0$, for $a < 0$.

Some elementary properties of F-barycenters, which follow easily from Definition 2.1, are recorded in the next lemma.

Lemma 2.4. *Fix $a < c$ such that $P[X \in [a,c]] > 0$, and let $b = b_F(a,c]$. Then*

(i) *$F(c) > F(a)$ if and only if $b > a$,*

(ii) *$(F(c) - F(a))b = (F(b) - F(a))\, b_F(a,b] + (F(c) - F(b))\, b_F(b,c]$,*

(iii) *$b_F(a,b] = b$ if and only if $b_F(b,c] = b$,*

(iv) *$b \ge b_F(a,x]$, for all $x \in (a,c]$.*

Definition 2.5. The **sequential barycenter array** (SBA) of F is the triangular array $\{m_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^n-1} = \{m_{n,k}(F)\} = M(F)$ defined inductively by

(2.1) $m_{1,1} = E[X] = \int x dF(x) = b_F(-\infty,\infty)$,

(2.2) $m_{n,2j} = m_{n-1,j}$, for $n \ge 1$ and $j = 1,\ldots,2^{n-1}-1$,

(2.3) $m_{n,2j-1} = b_F(m_{n-1,j-1}, m_{n-1,j}]$, for $j = 1,\ldots,2^{n-1}$, with the convention that $m_{n,0} = -\infty$ and $m_{n,2^n} = \infty$.

Example 2.6. Suppose X is uniformly distributed over $[0,1]$. Then

$$\{m_{n,k}(F)\} = \left\{\frac{k}{2^n}\right\}_{n=1}^{\infty}{}_{k=1}^{2^n-1}.$$

Example 2.7. Suppose X is binomially distributed with $n = 2$ and $p = \frac{1}{2}$. Then $m_{1,1} = 1$; $m_{2,1} = \frac{2}{3}$; $m_{2,3} = 2$; and, for $n \ge 3$,

$$m_{n,k} = \begin{cases} 0 & \text{for } k = 1,\ldots,2^{n-2}-1 \\ \frac{2}{3} & \text{for } k = 2^{n-2} \\ 1 & \text{for } k = 2^{n-2}+1,\ldots,2^{n-1} \\ 2 & \text{for } k = 2^{n-1}+1,\ldots,2^n-1. \end{cases}$$

As seen in Example 2.7, it may happen that the sequential barycenters of a given distribution are not distinct (i.e., $m_{n,k+1} = m_{n,k}$ for some n and k). The next example shows that monotonicity alone ($m_{n,k} \leq m_{n,k+1}$) is not enough to guarantee that an array be the SBA for some distribution. The additional condition needed — (2.6) in Theorem 2.12 below — is a martingale property. First, several useful properties of SBA's are noted, followed by an inversion formula (Theorem 2.10) to recover F from its SBA.

Example 2.8. Suppose $M = \{m_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^n-1}$ with $m_{2,1} = \frac{1}{2} = m_{2,2} \neq m_{2,3} = 1$. If $M = M(F)$ for some distribution function F, then, by (2.3),

$$m_{2,1} = b_F(-\infty, \frac{1}{2}] = \frac{1}{2} = b_F(-\infty, \infty) = m_{2,2}.$$

This implies F gives unit mass to $\frac{1}{2}$ and so, again by (2.3), $m_{2,3}(F) = \frac{1}{2}$. Thus, $M \neq M(F)$ for any F.

Notation. For SBA $\{m_{n,k}\}$, let $I_{n,k} = (m_{n,k-1}, m_{n,k}] \subset \mathbb{R}$.

Lemma 2.9. *Let* $\{m_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^n-1} = \{m_{n,k}(F)\}$ *be the SBA for distribution function* F*. Then*

(i) *if* $F(c) > F(a)$*, then there exist* n *and* j *with* $m_{n,j} \in [a, c]$*;*

(ii) $\{m_{n,k}(F)\}$ *is dense in the support of* F*;*

(iii) *for each* $n \geq 1$, $\{I_{n,k}\}_{k=1}^{2^n}$ *is a partition of* $\mathbb{R}$ *and* $\{I_{n+1,k}\}_{k=1}^{2^{n+1}}$ *is a refinement of* $\{I_{n,k}\}_{k=1}^{2^n}$*;*

(iv) $P[X \in [m_{n,k-1}, m_{n,k}]] > 0$*, for all* $n \geq 1$ *and* $k = 1, \ldots, 2^n$*.*

Proof. Parts (i) and (iii) are routine; (ii) is straightforward from (i); and (iv) follows by induction on n and Definition 2.1. □

Theorem 2.10. *F is completely determined by the values* $\{m_{n,k}(F)\}_{n=1}^{\infty}{}_{k=1}^{2^n-1}$*. In particular,* $F(m_{n,k})$ *is given inductively by* $F(m_{n,0}) = 0$, $F(m_{n,2^n}) = 1$*; by (2.2) for even k; and, for* $k = 2j - 1$*,*

(2.4)

$$F(m_{n,2j-1}) = F(m_{n-1,j-1}) + (F(m_{n-1,j}) - F(m_{n-1,j-1})) \frac{m_{n+1,4j-1} - m_{n+1,4j-2}}{m_{n+1,4j-1} - m_{n+1,4j-3}}$$

(with $\frac{0}{0} = 1$*).*

Proof. By Lemma 2.9 (ii) and (2.4), F is determined by $\{m_{n,k}(F)\}$. To see (2.4), note that Lemma 2.4 (ii) gives

$$F(m_{n,2j-1}) = F(m_{n,2j-2}) + (F(m_{n,2j}) - F(m_{n,2j-2})) \frac{b_F(m_{n,2j-1}, m_{n,2j}] - m_{n,2j-1}}{b_F(m_{n,2j-1}, m_{n,2j}] - b_F(m_{n,2j-2}, m_{n,2j-1}]}.$$

And, by (2.2) and (2.3), $m_{n-1,j-1} = m_{n,2j-2}$, $m_{n-1,j} = m_{n,2j}$, $m_{n+1,4j-1} = b_F(m_{n,2j-1}, m_{n,2j}]$, and $m_{n+1,4j-3} = b_F(m_{n,2j-2}, m_{n,2j-1}]$. □

Corollary 2.11. *$F_1 = F_2$ if and only if $m_{n,k}(F_1) = m_{n,k}(F_2)$, for all $n \geq 1$ and $1 \leq k \leq 2^n - 1$.*

Note: it is well known that certain other collections of barycenters — for example, $\{b_F(-\infty, t]\}_{t\in\mathbb{R}}$ — also determine F.

Theorem 2.12. *A triangular array $M = \{m_{n,k}\}_{n=1}^{\infty}\,_{k=1}^{2^n-1}$ is an SBA for some distribution function F if and only if M satisfies (2.2),*

(2.5) $$m_{n,k-1} \leq m_{n,k}, \qquad \textit{for all } n \geq 1 \textit{ and } k = 1, \ldots, 2^n, \textit{ and}$$

(2.6)
$$m_{n,4k-3} = m_{n,4k-2} \textit{ if and only if } m_{n,4k-1} = m_{n,4k-2}, \qquad \textit{for all } n \geq 2 \textit{ and } k = 1, \ldots, 2^{n-2}.$$

Proof. Given M is an SBA for some distribution function F, the necessity of (2.2) follows from Definition 2.5. Similarly, the necessity of (2.5) follows easily using induction on n and Definition 2.5. For the necessity of (2.6), note that by (2.2) and (2.3),

$$m_{n,4k-2} = m_{n-1,2k-1} = b_F(m_{n-1,2k-2}, m_{n-1,2k}] = b_F(m_{n,4k-4}, m_{n,4k}],$$

$$m_{n,4k-3} = b_F(m_{n,4k-4}, m_{n,4k-2}],$$

and

$$m_{n,4k-1} = b_F(m_{n-1,2k-1}, m_{n-1,2k}] = b_F(m_{n,4k-2}, m_{n,4k}],$$

Letting $a = m_{n,4k-4}$, $b = m_{n,4k-2}$, and $c = m_{n,4k}$, (2.6) follows by Lemma 2.4 (iii).

For the sufficiency portion of the proof, let $\{m_{n,k}\}_{n=1}^{\infty}\,_{k=1}^{2^n-1}$ be a triangular array satisfying (2.2), (2.5), and (2.6). Define random variables, $X_1, X_2, \ldots$, inductively as follows. $X_1 \equiv m_{1,1}$. For $n \geq 2$, X_n takes values in $\{m_{n,2j-1}\}_{j=1}^{2^{n-1}}$ with

(2.7)
$$P[X_n = m_{n,4j-3} | X_{n-1} = m_{n,4j-2}] = \begin{cases} 1 - P[X_n = m_{n,4j-1} | X_{n-1} = m_{n,4j-2}] \\ \qquad = \frac{m_{n,4j-2} - m_{n,4j-3}}{m_{n,4j-1} - m_{n,4j-3}} & \text{if } m_{n,4j-3} \neq m_{n,4j-2} \\ 1 & \text{if } m_{n,4j-3} = m_{n,4j-2} \end{cases}$$

Note that (2.5) ensures that (2.7) defines probabilities. And (2.6) yields $E[X_{n+1}||X_n] = X_n$ — that is, $X_1, X_2, \dots$ is a discrete martingale. Moreover, by (2.7), for all $n \geq N$ and $j = 1, \dots, 2^{n-1}$,

$$b_{F_n}(m_{N,2j-2}, m_{N,2j}] = m_{N,2j-1}. \tag{2.8}$$

To see that $X_n \to X$ a.s., where X is a random variable with the desired barycenters, first note that by the construction of $\{X_n\}$ above, with probability one

$$\{X_2 > m_{1,1}\} = \{X_n > m_{1,1}\},$$

for all $n > 1$. (These sets are not always equal. It can happen, for example, that on the set $\{X_2 > m_{1,1}\}$, the set $\{X_3 = m_{1,1}\} \neq \emptyset$. But, by Defininition 2.1 and Lemma 2.4 (iii), this can not happen with positive probability. This crucial part of the proof would not necessarily follow if the F-barycenters were defined on, say, closed intervals $[a, b]$ instead of on $(a, b]$.)

Next observe that conditioned on the set $\{X_2 > m_{1,1}\}$, $\{X_n\}_{n \geq 2}$ is a martingale which is bounded below by $m_{1,1}$. Thus it converges (on $\{X_2 > m_{1,1}\}$) to a random variable X^+ which, by (2.8), has the correct barycenters. A similar argument for the set $\{X_n \leq m_{1,1}\}$ completes the proof. □

Corollary 2.13. *If F is continuous then*

$$m_{n,k-1}(F) < m_{n,k}(F), \qquad \text{for all } n \geq 1 \text{ and } k = 1, \dots 2^n. \tag{2.9}$$

Corollary 2.14. *If $M = \{m_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^n-1}$ satisfies (2.2) and (2.9), then $M = M(F)$ for some distribution function F.*

Neither of the converses of Corollary 2.13 or 2.14 hold — as the next example shows for the former. On the other hand, under the conditions given in Proposition 2.17 below, continuity of F can be inferred from $M(F)$.

Example 2.15. Let $q_1, q_2, \dots$ be an ordering of the rationals in $\mathbb{R}$, and let F be the distribution function for the purely discrete measure which gives mass $\frac{1}{2^n}$ to q_n. Then, for any $a < b$, $a < b_F(a, b] < b$. Hence, $m_{n,k}(F) < m_{n,k+1}(F)$, for all $n \geq 1$ and $k = 0, \dots 2^n - 1$.

Definition 2.16. For $n \geq 0$ and distribution function F, the function $d_n = d_n^F : \mathbb{R} \to \mathbb{R}^+ \cup \{\infty\}$, is given by

$$d_n(x) = |I_{n,k_n}(x)|,$$

where $I_{n,k_n}(x)$ is the unique interval $(m_{n,k-1}(F), m_{n,k}(F)] = (m_{n,k_n(x)-1}, m_{n,k_n(x)}]$ containing $x \in \mathbb{R}$ (cf. Lemma 2.9 (iii)).

Proposition 2.17. *Let F be a distribution function with SBA $M(F) = \{m_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^n-1}$.*

(i) *If $x \notin M(F)$ and $d_n(x) = d_{n+1}(x)$, for some $n \geq 1$, then F is continuous at x.*

(ii) *If, for some $\epsilon > 0$, there exist infinitely many n such that*

$$\epsilon < \frac{d_{n+1}(x)}{d_n(x)}, \ \frac{d_{n+2}(x)}{d_{n+1}(x)} < 1 - \epsilon, \tag{2.10}$$

then F is continuous at x.

Proof. (i) By Definitions 2.5 and 2.16, if $d_n(x) = d_{n+1}(x)$ then either $b_F I_{n,k_n}(x) = m_{n,k}(F)$ or $b_F I_{n,k_n}(x) = m_{n,k-1}(F)$. In either case, this implies that the F-measure of the open interval $(m_{n,k-1}(F), m_{n,k}(F))$ is 0. Thus, if $x \notin M(F)$, then F is continuous at x.

(ii) By (2.2) and (2.4), for $n \geq 1$ and both $k_{n+1}(x)$ and $k_{n+2}(x)$ odd,

$$F(m_{n+1,k_{n+1}(x)}) - F(m_{n+1,k_{n+1}(x)-1}) = (F(m_{n,k_n(x)}) - F(m_{n,k_n(x)-1})\frac{m_{n+2,4k_n(x)-1} - m_{n+2,4k_n(x)-2}}{m_{n+2,4k_n(x)-1} - m_{n+2,4k_n(x)-3}}$$

and

$$\begin{aligned}\frac{m_{n+2,4k_n(x)-1} - m_{n+2,4k_n(x)-2}}{m_{n+2,4k_n(x)-1} - m_{n+2,4k_n(x)-3}} &= \frac{m_{n+2,4k_n(x)-1} - m_{n+2,4k_n(x)-2}}{(m_{n+2,4k_n(x)-1} - m_{n+2,4k_n(x)-2}) + (d_{n+1}(x) - d_{n+2}(x))} \\ &\leq \frac{d_n(x) - d_{n+1}(x)}{(d_n(x) - d_{n+1}(x)) + (d_{n+1}(x) - d_{n+2}(x))} \\ &= \frac{1 - \frac{d_{n+1}(x)}{d_n(x)}}{1 - \frac{d_{n+1}(x)}{d_n(x)}\frac{d_{n+2}(x)}{d_{n+1}(x)}}\end{aligned}$$

where the last equality holds since Theorem 2.12 and (2.10) imply (with the convention that $\frac{0}{0} = 1$) $d_n(x) \neq 0$ for any n. Similar inequalities hold for the other cases, and these along with (2.10) yield a constant $\rho < 1$ such that

$$F(m_{n+1,k_{n+1}(x)}) - F(m_{n+1,k_{n+1}(x)-1}) \leq \rho(F(m_{n,k_n(x)}) - F(m_{n,k_n(x)-1}))$$

for infinitely many n. Hence the F-measure of $\{x\}$ is 0 and F is continuous at x.
□

3. Random SBA Distributions.

This section describes the new method for generating random probability measures using sequential barycenter arrays. The description is given explicitly for random probability measures on $[0,1]$, but it is easy to extend this method to other supports. The basic idea can be seen in the following special case.

First pick (deteministically or at random) a point $m_{1,1}$ in $[0,1]$. This point will be the mean of the distribution. Given $m_{1,1}$, next pick a point $m_{2,1}$, uniformly in $[0, m_{1,1})$, which will be the conditional mean of the distribution given that its value is less than its mean. Similarly, pick the upper conditional mean $m_{2,3}$, uniformly in $(m_{1,1}, 1)$; and so on. By Theorem 2.12 and Corollary 2.13, the resulting (random) array $\{m_{n,k}(\omega)\}_{n=1}^{\infty}{}_{k=1}^{2^n-1}$ is the SBA for a unique Borel probability measure on $[0,1]$.

The requirement that the array be chosen by determining successive barycenters uniformly is not crucial and the more general method presented below will be based on two (possibly identical) Borel probability measures μ_0 and μ on $[0,1]$. That is, first choose $m_{1,1}$ according to μ_0, then choose the *distances* to the next two successive barycenters independently according to μ rescaled to $[0, m_{1,1}]$ and to $[m_{1,1}, 1]$, respectively. Continue in this manner using rescaled versions of μ from the first step on.

More formally, let μ_0 and μ be probability measures with support on $[0,1]$ and $[0,1)$ respectively. Denote by $\mathcal{P}([0,1])$ the set of all Borel probability measures on $[0,1]$. Let $\{X_{n,2j-1}\}_{n=1,j=1}^{\infty,2^{n-1}}$ be an array of independent random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ such that $X_{1,1}$ has distribution μ_0 and, for $n \geq 2$, each $X_{n,k}$ has distribution μ.

Define a random array $M = \{m_{n,k}\}_{n=1,k=1}^{\infty,2^n-1}$ inductively by

$$m_{1,1} = X_{1,1},$$

$$m_{n,2j} = m_{n-1,j}, \qquad \text{for } n > 1 \text{ and } j = 1, \dots 2^{n-1} - 1,$$

$$m_{n,4j-3} = \begin{cases} m_{n-1,2j-1} & \text{if } m_{n-1,2j-1} = m_{n-1,2j},\ m_{n-1,2j-2} = m_{n-1,2j-1}, \\ & \quad X_{n,4j-3} = 0, \text{ or } X_{n,4j-1} = 0 \\ m_{n-1,2j-1} - X_{n,4j-3}(m_{n-1,2j-1} - m_{n-1,2j-2}) & \text{otherwise} \end{cases}$$

and

$$m_{n,4j-1} = \begin{cases} m_{n-1,2j-1} & \text{if } m_{n,4j-3} = m_{n-1,2j-1} \\ m_{n-1,2j-1} + X_{n,4j-1}(m_{n-1,2j} - m_{n-1,2j-1}) & \text{otherwise} \end{cases}$$

(for all $n \geq 1$, $m_{n,0} = 0$ and $m_{n,2^n} = 1$).

Endow the set of triangular arrays $\mathcal{A} = [0,1] \times [0,1]^3 \times \cdots \times [0,1]^{2^n-1} \times \ldots$ with the standard product topology. Let $A \subset \mathcal{A}$ be the Borel subset of arrays which satisfy (2.2), (2.5), and (2.6). Notice that $M(\omega) \in A$ for all $\omega \in \Omega$. Let $Q_{(\mu_0,\mu)}$ be the distribution of M on $\mathcal{A}$. By Theorems 2.10 and 2.12, the mapping, T, (induced by (2.2) and (2.4)) which sends an array $\{m_{n,k}\} \in A$ to its associated distribution, $T(m)$, is Borel from A to $\mathcal{P}([0,1])$ given the weak* topology.

Definition 3.1. The **sequential barycenter array random probability measure** (SBA rpm) $B_{(\mu_0,\mu)}$ is the Borel measure $Q_{(\mu_0,\mu)}T^{-1}$ on $\mathcal{P}([0,1])$.

Proposition 3.2. *The distribution on the mean under* $B_{(\mu_0,\mu)}$ *is* μ_0. *That is,*

$$B_{(\mu_0,\mu)}\left(\{F: \int_0^1 x\, dF(x) \leq a\}\right) = \mu_0([0,a]).$$

Proof. Immediate by the definitions of Q, T and $B_{(\mu_0,\mu)}$. □

The SBA random probability measure construction thus provides a straightforward way to produce rpm's with any prescribed mean or distribution on the mean, whereas classical rpm constructions do not.

Throughout this article, a probability measure in $\mathcal{P}([0,1])$ will often be identified with its distribution function, and $B_{(\mu_0,\mu)}$ will often be regarded as a measure on the set of distribution functions.

Example 3.3. Suppose $\mu_0 = \mu = \delta_{\frac{1}{2}}$. Then, $Q_{(\mu_0,\mu)}$ gives probability 1 to the array $\left\{\frac{k}{2^n}\right\}_{n=1}^{\infty}{}_{k=1}^{2^n-1}$. Hence, by Example 2.6 and Theorem 2.10, $B_{(\mu_0,\mu)}$ gives probability one to the uniform distribution on $[0,1]$.

Remark. Sequential barycenter rpm's arise which can not be realized with a Dubins-Freedman construction — as indicated in the following two examples. Moreover, it is conjectured that unless μ_0 and μ both give unit mass to $\frac{1}{2}$, then $B_{(\mu_0,\mu)}$ is never a Dubins-Freedman rpm.

Example 3.4. Suppose $\mu_0 = \mu = \delta_z$, for $0 < z < 1$ and $z \neq \frac{1}{2}$. Then there does not exist a Dubins-Freedman base probability ν on the square $[0,1] \times [0,1]$ such that the corresponding Dubins-Freedman prior, P_ν, is equal to $B_{(\mu_0,\mu)}$. To see this, first note that, by construction, $B_{(\mu_0,\mu)} = \delta_{F_z}$ for some distribution function F_z. Moreover, by (2.4) of Theorem 2.10,

$$\begin{aligned} F_z(m_{1,1}(F_z)) &= F_z(z) \\ &= \frac{m_{2,3}(F_z) - m_{2,2}(F_z)}{m_{2,3}(F_z) - m_{2,1}(F_z)} \\ &= 1 - z \end{aligned}$$

and

$$\begin{aligned} F_z(m_{3,1}(F_z)) &= F_z(z(1-z)^2) \\ &= F(m_{2,1}(F_z)) \left(\frac{m_{4,3}(F_z) - m_{4,2}(F_z)}{m_{4,3}(F_z) - m_{4,1}(F_z)} \right) \\ &= z(1-z) \left(\frac{(z(1-z)^2 + z^3(1-z)) - z(1-z)^2}{(z(1-z)^2 + z^3(1-z)) - z(1-z)^3} \right) \\ &= z^2(1-z). \end{aligned}$$

So, the graph of F_z contains a point above the main diagonal of the unit square and a point below the diagonal.

However, if a probability measure ν on $[0,1]\times[0,1]$ has support solely on or above the main diagonal, then, by the Dubins-Freedman construction, P_ν has support on the set of distribution functions whose graphs do not contain any points below the diagonal. If the support of ν lies solely on or below the diagonal, then the support of P_ν is contained within the set of distribution functions whose graphs do not contain any points above the diagonal. If ν gives positive probability to regions above and below the diagonal, then by Dubins and Freedman (1967, Corollary 3.5) the support of P_ν is larger than the singleton set $\{F_z\}$. Hence, there does not exist a Dubins-Freeman rpm equal to $B_{(\mu_0,\mu)}$. □

Example 3.5. Let $\mu_0 = \mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\frac{1}{2}}$. Then, by construction, $B_{(\mu_0,\mu)}(\delta_0) = \frac{1}{2}$, $B_{(\mu_0,\mu)}(\delta_{\frac{1}{2}}) = \frac{3}{8}$, and $B_{(\mu_0,\mu)}(\{\delta_x : x \neq 0, \frac{1}{2}\}) = 0$. Suppose there exists a Dubins-Freedman base probability ν such that the corresponding Dubins-Freedman prior, P_ν, is equal to $B_{(\mu_0,\mu)}$. Then, in order for P_ν to give probability $\frac{3}{8}$ to $\delta_{\frac{1}{2}}$,

$$\nu\left(([\frac{1}{2}, 1) \times \{1\}) \cup ((0, \frac{1}{2}] \times \{0\}) \right) \geq \frac{3}{8}$$

must hold. However, if $P_\nu(\delta_{\frac{1}{2}}) > 0$ and $\nu\left([\frac{1}{2},1) \times \{1\}\right) > 0$, then, by the self-similarity of the Dubins-Freedman construction, it would follow that

$$P_\nu\left(\left\{\delta_x : \quad \frac{1}{4} \le x < \frac{1}{2}\right\}\right) > 0.$$

Analogously, if $\nu\left((0,\frac{1}{2}] \times \{0\}\right) > 0$, it would follow that

$$P_\nu\left(\left\{\delta_x : \quad \frac{1}{2} < x \le \frac{3}{4}\right\}\right) > 0.$$

Hence, there does not exist a Dubins-Freedman prior equal to $B_{(\mu_0,\mu)}$. □

What types of measures are in the support of $B_{(\mu_0,\mu)}$? If $\mu_0(\{0,1\}) = 0 = \mu(\{0\})$, then a straightforward argument using Proposition 2.17 (ii) and Borel-Cantelli shows that, for every $x \in [0,1]$, $B_{(\mu_0,\mu)}$-almost all distribution functions are continuous at x. Moreover, the stronger result, that $B_{(\mu_0,\mu)}$-almost all measures are continuous on $[0,1]$ also holds. This is similar to Dubins and Freedman (1967, Theorem 4.1), and contrasts to Dirichlet rpm's (Ferguson (1973)) which are almost surely discrete. Conversely, Theorem 3.8 below shows that if $\mu(\{0\}) > 0$ then $B_{(\mu_0,\mu)}$-almost all measures are discrete.

Theorem 3.6. *$B_{(\mu_0,\mu)}$-almost all measures are continuous on $[0,1]$ if and only if $\mu_0(\{0,1\}) = 0 = \mu(\{0\})$.*

Proof. The condition is clearly necessary from the definition of $B_{(\mu_0,\mu)}$. The sufficiency portion of the proof adapts a technique initiated by Dubins and Freedman (1967, Theorem 4.1). In particular, by Mauldin, Sudderth, Williams (1992, Lemma 5.2), it is enough to show that

$$\int\left(\int_D d(F\times F)(x,y)\right) dB_{(\mu_0,\mu)}(F) = 0, \tag{3.2}$$

for $D = \{(x,y) \in [0,1]\times[0,1] : \ x = y\}$.

For notational convenience, let F_M denote the distribution function associated with the SBA $M = \{m_{n,k}\}$ and, for a distribution function F with SBA $\{m_{n,k}(F)\}$, let $F(m_{n,k}) = F(m_{n,k}(F))$. Then (3.2) is obtained if

$$E_n = \int\left(\sum_{k=1}^{2^n}(F(m_{n,k}) - F(m_{n,k-1}))^2\right) dB_{(\mu_0,\mu)}(F)$$

converges to 0 as $n \to \infty$. By Theorem 2.10 and Definition 3.1,

$$
\begin{aligned}
E_n &= \int \left[\sum_{k=1}^{2^n} (F_M(m_{n,k}) - F_M(m_{n,k-1}))^2 \right] dQ_{(\mu_0,\mu)}(M) \\
&= \int \left[\sum_{j=1}^{2^{n-1}} (F_M(m_{n,2j-1}) - F_M(m_{n,2j-2}))^2 + (F_M(m_{n,2j}) - F_M(m_{n,2j-1}))^2 \right] dQ_{(\mu_0,\mu)}(M) \\
&= \int \left[\sum_{j=1}^{2^{n-1}} (F_{M(\omega)}(m_{n-1,j}(\omega)) - F_{M(\omega)}(m_{n-1,j-1}(\omega)))^2 \times \right. \\
&\qquad \left. \left(\left(\frac{m_{n+1,4j-1}(\omega) - m_{n+1,4j-2}(\omega)}{m_{n+1,4j-1}(\omega) - m_{n+1,4j-3}(\omega)} \right)^2 + \left(\frac{m_{n+1,4j-2}(\omega) - m_{n+1,4j-3}(\omega)}{m_{n+1,4j-1}(\omega) - m_{n+1,4j-3}(\omega)} \right)^2 \right) \right] dP(\omega) \\
&= \int \left[\sum_{j=1}^{2^{n-1}} (F_{M(\omega)}(m_{n-1,j}(\omega)) - F_{M(\omega)}(m_{n-1,j-1}(\omega)))^2 f(x_{n,2j-1}(\omega)) \right] dP(\omega),
\end{aligned}
$$

where

$$
f(x_{n,2j-1}) = \begin{cases} \int_{[0,1)} \int_{[0,1)} \left(\frac{z x_{n,2j-1}}{z x_{n,2j-1} + y(1-x_{n,2j-1})} \right)^2 + \left(\frac{y(1-x_{n,2j-1})}{z x_{n,2j-1} + y(1-x_{n,2j-1})} \right)^2 d\mu(y) d\mu(z) & \text{for } j \text{ odd} \\ \int_{[0,1)} \int_{[0,1)} \left(\frac{z(1-x_{n,2j-1})}{z(1-x_{n,2j-1}) + y x_{n,2j-1}} \right)^2 + \left(\frac{y x_{n,2j-1}}{z(1-x_{n,2j-1}) + y x_{n,2j-1}} \right)^2 d\mu(y) d\mu(z) & \text{for } j \text{ even.} \end{cases}
$$

The last equality holds by Definition 3.1 and the assumption that $\mu_0(\{0,1\}) = 0 = \mu(\{0\})$. Furthermore, by $\mu_0(\{0,1\}) = 0 = \mu(\{0\})$ and Lemma 3.7, for a fixed $\epsilon < \frac{1}{2}$, there exists a $K < 1$ and an interval $[\alpha, \beta] \subset (0,1)$, for which

$\mu_0([\alpha,\beta]^C), \mu([\alpha,\beta]^C) < \epsilon$, such that for all $n \geq 2$

$$\int \left[\sum_{j=1}^{2^{n-1}} (F_{M(\omega)}(m_{n-1,j}(\omega)) - F_{M(\omega)}(m_{n-1,j-1}(\omega)))^2 f(x_{n,2j-1}(\omega)) \right] dP(\omega)$$

$$= \sum_{j=1}^{2^{n-1}} \int_{\{\omega : x_{n,2j-1}(\omega)\in[\alpha,\beta]\}} \left[(F_{M(\omega)}(m_{n-1,j}(\omega)) - F_{M(\omega)}(m_{n-1,j-1}(\omega)))^2 f(x_{n,2j-1}(\omega)) \right] dP(\omega)$$
$$+ \sum_{j=1}^{2^{n-1}} \int_{\{\omega : x_{n,2j-1}(\omega)\in[\alpha,\beta]^C\}} \left[(F_{M(\omega)}(m_{n-1,j}(\omega)) - F_{M(\omega)}(m_{n-1,j-1}(\omega)))^2 f(x_{n,2j-1}(\omega)) \right] dP(\omega)$$

$$\leq \sum_{j=1}^{2^{n-1}} \int_{\{\omega : x_{n,2j-1}(\omega)\in[\alpha,\beta]\}} \left[(F_{M(\omega)}(m_{n-1,j}(\omega)) - F_{M(\omega)}(m_{n-1,j-1}(\omega)))^2 K \right] dP(\omega),$$
$$+ \sum_{j=1}^{2^{n-1}} \int_{\{\omega : x_{n,2j-1}(\omega)\in[\alpha,\beta]^C\}} \left[(F_{M(\omega)}(m_{n-1,j}(\omega)) - F_{M(\omega)}(m_{n-1,j-1}(\omega)))^2 \right] dP(\omega)$$

$$= \sum_{j=1}^{2^{n-1}} K \int \left[(F_{M(\omega)}(m_{n-1,j}(\omega)) - F_{M(\omega)}(m_{n-1,j-1}(\omega)))^2 \right] dP(\omega),$$
$$+ \sum_{j=1}^{2^{n-1}} (1-K) \int_{\{\omega : x_{n,2j-1}(\omega)\in[\alpha,\beta]^C\}} \left[(F_{M(\omega)}(m_{n-1,j}(\omega)) - F_{M(\omega)}(m_{n-1,j-1}(\omega)))^2 \right] dP(\omega)$$

$$= KE_{n-1}$$
$$+ (1-K) \sum_{j=1}^{2^{n-2}} \Bigg[\int_{\{\omega : x_{n,4j-1}(\omega)\in[\alpha,\beta]^C\}} (F_{M(\omega)}(m_{n-2,j}(\omega)) - F_{M(\omega)}(m_{n-2,j-1}(\omega)))^2 \times \left(\frac{m_{n,4j-1}(\omega) - m_{n,4j-2}(\omega)}{m_{n,4j-1}(\omega) - m_{n,4j-3}(\omega)} \right)^2 \, dP(\omega)$$
$$+ \int_{\{\omega : x_{n,4j-3}(\omega)\in[\alpha,\beta]^C\}} (F_{M(\omega)}(m_{n-2,j}(\omega) - F_{M(\omega)}(m_{n-2,j-1}(\omega))^2 \left(\frac{m_{n,4j-2}(\omega) - m_{n,4j-3}(\omega)}{m_{n,4j-1}(\omega) - m_{n,4j-3}(\omega)} \right)^2 dP(\omega) \Bigg]$$

$$\leq KE_{n-1} + (1-K) \sum_{j=1}^{2^{n-2}} \left[\int (F_{M(\omega)}(m_{n-2,j}(\omega)) - F_{M(\omega)}(m_{n-2,j-1}(\omega)))^2 (\epsilon + \epsilon) dP(\omega) \right]$$

$$= KE_{n-1} + 2\epsilon(1-K)E_{n-2}$$

Thus, letting $E_0 = 1$, we get $E_n \to 0$. □

Lemma 3.7. *Let*

$$f(x) = \int_{[0,1)} \int_{[0,1)} \left(\frac{xz}{xz + (1-x)y} \right)^2 + \left(\frac{(1-x)y}{xz + (1-x)y} \right)^2 d\mu(y) d\mu(z).$$

If $\mu(\{0,1\}) = 0$, then for all intervals $[\alpha, 1-\alpha] \subset (0,1)$ there exists a $K < 1$ such that $f(x) \le K$, for all $x \in [\alpha, 1-\alpha]$. The same result holds for $g(x) = f(1-x)$.

Proof. Let $p_x(y,z) = \frac{xz}{xz+(1-x)y}$ and let $\epsilon > 0$ be such that $\mu([\epsilon, 1-\epsilon]) > \frac{1}{2}$. Then

$$f(x) = \int_{[0,1)} \int_{[0,1)} (p_x(y,z))^2 + (1 - p_x(y,z))^2 \, d\mu(y) d\mu(z)$$

$$= 1 + 2 \Bigg[\int\limits_{[\epsilon,1-\epsilon]} \int\limits_{[\epsilon,1-\epsilon]} (p_x(y,z))^2 d\mu(y) d\mu(z) + \iint\limits_{([\epsilon,1-\epsilon]\times[\epsilon,1-\epsilon])^c} (p_x(y,z))^2 d\mu(y) d\mu(z)$$

$$- \int\limits_{[0,1)} \int\limits_{[0,1)} p_x(y,z) d\mu(y) d\mu(z) \Bigg]$$

$$\le 1 + 2 \Bigg[\frac{x(1-\epsilon)}{x(1-\epsilon) + (1-x)\epsilon} \int\limits_{[\epsilon,1-\epsilon]} \int\limits_{[\epsilon,1-\epsilon]} p_x(y,z) d\mu(y) d\mu(z) + \iint\limits_{([\epsilon,1-\epsilon]\times[\epsilon,1-\epsilon])^c} p_x(y,z) d\mu(y) d\mu(z)$$

$$- \int\limits_{[0,1)} \int\limits_{[0,1)} p_x(y,z) d\mu(y) d\mu(z) \Bigg]$$

$$= 1 - 2 \left[\frac{(1-x)\epsilon}{x(1-\epsilon) + (1-x)\epsilon} \int\limits_{[\epsilon,1-\epsilon]} \int\limits_{[\epsilon,1-\epsilon]} p_x(y,z) d\mu(y) d\mu(z) \right]$$

$$\le 1 - \frac{1}{2} \left(\frac{(1-x)\epsilon}{x(1-\epsilon) + (1-x)\epsilon} \right) \left(\frac{x\epsilon}{x\epsilon + (1-x)(1-\epsilon)} \right)$$

Hence, for $x \in [\alpha, 1-\alpha]$,

$$f(x) \le 1 - \frac{1}{2} \left(\frac{\alpha\epsilon}{(1-\alpha)(1-\epsilon) + \alpha\epsilon} \right)^2 < 1. \square$$

Theorem 3.8. *Let $\rho = \mu(\{0\})$, and let $N(F)$ be the number of jumps of F.*

(i) *If $\rho > 0$, then for all μ_0, $B_{(\mu_0,\mu)}$-almost all measures are discrete.*

(ii) *If $\rho \geq 1 - \frac{1}{\sqrt{2}}$, then for all μ_0, $B_{(\mu_0,\mu)}$-almost all measures have finite support.*

(iii) *If $\rho > 1 - \frac{1}{\sqrt{2}}$, then for all μ_0, $E_{B_{(\mu_0,\mu)}}[N] < \infty$.*

Proof. (i) Let $S(F)$ denote the sum of the jumps of a distribution function F. Let

$$J(m) = \int S(F)\; dB_{(\delta_m,\mu)}(F),$$

and set

$$J = \int S(F)\; dB_{(\mu_0,\mu)}(F) = \int J(m)\; d\mu_0(m).$$

To prove (i), it is enough to show that $J = 1$ and this obviously follows if $J(m) = 1$ for all $0 \leq m \leq 1$.

Clearly, $J(0) = 1 = J(1)$. Suppose $0 < m < 1$, and let

$$p_m(x,y) = \frac{(1-m)y}{(1-m)y + m(1-x)} = 1 - q_m(x,y).$$

Then, by Definition 3.1, Theorem 2.10 and the self-similarity of the sequential barycenter rpm construction,

$$J(m) = \rho + \rho(1-\rho) + \int_{(0,1)} \int_{(0,1)} J(1-x)p_m(1-x,y) + J(y)q_m(1-x,y)\; d\mu(x)d\mu(y).$$

Now set $R = \rho + \rho(1-\rho)$ and use induction to show that

$$J(m) \geq R + R(1-\rho)^2 + R(1-\rho)^4 + \cdots + R(1-\rho)^{2n}$$

for all $n \geq 1$. Thus, $J(m) \geq \frac{R}{1-(1-\rho)^2} = 1$. But, $J(m) \leq 1$; and so $J(m) = 1$, for all $0 \leq m \leq 1$.

(ii) Note that if $m_{n,2j-2} < m_{n,2j-1} < m_{n,2j}$ and either $X_{n+1,4j-3} = 0$ or $X_{n+1,4j-1} = 0$, then the $B_{(\mu_0,\mu)}$ rpm gives positive probability to the point $m_{n,2j-1}$ and probability zero to the set $(m_{2j-2}, m_{2j-1}) \cup (m_{2j-1}, m_{2j}]$. The idea is to use this fact in constructing a branching process whose extinction corresponds to the generation of a sequential barycenter measure with finite support. Specifically, let $\{Z_{i,n}\}$ be iid random variables such that

$$P[Z_{i,n} = 0] = \rho + \rho(1-\rho) = 1 - P[Z_{i,n} = 2].$$

Set $Y_1 \equiv 1$, and, for $n \geq 1$, let

$$Y_{n+1} = \sum_{i=1}^{Y_n} Z_{i,n}.$$

Then, $Y_1, Y_2, \ldots$ is a branching process and, by the sequential barycenter rpm construction,

$$B_{(\mu_0,\mu)}(\{\text{measures with finite support}\}) = \mu_0(\{0,1\}) + (1-\mu_0(\{0,1\})) \lim_{n\to\infty} P[Y_n = 0].$$

Standard results (Ross (1970, Theorem 4.12)) for branching processes yield $\lim_{n\to\infty} P[Y_n = 0] = 1$, if $\rho \geq 1 - \frac{1}{\sqrt{2}}$.

(iii) As indicated by the branching process constructed above, the number of points in the support of a generated sequential barycenter measure does not depend on the mean m of the measure, as long as $0 < m < 1$. That is, for any $0 < m_1, m_2 < 1$

$$E_{B_{(\delta_{m_1},\mu)}}[N] = E_{B_{(\delta_{m_2},\mu)}}[N].$$

Denote this common value by $E[N]$. Then

$$E[N] = 1 \cdot (\rho + \rho(1-\rho)) + (1-\rho)^2 2E[N].$$

Thus, for $\rho > 1 - \frac{1}{\sqrt{2}}$ and $R = \rho + \rho(1-\rho)$, $E[N] = \frac{R}{2R-1}$. Hence,

$$E_{B_{(\mu_0,\mu)}}[N] = 1 \cdot [\mu_0(\{0,1\}) + R(1-\mu_0(\{0,1\}))] + (1-\mu_0(\{0,1\}))(1-\rho)^2 \cdot 2E[N] < \infty,$$

if $\rho > 1 - \frac{1}{\sqrt{2}}$. □

Example 3.9. Suppose $\mu_0 = \mu = \frac{1}{2}\delta_{\frac{1}{2}} + \frac{1}{2}\delta_0$. Then $B_{(\mu_0,\mu)}$-almost all measures have finite support on the set of binary rationals and $E_{B_{\mu_0,\mu}}[N] = \frac{3}{2}$.

Often, a desirable property for random probability measures is that they have large or full support. Recall that a probability ν defined on a compact Hausdorff space $\mathcal{H}$ has **full support** if every nonempty open subset of $\mathcal{H}$ has positive ν-measure. Note that this is equivalent to $\mathcal{H}$ being the smallest compact set which has ν-measure one. The next theorem gives conditions on μ_0 and μ which ensure that $B_{\mu_0,\mu}$ has full support. Let $supp(\nu)$ denote the support of measure ν.

Theorem 3.10. *If μ_0 and μ have full support on $[0,1]$, then $B_{(\mu_0,\mu)}$ has full support on $\mathcal{P}([0,1])$.*

Proof. It is enough to show that each set in a base for the weak* topology of $\mathcal{P}([0,1])$ has positive $B_{(\mu_0,\mu)}$ measure. One such base (see Billingsley (1968, Appendix III)) consists of sets of the form

$$\{\sigma \in \mathcal{P}([0,1]) :\ \sigma(O_i) > \tau(O_i) - \epsilon, \quad i = 1, \ldots, k\}$$

where each O_i is an open subset of $[0,1]$, τ is a measure in $\mathcal{P}([0,1])$, and $\epsilon > 0$. Each open set in $[0,1]$ is a countable disjoint union of open intervals. Thus, another base for the weak* topology of $\mathcal{P}([0,1])$ is

$$A = \{\sigma \in \mathcal{P}([0,1]) :\ \sigma(\cup_{j=1}^{r} O_{i,j}) > \tau(\cup_{j=1}^{r} O_{i,j}) - \epsilon, \quad i = 1, \ldots, k\} \tag{3.3}$$

where, for each i, the $O_{i,j}$ are disjoint open intervals in $[0,1]$. So it suffices to show that $B_{(\mu_0,\mu)}(A) > 0$ for sets A of form (3.3). This will follow if $B_{(\mu_0,\mu)}(C) > 0$ for C equal to

$$C = \cap_{i,j}\{\sigma \in \mathcal{P}([0,1]) :\ \sigma(O_{i,j}) > \tau(O_{i,j}) - \frac{\epsilon}{r}, \quad i = 1, \ldots, k\}, \tag{3.4}$$

since $C \subset A$.

Let F_τ and $\{m_{n,k}(F_\tau)\}$ denote the distribution function of τ and SBA of F_τ, respectively. If $O_{i,j} \cap supp(\tau) = \emptyset$ then

$$\{\sigma \in \mathcal{P}([0,1]) :\ \sigma(O_{i,j}) > \tau(O_{i,j}) - \frac{\epsilon}{r}, \quad i - 1, \ldots, k\} = \mathcal{P}([0,1]).$$

So, without loss of generality, assume that the intersection in (3.4) is over open intervals $O_{i,j}$ such that $O_{i,j} \cap supp(\tau) \neq \emptyset$. With this assumption and since $\{m_{n,k}(F_\tau)\}$ is dense in the support of τ, it follows that for each $O_{i,j}$ there exists $n_{i,j}, k_{i,j}$, and $l_{i,j}$ such that

$$(m_{n_{i,j},k_{i,j}}(F_\tau), m_{n_{i,j},l_{i,j}}(F_\tau)] \subset O_{i,j} \tag{3.5}$$

and

$$\tau((m_{n_{i,j},k_{i,j}}(F_\tau), m_{n_{i,j},l_{i,j}}(F_\tau)]) > \tau(O_{i,j}) - \frac{\epsilon}{2r} \tag{3.6.}$$

(Note that $n_{i,j}, k_{i,j}$, and $l_{i,j}$ exist such that (3.5) and (3.6) hold even if τ gives positive mass to some $m = m_{n,k}(F_\tau) \in O_{i,j}$, since the interval $(m, m] = \{m_{n,k}(F_\tau)\}$ will be in the partition $\{I_{n+1,k}\}_{k=1}^{2^{n+1}}$.)

Let $N = \max_{i,j} n_{i,j}$. By (2.4) of Theorem 2.10, there exists a δ such that

$$D = \{\sigma \in \mathcal{P}([0,1]) : \ |m_{N,k}(F_\sigma) - m_{N,k}(F_\tau)| < \delta, \text{ for all } k = 1, \ldots, 2^N - 1\} \subset C.$$

Now, given μ_0, μ have full support on $[0,1]$, it follows that $B_{(\mu_0,\mu)}(D) > 0$. $\square$

It is straightforward to modify the above proof to show the following.

Proposition 3.11. *If μ has full support on $[0,1)$, then $supp(B_{\delta_m,\mu}) = \{\sigma \in \mathcal{P}([0,1]) : \int x d\sigma = m\}$, for $0 \le m \le 1$.*

4. Applications.

Models for Mass Distribution. A basic problem in physics and biology is to describe various (random) distributions of mass in space, such as charged particles on a line, dispersion of species in a region, or galaxies in the universe. It is widely accepted, for example, that "the universe consists of clusters of galaxies containing clusters of solar systems containing clusters of planetary systems containing ..." and so on. The prevailing models for describing this distribution of mass in the universe (e.g., Peebles (1993)) are random distributions which specify the probability of finding a galaxy in a certain region of space.

For example, the *Poisson model* states that the probability P of finding a galaxy in differential volume dV at location r satisfies

$$dP = f(r)dV \tag{4.1}$$

for some given function f of position. This model, however, does not yield any of the clustering which has been observed experimentally, and a *clustering-hierarchy* (or fractal) *model* has been developed which does exhibit this phenomenon. In the clustering-hierarchy model, the probability $\hat{P}$ of finding a galaxy at distance r from a given galaxy satisfies

$$d\hat{P} = n(r)dV \tag{4.2}$$

and similarly for higher correlations (e.g., $d\hat{P} = n(r_1, r_2)dV_1 dV_2$ for finding galaxies in dV_1 and dV_2 at vector positions r_1 and r_2 from a given galaxy).

Both these models, and all similar "probability of finding a galaxy" models ignore the fundamental problem of precise definition of galaxy. In addition, the Poisson model has no clustering, and the clustering-hierarchy model is *inexact* in that it only specifies correlations up to three or four neighboring galaxies at the most, whereas true distributions reflect interactions between all groups of galaxies.

An alternative model for mass distribution which does not have these shortcomings can be obtained by using an SBA rpm $B_{\mu_0,\mu}$ for various μ_0 and μ. The description given below will specialize to the 1-dimensional case of mass along a line; for analogous models in two and three dimensions the reader is referred to [Hill and Monticino, to appear]. For an SBA rpm $B_{\mu_0,\mu}$, notice that if μ has most of its mass close to 0, then $B_{\mu_0,\mu}$ will automatically have the clustering property, since barycenters (and hence masses) will tend to be concentrated in clumps. The next two examples are representative of concrete mass-dispersion models which can be obtained in this manner; the two base distributions selected for these models were chosen because of the importance of power laws and the exponential distribution in current mass-distribution theories.

Example 4.1. (Power-law mass distribution model) Let μ_0 be arbitrary and let $\mu = \mu_\alpha$ have distribution

$$\mu[0, x] = x^\alpha \qquad \text{for some } 0 < \alpha \leq 1.$$

Then for SBA rpm B_{μ_0,μ_α}, "tighter" clustering occurs for smaller values of α, and in the limiting case $\alpha = 0$ there is total clustering (dirac measure).

Example 4.2. (Exponential-law mass distribution model) Let μ_0 be arbitrary, and let $\mu = \mu_\beta$ have distribution

$$\mu[0, x] = (1 - e^{-\beta x})/(1 - e^{-\beta}) \qquad \text{for } \beta > 0.$$

For this SBA rpm $B_{\mu_0,\mu}$, clustering increases as β increases.

Note that purely atomic analogs of these two models can easily be obtained by placing positive mass at zero also, i.e., letting $\mu = q\delta_0 + (1-q)\mu_\alpha$ for some $q \in (0, 1)$.

Such SBA rpm models for mass-distribution are not only easy to construct and implement, but they also have several other significant advantages.

1. *Richness.* Various choices of base measures μ_0 and μ can model a great variety of clustering (and even "anti-clustering" if μ is concentrated near $\frac{1}{2}$), and can easily be extended to constructions where the successive barycenters are dependent probabilistically, or in time, or position (or all three).

2. *Scale-invariance.* The SBA rpm's $B_{\mu_0,\mu}$ apply equally well at all scales from galactic to atomic (and beyond), and avoid the problem of a precise definition of galaxy or solar system.

3. *Exactness.* Unlike the clustering hierarchy and similar models which specify only partial (e.g., two- or three-point correlations), the $B_{\mu_0,\mu}$ model is exact.

4. *Gravitational stability.* Constructions based on the Poisson model (4.1) and clustering-hierarchy model (4.2) are gravitationally unstable in the sense that placing a new galaxy in a given configuration will alter the gravitational forces (and in general will result in a change of positions of the previously-placed galaxies). An SBA rpm $B_{\mu_0,\mu}$, on the other hand, has a gravitational stability not shared by those models, since in the $B_{\mu_0,\mu}$ construction, once a center of gravity (barycenter) for a particular subinterval is specified, the gravitational force between the mass on that subinterval and mass outside that subinterval remains constant for the rest of the construction, regardless of which random distribution is ultimately generated. This is a consequence of determining the centers of gravity rather than the masses directly (whose distribution can be calculated *a posteriori* by the inversion formula Theorem 2.10).

Experimental Approximation of Universal Constants. Given a continuous function $f : \mathcal{P}[0,1] \to \mathbb{R}$, suppose the universal bound

$$\phi(f,m) := \sup\{f(F) : F \in \mathcal{P}[0,1], b_F = m\}$$

is to be determined. By the continuity of f (convergence in distribution) and Proposition 3.11, the following proposition gives an experimental method to approximate ϕ. Let λ denote Lebesgue measure on $[0,1]$.

Proposition 4.3. *Fix $m \in (0,1)$ and let $F_1, F_2, \dots$ be iid $B_{\delta_m,\lambda}$. Then*

$$\max_{1\le i\le n} f(F_i) \nearrow \phi(f,m) \quad a.s. \tag{4.3}$$

Example 4.4. Suppose the sharp bound $c_{m,h}$ is desired for the inequality

$$E[h(X-m)] \leq c_{m,h} \quad \text{for all} \quad 0 \leq X \leq 1 \quad \text{with} \quad E[X] = m$$

for some continuous $h : \mathbb{R} \to \mathbb{R}$. (E.g., if $h(x) = x^2$, then $c_{m,h} = m - m^2$, which is simply the familiar inequality Var $X \leq m - m^2$ if $0 \leq X \leq 1$ and $E[X] = m$.)

Letting $f(F) = \int h(x-m)dF(x)$, it follows from Proposition 4.3 that if $F_1, F_2, \ldots$ are constructed independently with distribution $B_{\delta_m,\lambda}$, then

$$\max_{1 \leq i \leq n} \int h(x-m)dF_i(x) \nearrow c_{m,h} \quad \text{a.s.}$$

Remarks. The assumption that $0 \leq X \leq 1$ is not essential, and generalization to more general supports can be accomplished by replacing λ by any measure with full support in the desired range. Note that this method also approximates extremal distributions (those nearly attaining c_m) by simply keeping track of which F_i attain $\max_{1 \leq i \leq n} f(F_i)$.

The reason for use of an SBA rpm $B_{\delta_m,\lambda}$ in Proposition 4.3 (or any other $B_{\delta_m,\mu}$ with $supp(\mu) = [0,1]$) is simply that such constructions produce rpm's with the prescribed mean, whereas previous standard constructions do not.

Average-Optimal Control Problems. Suppose a function $g : \mathcal{P}([0,1]) \times \mathbb{R} \to \mathbb{R}$ is given, and the objective is to choose c (the control parameter) so as to make $g(F,c)$ as large as possible, on the average, over all distributions F on $[0,1]$ with given mean m. The SBA rpm $B_{\delta_m,\lambda}$ is a natural prior for randomly choosing elements of $\mathcal{P}([0,1])$ with mean m, since it chooses the successive barycenters uniformly at each stage. Under this prior, the above average-optimal control problem simply becomes

$$\text{choose } c^* \text{ to } \underset{c}{\text{maximize}} \int g(F,c)dB_{\delta_m,\lambda}(F).$$

Typical control problem objectives of this type include: picking the control to keep a process (or random variable) within a certain range with high probability, e.g., find c^* to make $P(a \leq X + c^* \leq b)$ as large as possible, on the average, over all distributions in $P([0,1])$ with mean m; and the following control problem from optimal stopping theory.

Example 4.5. Suppose a stop rule t is to be chosen for stopping a sequence of three random variables X_1, X_2, X_3, knowing only that the $\{X_i\}$ are independent, take values in $[0,1]$, and have identical means m. What stop rule will make EX_t as large as possible, on the average, over all such $\{X_i\}$? By standard backward induction (Chow, Robbins and Siegmund, 1971), it is clear that there is an optimal stop rule t_c of the form

$$\{t_c = 2\} \Leftrightarrow \{t_c > 1\} \cap \{X_i > m\}$$

and

$$\{t_c = 1\} \Leftrightarrow \{X_1 > c\}.$$

Using this stop rule t_c on X_1, X_2, X_3 yields expected reward

$$EX_{t_c} = \int_{X_1 > c} X_1 + P(X_1 \leq c) \int_{X_2 > m} X_2 + mP(X_1 \leq c)P(X_2 \leq m),$$

and the objective now is reduced to finding c^* which will make EX_{t_c} as large as possible, on the average, over all independent $\{X_i\}$ taking values in $[0,1]$ and having mean m. As in classical optimal stopping theory, the optimal value for c is exactly the expected reward of continuing past time one using t_c (e.g., if the distributions of X_2 and X_3 are known, then

$$c = \int_{X_2 > m} X_2 + P(X_2 \leq m)m). \tag{4.4}$$

In the present setting where only the means and bounds for the $\{X_i\}$ are known, however, the optimal c depends on the prior for X_1, X_2, X_3, which in the case of $B_{\delta_m,\lambda}$ would mean the optimal value of c is the $B_{\delta_m,\lambda}$-average of (4.4), namely

$$c^* = \int \left[\int_{x > m} x dF(x) + mF(m) \right] dB_{\delta_m,\lambda}(F).$$

Using the definition of $B_{\delta_m,\lambda}$, it can be seen that in this case

$$c^* = c^*_m = \int_0^1 \int_0^1 \left(\frac{m(1-m)y}{mx + (1-m)y} + \frac{(m + (1-m)y)(mx + (1-m)y)}{mx} \right) dx\, dy.$$

References

Billingsley, P., *Probability and Measure (2nd edition)*, John Wiley & Sons, New York, 1986.

Billingsley, P., *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

Chow, Y. S., Robbins, H. and Siegmund, D., *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin Company, Boston, 1971.

Cifarelli, D. M. and Regazzini, E., *Distribution functions of means of Dirichlet process*, Ann. Stat. **18** (1990), no. 1, 429-442.

Dubins, L. E. and Freedman, D. A., *Random distribution functions*, Proc. Fifth Berkeley Symp. Math. Statist. Prob. **2** (1967), 183–214.

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

Ferguson, T. S., *Prior distributions on spaces of probability measures*, Ann. Stat. **2** (1974), no. 4, 615–629.

Graf, S., Mauldin, R. D., and Williams, S. C., *Random homeomorphisms*, Advances in Math. **60** (1986), 239–359.

Hill, T. and Monticino, M., *Barycenter models for distribution of mass in 2 and 3 dimensions*, in preparation.

Mauldin, R. D., Sudderth, W. D., and Williams, S. C., *Polya trees and random distributions*, Ann. Stat. **20** (1992), no. 3, 1203–1221.

Mauldin, R. D. and Monticino, M. G., *Randomly generated distributions*, Israel Journal of Mathematics **91** (1995), 215–237.

Monticino, M., *Constructing prior distributions with trees of exchangeable processes*, to appear, Journal of Statistical Planning and Inference (1996).

Monticino, M., *A note on the moments of the mean for a Dubins-Freedman prior*, University of North Texas, Department of Mathematics Technical Report (1995).

Peebles, P. J. E., *Principles of Physical Cosmology*, Princeton University Press, Princeton, New Jersey, 1993.

Ross, S. M., *Applied Probability Models with Optimization Applications*, Dover Publications, Inc., New York, 1970.