Random Probability Measures with Given Mean and Variance

Running title: Random Probability Measures

by Lisa Bloomer and Theodore P. Hill*
School of Mathematics, Georgia Institute of Technology
Atlanta, GA 30332 USA. E-mail: hill@math.gatech.edu

Summary

This article describes several natural methods of constructing random probability measures with prescribed mean and variance, and focuses mainly on a technique which constructs a sequence of simple (purely discrete, finite number of atoms) distributions with the prescribed mean and with variances which increase to the desired variance. Basic properties of the construction are established, including conditions guaranteeing full support of the generated measures, and conditions guaranteeing that the final measure is discrete. Finally, applications of the construction method to optimization problems such as Plackett's Problem are mentioned, and to experimental determination of average-optimal solutions of certain control problems.

Key words and phrases: random distribution, random homeomorphisms, random probability measures, variance split, variance split array.

§1. Introduction

Suppose you want to construct a probability measure on $[0,1]$, at random, with given mean m and variance σ. The underlying objective might be to generate conjectures to the solution of a stochastic optimization problem like Plackett's' Problem (see below) or to an average-optimal control problem, or it might be to identify a suitable prior for a Bayesian problem involving distributions with given mean and variance. In such problems, it is usually desired that the random probability construction be natural, be easy to implement, and have support which is dense in the set of all probability measures with that mean and variance. The purpose of this article is to introduce several such constructions, with primary focus on a "variance-split array" method, and to discuss several applications. As such, these results complement earlier construction methods of Dubins and Freedman (1967), Ferguson (1974), Graf *et al.* (1986), Mauldin *et al.* (1992), Monticino (1998) and others, which did not generate distributions with given mean, and of Hill and Monticino (1998) which generated random distributions with given means but not variances. In each of these previous methods, the set of probability measures with given mean *and* variance is a null set, and even the calculation of the distributions of those unknown means and variances is difficult.

One way to pick a probability at random on $[0,1]$ is to fix a (large) integer n, and generate a probability with support on $n+1$ given points, say $\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\}$, by generating the *probabilities* $p_i = \text{Prob}(\{i/n\})$, $i = 0, \ldots, n$ uniformly at random. The set of values $\vec{p} = (p_0, p_1, \ldots, p_n)$ which yield a distribution with given mean m and variance σ^2 is the polyhedron formed by the intersection of solutions to linear equations of the form $\Sigma p_i = 1$, $\Sigma i p_i = a$, $\Sigma i^2 p_i = b$, and the positive cone $p_i \geq 0$, $i = 0, 1 \ldots, n$. Thus to pick a point in this region at random, one can (cf. Bloomer (2000)) calculate the extreme points, triangulate, pick a simplex at random proportionately according to size, and then pick a point in the simplex using an efficient algorithm. Drawbacks of this method are that it is computationally intensive, and it does not allow easy improvement from n to $n+1$.

Alternatively, one could fix the probabilities $p_0, p_1, \ldots, p_n$ (say uniformly $(n+1)^{-1}$), and then pick the distinct *locations* $a_0, a_1, \ldots, a_n$ of these masses at random, and so that the resulting distribution has correct mean and variance. Drawbacks for this method (cf.

Bloomer 2000) are the inefficiency of the standard method for picking a point uniformly in the intersection of the circle $\Sigma a_i^2 = c$ and the unit cube, and the existence of means and moments (e.g., $m = \frac{1}{2}$, $\sigma^2 = \frac{1}{4}$) for which no solution exists for even n.

Yet another method to pick a distribution at random on $[0,1]$ with given mean and variance, is to pick the higher order moments randomly, since the moments $\{EX^n\}$ of a compactly-supported distribution determine the distribution. Given the first n moments of a distribution on $[0,1]$, sharp upper and lower bounds are known (cf. Skibinsky (1968)) for the $(n+1)st$ moment, so starting with given mean and variance, one may choose the third moment uniformly in its range interval, and given that, next choose the fourth moment in its range interval (given the first three moments) and so on (cf. Bloomer 2000). The main drawback of this method is the inversion process for large n – reconstructing the measure from its moments. A thorough investigation of this technique remains to be done.

The focus of this article is a martingale-like method which constructs a sequence of simple distributions with the prescribed mean and with variances which increase to the desired variance. This will be done by introducing the notions of variance splits of a distribution, variance split arrays, and mean-variance arrays.

§2. Variance Splits

This section introduces the notion of a *variance-split* of a probability measure μ, that is, expressing μ as a convex combination of two other probability measures with equal variance, each strictly less than the variance of μ. This technique will be used in later sections to randomly generate a sequence of splittings which in turn lead to a random distribution with the desired mean and variance.

Here, and throughout, $\mathcal{P}$ denotes the set of Borel probability measures on $\mathbb{R}$, $\mathcal{P}([0,1])$ is the set of Borel probability measures on $[0,1]$, and $\mathcal{P}_{m,V}[0,1])$ is the set of Borel probability measures on $[0,1]$ with mean m and variance V; $V(\mu)$ is the variance of $\mu \in \mathcal{P}$; and δ_a is the Dirac point mass on a.

Definition 2.1. A pair of probability measures (μ_1, μ_2) is a *variance split* of the probability measure μ if $V(\mu_1) = V(\mu_2) < V(\mu)$ and there is a $p \in (0,1)$ so that $\mu = p\mu_1 + (1-p)\mu_2$.

The number p will be referred to as the *splitting probability.* The variance split (μ_1, μ_2) is called *simple* if ess sup $\mu_1 \leq$ ess inf μ_2.

Example 2.2. The pair (δ_0, δ_1) is a simple variance split for the Bernoulli measure $p\delta_0 + (1-p)\delta_1$, with splitting probability p.

The following example gives two different variance splits for the uniform measure on the unit interval, demonstrating that a variance split may not be unique.

Example 2.3. Suppose μ is the uniform measure on the unit interval. If μ_1 is the uniform measure on $[0, 1/2]$ and μ_2 is the uniform measure on $[1/2, 1]$, then (μ_1, μ_2) is a simple variance split of μ with splitting probability $1/2$, since $\mu = 1/2\mu_1 + 1/2\mu_2$ and $V(\mu_1) = V(\mu_2) = 1/48 < 1/12 = V(\mu)$.

A non-simple variance split of μ is the pair $(\hat{\mu}_1, \hat{\mu}_2)$, where $\hat{\mu}_1$ is the probability measure on the unit interval with cumulative distribution function $\hat{F}_1(x) = x^2$, $0 \leq x \leq 1$, and the c.d.f. for $\hat{\mu}_2$ is $\hat{F}_2(x) = 2x - x^2$, $0 \leq x \leq 1$. Here the splitting probability is $1/2$, and the variances of μ_1 and μ_2 are easily calculated to be $1/18$.

Theorem 2.4. *Every Borel probability measure with compact support has a simple variance split.*

Proof. Fix μ with compact support, let $F(x)$ be c.d.f. for μ, and let $p \in (0,1)$. For any x in $\mathbb{R}$, define $F_{Y_p}(x)$ and $F_{Z_p}(z)$ as follows:

$$F_{Y_p}(x) = \begin{cases} \frac{F(x)}{p} & \text{if } F(x) < p \\ 1 & \text{if } F(x) \geq p \end{cases} \tag{1}$$

and

$$F_{Z_p}(x) = \begin{cases} 0 & \text{if } F(x) < p \\ \frac{F(x)-p}{1-p} & \text{if } F(x) \geq p \end{cases}$$

Then

$$\begin{aligned}(p\mathcal{L}(Y_p) + (1-p)\mathcal{L}(Z_p))([0,x]) &= p\mathcal{L}(Y_p)([0,x]) + (1-p)\mathcal{L}(Z_p)([0,x]) \\ &= pF_{Y_p}(x) + (1-p)F_{Z_p}(x) = \begin{cases} p(\frac{F(x)}{p}) + (1-p)\cdot 0 & \text{if } F(x) < p \\ p \cdot 1 + (1-p)\frac{F_X(x)-p}{1-p} & \text{if } F(x) \geq p \end{cases} = F(x).\end{aligned} \tag{2}$$

Next note that

$$V(\mu) = V(Y_p)p + V(Z_p)(1-p) + (EY_p)^2 p + (EZ_p)^2(1-p) - (E\mu)^2.$$

Clearly the function $p \mapsto EY_p$ is continuous for any $n \in \mathbb{N}$, so $p \mapsto V(Y_p) = EY_p^2 - (EY_p)^2$ is continuous. Similarly, the function $p \mapsto V(Z_p)$ is continuous.

By (1), $F_{Y_p} \to F$ as $p \to 1$, and therefore $V(Y_p) \to V(\mu)$. Also, the support of Z_p decreases to a single point as $p \to 1$, which implies that $V(Z_p) \to 0$. Similarly, as $p \to 0$, the values $V(Z_p) \to V(\mu)$ and $V(Y_p) \to 0$. Therefore, the continuity of these functions implies that there must be some value $\hat{p}$ for which $V(Y_{\hat{p}}) = V(Z_{\hat{p}})$. Let $\mu_1 = \mathcal{L}(Y_{\hat{p}})$ and $\mu_2 = \mathcal{L}(Z_{\hat{p}})$. Then, (μ_1, μ_2) is a variance split for μ, which is simple by (1). □

(Note that the proof of Theorem 2.4 is easy for continuous μ, by conditioning μ to the right and left of a point b chosen so that the two variances are equal.)

The following example shows that a measure with non-compact support may not have a simple variance split.

Example 2.5. Let μ be the exponential distribution with mean 1. If $\mu = p\mu_1 + (1-p)\mu_2$, where $p \in (0,1)$ and ess sup $\mu_1 \leq$ ess inf μ_2, then clearly μ_2 is a shifted mean-1 exponential (density $f(x) = e^{-x+a}$ for $x \geq a$), so $V(\mu_2) = V(\mu)$ (and $V(\mu_2) > V(\mu_1)$) so (μ_1, μ_2) is not a variance split. (The exponential distribution does have non-simple variance splits, however, as can be seen by taking μ_1 with (renormalized) density e^{-x} on $S = \bigcup_{n=0}^{\infty} [2n, 2n+1)$ and zero elsewhere, and, similarly, μ_2 with renormalized density e^{-x} on $\mathbb{R} \backslash S$.)

§3 Variance Split Arrays

Definition 3.1. A triangular array $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is a *variance split array* for the probability measure μ if, for each $n \in \mathbb{N}$,

$$\mu = \sum_{k=1}^{2^{n-1}} \mu_{n,k} p_{n,k}; \tag{3}$$

and

for each $k \in \{1, \ldots, 2^{n-1}\}$, either $(\mu_{n+1,2k-1}, \mu_{n+1,2k})$ is a variance split of $\mu_{n,k}$ with splitting probability $p_{n+1,2k-1}/p_{n,k}$, or $\mu_{n,k} = \mu_{n+1,2k-1} = \mu_{n+1,2k} = \delta_{E\mu_{n,k}}$. (4)

The array is called *uniform* if $\max_k V(\mu_{n,k}) \to 0$ as $n \to \infty$, and is called *canonical* if for each $n \in \mathbb{N}$ and $k \in \{1, \ldots, 2^{n-1}\}$, $(\mu_{n+1,2k-1}, \mu_{n+1,2k})$ is a simple variance split of $\mu_{n,k}$ if $V(\mu_{n,k}) > 0$.

Example 3.2. If $\mu_{n,k}$ is the uniform distribution on $[\frac{k-1}{2^{n-1}}, \frac{k}{2^{n-1}}]$ and $p_{n,k} = 1/2^{n-1}$, then $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is a canonical variance split array for the uniform distribution on the unit interval. And, since $V(\mu_{n,k}) = (1/2^{n-1})^2(1/12) \to 0$ as $n \to \infty$, it is uniform.

Example 3.3. If $\mu_{n,k}$ is the uniform distribution on the set $[\frac{k-1}{2^n}, \frac{k}{2^n}] \cup [\frac{1}{2} + \frac{k-1}{2^n}, \frac{1}{2} + \frac{k}{2^n}]$, and $p_{n,k} = 1/2^{n-1}$, then $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is a variance split array for the uniform distribution on $[0, 1]$. However, $V(\mu_{n,k}) = 1/16 + (1/2^n)(1/12) \to 1/16$ as $n \to \infty$, so this array is not uniform. It is also not canonical, since the overlapping supports show the variance splits are not simple.

Theorem 3.4. *Every probability measure with compact support has a canonical variance split array, and every such array is uniform.*

Proof. If $V(\mu) = 0$, then $\{\delta_{E\mu}, 1/2^{n-1}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is a canonical variance split array for μ.

If $V(\mu) > 0$, let $\mu_{1,1} = \mu$ and $p_{1,1} = 1$.

For induction, assume that a canonical variance split array has been defined for μ up to row N. That is, there is an array $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{N}{}_{k=1}^{2^{n-1}}$ that satisfies conditions (3) and (4) for each $n \in \{1, \ldots, N\}$ and $k \in \{1, \ldots, 2^{n-1}\}$.

Now, let $k \in \{1, \ldots, 2^{N-1}\}$. If $V(\mu_{N,k}) = 0$, let $\mu_{N+1,2k-1} = \mu_{N+1,2k} = \delta_{E\mu_{N,k}}$ and $p_{N+1,2k-1} = p_{N+1,2k} = p_{N,k}/2$.

Suppose $V(\mu_{N,k}) > 0$. Since μ has compact support, and the support of $\mu_{N,k}$ is contained in $[\text{ess inf}\, \mu, \text{ess sup}\, \mu]$, $\mu_{N,k}$ has compact support. So, by Theorem 2.4 there is a simple variance split (μ_1, μ_2) for $\mu_{N,k}$ with splitting probability p. Let $\mu_{N+1,2k-1} = \mu_1$, $p_{N+1,2k-1} = p \cdot p_{N,k}$, $\mu_{N+1,2k} = \mu_2$, and $p_{N+1,2k} = (1-p) \cdot p_{N,k}$. Then

$$\begin{aligned}\sum_{i=1}^{2^N} p_{N+1,i}\mu_{N+1,i} &= \sum_{k=1}^{2^{N-1}} (p_{N+1,2k-1}\mu_{N+1,2k-1} + p_{N+1,2k}\mu_{N+1,2k}) \\ &= \sum_{k=1}^{2^{N-1}} p_{N,k}\mu_{N,k} = \mu,\end{aligned}$$

so this defines a canonical variance split array up to row $N+1$, which completes the induction, implying the existence of a canonical variance split array.

Next, it must be shown the array is uniform. For any variance split array, if $\mu_{n,k}$ has positive variance, then the variance of $\mu_{n+1,2k}$ is strictly smaller than that of $\mu_{n,k}$. This implies that $\max_k V(\mu_{n,k})$ is decreasing in n, and since this is a positive sequence, it has a limit.

Let $l(\mu) = \text{ess sup}(\mu) - \text{ess inf}(\mu)$, which is strictly positive since $V(\mu) > 0$. Let $s(\mu_{n,k})$ be the "sibling" of $\mu_{n,k}$, defined by $s(\mu_{n,k}) = \mu_{n,k-1}$ if k is even, and $s(\mu_{n,k}) = \mu_{n,k+1}$ if k is odd. Finally, let $a_i(\mu_{n,k})$ be the "i^{th} immediate ancestor" of $\mu_{n,k}$, defined by $a_i(\mu_{n,k}) = \mu_{n-1,\lceil k/2^i \rceil}$.

Suppose, by way of contradiction, that $\lim_{n\to\infty} \max_k V(\mu_{n,k}) = d > 0$. Since $\max_k V(\mu_{n,k})$ strictly decreases in n to d, for each $n \in \mathbb{N}$ there is a k_n so that $V(\mu_{n,k_n}) > d$. Let n be fixed. Since the largest possible variance for a probability measure on an interval with length l is $l^2/4$, it follows that $l(\mu_{n,k_n}) > 2\sqrt{d}$. Note, however, that $s(\mu_{n,k_n})$ has the same variance as μ_{n,k_n}, so $l(s(\mu_{n,k_n})) > 2\sqrt{d}$ as well. Since this is a canonical variance split array, $l(\mu_{n,k}) \geq l(\mu_{n+1,2k-1}) + l(\mu_{n+1,2k})$ for each n and k. So, $l(a_1(\mu_{n,k_n})) > 4\sqrt{d}$. Similarly, it is easy to see that $l(\mu) > 2n\sqrt{d}$. But since n was arbitrary, this implies the support of μ is unbounded, a contradiction. Hence, $\max_k V(\mu_{n,k}) \to 0$, and this variance split array is uniform. □

Proposition 3.5. *If $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is a canonical variance split array for a compactly supported probability measure μ, then for each n, there are numbers $A_{n,0} \leq A_{n,1} \leq \cdots \leq A_{n,2^{n-1}}$ so that* $\text{supp}(\mu_{n,k}) \subset [A_{n,k-1}, A_{n,k}]$.

Proof. Routine by the definition of simple variance split, and induction. □

The next result, that the means are dense in the support of the original measure, will be used in the main construction to follow.

Theorem 3.6. *If* $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ *is a uniform variance split array for* μ*, then* $\{E\mu_{n,k}\}$ *is dense in the support of* μ*.*

Proof. Let $x \in \text{supp}(\mu)$, let $\epsilon > 0$, and let $I_{x,\epsilon} = \{a \in \mathbb{R} : |x-a| < \epsilon/2\}$. Let k_n be defined so that $\mu_{n,k_n}(I_{x,\epsilon}) = \max_k \mu_{n,k}(I_{x,\epsilon})$. Since $x \in \text{supp}(\mu)$, $\mu(I_{x,\epsilon}) = \sum_k p_{n,k}\mu_{n,k}(I_{x,\epsilon}) > 0$, so $\mu_{n,k_n}(I_{x,\epsilon}) > 0$. Further, since each $\mu_{n,k}(I_{x,\epsilon})$ can be written as $\mu_{n,k}(I_{x,\epsilon}) = \mu_{n+1,2k-1}(I_{x,\epsilon})p_{n+1,2k-1}/p_{n,k} + \mu_{n+1,2k}(I_{x,\epsilon})p_{n+1,2k}/p_{n,k} = \mu_{n+1,2k-1}(I_{x,\epsilon})p' + \mu_{n+1,2k}(I_{x,\epsilon})(1-p')$, the number $\mu_{n,k}(I_{x,\epsilon})$ is a weighted average of two other numbers. Therefore, the maximum of these two numbers can be no smaller than $\mu_{n,k}(I_{x,\epsilon})$, so $\mu_{n,k_n}(I_{x,\epsilon})$ is increasing in n.

Consider the set $B_n = \{b \in \mathbb{R} : |E\mu_{n,k_n} - b| < \epsilon/2\}$. By Chebyshev's inequality, $\mu_{n,k_n}(B_n) \geq 1 - 4V(\mu_{n,k_n})/\epsilon^2$. Since the array is uniform, $\max_k V(\mu_{n,k})$ decreases to zero as $n \to \infty$, so $\mu_{n,k_n}(B_n)$ increases to 1.

This implies there exists an N so that $I_{x,\epsilon} \cap B_N \neq \emptyset$. Then for every point y in $I_{x,\epsilon} \cap B_N$, $|E\mu_{N,k_N} - x| \leq |E\mu_{N,k_N} - y| + |y-x| \leq \epsilon/2 + \epsilon/2 = \epsilon$. Hence, $\{E\mu_{n,k}\}$ is dense in the support of μ. □

If the variance split array is canonical, $A_{n,0} \leq E\mu_{n,1} \leq A_{n,1} \leq E\mu_{n,2} \leq \cdots \leq A_{n,2^{n-1}}$, but in general, it is not true that the sequence $\{A_{n,k}\}$ is dense in $\text{supp}(\mu)$, as the following example shows.

Example 3.7. Let μ be the probability measure $(\nu_1 + \delta_{1/3} + \delta_{2/3} + \nu_2)/4$, where ν_1 is uniform on $[0, 1/6]$ and ν_2 is uniform on $[5/6, 1]$. A simple variance split for μ is $((\nu_1 + \delta_{1/3})/2, (\delta_{2/3} + \nu_2)/2)$ with $A = 1/2$. If this simple split is used as the basis for a canonical variance split array, then $A_{n,2^{n-2}} = A_{2,1} = 1/2$. It can be seen that any further divisions of $(\nu_1 + \delta_{1/3})/2$ must contain a measure that places mass on 1/3 and on some subinterval of $[0, 1/6]$. Therefore, the interval $[A_{n,2^{n-2}-1}, A_{n,2^{n-2}}] \supset [1/6, 1/2]$, so there is no $A_{n,k}$ within 1/12 of 1/3.

On the other hand, the following theorem shows that the numbers $\{A_{n,k}\}$ are "almost dense," in that their closure contains every point in the support of μ that is not isolated.

Theorem 3.8. *Let* $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ *be a canonical variance split array for a compactly supported probability measure* μ*, and let* $\{A_{n,k}\}_{n=1}^{\infty}{}_{k=0}^{2^{n-1}}$ *be the sequence of numbers guaranteed by* Proposition 3.5. *Then, for every* $\epsilon > 0$*, and* $x, y \in \operatorname{supp}(x)$*,* $x < y$*, there are an* n *and a* k *so that* $A_{n,k} \in (x - \epsilon, y + \epsilon)$*.*

Proof. Let $0 < \hat{\epsilon} < \min\{\epsilon, y - x\}$. By Theorem 3.6, there exist integers N, M, i_N, and j_M such that $|E\mu_{N,i_N} - x| \leq \hat{\epsilon}/2$ and $|E\mu_{M,j_M} - y| \leq \hat{\epsilon}/2$. Without loss of generality, $N = M$. Clearly, $\mu_{N,i_N} \neq \mu_{N,j_N}$, since otherwise $|x - y| \leq |E\mu_{N,i_N} - x| + |E\mu_{N,i_N} - y| = |E\mu_{N,i_N} - x| + |E\mu_{N,j_N} - y| \leq \hat{\epsilon}/2 + \hat{\epsilon}/2 = \hat{\epsilon} < |x - y|$. Furthermore, since $|E\mu_{N,i_N} - x| \leq \epsilon/2$, then $E\mu_{N,i_N} > x - \epsilon$. Similarly, $E\mu_{N,j_N} < y + \epsilon$. Hence, $x - \epsilon < E\mu_{N,i_N} \leq A_{N,i_N} \leq E\mu_{N,j_N} < y + \epsilon$. □

§4. Constructions of Random Probabilities Using Variance Split Arrays

The purpose of this section is to use variance split arrays to construct a random probability measure with given mean and variance. Once it is shown that the mean-variance array parameters characterize a probability distribution, the idea is to randomly construct this array, inductively, by constructing two-point variance splits of previous measures at each stage. Continuing this process, and controlling the rate of increase of the variance, yields in the limit a random distribution with the correct mean and variance.

Definition 4.1. If $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is a uniform variance split array, then the array $\{E(\mu_{n,k}), V(\mu_{n,k}), p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is called the associated *mean-variance array.*

Theorem 4.2. *If* $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ *is a uniform variance split array for* μ*, then* μ *is uniquely determined by its mean-variance array.*

Let ρ denote the *Prohorov metric* on $\mathcal{P}$, that is $\rho(\mu_1, \mu_2) = \inf\{\delta > 0 : \mu_1(C) \leq \mu_2(C_\delta) + \delta$ and $\mu_2(C) \leq \mu_1(C_\delta) + \delta$ for all compact $C \subset [0,1]\}$, where $C_\delta = \{x : d(x, C) \leq \delta\}$.

The following two lemmas, whose proofs are routine, will be used in the proof of the theorem.

Lemma 4.3. *If μ has mean m and variance V, then $\rho(\mu, \delta_m) \leq \sqrt[3]{V}$.*

Lemma 4.4. *If $\mu = p\mu_1 + (1-p)\mu_2$, where $0 \leq p \leq 1$, and if μ_1' and μ_2' are probability measures with $\rho(\mu_1, \mu_1') = \epsilon_1$ and $\rho(\mu_2, \mu_2') = \epsilon_2$, then $\rho(\mu, \mu') \leq \max\{\epsilon_1, \epsilon_2\}$, where $\mu' = p\mu_1' + (1-p)\mu_2'$.*

Proof of Theorem 4.2. Fix a uniform variance split array $\{\mu_{n,k}, p_{m,k}\}$ for μ, and define random variables $X_1, X_2, \ldots$ as follows. $P(X_1 = E\mu_{1,1}) = p_{1,1} = 1$ and for $n \geq 1$,

$$P(X_{n+1} = E\mu_{n+1,2k-1} | X_n = E\mu_{n,k}) = \frac{p_{n+1,2k-1}}{p_{n,k}},$$

and

$$P(X_{n+1} = E\mu_{n+1,2k} | X_n = E\mu_{n,k}) = \frac{p_{n+1,2k}}{p_{n,k}}.$$

Then $E(X_{n+1}|X_n = E\mu_{n,k}) = E\mu_{n+1,2k-1}\frac{p_{n+1,2k-1}}{p_{n,k}} + E\mu_{n,k}\frac{p_{n+1,2k}}{p_{n,k}} = \frac{E\mu_{n,k}p_{n,k}}{p_{n,k}} = E\mu_{n,k}$, where the second equality follows since either $(\mu_{n+1,2k-1}, \mu_{n+1,2k})$ is a variance split of $\mu_{n,k}$, or $\mu_{n,k} = \mu_{n+1,2k-1} = \mu_{n+1,2k} = \delta_{E\mu_{n,k}}$. Thus, $E(X_{n+1}|X_n) = X_n$ a.s., so $\{X_n\}_{n=1}^\infty$ is a martingale. Then

$$EX_n^2 = \sum_{k=1}^{2^{n-1}} (E\mu_{n,k})^2 p_{n,k} = V(\mu) + (E\mu)^2 - \sum_{k=1}^{2^{n-1}} V(\mu_{n,k})p_{n,k} < V(\mu) + (E\mu)^2,$$

where the second equality follows since $V(\Sigma p_i\mu_i) = \Sigma p_i V(\mu_i) + V(\Sigma p_i \delta_{E\mu_i})$. Since $E|X_n| \leq (EX_n^2)^{1/2}$, $\{X_n\}_{n=1}^\infty$ is an L^1 bounded martingale and so it converges a.e. to a finite limit X.

By Lemmas 4.3 and 4.4, $\rho(\mu, \mathcal{L}(X_n)) \leq \max_k \sqrt[3]{V(\mu_{n,k})}$. Since $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is uniform, $\max_k \sqrt[3]{V(\mu_{n,k})} = \sqrt[3]{\max_k V(\mu_{n,k})}$ converges to zero, and so X has distribution μ, which has thus been uniquely determined by its mean-variance array. $\square$

As the following example shows, if the array is not uniform, it is possible for the approximating measures to converge to a measure other than the desired measure.

Example 4.5. Consider the array for the uniform measure on the unit interval described in Example 3.3. In this array, $\min_k E\mu_{n,k} = E\mu_{n,1} = 1/4 + 1/(2^{n+1})$, which converges to 1/4. Therefore, the weak limit of the measures $\sum_k \delta_{E\mu_{n,k}}$ cannot place positive measure on the set $[0, 1/4)$. But the measure of this set under the uniform measure on the unit interval is 1/4. So, while the array of means, variances, and probabilities does define a measure, the measure it defines is not the uniform measure.

Not all arrays of numbers $\{m_{n,k}, V_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ are mean-variance arrays, and the following theorem identifies exactly which are.

Theorem 4.6. *The array $\{m_{n,k}, V_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is a mean-variance array for some probability measure μ if and only if it satisfies, for all $n > 1$ and $k \in \{1, \ldots, 2^{n-1}\}$,*

$$p_{1,1} = 1 \text{ and } 0 < p_{n,k} < 1; \tag{5a}$$

$$V_{n,k} \geq 0; \tag{5b}$$

$$p_{n+1,2k-1} + p_{n+1,2k} = p_{n,k}; \tag{5c}$$

$$m_{n+1,2k-1} \cdot p_{n+1,2k-1} + m_{n+1,2k} \cdot p_{n+1,2k} = m_{n,k} \cdot p_{n,k}; \tag{5d}$$

$$\begin{aligned} &V_{n+1,2k-1} = V_{n+1,2k} \text{ and} \\ &(V_{n+1,2k-1} + m^2_{n+1,2k-1}) \cdot p_{n+1,2k-1} + (V_{n+1,2k} + m^2_{n+1,2k}) \cdot p_{n+1,2k} \\ &\qquad = (V_{n,k} + m^2_{n,k}) \cdot p_{n,k}; \text{ and} \end{aligned} \tag{5e}$$

$$\max_k V_{n,k} \to 0 \text{ as } n \to \infty. \tag{5f}$$

A limit-representation result, similar to Theorem 4.2 (and also based on a martingale argument), will be useful in the proof of Theorem 4.6.

Lemma 4.7. *If $\{m_{n,k}, V_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ satisfies (5a–f), then the weak limit of $\sum_k \delta_{m_{n,k}} p_{n,k}$ as $n \to \infty$ is a probability measure μ with mean $m_{1,1}$ and variance $V_{1,1}$.*

Proof. Define random variables $X_1, X_2, \ldots$ as follows. $P(X_1 = m_{1,1}) = p_{1,1} = 1$ and for $n \geq 1$,

$$P(X_{n+1} = m_{n+1,2k-1} | X_n = m_{n,k}) = \frac{p_{n+1,2k-1}}{p_{n,k}}, \text{ and}$$
$$P(X_{n+1} = m_{n+1,2k} | X_n = m_{n,k}) = \frac{p_{n+1,2k}}{p_{n,k}}.$$

As in the proof of Theorem 4.2, it is easy to check (using (5d)) that $\{X_n\}$ is a martingale, and (using (5e)) that it is L^1 bounded, so X_n converges a.e. to a finite limit X. Let μ be the law of X. Since almost everywhere convergence implies weak convergence, μ is the weak limit of $\sum_k \delta_{m_{n,k}} p_{n,k}$.

To see that μ has mean $m_{1,1}$ note that by (5d) and induction on n, $EX_n = \sum_k m_{n,k} p_{n,k} = m_{1,1}$ for all n; also $V(X_n) = \sum_{k=1}^{2^{n-1}} m_{n,k}^2 p_{n,k} - m_{1,1}^2 = V_{1,1} - \sum_{k=1}^{2^{n-1}} V_{n,k} p_{n,k}$, where the first equality follows by the definition of X_n and the second by (5e) and induction on n. Since, by (5f), $\max_k V_{n,k} \to 0$ as $n \to \infty$, $V(X_n)$ approaches $V_{1,1}$, so $V(X) = V_{1,1}$. $\square$

Proof of Theorem 4.6. First, let $\{E\mu_{n,k}, V(\mu_{n,k}), p_{n,k}\}$ be a mean-variance array for the measure μ. Conditions (5a) and (5b) follow from the fact that for each $n \in \mathbb{N}$ and $k \in \{1, \ldots, 2^{n-1}\}$, $p_{n,k}$ is a probability and $V(\mu_{n,k})$ is a variance. Since $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is a variance split array for μ,

$$\mu_{n,k} = \mu_{n+1,2k-1} \frac{p_{n+1,2k-1}}{p_{n,k}} + \mu_{n+1,2k} \frac{p_{n+1,2k}}{p_{n,k}}. \tag{6}$$

This implies that $p_{n+1,2k-1}/p_{n,k} + p_{n+1,2k}/p_{n,k} = 1$, since otherwise the measures in (6) could not all be probability measures. Thus, (5c) is satisfied. Condition (5d) follows by taking expectations in (6), (5e) by evaluating the second moments of the measures in (6), and (5f) follows since the variance split array is uniform.

For the converse, let $\{m_{m,k}, V_{n,k}, p_{n,k}\}$ satisfy (5a–f).

Define the array $d(n, k)$ as $\{d_{i,j}(m_{n,k}), d_{i,j}(V_{n,k}), d_{i,j}(p_{n,k})/p_{n,k}\}_{i=1}^{\infty}{}_{j=1}^{2^{i-1}}$ where $d_{i,j}(x_{n,k}) = x_{n+i-1, j+2^{i-1}(k-1)}$, and x is m, V, or p. It will next be shown that

$$\text{the array } d(n,k) \text{ satisfies (5a–f),} \tag{7}$$

which simply says that the array "descended" from the $(n,k)^{th}$ member of a mean-variance array is also a mean-variance array.

Conditions (5a) and (5b) are trivial, since $\{m_{n,k}, V_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ satisfies these requirements. Conditions (5c), (5d), and (5e) follow since $d_{i+1,2j-1}(x_{n,k}) = x_{n+i,2j-1+2^i(k-1)} = x_{(n+i-1)+1,2(j+2^{i-1}(k-1))-1}$ and $d_{i+1,2j}(x_{n,k}) = x_{(n+i-1)+1,2(j+2^{i-1}(k-1))}$, and since the original array satisfies these conditions. Finally, (5f) follows since $\max_j d_{i,j}(V_{n,k}) \leq \max_k V_{n+i-1,k}$, where the inequality follows since the maximum is taken over a smaller set. This establishes (7).

Lemma 4.7 and (7) imply the existence of measures $\mu_{n,k}$ with means $m_{n,k}$ and variances $V_{n,k}$. To see that $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is a uniform variance split array, note that for each n and k, $\mu_{n,k}$ is the weak limit of $\sum_j \delta_{d_{i,j}(m_{n,k})} d_{i,j}(p_{n,k})/p_{n,k}$. Then,

$$
\begin{aligned}
&\left(\frac{p_{n+1,2k-1}}{p_{n,k}}\right) \sum_{j=1}^{2^{i-1}} \left(\frac{d_{i,j}(p_{n+1,2k-1})}{p_{n+1,2k-1}}\right) \delta_{d_{i,j}(m_{n+1,2k-1})} \\
&\quad + \left(\frac{p_{n+1,2k}}{p_{n,k}}\right) \sum_{j=1}^{2^{i-1}} \left(\frac{d_{i,j}(p_{n+1,2k})}{p_{n+1,2k}}\right) \delta_{d_{i,j}(m_{n+1,2k})} \\
&= \sum_{j=1}^{2^{i-1}} \left(\frac{d_{i,j}(p_{n+1,2k-1})}{p_{n,k}}\right) \delta_{d_{i,j}(m_{n+1,2k-1})} + \sum_{j=1}^{2^{i-1}} \left(\frac{d_{i,j}(p_{n+1,2k})}{p_{n,k}}\right) \delta_{d_{i,j}(m_{n+1,2k})} \qquad (8) \\
&= \sum_{j=1}^{2^{i-1}} \left(\frac{d_{i+1,j}(p_{n,k})}{p_{n,k}}\right) \delta_{d_{i+1,j}(m_{n,k})} + \sum_{j=1}^{2^{i-1}} \left(\frac{d_{i+1,j+2^{i-1}}(p_{n,k})}{p_{n,k}}\right) \delta_{d_{i+1,j+2^{i-1}}(m_{n,k})} \\
&= \sum_{j=1}^{2^{i}} \left(\frac{d_{i+1,j}(p_{n,k})}{p_{n,k}}\right) \delta_{d_{i+1,j}(m_{n,k})}.
\end{aligned}
$$

where the second equality in (8) follows since $d_{i,j}(x_{n+1,2k-1}) = d_{i+1,j}(x_{n,k})$ and $d_{i,j}(x_{n+1,2k}) = d_{i+1,j+2^{i-1}}(x_{n,k})$, and the last equality by changing the index of the second term and combining the sums. Taking $i \to \infty$ yields $\mu_{n,k} = (p_{n+1,2k-1}/p_{n,k})\mu_{n+1,2k-1} + (p_{n+1,2k}/p_{n,k})\mu_{n+1,2k}$. This, with (5e), shows that $(\mu_{n+1,2k-1}, \mu_{n+1,2k})$ is a variance split of $\mu_{n,k}$, and so $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ is a variance split array for μ. Finally, (5f) implies that this array is uniform. $\square$

The above theorem allows a mean-variance array to be chosen randomly in the following way. First, $m_{1,1}$ and $V_{1,1}$ are set to the desired mean and variance, respectively, and $p_{1,1}$ is set to 1. Then, given $(m_{n,k}, V_{n,k}, p_{n,k})$, the six values representing the mean, variance, and splitting probability of $\mu_{n,k}$ will be chosen. If $V_{n+1,2k-1}$ and $V_{n+1,2k}$ are set to a random proportion of $V_{n,k}$ and the location of $m_{n+1,2k-1}$ is chosen, the other values, $m_{n+1,2k}$, $p_{n+1,2k-1}$ and $p_{n+1,2k}$ are specified by (5). So the general algorithm will be to choose two numbers, s and t, and assign $V_{n+1,2k-1} = V_{n+1,2k} = (1-s)V_{n,k}$ and $m_{n+1,2k-1} = m_{n,k} + t$. Then if $m_{n+1,2k} = m_{n,k} - \frac{s}{t}V_{n,k}$, $p_{n+1,2k-1} = (\frac{sV_{n,k}}{t^2+sV_{n,k}})p_{n,k}$, and $p_{n+1,2k} = p_{n,k} - p_{n+1,2k-1}$, (5) will be satisfied.

To ensure that $(1-s)V_{n,k}$ is strictly smaller than $V_{n,k}$, the value s must be chosen in the interval $(0, 1]$. For a general measure, the only condition on t is that $t \neq 0$. However, choosing a measure on the unit interval places further conditions on t, since then the means must all lie in $[0, 1]$ and the variances must all lie in $[0, m_{n,k}(1 - m_{n,k})]$.

The next proposition identifies the necessary restriction on the proportion t.

Proposition 4.8. *Let m, V, and s be such that $0 \leq m \leq 1$, $0 \leq V \leq m(1-m)$ and $0 < s \leq 1$. Let $K_1 = -m+1/2$, let $K_2 = \sqrt{1/4 - (1-s)V}$, and let $K_3 = (sV)/(m-m^2-(1-s)V)$. Then*

i. $0 \leq m + t \leq 1$;
ii. $0 \leq m - c\frac{s}{t}V \leq 1$;
iii. $0 \leq (1-s)V \leq (m+t)(1-(m+t))$; *and*
iv. $0 \leq (1-s)V \leq (m - \frac{s}{t}V)(1-(m-\frac{s}{t}V))$

if and only if $K_1 - K_2 \leq t \leq K_3(K_1 - K_2)$ *or* $K_3(K_1 + K_2) \leq t \leq K_1 + K_2$.

Proof. Routine, using basic algebra and the fact that $m(1-m) \geq V \geq (1-s)V$. □

If $K_1 - K_2 \leq t \leq K_3(K_1 - K_2)$, then $m_{n+1,2k-1}$ will be less than $m_{n,k}$, and $m_{n+1,2k}$ will be greater. If $K_3(K_1+K_2) \leq t \leq K_1+K_2$, then $m_{n+1,2k-1}$ will be greater than $m_{n,k}$, and $m_{n+1,2k}$ will be smaller. A reasonable condition on the law of t, therefore, is that these intervals have equal probability.

Theorem 4.6 and Proposition 4.8 will now be used to construct a random Borel probability on $[0,1]$ with given mean m and variance V. That is, a probability $B = B^{m,V} = B^{m,V}_{\mu_0,\mu_1}$ on the measurable space $(\mathcal{P}([0,1]),\Sigma)$ will be constructed so that $B(\{\mu \in \mathcal{P}[0,1] :$ mean of $\mu = m$ and variance of $\mu = V\}) = 1$, where $\mathcal{P}([0,1])$ is the set of all Borel probability measures on $[0,1]$, Σ is the smallest sub σ-algebra of subsets of $\mathcal{P}([0,1])$ that includes the weak* topology (cf. Bloomer 2000, Dubins and Freedman 1967), and μ_0 and μ_1 are arbitrary base measures in $\mathcal{P}([0,1])$.

Let $\{S_{n,k}, X_{n,k}, B_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ be an array of independent random variables on a probability space $(\Omega,\mathcal{F},P)$ such that for each n and k: $S_{n,k}$ has distribution μ_0 on $(0,1]$; $X_{n,k}$ has distribution μ_1 on $[0,1]$; and $B_{n,k}$ is Bernoulli with mean $1/2$.

Next define a random array $A_{m,V} = \{x_{n,k},\ y_{n,k},\ p_{n,k}\}_{n=1}^{\infty}{}_{k=1}^{2^{n-1}}$ inductively by $m_{1,1} = m$, $V_{1,1} = V$, and $p_{1,1} = 1$. Given $(m_{n,k}, V_{n,k}, p_{n,k})$, set $V_{n+1,2k-1} = V_{n+1,2k} = (1 - S_{n,k})V_{n,k}$ and let $K_{1,n,k} = -m_{n,k} + 1/2$, let $K_{2,n,k} = \sqrt{1/4 - (1 - S_{n,k})V_{n,k}}$, and let $K_{3,n,k} = (S_{n,k}V_{n,k})/(m_{n,k} - m_{n,k}^2 - (1 - S_{n,k})V_{n,k})$. Next define

$$\begin{aligned} T_{n,k} &= \begin{cases} (K_{3,n,k} - X_{n,k}(1 - K_{3,n,k}))(K_{1,n,k} - K_{2,n,k}) & \text{if } B_{n,k} = 0 \\ (K_{3,n,k} - X_{n,k}(1 - K_{3,n,k}))(K_{1,n,k} + K_{2,n,k}) & \text{if } B_{n,k} = 1 \end{cases} \\ &= (K_{3,n,k} - X_{n,k}(1 - K_{3,n,k}))(K_{1,n,k} + (2B_{n,k} - 1)K_{2,n,k}), \end{aligned}$$

and set

$$\begin{aligned} m_{n+1,2k-1} &= m_{n,k} + T_{n,k}; \\ m_{n+1,2k} &= m_{n,k} - \frac{S_{n,k}}{T_{n,k}} V_{n,k}; \\ p_{n+1,2k-1} &= \left(\frac{S_{n,k}V_{n,k}}{T_{n,k}^2 + S_{n,k}V_{n,k}} \right) p_{n,k}; \text{ and} \\ p_{n+1,2k} &= p_{n,k} - p_{n+1,2k-1}. \end{aligned}$$

Letting $\mathcal{X} = [0,1] \times [0,1/4] \times [0,1]$, endow $\mathcal{A} = \mathcal{X} \times \mathcal{X}^2 \times \cdots \times \mathcal{X}^{2^{n-1}} \times \cdots$ with the standard product topology. Let $S \subset \mathcal{A}$ be the set of all arrays that satisfy (5) and $0 \le E(\mu_{m,k}) \le 1$, $0 \le V(\mu_{n,k}) \le E(\mu_{n,k})(1 - E(\mu_{n,k}))$ for all $n \in \mathbb{N}$ and $k \in \{1, 2, \ldots, 2^{n-1}\}$. Proposition 4.8 shows that $A_{m,V}(\omega) \in S$ for all $\omega \in \Omega$. Let $Q^{m,V}_{\mu_0,\mu_1}$ be the distribution of A on S. By Theorems 4.2 and 4.6, $T_A : S \to \mathcal{P}([0,1])$, the map associating a mean-variance array with the measure it defines, is Borel given the weak* topology.

Definition 4.9. The probability measure $B^{m,V}_{\mu_0,\mu_1}$ on $(\mathcal{P}([0,1]),\Sigma)$ is defined by $B^{m,V}_{\mu_0,\mu_1} = Q^{m,V}_{\mu_0,\mu_1} \circ T_A^{-1}$.

Since the measure defined by a mean-variance array has mean $m = m_{1,1}$ and variance $V = V_{1,1}$, and since the above algorithm sets these values to the desired mean and variance, the measure chosen in this manner has the same mean and variance.

Proposition 4.10. *$B^{m,V}_{\mu_0,\mu_1}$ has support on a subset of $\mathcal{P}_{m,V}([0,1])$.*

Proof. Immediate by the definitions of $Q^{m,V}_{\mu_0,\mu_1}$, T_A and $B^{m,V}_{\mu_0,\mu_1}$ and by Theorem 4.6. □

If at any step in the algorithm described above, $V_{n,k}$ is set to zero, then the measure $\mu_{n,k}$ is a Dirac measure, and the measure μ has a jump. The following theorem uses this fact to allow measures to be chosen discrete almost surely.

Theorem 4.11. *If $\mu_0(\{1\}) > 0$, then $B^{m,V}_{\mu_0,\mu_1}$ is discrete almost surely.*

Proof. If $V_{n,k} = 0$, the resulting distribution has a jump at $m_{n,k}$. For the distribution to have a continuous component requires that for each n, there is at least one k so that $V_{n,k} \neq 0$.

By the method of constructing the mean-variance array, $V_{n,k} = \Pi_{i=1}^{n-1}(1-S_i)V$. This will be non-zero if and only if $S_i \neq 0$ for $i = 1, \ldots, n-1$. The probability of this happening is

$$P(S_1 \neq 1, S_2 \neq 1, \ldots S_{n-1} \neq 1) = (1 - \mu_0(\{1\})^{n-1},$$

which tends to zero as $n \to \infty$, since $\mu_0(\{1\}) > 0$. □

Ferguson (1974) gave as a desirable property of a measure on the space of probability measures that the measure should have full support in the class of interest. The next theorem gives conditions under which the measure described by this algorithm has full support in the space of probability measures with mean m and variance V.

Theorem 4.12. *If μ_0 has full support on $(0,1]$ and μ_1 has full support in $[0,1]$, then $B^{m,V}_{\mu_0,\mu_1}$ has full support in $\mathcal{P}_{m,V}([0,1])$.*

Sketch of Proof. Following the main idea in Theorem 3.10 of Hill and Monticino (1997) (cf. Bloomer 2000), it is enough to show that each set in a base for the weak* topology of $\mathcal{P}_{m,V}([0,1])$ has positive $B^{m,V}_{\mu_0,\mu_1}$ measure. One such base (cf. Billingsley (1968, Appendix III)) consists of sets of the form

$$\{\sigma \in \mathcal{P}_{m,V}([0,1]) : \sigma(\mathcal{O}_i) > \tau(\mathcal{O}_i) - \epsilon,\ i = 1, \ldots, k\},$$

where each $\mathcal{O}_i$ is an open subset of $[0,1]$, τ is a measure in $\mathcal{P}_{m,V}([0,1])$, and $\epsilon > 0$. To show that each such set has positive measure, restrict to open intervals, apply Proposition 3.5 to find $\{A_{n,k}\}$ with $\text{supp}(\tau_{m,k}) \subset [A_{n,k-1}, A_{n,k}]$, and use the "almost-dense" conclusion of Theorem 3.8 and the uniformity of the canonical variance split array for τ guaranteed by Theorem 3.4, to conclude that $\sigma(\mathcal{O}) \geq \tau(\mathcal{O}) - \hat{\epsilon} > 0$. □

The intuition behind Theorem 4.12 is this. Fix any $\mu \in \mathcal{P}_{m,V}([0,1])$ and calculate its mean-variance array; since μ_0 and μ_1 have full support, $B^{m,V}_{\mu_0,\mu_1}$ will eventually generate an array close to that, and that array then determines a $\hat{\mu}(w)$ close to μ.

§5 Applications

The mean-variance array construction procedure for generating random probability measures on $[0,1]$ with fixed mean m and variance V is fast, efficient, and easy to implement. Figure 1 gives two typical sample random measures with mean 1/2 and variance 1/100 using the mean-variance array technique (with $\mu_0 = \mu_1 =$ Lebesgue measure) and Figure 2 gives the average of 500 random measures with that same mean and variance. The equation for the limiting S-shaped curve in Figure 2 is not known to the authors.

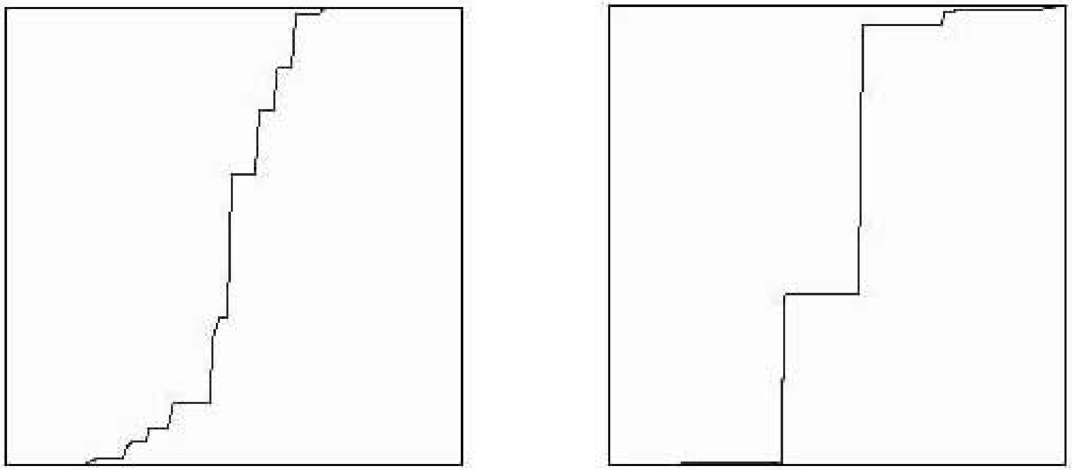

Figure 1. Sample random measures with mean 1/2 and variance 1/100.

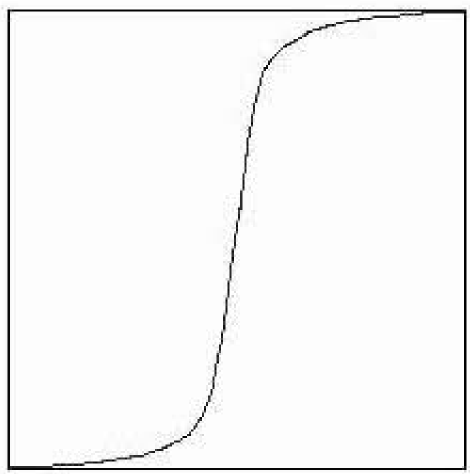

Figure 2. The average of 500 random measures with mean 1/2 and variance 1/100.

One concrete application of the mean-variance array method is to generate conjectured solutions to stochastic extremal problems, e.g., worst-case distributions or sharp constants in stochastic inequalities. For example, given a continuous function $f : \mathcal{P}([0,1]) \to \mathbb{R}$, suppose the sharp constant $k(m,V)$ and extremal distribution are sought for the inequality $f(F) \leq k(m,V)$ for all $F \in \mathcal{P}_{m,V}([0,1])$.

By the continuity of f, (convergence in distribution) and full support of the mean-variance array (Theorem 4.12), the following proposition gives an experimental method to approximate k and the extremal distribution $\hat{F}$.

Proposition 5.1. *Fix $m \in (0,1)$ and $V \in (0, m-m^2)$, and let $F_1, F_2 \dots$ be iid $B^{m,V}_{\lambda,\lambda}$, where λ is Lebesgue measure on $[0,1]$. Then*

$$\max_{1\le i\le n} f(F_i) \nearrow k(m,V) \text{ a.s.}$$

and if $\hat{F}_n$ is defined by $f(\hat{F}_n) = \max_{1\le i\le n} f(F_i)$, then $\hat{F}_n \xrightarrow{\mathcal{D}} \hat{F}$.

Example 5.2. A Generalization of Plackett's Problem. In 1947, Plackett (see Mattner (1993)) considered the problem of finding the maximum expected distance between two identically distributed random variables with mean 0 and variance 1, i.e., find $\max\{E|X-Y| : X \text{ and } Y \text{ are iid}, EX = 0, EX^2 = 1\}$. Rewriting the expected value as

$$E|X-Y| = 2\int_{-\infty}^{\infty} F(x)(1-F(x))dx$$

reduces the problem to the form described above. With unbounded support, the solution is known: the extremal measure is the uniform measure on $[-\sqrt{3}, \sqrt{3}]$.

Figure 3 shows the extremal distribution for $E|X-Y|$ among 10,000 simulations of distributions with $m = 1/4$ and $V = 1/100$ and $V = 1/12$. By Proposition 5.1, the simulation suggests that the extremal distribution for $V = 1/100$ is uniform (which is known to be the case by the solution to the unbounded case of Plackett's problem), and suggests that the extremal distribution for $V = 1/12$ is a convex combination of point mass at zero and the uniform distribution (this has not been verified analytically by the authors).

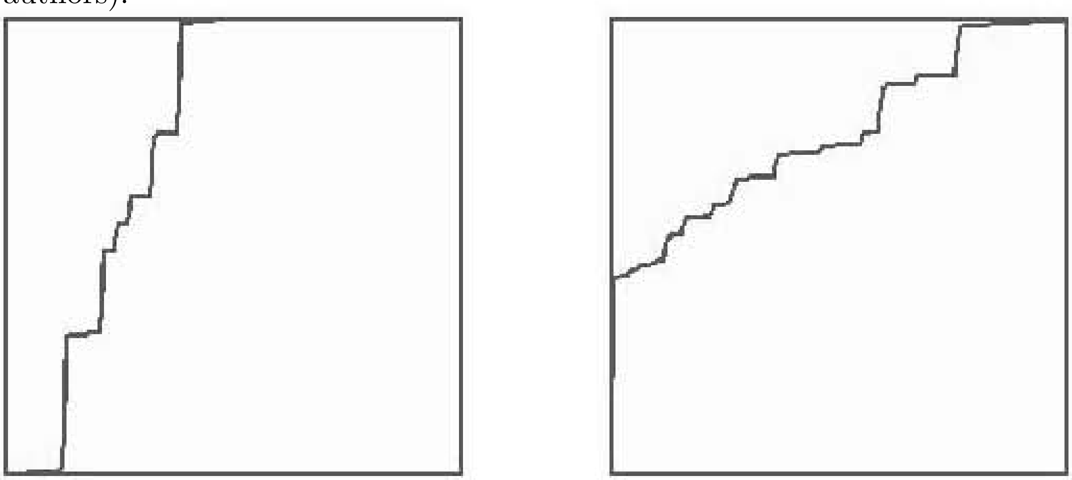

Figure 3. Extremal measures for $E|X-Y|$ with mean 1/4, and variance 1/100 (left) and variance 1/12 (right), based on 10,000 simulations of each.

Even for problems which are intractable analytically, such as maximizing nonlinear continuous functions of higher moments, the technique given by Proposition 5.1 will easily generate approximate solutions.

Another application is to 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 find a value $c^*_{m,V}$ that maximizes $g(F,c)$ on the average over all distributions F on $[0,1]$ with given mean $m \in [0,1]$ and variance $V \in [0, m-m^2]$. (In many applications in statistics, m and V are assumed known – for example, random errors are often assumed to have mean zero, and known variance which depends on the measuring device.) Since the mean-variance array prior $B^{m,V}_{\lambda,\lambda}$, where again λ is Lebesgue measure on $[0,1]$ is a natural prior for picking elements of $\mathcal{P}_{m,V}([0,1])$ randomly, under this prior, the average-optimal control problem becomes

$$\text{choose } c^*_{m,V} \text{ to maximize } \int g(F,c)dB^{m,V}_{\lambda,\lambda}(F).$$

Example 5.3. Suppose a stopping rule t is to be chosen for 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 each have mean m and variance V. The goal is to maximize EX_t on average. By backward induction, it is clear that there is an optimal stop rule t_c of the form $t_c = 1$ if $X_1 > c$, $t_c = 2$ if $X_1 \le c$ and $X_2 > m$, and $t_c = 3$ otherwise, so the goal is to find the value c which maximizes EX_{t_c}. Since only partial information is known, the optimal c depends on the distributions of X_1, X_2, and X_3. Simulations for this problem suggest that with small variances, a small error in the choice of c lowers the expected return by a large margin. However, for large variances, the choice of c is not as important, in that large changes in c yield only small changes in the expected return.

Acknowledgement

This research was supported in part by NSF Grant DMS-9971146.

References

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

Bloomer, L. (2000) Random Probability Measures with Given Mean and Variance. Doctoral Dissertation, Georgia Institute of Technology.

Dubins, L. and Freedman, D. (1967) Random distribution functions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **2**, 183–214, University of California Press Berkeley.

Ferguson, T. S. (1974) Prior distributions on spaces of probability measures. *Annals of Statistics* **2** (4), 615–629.

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

Hill, T. P. and Monticino, M. (1997) Sequential barycenter arrays and random probability measures. Technical report, School of Mathematics, Georgia Institute of Technology.

Hill, T. P. and Monticino, M. (1998) Constructions of random distributions via sequential barycenters. *Annals of Statistics* **26** (4), 1242–1253.

Mattner, L. (1993) *Extremal Problems of Probability Distributions: A General Method and Some Examples*, volume 22 of *IMS Lecture Notes Monograph*, pages 274–283. Hayward, CA.

Mauldin, R. D., Sudderth, W. D. and Williams, S. C. (1992) Polya trees and random distributions. *Annals of Statistics* **20** (3), 1203–1221.

Monticino, M. (1998) Constructing prior distributions with trees of exchangeable processes, *J. Statist. Planning and Inference* **73**, 113–133.

Skibinsky, M. (1968) Extreme nth moments for distributions on $[0,1]$ and the inverse of a moment space map. *J. Appl. Prob.* **5**, 693–701.