Moments of the mean of Dubins-Freedman random probability distributions

Pieter Allaart
Mathematics Department
University of North Texas

May 14, 2001

Abstract

Explicit formulas are given to recursively generate the moments of the mean M for Dubins-Freedman random distribution functions with arbitrary base measure μ. Using a standard inversion formula for moments of a distribution on the unit interval, the distribution of M is approximated for several natural choices of μ. The support of the mean is also considered. It is shown that the support of M is connected whenever μ is concentrated on the vertical bisector of the unit square S, but may have arbitrarily many gaps otherwise.

1 Introduction

In their 1967 paper, Dubins and Freedman [3] introduced the following method for constructing a probability distribution on the unit interval $[0,1]$ at random. Fix a base measure μ on the unit square $S=[0,1]^2$. Pick a point (x,y) in S at random according to this measure μ. The chosen point divides S into four rectangles, each having a vertex at (x,y). Call the lower left rectangle L, and the upper right one R. The unique affine transformations that map S into L and R, respectively, scale the base measure μ into measures μ_L and μ_R, concentrated on L and R, respectively. Next, select a point (x_L,y_L) in L according to μ_L, and a point (x_R,y_R) in R according to μ_R. These points now determine lower left and upper right rectangles within L and R (four rectangles in all), and respective scaled versions of μ. Continuing in this manner one obtains, at the nth stage, a set of 2^n closed rectangles. Dubins and Freedman show that, if μ assigns mass 0 to the corners $(0,0)$ and $(1,1)$ of S, with probability 1 the intersection of these sets of rectangles is the solid graph of a distribution function F.

The above procedure has been generalized by Kraft [8], Graf, Mauldin and Williams [6], and Mauldin and Monticino [11]. Other methods for constructing a probability measure at random were proposed by Ferguson [5], Mauldin and Williams [10], and Mauldin, Sudderth and Williams [9]. Hill and Monticino [7] used barycenter arrays to construct a distribution at random with a given mean or distribution of the mean. Their work was recently extended by Bloomer [1] to generate distributions with a given mean and variance.

None of the other methods mentioned above generate distributions with a fixed mean and, in general, the distribution of the mean is difficult to compute. Cifarelli and Regazzini [2] calculated the distribution of the mean for Dirichlet process priors (see [5]) using advanced analytical tools. For Dubins-Freedman priors, some work in this direction was done for the case where the base measure μ is the uniform distribution on the vertical bisector $D_1=\{(x,y)\in S|x=1/2\}$. Mauldin and Williams [10] calculated the variance of the mean M, as an application of their Polya tree construction. Generalizing their method, Monticino [12] obtained the complete moment sequence of the mean (given implicitly by a difference equation).

The distribution of the mean is a very useful tool in nonparametric Bayesian statistics. Specifically, a decisionmaker trying to produce inference concerning the mean of a population may have prior beliefs about the possible values of the mean. If a nonparametric prior is assumed, it should be compatible with these prior beliefs. Hence to identify appropriate nonparametric priors, it is important to assess their induced distribution of the mean.

The main purpose of this paper is to show that the moments of the mean can be calculated directly, using only the statistical self-similarity of the Dubins-Freedman construction (and without reference to Polya tree schemes). In fact, the method outlined below works for any base measure μ on the square, though calculation of the moments in most cases requires the numerical approximation of certain integrals with respect to μ. However, for several natural choices of μ, such as the uniform distribution on S or the uniform distribution on D_1, the moment sequence can be generated exactly to arbitrary depth, and in fact in these cases all of the moments turn out to be rational.

Theorem 2.1 below provides a recursive relation that determines the moments of the mean M_μ implicitly for any base measure μ. Since a distribution on a compact interval is uniquely determined by its moment sequence, the distribution of M_μ can then be approximated using a standard inversion formula. Examples are given for several natural choices of μ. Some special attention is given to base measures concentrated on the vertical bisector D_1, and it is shown that for such measures μ, the variance of M_μ is tightly connected to the variance of μ (see Corollary 2.2).

In Section 3, the support of M_μ is considered. As a first result, it is shown that M_μ is degenerate if and only if μ is supported on a certain curve with endpoints $(0,0)$ and $(1,1)$. This surprising fact follows immediately from Theorem 2.1. Its consequences should not be overestimated, since the class of distributions that can be obtained using such a base measure is too small to induce a useful prior. As an extreme case, the only way to generate distributions with almost-sure mean $1/2$ is to generate the uniform distribution on $[0,1]$ almost surely.

Next, it is shown that whenever the base measure μ is supported on the vertical bisector D_1, the support of M_μ is connected (that is, a point or an interval). That this may fail for more general base measures is illustrated in Example 3.3, which shows that the number of gaps in the support of M_μ may be arbitrarily large. Some sufficient conditions for the support of M_μ to be connected are given next. The author does not know an abstract characterization of those base measures for which the mean has a connected support.

In Section 4 it is shown that the mapping $\mu \to M_\mu$ reverses the stochastic ordering of probability measures. That is, if μ and ν are (one-dimensional) base measures supported on the vertical bisector D_1 and μ is stochastically less than ν, then M_μ is stochastically larger than M_ν.

Section 5, finally, contains a list of open problems.

2 Moments of the mean

The construction described in the Introduction can be formalized as follows. Call a probability measure μ on S a *base measure* if it assigns mass 0 to the corners $(0,0)$ and $(1,1)$. Fix such a base measure μ, and set $U_{1,0} = V_{1,0} = 0$, and $U_{1,2} = V_{1,2} = 1$. Next, choose a point $P_{1,1} = (X_{1,1}, Y_{1,1})$ in S according to μ, and set $(U_{1,1}, V_{1,1}) = (X_{1,1}, Y_{1,1})$. This completes stage 1. Proceeding inductively, assume that after stage n, a set of points $\{(U_{k,j}, V_{k,j}) : 1 \leq k \leq n, 0 \leq j \leq 2^k\}$ has been constructed. Select points $P_{n+1,j} = (X_{n+1,j}, Y_{n+1,j})$, $j = 1, 2, \ldots, 2^n$ in S according to μ, independently of each other and of all the points $P_{k,j}$, $1 \leq k \leq n$, $j = 1, 2, \ldots, 2^{k-1}$, chosen at earlier stages. Define

$$(U_{n+1,2j}, V_{n+1,2j}) = (U_{n,j}, V_{n,j}), \qquad j = 0, 1, \ldots, 2^n,$$

and

$$U_{n+1,2j+1} = U_{n,j} + X_{n+1,j+1}(U_{n,j+1} - U_{n,j}),$$
$$V_{n+1,2j+1} = V_{n,j} + X_{n+1,j+1}(V_{n,j+1} - V_{n,j}),$$

$j = 0, 1, \ldots, 2^n - 1$.

For each n, consider the closed rectangles parallel to the axes whose lower left and upper right vertices are $(U_{n,j}, V_{n,j})$ and $(U_{n,j+1}, V_{n,j+1})$, respectively ($j = 0, 1, \ldots, 2^n - 1$). There are 2^n such rectangles. Let A_n be their union. Then, as proved by Dubins and Freedman ([3], Section 2), with probability 1 the limit $\bigcap_{n=1}^{\infty} A_n$ is the solid graph of some distribution function F.

Assume from now on that F is a random probability distribution function constructed by the above procedure, using a base measure μ. The mean of F is the random variable

$$M_\mu := \int_{[0,1]} x \, dF(x).$$

When no confusion about the underlying base measure is possible, the subscript μ will be dropped and M_μ will be simply denoted by M.

For $n \geq 0$, let a_n be the n-th moment of M:

$$a_n = EM^n, \qquad n = 0, 1, 2, \ldots.$$

Of course, $a_0 = 1$.

Theorem 2.1 *The sequence $\{a_n\}_n$ is determined implicitly by the recursive equations*

$$a_n = \sum_{j=0}^{n} \sum_{k=0}^{n-j} \binom{n}{j} \binom{n-j}{k} c_{n,j,k} a_j a_k, \qquad n = 0, 1, 2, \ldots, \tag{1}$$

where

$$c_{n,j,k} = \int\int_S x^{n-k}(1-x)^k y^j (1-y)^{n-j} \, d\mu(x,y).$$

Note that the right hand side of (1) depends on a_n. Subtracting the terms containing a_n and solving for a_n gives

$$a_n = (1 - c_{n,n,0} - c_{n,0,n})^{-1} \left(\sum_{j=1}^{n-1} \sum_{k=0}^{n-j} \binom{n}{j} \binom{n-j}{k} c_{n,j,k} a_j a_k + \sum_{k=0}^{n-1} \binom{n}{k} c_{n,0,k} a_k \right), \tag{2}$$

a form more suitable for direct computation of the sequence $\{a_n\}$.

An important class of base measures are those supported on the vertical bisector of S, the line segment $D_1 = \{(x,y) \in S | x = 1/2\}$. Any measure μ in this class can be written as a product measure of point mass at $1/2$ and some probability measure ν on $[0,1]$. To keep notation simple, μ and ν will be identified. Observe that if μ is concentrated on D_1, the equations (1) simplify to

$$a_n = \left(\frac{1}{2}\right)^n \sum_{j=0}^{n} \sum_{k=0}^{n-j} \binom{n}{j} \binom{n-j}{k} d_{n,j} a_j a_k, \tag{3}$$

where

$$d_{n,j} = \int_{[0,1]} y^j (1-y)^{n-j} \, d\mu(y).$$

Corollary 2.2 *Suppose μ is supported on D_1. Let $m_i := \int x^i \, d\mu(x)$, $i = 1, 2, \dots$, and let $\sigma^2 = m_2 - m_1^2$ denote the variance of μ. Then the variance of M is given by*

$$Var(M) = \frac{\sigma^2}{3 + 2(m_1 - m_2)} = \frac{\sigma^2}{3 + 2\{m_1(1 - m_1) - \sigma^2\}}.$$

As an immediate consequence,

$$\frac{\sigma^2}{\frac{7}{2} - 2\sigma^2} \le Var(M) \le \frac{\sigma^2}{3}, \tag{4}$$

where the lower bound is attained whenever $m_1 = 1/2$.

Proof. Calculate a_1 and a_2 using (2), and simplify the expression $Var(M) = a_2 - a_1^2$. $\square$

Note that (4) shows that the variances of μ and M are intimately related: bounding the left side further and assuming $\sigma^2 \neq 0$ one obtains $2/7 \le Var(M)/\sigma^2 \le 1/3$.

Several examples now follow. The density graphs in Figures 1, 2 and 3 were obtained by approximating the distribution function of M using the standard inversion formula (cf. Feller [4], page 223)

$$P(M \le t) = \lim_{n \to \infty} \sum_{k \le nt} \sum_{j=0}^{n-k} \binom{n}{k} \binom{n-k}{j} (-1)^j a_{k+j}.$$

Example 2.3 If μ is the uniform distribution on S, then

$$a_n = \frac{1}{(n+1)^2} \sum_{j=0}^{n} \sum_{k=0}^{n-j} \frac{\binom{n-j}{k}}{\binom{n}{k}} a_j a_k.$$

Thus, $a_1 = 1/2$, $a_2 = 17/56$, $a_3 = 23/112$, $a_4 = 2801/18816$, etc., and $Var(M) = 3/56$. An estimate of the density of M using the first 128 moments is given in Figure 1.

Example 2.4 If μ is the uniform distribution on D_1, then

$$a_n = \frac{1}{2^n(n+1)} \sum_{j=0}^{n} \sum_{k=0}^{n-j} \binom{n-j}{k} a_j a_k. \tag{5}$$

Thus, $a_1 = 1/2$, $a_2 = 11/40$ (as found by Mauldin and Williams [10]), $a_3 = 13/80$, $a_4 = 12661/124800$, etc., and $Var(M) = 1/40$. Equation (5) is significantly simpler and more efficient than the equations given by Monticino [12], who computed the a_n for this special case using a completely different technique.

An estimate of the density of M appears in Figure 2.

Example 2.5 If μ is uniform on the diagonal from $(0, 1)$ to $(1, 0)$, then

$$a_n = \frac{1}{2n+1} \sum_{j=0}^{n} \sum_{k=0}^{n-j} \frac{\binom{n}{j}\binom{n-j}{k}}{\binom{2n}{j+k}} a_j a_k.$$

Thus, $a_1 = 1/2$, $a_2 = 19/56$, $a_3 = 29/112$, $a_4 = 207107/984704$, etc., and $Var(M) = 5/56$.

An estimate of the density of M is given in Figure 3.

Example 2.6 Let μ be supported this time on the horizontal bisector $D_2 = \{(x, y) \in S | y = 1/2\}$, and denote by μ^* the reflection of μ in the main diagonal of S. Thus μ^* is a base measure supported on D_1. Note that choosing F at random with base measure μ is equivalent to choosing F^{-1} at random using the base measure μ^*. From the relationship

$$M(F) = \int_{[0,1]} x \; dF(x) = \int_{[0,1]} F^{-1}(t) \; dt = 1 - M(F^{-1}),$$

it follows that $M_\mu \stackrel{d}{=} 1 - M_{\mu^*}$. In particular, if μ is symmetric, $M_\mu \stackrel{d}{=} 1 - M_\mu$ so both μ and μ^* induce the same distribution of the mean. For example, if μ is the uniform distribution on D_2, the moments of M are given by Example 2.4, and its density appears in Figure 2.

Corollary 2.7 *If μ is concentrated on D_1 or on D_2, then μ is uniquely determined by the distribution of M.*

Proof. Using (3), and expressing $d_{n,j}$ in terms of the moments of μ, it follows easily that the a_n's determine the moments of μ. Since μ is supported on a bounded interval, μ is determined by its moments. □

Proof of Theorem 2.1. For brevity, let $(X, Y) = (X_{1,1}, Y_{1,1})$, the first point in the random construction of F. Define

$$B_1 = Y^{-1} \int_{[0,X]} x \; dF(x),$$
$$B_2 = (1 - Y)^{-1} \int_{(X,1]} x \; dF(x).$$

Then B_1 and B_2, if well defined, are the F-barycenters of $[0, X]$ and $(X, 1]$, respectively, and

$$M = \begin{cases} B_2, & \text{if} \quad Y = 0, \\ B_1, & \text{if} \quad Y = 1, \\ YB_1 + (1 - Y)B_2, & \text{if} \quad 0 < Y < 1. \end{cases}$$

From the construction of F it is clear that B_1 and B_2 are conditionally independent given (X, Y). Moreover, provided $0 < Y < 1$,

$$(B_1 | (X, Y)) \stackrel{d}{=} XM, \qquad \text{and} \qquad (B_2 | (X, Y)) \stackrel{d}{=} X + (1 - X)M.$$

Thus,

$$\begin{aligned} E[M^n | (X, Y)] &= E[(YB_1 + (1 - Y)B_2)^n I_{\{0<Y<1\}} | (X, Y)] \\ &\qquad + E[B_1^n I_{\{Y=1\}} | (X, Y)] + E[B_2^n I_{\{Y=0\}} | (X, Y)] \\ &= \sum_{j=0}^{n} \binom{n}{j} Y^j (1 - Y)^{n-j} E[B_1^j | (X, Y)] E[B_2^{n-j} | (X, Y)] \\ &= \sum_{j=0}^{n} \binom{n}{j} Y^j (1 - Y)^{n-j} X^j a_j \sum_{k=0}^{n-j} \binom{n-j}{k} X^{n-j-k} (1 - X)^k a_k \\ &= \sum_{j=0}^{n} \sum_{k=0}^{n-j} \binom{n}{j} \binom{n-j}{k} Y^j (1 - Y)^{n-j} X^{n-k} (1 - X)^k a_j a_k. \end{aligned}$$

Taking expectations on both sides completes the proof. □

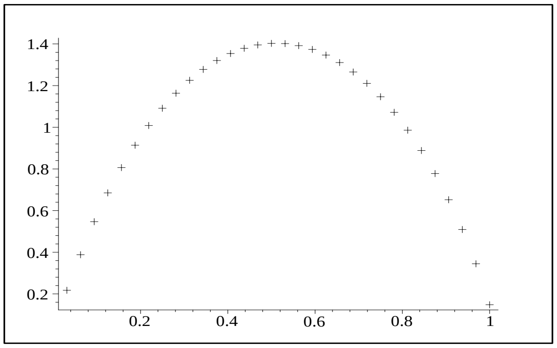

Figure 1: Estimated density of M for the base measure μ in Example 2.3, based on 128 moments.

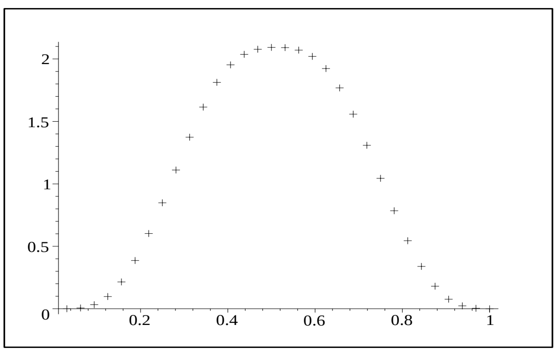

Figure 2: Estimated density of M for the base measure μ in Example 2.4, based on 128 moments.

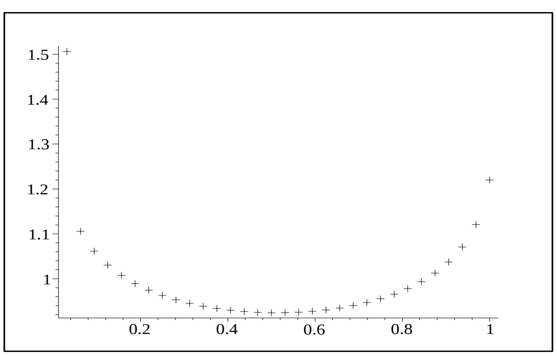

Figure 3: Estimated density of M for the base measure μ in Example 2.5, based on 128 moments.

3 The support of M

This section considers the support of the mean M. The following notation will be used in this section. First, recall from the beginning of Section 2 the definition of the random points $P_{n,j} = (X_{n,j}, Y_{n,j})$, used in the construction of F. Observe particularly that these points are i.i.d. with common distribution μ. For a base measure μ, $\operatorname{supp}(\mu)$ denotes the support of μ (that is, the smallest closed subset of S to which μ assigns measure 1). Similarly, $\operatorname{supp}(M)$ is the support of M, the smallest closed subset A of $[0, 1]$ such that $P(M \in A) = 1$. For sets A and B of real numbers, $A \oplus B$ denotes the set $\{a + b | a \in A \text{ and } b \in B\}$. The length of an interval I is denoted by $|I|$, but if I is a finite set of points then $|I|$ denotes the cardinality of I. Finally, for a point p in S and $r > 0$, $B(p, r)$ denotes the open disk with center p and radius r.

The first result in this section describes the set of base measures μ for which the support of M is degenerate. To gain some insight in this surprising theorem, assume for the moment that μ assigns probability 1 to a single point (r, s) in the interior of S. Then $(X_{n,j}, Y_{n,j}) = (r, s)$ for all $n \geq 1$ and $j = 1, 2, \ldots, 2^{n-1}$. Hence, the portions of the graph of F lying inside the lower left and upper right rectangles determined by (r, s) are scaled copies of the entire graph of F. (In other words, the graph of F is a self-similar set). It follows that the barycenter of F on $[0, r]$ is $b_1 = rM$, and the barycenter on $(r, 1]$ is $b_2 = r + (1 - r)M$. Thus,

$$\begin{aligned} M &= sb_1 + (1 - s)b_2 = srM + (1 - s)\{r + (1 - r)M\} \\ &= \{rs + (1 - r)(1 - s)\}M + r(1 - s), \end{aligned}$$

and hence $M = r(1 - s)/(r + s - 2rs)$.

Conversely, suppose a "target" mean m is given, where $0 < m < 1$. Then $M = m$ if and only if $r(1 - s) = m(r + s - 2rs)$. Solving for s gives

$$s = \frac{(1 - m)r}{m + (1 - 2m)r}. \tag{6}$$

It is then natural to guess that for *any* base measure μ concentrated on the set of points (r, s) satisfying (6), $M = m$ almost surely. Theorem 3.1 below shows that this is indeed the case.

Define the family of curves

$$C(m) := \left\{ (x, y) \in S : y = \frac{(1 - m)x}{m + (1 - 2m)x} \right\}, \qquad 0 < m < 1.$$

Let $C(0) := \{(x, y) \in S : x = 0 \text{ or } y = 1\}$, and $C(1) := \{(x, y) \in S : x = 1 \text{ or } y = 0\}$. For $0 < m < 1$, $C(m)$ is a curve whose endpoints are $(0, 0)$ and $(1, 1)$. Moreover, $C(m)$ is convex if $m > 1/2$, concave if $m < 1/2$, and degenerates to the main diagonal of S if $m = 1/2$.

Say a set A is *bounded above* by $C(m)$ if for each point (x, y) in A there exists a $z \geq y$ such that $(x, z) \in C(m)$. Similarly, A is *bounded below* by $C(m)$ if for every (x, y) in A there exists a $z \leq y$ such that $(x, z) \in C(m)$.

Theorem 3.1 *Let* $0 \leq m \leq 1$.

(i) $M = m$ *a.s. if and only if* μ *is supported on* $C(m)$.

(ii) If $\operatorname{supp}(\mu)$ *is bounded above by* $C(m)$, *then* $M \geq m$ *a.s.*

(iii) If $\operatorname{supp}(\mu)$ *is bounded below by* $C(m)$, *then* $M \leq m$ *a.s.*

Proof. To prove the "if" part of (i), it is clearly sufficient to show that if $\mathrm{supp}(\mu) \subset C(m)$, then $EM = m$ and $Var(M) = 0$. This is a straightforward exercise, using Theorem 2.1. The "only if" part follows since if $\mathrm{supp}(\mu)$ intersects $C(m_1)$ and $C(m_2)$ for $m_1 \neq m_2$, then M can be arbitrarily close to both m_1 and m_2 with positive probability. The other statements follow easily from (i) by monotonicity considerations. □

It is worth noting that if $m = 1/2$, $C(m)$ is the main diagonal of S, and thus the distribution generated is the uniform distribution almost surely. But if $m \neq 1/2$, truely random distributions are generated, though their variability is small. To generate distributions with a fixed mean, it is much better to use the method described by Hill and Monticino [7].

Theorem 3.2 *Suppose μ is concentrated on the vertical bisector D_1. If μ is not point mass, then* $\mathrm{supp}(M)$ *is the interval* $[1-b, 1-a]$, *where*

$$a = \inf\{y : \mu([0,y]) > 0\}, \qquad b = \sup\{y : \mu([y,1]) > 0\}.$$

(So a and b are the left and right endpoints of $\mathrm{supp}(\mu)$, *respectively.)*

Proof. From the definition of a and b, it follows that for every $\varepsilon > 0$, $\mu[a, a+\varepsilon] > 0$ and $\mu[b-\varepsilon, b] > 0$. The proof involves the construction of binary trees with intervals as their nodes, of the form

$$T = \left\{J_{k,j} : 1 \le k \le n, j = 1, 2, \ldots, 2^{k-1}\right\}. \tag{7}$$

For such a tree, denote $J_{1,1}$ by $\mathrm{root}(T)$, and let

$$T^l = \left\{J_{k,j} : 2 \le k \le n, j = 1, 2, \ldots, 2^{k-2}\right\}$$

and

$$T^r = \left\{J_{k,j} : 2 \le k \le n, j = 2^{k-2}+1, \ldots, 2^{k-1}\right\}$$

denote the left and right subtrees of T, respectively. Thus T is uniquely determined by the triple $(\mathrm{root}(T), T^l, T^r)$.

If T is given by (7), let $\{Y \in T\}$ denote the event that $Y_{k,j} \in J_{k,j}$ for all $1 \le k \le n$, $j = 1, 2, \ldots, 2^{k-1}$. Define

$$m^-(T) = \sup\{u : P(M > u | Y \in T) = 1\},$$

and

$$m^+(T) = \inf\{v : P(M < v | Y \in T) = 1\}.$$

Thus $P(m^-(T) \le M \le m^+(T) | Y \in T) = 1$.

We will now construct an array of intervals $\{I_{n,k} : n \ge 0, 0 \le k \le 2^n - 1\}$ and an array of trees $\{T_{n,k} : n \ge 0, 0 \le k \le 2^n - 1\}$ with the following properties:

$$|I_{n,k}| \le (3/4)^n, \tag{8}$$

$$\bigcup_{k=1}^{2^n-1} I_{n,k} = [1-b, 1-a], \tag{9}$$

$$I_{n,k} \supset [m^-(T_{n,k}), m^+(T_{n,k})], \tag{10}$$

and

$$P(Y \in T_{n,k}) > 0. \tag{11}$$

The proof is then completed by the following argument. Given $1-b \le c < d \le 1-a$, there exist, by virtue of (8) and (9), integers n and k $(0 \le k \le 2^n - 1)$ such that $I_{n,k} \subset [c,d]$. Using (10) and (11) it then follows that

$$\begin{aligned} P(c < M < d) &\ge P(M \in I_{n,k}) \\ &\ge P(M \in I_{n,k} | Y \in T_{n,k}) P(Y \in T_{n,k}) \\ &\ge P(m^-(T) \le M \le m^+(T) | Y \in T_{n,k}) P(Y \in T_{n,k}) \\ &= P(Y \in T_{n,k}) > 0. \end{aligned}$$

To construct $I_{n,k}$ and $T_{n,k}$, observe that for any binary tree of intervals T, if $\text{root}(T) = [a, a+\varepsilon]$, then

$$\begin{aligned} m^-(T) &\ge \frac{a+\varepsilon}{2} m^-(T^l) + \frac{1-a-\varepsilon}{2}\{m^-(T^r)+1\}, \\ m^+(T) &\le \frac{a}{2} m^+(T^l) + \frac{1-a}{2}\{m^+(T^r)+1\}, \end{aligned} \tag{12}$$

while if $\text{root}(T) = [b-\varepsilon, b]$,

$$\begin{aligned} m^-(T) &\ge \frac{b}{2} m^-(T^l) + \frac{1-b}{2}\{m^-(T^r)+1\}, \\ m^+(T) &\le \frac{b-\varepsilon}{2} m^+(T^l) + \frac{1-b+\varepsilon}{2}\{m^+(T^r)+1\}. \end{aligned} \tag{13}$$

These inequalities follow since, when considered as a function of $\{Y_{n,j}\}_{n,j}$, M is nonnondecreasing in each of its arguments. The above inequalities suggest the following construction. For brevity, let $\varepsilon_n := (1/2)^n$. Set $I_{0,0} = [1-b, 1-a]$ and, for $n \ge 0$ and $0 \le k \le 2^n - 1$, construct $I_{n+1,2k}$ and $I_{n+1,2k+1}$ from $I_{n,k} = [l, r]$ by

$$\begin{aligned} I_{n+1,2k} &= [l/2 + (1-a-\varepsilon_{n+1})/2, r/2 + (1-a)/2], \\ I_{n+1,2k+1} &= [l/2 + (1-b)/2, r/2 + (1-b+\varepsilon_{n+1})/2]. \end{aligned}$$

It is straightforward to verify inductively that the array $\{I_{n,k}\}$ satisfies (8) and (9). Next, set $T_{1,0} = \{[a, a+1/2]\}$ and $T_{1,1} = \{[b-1/2, b]\}$, and for $n \ge 1$ and $0 \le j \le 2^{n+1} - 1$, define $T_{n+1,j}$ by

(i) $\text{root}(T_{n+1,j}) = \begin{cases} [a, a+\varepsilon_{n+1}], & j \quad \text{even}, \\ [b-\varepsilon_{n+1}, b], & j \quad \text{odd}. \end{cases}$

(ii) $T^l_{n+1,j} = T^r_{n+1,j} = T_{n,[j/2]}$.

Since each interval in $T_{n,k}$ is of the form $[a, a+\varepsilon]$ or $[b-\varepsilon, b]$ for some $\varepsilon > 0$, (11) follows immediately by the independence of the $Y_{k,j}$'s. Moreover, a routine induction proof using (12) and (13) establishes (10). This completes the construction, and thereby the proof. □

Theorem 3.2 implies that for any base measure μ concentrated on D_1, $\text{supp}(M)$ is connected. This may fail for more general base measures. As an extreme example, suppose μ gives mass 1/2 to each of the points $(0,1)$ and $(1,0)$. Then F is point mass at either 0 or 1, so $\text{supp}(M) = \{0,1\}$. As the following example shows, the support of M can, in general, have an arbitrarily large number of gaps.

Example 3.3 Let $0 < r < 1/2$, and consider the base measure μ that gives mass $1/2$ to each of the points $(r, 1-r)$ and $(1-r, r)$. Let $\mathcal{T}_n$ be the set of all binary trees T with points at their nodes, of the form

$$T = \{p_{k,j} : 1 \le k \le n, j = 1, 2, \ldots, 2^{k-1}\}, \tag{14}$$

where each $p_{k,j}$ is either $(r, 1-r)$ or $(1-r, r)$. By convention, $\mathcal{T}_0 = \{\emptyset\}$, the set consisting only of the empty tree. For a tree T given by (14), define $\text{root}(T) := p_{1,1}$, and define the left subtree T^l and right subtree T^r by $T^l := \{p_{k,j} : 2 \le k \le n,\ j = 1, 2, \ldots, 2^{k-2}\}$, and $T^r := \{p_{k,j} : 2 \le k \le n,\ j = 2^{k-2}+1, 2^{k-2}+2, \ldots, 2^{k-1}\}$. Write $(X, Y) \in T$ if $(X_{k,j}, Y_{k,j}) = p_{k,j}$ for all $1 \le k \le n$ and $j = 1, 2, \ldots, 2^{k-1}$. Define

$$m^-(T) = \sup\{u : P(M > u | (X, Y) \in T) = 1\},$$
$$m^+(T) = \inf\{v : P(M < v | (X, Y) \in T) = 1\}.$$

Since M takes its minimum value when $P_{n,j} = (r, 1-r)$ for all n and j, it follows that $m^-(\emptyset)$ is the solution of $x = r(1-r)x + (1-r)rx + r^2$, so $m^-(\emptyset) = r^2[1 - 2r(1-r)]^{-1}$. By symmetry, $m^+(\emptyset) = 1 - m^-(\emptyset) = (1-r)^2[1-2r(1-r)]^{-1}$. If T is a non-empty tree (of depth $n \ge 1$), then $\text{root}(T)$ can take on two different values. If $\text{root}(T) = (r, 1-r)$, then

$$\begin{aligned} m^-(T) &= r(1-r)\{m^-(T^l) + m^-(T^r)\} + r^2, \\ m^+(T) &= r(1-r)\{m^+(T^l) + m^+(T^r)\} + r^2. \end{aligned} \tag{15}$$

On the other hand, if $\text{root}(T) = (1-r, r)$, then

$$\begin{aligned} m^-(T) &= r(1-r)\{m^-(T^l) + m^-(T^r)\} + (1-r)^2, \\ m^+(T) &= r(1-r)\{m^+(T^l) + m^+(T^r)\} + (1-r)^2. \end{aligned} \tag{16}$$

Since $m^+(\emptyset) - m^-(\emptyset) = (1-2r)/(1-2r(1-r))$, (15) and (16) imply that for every tree $T \in \mathcal{T}_n$,

$$\begin{aligned} m^+(T) - m^-(T) &= \frac{1-2r}{1-2r(1-r)}(2r(1-r))^n \\ &=: \lambda_n. \end{aligned}$$

Note that $\lambda_n \to 0$ as $n \to \infty$. Thus, the support of M can be expressed as

$$\text{supp}(M) = \bigcap_{n=1}^{\infty} \Delta_n,$$

where

$$\Delta_n = \bigcup_{T \in \mathcal{T}_n} [m^-(T), m^+(T)].$$

For $n \ge 0$, let $L_n = \{m^-(T) : T \in \mathcal{T}_n\}$, and define $d_n := (1-2r)(r(1-r))^n$. From (15) and (16) it follows that $L_{n+1} = A \cup B$, where $A = r(1-r)(L_n \oplus L_n) + r^2$ and $B = r(1-r)(L_n \oplus L_n) + (1-r)^2$. Notice that $(1-r)^2 - r^2 = d_0$. A routine induction argument now shows that

$$L_n = \{m^-(\emptyset) + \sum_{i=0}^{n-1} k_i d_i : 0 \le k_i \le 2^i, 0 \le i \le n-1\}. \tag{17}$$

Define

$$n^* := \max\{n : \lambda_n < d_{n-1}\} = \max\left\{n : 2^n < \frac{1-2r(1-r)}{r(1-r)}\right\}.$$

Claim. If $n < n^*$, then for all $x \neq y \in L_n$, $|y - x| > \lambda_n$.

Proof. Let $x = m^-(\emptyset) + \sum_{i=0}^{n-1} k_i d_i$, and $y = m^-(\emptyset) + \sum_{i=0}^{n-1} l_i d_i$. Let $i_0 := \min\{i : k_i \neq l_i\}$. Assume w.l.o.g. that $k_{i_0} < l_{i_0}$. Let $q = 2r(1-r)$. Then

$$\begin{aligned}
y - x &= \sum_{i=i_0}^{n-1} (l_i - k_i) d_i \geq d_{i_0} - \sum_{i=i_0+1}^{n-1} 2^i d_i = d_{i_0} - \sum_{i=i_0+1}^{n-1} (1-2r) q^i \\
&= (1-2r) \left[(r(1-r))^{i_0} - \frac{q^{i_0+1}}{1-q} (1 - q^{n-i_0-1}) \right] \\
&= \frac{1-2r}{1-2r(1-r)} \left[(r(1-r))^{i_0} \{1 - 2r(1-r) - 2^{i_0+1} r(1-r)\} + (2r(1-r))^n \right] \\
&> \frac{1-2r}{1-2r(1-r)} (2r(1-r))^n = \lambda_n. \qquad \square
\end{aligned}$$

Thus, for $n \leq n^*$, Δ_n consists of $|L_n|$ disjoint, non-touching intervals. For $n > n^*$, $d_{n-1} \leq \lambda_n$ and therefore $\Delta_n = \Delta_{n-1}$, since no more new gaps between the intervals are created. It follows that $\text{supp}(M) = \Delta_{n^*}$. The number of intervals in $\text{supp}(M)$ is $|L_n| = 2 \prod_{k=1}^{n-1} (2^k + 1)$. For instance, if $r = 1/10$, then $n^* = 3$, and $\text{supp}(M)$ consists of $|L_3| = 30$ disjoint intervals. $\square$

The above example raises the natural question when the support of M is a connected set. One sufficient condition was given in Theorem 3.2; another is given by the following proposition. Define

$$D(\mu) = \{m \in [0,1] | \, \text{supp}(\mu) \cap C(m) \not\subseteq \{(0,0), (1,1)\}\},$$

and let $m_l := \inf D(\mu)$ and $m^u := \sup D(\mu)$.

Proposition 3.4

$$D(\mu) \subset \text{supp}(M) \subset [m_l, m^u].$$

Proof. The second inclusion is an immediate consequence of Theorem 3.1. The first inclusion can be seen as follows. For each $m \in D(\mu)$ and each $\delta > 0$, there is a point $p = p(m) \in C(m)$ such that $\mu(B(p, \delta) \cap S) > 0$. Given $\varepsilon > 0$, we can choose $\delta > 0$ and n sufficiently large so that

$$(P_{k,j} \in B(p(m), \delta) \text{ for all } 1 \leq k \leq n, j = 1, 2, \ldots, 2^{k-1}) \Rightarrow m - \varepsilon < M < m + \varepsilon. \tag{18}$$

Since the event on the left side of (18) has positive probability, this completes the proof. $\square$

Corollary 3.5 *If for every $m_l \leq m \leq m^u$, $\text{supp}(\mu)$ intersects $C(m)$ at a point other than $(0,0)$ or $(1,1)$, then $\text{supp}(M) = [m_l, m^u]$.*

Corollary 3.6 *If $\text{supp}(\mu) - \{(0,0), (1,1)\}$ is connected, then $\text{supp}(M)$ is connected.*

Proof. For each $m_l \leq m \leq m^u$, $\text{supp}(\mu)$ intersects $C(m)$ at a point other than $(0,0)$ or $(1,1)$. $\square$

4 The stochastic ordering reversed

The final result of this paper concerns base measures concentrated on the vertical bisector D_1. For such measures, the mapping $\mu \to M_\mu$ reverses the stochastic ordering $\leq^{st}$, as will be shown below. Recall that if X and Y are random variables, then X is stochastically less than Y, denoted $X \leq^{st} Y$, iff $P(X > t) \leq P(Y > t)$ for all t. Similarly, if μ and ν are probability measures on $[0,1]$, then $\mu \leq^{st} \nu$ iff $\mu((t,1]) \leq \nu((t,1])$ for all t.

Theorem 4.1 *If μ and ν are base measures supported on D_1 and $\mu \leq^{st} \nu$, then $M_\nu \leq^{st} M_\mu$.*

Proof. For an event A in the sigma algebra of the array $\{Y_{n,j} : n \geq 1, j = 1, \ldots, 2^{n-1}\}$, let $P_\mu(A)$ denote the probability of A when the distribution of the $Y_{n,j}$ is μ. Define $P_\nu(A)$ similarly. For $n \geq 1$, define the vector

$$\vec{Y}_n = \left(Y_{1,1}, Y_{2,1}, Y_{2,2}, Y_{3,1}, Y_{3,2}, Y_{3,3}, Y_{3,4}, \ldots, Y_{n,1}, \ldots, Y_{n,2^{n-1}}\right) \in \mathbb{R}^{2^n - 1}.$$

Given $\vec{Y}_n = \vec{x}$, the values of the random distribution function F at the points $k/2^n$, $k = 0, 1, \ldots, 2^n$, are completely determined. Say $F(k/2^n) = z_{n,k}$, where the $z_{n,k}$ are easily expressible in terms of the components of $\vec{x}$. Define

$$F^-(t) = \sum_{k=0}^{2^n - 1} z_{n,k} \chi_{[k/2^n, (k+1)/2^n)}(t),$$
$$F^+(t) = \sum_{k=0}^{2^n - 1} z_{n,k+1} \chi_{[k/2^n, (k+1)/2^n)}(t).$$

Let $m^-(\vec{x})$ and $m^+(\vec{x})$ be the means associated with the distributions F^- and F^+ respectively. That is, $m^-(\vec{x}) = \int t dF^-(t)$, and $m^+(\vec{x}) = \int t dF^+(t)$. It is easy to see that m^- and m^+ are decreasing functions from $\mathbb{R}^{2^n - 1}$ to $\mathbb{R}$. Define $M_n^- = m^-(\vec{Y}_n)$, and $M_n^+ = m^+(\vec{Y}_n)$. Clearly, M_n^- increases almost surely and M_n^+ decreases almost surely. Moreover,

$$M_n^- \leq M \leq M_n^+,$$

and

$$M_n^+ - M_n^- = (1/2)^n,$$

where the last equality follows since the mean of a distribution function G on $[0, 1]$ is the area of the region in S that lies above the graph of G. Thus it follows that

$$P_\mu(M \leq t) = \lim_{n \to \infty} P_\mu(M_n^- \leq t), \tag{19}$$

and likewise with P_μ replaced by P_ν. Hence the proof will be finished once it is shown that

$$P_\mu(m^-(\vec{Y}_n) \leq t) \leq P_\nu(m^-(\vec{Y}_n) \leq t).$$

Since $m^- : \mathbb{R}^{2^n - 1} \to \mathbb{R}$ is decreasing, this follows immediately from Theorem 1.A.3.(b) of Shaked and Shanthikumar [13]. □

5 Open problems

This paper ends with a list of open problems that deserve further attention.

1. Which distributions on $[0, 1]$ are the distribution of the mean for some base measure μ, concentrated (a) on S? (b) on D_1?

2. Prove or disprove that the distribution of M is either degenerate or continuous. Under what conditions does M have a density?

3. Find necessary and sufficient conditions on μ such that the support of M is connected.

4. Determine "natural" base measures for which the distribution of the mean can be calculated in explicit form.

5. Prove (or disprove) that if μ is concentrated on D_1 and μ is unimodal, then the distribution of M is unimodal.

References

[1] Bloomer, L. (2000). Random probability measures with given mean and variance. Ph.D. dissertation, Georgia Institute of Technology.

[2] Cifarelli, D. M. and Regazzini, E. (1990). Distribution functions and means of Dirichlet processes. *Ann. Statist.* **18**, no. 1, 429-442.

[3] Dubins, L. E. and Freedman, D. A. (1967). Random distribution functions. *Proc. Fifth Berkeley Symposium Math. Statist. Probl.* **2**, 183-214.

[4] Feller, W. (1966). *An introduction to probability theory and its applications*, Vol. 2. Wiley, New York.

[5] Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.

[6] Graf, S., Mauldin, R. D. and Williams, S. C. (1986). Random homeomorphisms. *Adv. Math.* **60**, 239-359.

[7] Hill, T. P. and Monticino, M. G. (1998). Constructions of random distributions via sequential barycenter arrays. *Ann. Statist.* **26**, no. 4, 1242-1253.

[8] Kraft, C. H. (1964). A class of distribution function processes which have derivatives. *J. Appl. Prob.* **1**, 385-388.

[9] Mauldin, R. D., Sudderth, W. D. and Williams, S. C. (1992). Polya trees and random distributions. *Ann. Statist.* **20**, no. 3, 1203-1221.

[10] Mauldin, R. D. and Williams, S. C. (1990). Reinforced random walks and random distributions. *Proc. Amer. Math. Soc.* **110**, no. 1, 251-258.

[11] Mauldin, R. D. and Monticino, M. G. (1995). Randomly generated distributions. *Israel J. Math.* **91**, 215-237.

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

[13] Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic orders and their applications.* Academic Press, London.

Appendix: Calculation for the proof of Theorem 3.1

Claim: If μ is supported on $C(m) = \left\{(x,y) \in S : y = \frac{(1-m)x}{m+(1-2m)x}\right\}$, then $EM = m$ and $Var(M) = 0$.

Proof. Note that $(x,y) \in C(m)$ is equivalent to

$$m\{xy + (1-x)(1-y)\} = m - x(1-y).$$

Hence if (X,Y) is a random point with distribution μ, then

$$m\{XY + (1-X)(1-Y)\} = m - X(1-Y) \quad a.s. \tag{20}$$

Recall $a_1 = EM$ and $a_2 = EM^2$, so it must be shown that $a_1 = m$ and $a_2 = m^2$. By Eq. (2),

$$a_1 = \frac{c_{1,0,0}}{1 - c_{1,1,0} - c_{1,0,1}},$$

so $a_1 = m$ iff $m(1 - c_{1,1,0} - c_{1,0,1}) = c_{1,0,0}$. Indeed, using (20):

$$\begin{aligned} m(1 - c_{1,1,0} - c_{1,0,1}) &= m(1 - E[XY + (1-X)(1-Y)]) \\ &= m - E[m - X(1-Y)] \\ &= E[X(1-Y)] \\ &= c_{1,0,0}. \end{aligned}$$

Next, substituting $n = 2$ and $a_1 = m$ in Eq. (2),

$$\begin{aligned} a_2 &= (1 - c_{2,2,0} - c_{2,0,2})^{-1} \left(\sum_{k=0}^{1} \binom{2}{1} c_{2,1,k} a_1 a_k + \sum_{k=0}^{1} \binom{2}{k} c_{2,0,k} a_k \right) \\ &= (1 - c_{2,2,0} - c_{2,0,2})^{-1} (2c_{2,1,0} m + 2c_{2,1,1} m^2 + c_{2,0,0} + 2c_{2,0,1} m). \end{aligned}$$

Hence $a_2 = m^2$ iff

$$m^2(1 - c_{2,2,0} - c_{2,0,2}) = 2m(c_{2,1,0} + c_{2,0,1}) + 2m^2 c_{2,1,1} + c_{2,0,0}.$$

Equivalently,

$$m^2(1 - c_{2,2,0} - c_{2,0,2} - 2c_{2,1,1}) = 2m(c_{2,1,0} + c_{2,0,1}) + c_{2,0,0}. \tag{21}$$

Indeed,

$$\begin{aligned} m^2(1 - c_{2,2,0} - c_{2,0,2} - 2c_{2,1,1}) &= m^2(1 - E[X^2Y^2 - (1-X)^2(1-Y)^2 - 2XY(1-X)(1-Y)]) \\ &= m^2(1 - E[XY + (1-X)(1-Y)]^2) \\ &= m^2 - E[m\{XY + (1-X)(1-Y)\}]^2 \\ &= m^2 - E[m - X(1-Y)]^2 \\ &= 2mE[X(1-Y)] - E[X^2(1-Y)^2], \end{aligned}$$

and

$$\begin{aligned} 2m(c_{2,1,0} + c_{2,0,1}) + c_{2,0,0} &= 2mE[X^2Y(1-Y) + X(1-X)(1-Y)^2] + E[X^2(1-Y)^2] \\ &= 2E[X(1-Y)m\{XY + (1-X)(1-Y)\}] + E[X^2(1-Y)^2] \\ &= 2E[X(1-Y)(m - X(1-Y))] + E[X^2(1-Y)^2] \\ &= 2mE[X(1-Y)] - E[X^2(1-Y)^2]. \end{aligned}$$

Hence (21) holds, and the proof is complete. □