

CONSTRUCTING RANDOM PROBABILITY DISTRIBUTIONS

THEODORE P. HILL*

*School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332-0160 USA
hill@math.gatech.edu*

DAVID E. R. SITTON†

*Planning System Incorporated
115 Christian Lane
Slidell LA 70458 USA
dsitton@psisidell.com*

Abstrac: This article surveys several classes of iterative methods for constructing random probability distributions (or random convex functions, or random homeomorphisms), and includes illustrative applications in statistics, optimal-control theory, and game theory. Computer simulations of these methods are fast and easy to implement.

1. Introduction

The main purpose of this article is to provide a short survey of methods for constructing random probability distributions (or, equivalently [10, 12, 16], for constructing random convex functions or random homeomorphisms). As such, this paper complements and extends the excellent overviews of constructions of random probability measures by Ferguson [14], by Diaconis and Freedman [9], and Monticino [34]. Constructions of random probability measures are not only intrinsically interesting, but also have useful applications in such areas as game theory, statistics, optimal control theory, and analysis of algorithms [7, 8, 25, 29].

*Research partially supported by U.S. National Science Foundation Grant DMS-9971146.

†Research partially supported by a graduate Presidential Fellowship at the Georgia Institute of Technology.

For example, in game theory, basic minimax theorems imply that in wide classes of games the optimal strategies are random, as opposed to pure or deterministic strategies. Often the solution is known to be a probability distribution on several points, or on an interval. If no analytical solution for the game is known, then the optimal strategy (probability distribution) can often be approximated numerically, by constructing probability distributions at random, and keeping track of the distribution attaining the extremal expected payoffs. Provided that the construction method produces random distributions which are dense in the set of all probability distributions in the class of interest, and that the underlying game satisfies general continuity conditions, the simulation will converge to the game's optimal strategy. As a concrete example, Gal [15] describes the following single 2-person search game on three arcs, for which the general optimal strategy is still unknown. (Optimal strategies within certain classes are known; see [5, 15].) Two cities A and B are connected by 3 non-intersecting roads of equal lengths. Player I places a landmine at some location on one of the three roads, and Player II, starting at A , searches along the roads at unit speed until he finds the landmine. Player I's objective is to make the time to discovery as large as possible, and Player II's is to minimize. It is known that the optimal solution for Player I is a single continuous probability distribution on the interval (i.e., concatenation of the three roads), which can be approximated numerically by constructing probability measures at random and identifying the best-case distributions, as is seen in §5 below.

Another application of constructions of random probability distributions is to determine average-optimal strategies, or the average-optimal distribution of errors in certain numerical algorithms [17, 36]. For example, in optimal stopping theory, the observer may not know the distributions of the observations completely, but instead may only have partial information, such as knowing the means, or means and variances. The objective then is to find a stop rule t which is best *on the average* among random variables in the given class, in which case an appropriate probability on that class of distributions needs to be identified. Similarly, in many numerical algorithms, it is known that the worst-case errors are actually quite rare [35, 36], and in analyzing errors (in order to select between two different algorithms, for example), average error is a better criterion than worst-case error. Thus performance analysis of such algorithms hinges on identification of an appropriate model for the random input (or error), i.e., constructions of appropriate random probability distributions.

In addition, constructions of random probability distributions have basic

applications in classical Bayesian statistics to produce natural priors (e.g., random distribution of a species in a region), in probability theory and analysis to numerically generate sharp constants or optimal distributions such as in Plackett's problem [28] (see §5 below), and in theories underlying new statistical tests for the detection of fabricated numerical data [19].

Important properties for constructions of random probability distributions to have are that they are natural, they are easy to implement, and they have dense support in the desired class of probability distributions. All the constructions discussed below share these three properties, and all are essentially non-parametric statistical methods. As noted by Monticino [34], "non-parametric priors may avoid biases potentially introduced by the selection of a particular parametric family," and the constructions given here all have wide support.

Section 2 contains the basic definitions and framework, and descriptions of several classical methods for constructing general probability distributions; Section 3 surveys methods for constructing random distributions with given moments (such as means and/or variances); Section 4 describes several methods for generating random probability distributions which are absolutely continuous (i.e., which have densities – the aforementioned constructions all yield purely singular distributions); and Section 5 gives typical applications to several concrete problems.

2. Constructions of General Random Probability

The basic idea of Dubins and Freedman [10, 12], to construct a random probability measure by constructing its distribution function at random, underlies each of the constructions below, where all distributions are taken to have support in $[0, 1]$. The measure-theoretic setting [11] is this: Λ is the space of all distribution functions on $[0, 1]$, that is $\Lambda = \{F : [0, 1] \rightarrow [0, 1], F \text{ is non-decreasing and right continuous, } F(0^-) = 0, F(1) = 1\}$, where for $F \in \Lambda$, the Borel probability measure on $[0, 1]$ defined by F is determined by $P([0, t]) = F(t)$, $t \in [0, 1]$; and Σ^* is the smallest σ -algebra of subsets of Λ such that for every $A \in \mathbb{B}[0, 1]$, the function $F \mapsto \int_A dF(x)$ is Borel measurable. Thus a random probability distribution (r.p.d.) \mathbb{F} on $[0, 1]$ is a measurable function from a probability space (Ω, \mathcal{F}, P) to (Λ, Σ^*) . That is, \mathbb{F} is a probability-distribution-valued random variable, and \mathbb{F}_ω will denote its value (d.f.) for each ω in Ω .

Dubins and Freedman [12] give a natural iterative method for constructing r.p.d.'s \mathbb{F} via a given base measure μ on $S = [0, 1]^2$. For example, if

μ is uniform on the vertical bisector $\{(\frac{1}{2}, y) : 0 \leq y \leq 1\}$, their construction proceeds as follows. Recall that $F(0^-) = 0$ and $F(1) = 1$ for all $F \in \Lambda$, and let $\mathbb{F} = \mathbb{F}_\mu$ be the r.p.d. defined inductively on the dyadic rationals in $[0, 1]$ by $\mathbb{F}(0) = 0$, $\mathbb{F}(1) = 1$, $\mathbb{F}(\frac{1}{2}) = X_1$, $\mathbb{F}(\frac{1}{4}) = X_2\mathbb{F}(\frac{1}{2})$, $\mathbb{F}(\frac{3}{4}) = \mathbb{F}(\frac{1}{2}) + X_3(1 - \mathbb{F}(\frac{1}{2}))$, \dots , where X_1, X_2, \dots are i.i.d. $U[0, 1]$ random variables independent of \mathbb{F} . (So $\mathbb{F}(\frac{1}{2})$ is uniformly distributed on $[0, 1]$, and, given $\mathbb{F}(\frac{1}{2})$, $\mathbb{F}(\frac{1}{4})$ and $\mathbb{F}(\frac{3}{4})$ are uniformly distributed on their possible ranges $[0, \mathbb{F}(\frac{1}{2})]$ and $[\mathbb{F}(\frac{1}{2}), 1]$, respectively, and so on for $\mathbb{F}(\frac{1}{8}), \mathbb{F}(\frac{3}{8}), \dots$)

As another example of a natural base measure, let μ be uniformly distributed on S , and define the random sequence $\mathbb{F}(0) = 0$, $\mathbb{F}(1) = 1$, $\mathbb{F}(X_1) = Y_1$, $\mathbb{F}(X_2) = Y_2$, $\mathbb{F}(X_3) = Y_3, \dots$ as follows. X_1 and Y_1 are i.i.d. $U[0, 1]$, X_2 and Y_2 are independent and uniformly distributed on $[0, X_1]$ and $[0, Y_1]$, respectively, X_3 and Y_3 are independent and uniformly distributed on $[X_1, 1]$ and $[Y_1, 1]$, respectively, and so on.

For these constructions and a much wider class, Dubins and Freedman [12] establish many basic results including: these random sequences determine a r.d.f. \mathbb{F} a.s.; \mathbb{F} is dense in Λ ; and \mathbb{F} is a.s. singular (with respect to Lebesgue measure) – essentially since the process is “self-similar.” For more general base measures μ on S they show that: almost all distributions generated are continuous if and only if μ assigns probability 0 to the vertical edges of S and μ assigns positive probability to the interior of S ; almost all distributions generated are purely discrete if either μ assigns probability 1 to the horizontal edges of S or μ assigns positive probability to the vertical edges of S ; and if μ does not give probability 1 to the main diagonal of S , then almost all the generated distributions are singular.

Special cases of the Dubins-Freedman construction, and extensions to more general settings, including changing base measures at each stage of the construction, are found in [16, 24, 30, 31]. In an effort to use r.p.d.’s as priors in Bayesian statistics, Ferguson [13, 14] developed *Dirichlet priors*, which are a.s. discrete, have full support under fairly general conditions, and, in contrast to the Dubins-Freedman constructions, have easily describable posterior distributions.

Another method for constructing r.p.d.’s uses a Pólya-urn scheme technique to generate a sequence of exchangeable random variables. Via de Finetti’s theorem, every infinite exchangeable sequence is a random (possibly continuous) mixture of sequences of i.i.d. random variables, which therefore yields a random probability measure; see [2, 26, 27, 29, 33]. Mauldin, Sudderth, and Williams [29] show that the set of Pólya tree priors forms a conjugate class, and find conditions under which a Pólya tree prior

is a.s. continuous, or has full support on Λ . Monticino [33] establishes connections between Pólya tree constructions and “random rescaling” r.p.d.’s, and shows that trees of arbitrary exchangeable processes can be used in place of Pólya urn schemes.

3. Construction of Distributions with Given Moments

In many applications involving unknown or random distributions, one or more of the moments of the distribution are assumed to be known. For example, in algorithms involving random error, the error is often unbiased, that is, has mean zero. Similarly, in many experiments involving measurements, the error may also have known standard deviation, hence second moment, based on the known variability of the measuring device. In trying to model these random distributions, the constructions mentioned above are not useful, since the distributions generated do not have fixed means or variances, and the sets of distributions with prescribed means or variances are null sets in the underlying probability space. In fact, even the calculation of the distribution of the means, except in some trivial cases, is difficult (see [6, 32]).

Several methods for constructing r.p.d.’s with given moments have evolved. The method in [21, 22] generates a random distribution by generating its sequential barycenter array, or successive conditional expectations, at random. For a distribution $F \in \Lambda$, the F -barycenter of $(a, c]$, $b_F(a, c]$, is the conditional expectation of F over the interval $(a, c]$, that is

$$b_F(a, c] = \begin{cases} \int_{(a,c]} x dF(x) / (F(c) - F(a)) & \text{if } F(c) > F(a) \\ a & \text{if } F(c) = F(a), \end{cases}$$

and the sequential barycenter array $\{m_{n,k}\}_{n=1}^{\infty} \binom{2^n-1}{k=1}$ of F is defined inductively as follows:

$$\begin{aligned} m_{1,1} &= \int x dF(x) \\ m_{n,2j} &= m_{n-1,j} \text{ for } n \geq 1 \text{ and } j = 1, \dots, 2^{n-1} - 1 \\ m_{n,2j-1} &= b_F(m_{n-1,j-1}, m_{n-1,j}) \text{ for } j = 1, \dots, 2^{n-1} \\ &\quad (\text{with } m_{n,0} := 0, \text{ and } m_{n,2^n} := 1). \end{aligned}$$

In [22], it is shown that the sequence $\{m_{n,k}(F)\} = \{m_{n,k}\}$ uniquely deter-

mines the distribution F , via the inversion formula

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

The main idea in [22] is to use these barycenter characterizations of a distribution to generate a sequential barycenter array randomly, and then to recover the observed value F_ω of the random distribution from the array using the inversion formula. Since the distribution of the initial barycenter $m_{1,1}$ may be specified, this construction can generate r.p.d.'s with any prescribed mean or distribution of the means. As with the Dubins-Freedman construction above, the generation of the random barycenter array depends on a base measure μ which may be chosen to fit the given model desired. For example, if μ is uniform and a r.p.d. with mean $1/3$ is desired, first fix $m_{1,1} = 1/3$, then generate the random conditional mean less than $1/3$, $m_{2,1}$, uniformly in $(0, 1/3)$, and the conditional mean above $1/3$, $m_{2,3}$, uniformly in $(1/3, 1)$, and so on, at each step generating the new barycenters uniformly between the previous ones. (See [34] for a graphic "mobile" description of this process.) By using non-uniform bases, "clumping" or "anti-clumping" constructions may be attained, where mass in the distribution is more (or less) likely to be near other masses. The results in [22] include conditions on the base measure μ so that the construction has full support (in the subset of Λ with given mean or distribution of the mean), and conditions on μ so that the r.p.d.'s generated are a.s. continuous, or a.s. discrete, or have finite support a.s.

Although the sequential barycenter construction allows one to specify the mean of the r.p.d., it does not generate distributions with fixed higher moments, such as given mean and variances simultaneously. One approach to solving this problem is in [3, 4], which generates r.p.d.'s with given mean and variance via *variance split arrays*.

Definition. A pair of probability measures (μ_1, μ_2) is a *variance split* of the probability measure μ if $\text{Var}(\mu_1) = \text{Var}(\mu_2) < V(\mu)$, and there is a $p \in (0, 1)$ so that $\mu = p\mu_1 + (1 - p)\mu_2$. A triangular array $\{\mu_{n,k}, p_{n,k}\}_{n=1}^{\infty} \sum_{k=1}^{2^{n-1}}$ is a *canonical variance split array* for the probability measure μ if for each $n \in \mathbb{N}$, $\mu = \sum_{k=1}^{2^{n-1}} \mu_{n,k} p_{n,k}$, and for each k , $(\mu_{n+1,2k-1}, \mu_{n+1,2k})$ is a non-degenerate variance split of $\mu_{n,k}$ with splitting probability $p = p_{n+1,2k-1}/p_{n,k}$. The array is called *uniform* if $V(\mu_{n,k}) \rightarrow 0$ as $n \rightarrow \infty$.

Theorem 1. *Every Borel probability measure with compact support has a canonical variance split array, and every such array is uniform. Moreover, every such array uniquely determines the distribution.*

Analogously to the sequential barycenter array construction, a random distribution may also be constructed via a base measure μ by constructing this variance-split array, or associated mean-variance array, at random (see Figure 1). In [3, 4], necessary and sufficient conditions are obtained for an array to be a mean-variance array, and conditions are given which guarantee that the generated distributions are a.s. discrete, and that they have full support in the subset of Λ with given mean and variance.



Figure 1. Left is a sample random probability distribution with mean 0.5 and variance 0.01 constructed using the variance-split array method. Right is the average of 500 r.d.f.'s with the same mean and variance.

Another method for constructing r.p.d.'s on $[0, 1]$ with given mean and variance, or in fact given moments of any orders, is to pick the moments at random (e.g., using a natural base measure μ , such as uniform), since the moments $\{EX^n\}$ of a compactly supported distribution uniquely determine the distribution. Given the first n moments M_1, \dots, M_n of a distribution $F \in \Lambda$ (i.e., $M_j = \int x^j dF(x)$), sharp lower and upper bounds \underline{M}_{n+1} and \overline{M}_{n+1} are known for the $(n+1)$ st moment [38]. That is, given the first n moments $M_1(F), \dots, M_n(F)$, the $(n+1)$ st moment $M_{n+1}(F)$ lies in the closed interval $[\underline{M}_{n+1}(M_1, \dots, M_n), \overline{M}_{n+1}(M_1, \dots, M_n)]$. These bounds are sharp, and attained (by discrete distributions), and are given in easily-calculated recursive form.

Thus, for example, to generate a random distribution with given first, second and third moments, M_1, M_2, M_3 , generate the sequence of higher moments randomly as follows. First pick $M_4(F)$ in $[\underline{M}_4(M_1, M_2, M_3), \overline{M}_4(M_1, M_2, M_3)]$, say uniformly, and then pick $M_5(F)$ uniformly in $[\underline{M}_5(M_1, M_2, M_3, M_4(F)), \overline{M}_5(M_1, M_2, M_3, M_4(F))]$, and so on. The main drawback of this method is that the inversion process – recovering the distribution from its moments – seems calculation-intensive for large n . Perhaps new advances in inversion algorithms will make this technique more attractive.

Methods for generating r.p.d.'s in higher dimension are useful in various statistical problems such as describing distributions of mass in space, with fixed center of mass. Some of the above construction ideas carry over to higher dimensional settings [20], and complement other known methods including Kolmogorov's "rock-crushing" model [23], multi-type branching processes, and embeddings of random graphs [1].

4. Construction of Absolutely Continuous Distributions

All of the above methods for constructing random probability distributions yield singular distributions almost surely, either purely discrete measures or continuous measures which are singular with respect to Lebesgue measure, i.e., which live on a null set. Since nearly all the continuous distributions encountered in practice and in theoretical statistics (e.g., gaussian, exponential, uniform) are absolutely continuous with respect to Lebesgue measure, that is, have densities or Radon-Nikodym derivatives, it is useful to have constructions of r.p.d.'s which generate a.c. measures almost surely. Kraft [24] introduced a generalization of the Dubins-Freedman construction in which the base measure changes for each successive point in the construction of the r.d.f. Under certain conditions on the changing base measure $\{\mu_{i,j}\}$, Kraft showed that the generated r.d.f.'s are almost surely absolutely continuous, but prescribed no structure to the changing base measures, nor established density of support.

Complementing Kraft's construction, Sitton [37] established several natural methods for constructing r.d.f.'s which are almost surely a.c. One method is based on the fact that convolution of measures increases smoothness (or decreases concentration [18]). For example, if F_1 is a.c., then so is $F_1 * F_2$ for all $F_2 \in \Lambda$, and if F_1 is α -Lipschitz, then so is $F_1 * F_2$. Sitton defines the convolution $\mathbb{F}_1 * \mathbb{F}_2$ of two r.d.f.'s, and gives a natural method of randomly rescaling them (prior to performing the convolution) via an independent $(0, 1)$ -valued random variable X , so that the result $\hat{\mathbb{F}}|_X(\mathbb{F}_1, \mathbb{F}_2)$ is a r.d.f. (on $[0, 1]$). He then shows that if \mathbb{F}_1 is any r.d.f., and \mathbb{F}_2 is a r.d.f. which is a.s. absolutely continuous, then $\hat{\mathbb{F}}|_X(\mathbb{F}_1, \mathbb{F}_2)$ is an a.s. absolutely continuous r.d.f., and proves criteria useful for establishing full support of the r.d.f.

Theorem 2. [37] *Let X be a r.v. in $[0, 1]$ for which $P(X \in (0, 1)) = 1$, and $0 \in \text{supp}(X)$. If \mathbb{F}_1 and \mathbb{F}_2 are r.d.f.'s, independent of X , with support in Λ , then the support of the r.d.f. $\hat{\mathbb{F}}|_X(\mathbb{F}_1, \mathbb{F}_2)$ contains the support of \mathbb{F}_1 .*

Corollary 1. *If \mathbb{F}_2 is the Dubins-Freedman r.d.f. (with uniform base measure μ), \mathbb{F}_1 is the uniform d.f. on $[0, 1]$, and $U \sim U(0, 1)$ is independent of \mathbb{F}_1 , then the r.d.f. $\hat{\mathbb{F}} = \hat{\mathbb{F}}|_U(\mathbb{F}_1, \mathbb{F}_2)$ is a.s. absolutely continuous, and has*

full support in Λ , hence full support among all d.f.'s in Λ which are a.c.

In addition, [37] contains bounds on the density of the r.d.f.'s, and a general form for the average d.f. $\bar{\mathbb{F}}$ of a r.d.f. \mathbb{F} , where $\bar{\mathbb{F}}$ is the function $\bar{\mathbb{F}} : \mathbb{R} \rightarrow [0, 1]$ defined by

$$\bar{\mathbb{F}}(x) = \int_{G \in \Lambda} G(x) d\mathbb{F}(G) = \int_{\omega \in \Omega} \mathbb{F}_\omega(x) dP(\omega) =: E(\mathbb{F}(x)).$$

Another method for constructing r.d.f.'s with densities a.s. is by piecewise linear interpolation of a randomly stopped r.d.f. construction process such as the Dubins-Freedman (or Hill-Monticino, or Bloomer) r.d.f.'s. The idea in [37] is this. The Dubins-Freedman r.d.f. is the limit of an infinitely-many-stage process, which in practice is terminated after a fixed number of steps. To create a r.d.f. which generates distributions with densities (in fact piecewise constant densities), stop the Dubins-Freedman construction at a random time T , and interpolate linearly between the $T + 1$ points on the graph of \mathbb{F}_ω , to obtain the r.d.f. \mathbb{F}_T . If T is a constant (fixed stopping time n for interpolation), then the support of \mathbb{F}_n will not in general contain the support of \mathbb{F} . However, if T is an \mathbb{N} -valued r.v. independent of \mathbb{F} , with unbounded range, then the support of \mathbb{F}_T will contain the support of \mathbb{F} . A proof of this result in [37] is included to illustrate some of the techniques used.

Theorem 3. *If \mathbb{F} is a Dubins-Freedman r.d.f. (associated with some measure μ), T is an unbounded \mathbb{N} -valued r.v., and \mathbb{F} and T are independent, then the support of \mathbb{F}_T contains the support of \mathbb{F} .*

Proof. A d.f. $G \in \Lambda$ is in the support of \mathbb{F} if and only if for any $\epsilon > 0$, $P(d(\mathbb{F}, G) \leq \epsilon) > 0$, where $d(F, G)$ is the Levy distance given by $d(F, G) = \inf\{\epsilon > 0 : G(x - \epsilon) - \epsilon \leq F(x) \leq G(x + \epsilon) + \epsilon \forall x \in \mathbb{R}\}$.

Fix $\epsilon > 0$ and $G \in \text{supp}(\mathbb{F})$. Then $G \in \text{supp}(\mathbb{F}_T)$ if $P(d(\mathbb{F}_T, G) < \epsilon) > 0$. For $\omega \in \Omega$, let $J(\omega)$ be the set of all points selected in the Dubins and Freedman method.

In [12], Dubins and Freedman showed that for almost every ω , $J(\omega)$ uniquely defines \mathbb{F}_ω . Fix ω so that $J(\omega)$ uniquely defines \mathbb{F}_ω .

For all $(x, \mathbb{F}_\omega(x)) \in J(\omega)$, $\mathbb{F}_{n,\omega}(x)$ converges to $\mathbb{F}_\omega(x)$. The points in J_ω completely define \mathbb{F}_ω , so for any x in \mathbb{R} , $\mathbb{F}_{n,\omega}(x) \rightarrow \mathbb{F}_\omega(x)$ as $n \rightarrow \infty$. Thus, $\mathbb{F}_{n,\omega}$ converges pointwise to \mathbb{F}_ω . The sequence $\{\mathbb{F}_{n,\omega}\}$ is tight since all mass lies in $[0, 1]$ for each $\mathbb{F}_{n,\omega}$. It follows that $\mathbb{F}_{n,\omega}$ converges vaguely to \mathbb{F}_ω , and \mathbb{F}_n almost surely vaguely converges to \mathbb{F} .

Let $\{n_j\}_{j \in \mathbb{N}}$ be an enumeration of the support of T , which is unbounded by hypothesis. Then, \mathbb{F}_{n_j} almost surely vaguely converges to \mathbb{F} as $j \rightarrow \infty$.

Since vague convergence is convergence in the Levy metric, i.e. weak* convergence, $\mathbb{F}_{n_j,\omega}$ converges to \mathbb{F}_ω in the Levy metric. So, there exists a

$K_{\epsilon, \omega} \in \mathbb{N}$ large enough that for all $j > K_{\epsilon, \omega}$, $d(\mathbb{F}_{n_j, \omega}, \mathbb{F}_\omega) < \epsilon$. For every $j \in \mathbb{N}$, define the sets $A_j, A \subset \Omega$ by $A_j = \{\omega \mid d(\mathbb{F}_{n_j, \omega}, G) < \epsilon\} \in \mathcal{F}$ and $A = \{\omega \mid d(\mathbb{F}_\omega, G) < \epsilon\} \in \mathcal{F}$.

Assume $\omega \in A$. Then, there is a $K' = K_{\epsilon - d(\mathbb{F}_\omega, G), \omega} \in \mathbb{N}$ large enough that for all $j > K'$, $0 \leq d(\mathbb{F}_{n_j, \omega}, \mathbb{F}_\omega) < \epsilon - d(\mathbb{F}_\omega, G)$, because $\epsilon > d(\mathbb{F}_\omega, G)$. By the triangle inequality, $0 \leq d(\mathbb{F}_{n_j, \omega}, G) \leq d(\mathbb{F}_{n_j, \omega}, \mathbb{F}_\omega) + d(\mathbb{F}_\omega, G) < \epsilon$. Therefore, $\omega \in A_{n_j}$ for all $j > K'$. The sequence $\{n_j\}$ is unbounded and positive so $\{\omega \mid d(\mathbb{F}_\omega, G) < \epsilon\} \subset \bigcup_j A_{n_j}$. Thus, $0 < P(d(\mathbb{F}, G) < \epsilon) \leq P\left(\bigcup_{j \in \mathbb{N}} A_{n_j}\right)$ because by hypothesis, G is in the support of \mathbb{F} . By the subadditivity of probability measures,

$$0 < P\left(\bigcup_{j \in \mathbb{N}} A_{n_j}\right) \leq \sum_{j=1}^{\infty} P(A_{n_j}).$$

It follows that there exists a $M \in \mathbb{N}$ for which $P(A_{n_M}) > 0$. In particular,

$$0 < P(d(\mathbb{F}_{n_M}, G) < \epsilon).$$

Finally,

$$\begin{aligned} P(d(\mathbb{F}_T, G) < \epsilon) &= \sum_{j=1}^{\infty} P(d(\mathbb{F}_{n_j}, G) < \epsilon \mid T = n_j) P(T = n_j) \\ &= \sum_{j=1}^{\infty} P(d(\mathbb{F}_{n_j}, G) < \epsilon) P(T = n_j) \\ &= \sum_{j=1}^{\infty} P(A_{n_j}) P(T = n_j) \geq P(A_{n_M}) P(T = n_M) > 0. \end{aligned}$$

The first equality is Bayes' Formula; the second follows because \mathbb{F} and T are independent; and the third follows from the definition of A_{n_j} . The first inequality follows because all terms in the sum are non-negative, and the last inequality comes from the hypothesis on M and because $n_M \in \text{supp}(T)$.

Corollary 2. *If T is a geometric random variable, \mathbb{F} is the Dubins-Freedman r.d.f. with base measure μ uniform on $[0, 1]^2$, and T and F are independent, then \mathbb{F}_T is a.s. absolutely continuous, and has full support (i.e., $\text{supp } \mathbb{F}_T = \Lambda$).*

In some applications such as queueing problems or renewal processes, the unknown (random) distribution may not only be known to be absolutely continuous, but also be known to have a monotone density. In [37], several constructions of r.d.f.'s of this type are given. For example, the sequential

barycenter method in §3 can be used to generate a r.d.f. \mathbb{F} with constant mean $1/2$, using the fact that $EX = \int_0^1 (1 - F(x)) dx$ (by Fubini's Theorem). Since \mathbb{F} is continuous a.s., $2(1 - \mathbb{F})$ will generate a random non-increasing continuous function on $[0, 1]$ with integral 1, that is, it will generate continuous monotone non-increasing densities on $[0, 1]$ almost surely. Sitton [37] establishes conditions under which given methods for constructing r.d.f.'s generate distributions which are dense in the set of all distributions with monotone densities, or with bounded monotone densities, respectively.

5. Applications

The purpose of this section is to illustrate application of some of the constructions described above to several concrete problems.

5.1. Generation of models

Example 1. (Power-law distribution of mass). For some physical problems, mass (or charge, etc.) is randomly distributed according to a power law $\mu[0, x] = x^\alpha$. Note that for the sequential barycenter method, in which the base measure defines the random *distance* from the n -th stage barycenter to the $(n + 1)$ -st barycenter, tighter clustering occurs for smaller values of α , and in the limiting case $\alpha = 0$, there is total clustering (Dirac measure) at the center of mass. Figure 2 shows three sample simulations for distribution of mass, with fixed center of mass $1/2$, using the sequential barycenter method for $\alpha = 1$ (uniform), $\alpha = 5$ (anti-clustering), and $\alpha = 0.5$ (clustering).



Figure 2. Simulations of power-law mass distribution models constructed using a symmetric sequential barycenter method. All have mean (center of mass) at 0.5 ; the left figure is for the case $\alpha = 1.0$, middle figure is for $\alpha = 5$ (anti-clustering), and right is $\alpha = 0.5$ (clustering).

Numerical Approximation of Universal Constants and Extremal Distributions

Example 2. [3, 4] (Generalization of Plackett's Problem). Plackett (see [28]) considered the problem of finding the maximum distance between two identically distributed random variables with given mean and variance, i.e., find $\max\{E|X - Y| : X \text{ and } Y \text{ are i.i.d. with } EX = m, \text{Var } X = \sigma^2\}$. Rewriting the expected value as

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

reduces the problem to finding a single unknown extremal distribution F , with given mean m and variance σ^2 . Using the above variance-split array method for constructing a r.d.f. (rescaled to $[0, 1]$ with mean $1/2$, variance $1/10$), and keeping track of the extremal distribution up to time n , simulations suggest convergence to the known optimal distribution, which is uniform. Similarly, simulations for the problem $\max\{E|X - Y| : X \text{ and } Y \text{ are i.i.d., } 0 \leq X \leq 1, EX = 1/4, \text{Var}(X) = 1/12\}$, the solution of which is not known to the authors, suggest that the extremal distribution is a convex combination of point mass at 0 and a uniform distribution (see Figure 3). Since the variance-split r.d.f. is dense in the support of d.f.'s in Λ with given mean and variance, and since the objective function $E|X - Y|$ is continuous (convergence in distribution) in the distribution of X , the extremal distributions up to time n in the r.d.f. construction will converge (weakly) to the unknown extremal distribution.



Figure 3. Extremal measures for $E|X - Y|$ with mean 0.25 and variance 0.01 (left) and variance 0.08 (right), based on 10,000 simulations of each constructed using the variance-split method.

Similar examples of applications to optimal stopping theory with partial information are found in [3, 4, 20, 34]. These include the problem of finding a stop rule t which is optimal, on the average, for stopping a sequence of 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 , or each have mean m and variance σ^2 .

As one final example, consider the still-open 2-person game problem mentioned in the introduction.

Example 3. [37] (Hide and Seek Game [15]) Two players, a “Hider” and a “Seeker,” play the following game on a graph consisting of 2 vertices A and B , and 3 edges (paths) between the vertices, of lengths ℓ_1, ℓ_2, ℓ_3 respectively. First Hider places a marker (coin or landmine) somewhere along one of the 3 paths, and then Seeker walks a continuous path along the graph, starting at A , and ending at the location of the marker. Then Seeker pays Hider D dollars, where D is the total distance travelled by Seeker. It is known that the optimal solution for Hider is an a.c. random distribution on the paths, but even in the special case $\ell_1 = \ell_2 = \ell_3 = 1$, the optimal strategy (probability distribution) for Hider is not known (although a particular distribution, which has been shown to be optimal among a large class of optimal strategies for the Seeker, is believed to be optimal in general [5, 15]). For the unequal path problem, no optimal solutions have even been conjectured.

Since the optimal solution for the Hider is an a.c. probability distribution on each path, one of the methods in Sect. 4 may be used to approximate the solution via simulation, by generating a.c. distributions F at random, calculating the expected distance $E_F(D)$, and tracking the extremal F , i.e. $\max_n \{E_{F_n}(D)\}$. Since the constructions discussed in Sect. 4 are dense in the set of all a.c. d.f.’s (in the figures shown later, the linear interpolation method with random time T was used), this maximum will converge to the optimal value (and its argument to the extremal F^*) a.s. Figures 4 and 5

show simulations for the equal-length-path problem and the unequal-path problem with lengths 3, 4, and 5, respectively.



Figure 4. Experimental simulations were run, using convolution r.d.f.'s, for the search game on 3 equal length arcs. On left is the d.f. with the highest expected payoff to the Hider, 1.55165, out of 500 observations. Right is the actual d.f. of the Hider's analytically conjectured optimal strategy for the equal-length arc search game. Its value to the Hider is 1.56438.



Figure 5. This is the approximate solution of the Search game on 3 arcs of lengths 3, 4, and 5 respectively. From left to right, these are the distributions of the marker given it is on the arc of length 3, 4, or 5 respectively. The empirically optimal probabilities for each arc are .305, .235, and .46, and the value of this approximation is 5.44.

Acknowledgement

The first author thanks the organizers of ICAAA 2002 for the invitation to present these ideas at the ICM Satellite Meeting in Hanoi in August 2002, and for financial support. He also thanks the Hanoi Institute of Mathematics, and in particular Professor Nguyen Van Thu, for their hospitality during the conference.

References

[1] Aldous, D. J. (1993). Exchangeability and related topics, Ecole d'Ete de Probabilité de Saint-Flour XIII. *Lecture Notes in Math.*, 1117, 1-197.
 [2] Blackwell, D. and Kendall, D. (1964). The Martin boundary for Pólya's urn scheme and an application to stochastic population growth. *J. Appl. Prob.*, 1, 284-296.

- [3] Bloomer, L. (2000) Random probability measures with given mean and variance, PhD Thesis, Georgia Institute of Technology.
- [4] Bloomer, L. and Hill, T. (2002). Random probability measures with given mean and variance, *J. Theoretical Probability* **15**, 919–937.
- [5] Bostock, F. (1984). On a search problem on three arcs. *SIAM Journal of Algorithmic Discrete Methods*, **5**(1), 94–100. Cifarelli, D. M. and Regazzini, E. (1990). Distribution functions and means of Dirichlet process. *Ann. Statist.*, **18**(1), 429–442.
- [6] Diaconis, P. and Freedman, D. (1986a). On the consistency of Bayes estimates. *Ann. Statist.*, **14**(1), 1–67, (with discussion).
- [7] Diaconis, P. and Freedman, D. (1986b). On inconsistent Bayes estimates of location. *Ann. Statist.*, **14**(1), 68–87.
- [8] Diaconis, P. and Freedman, D. (1999). Iterated Random Functions. *SIAM Review*, **41**(1), 45–76.
- [9] Dubins, L. and Freedman, D. (1963). Random distribution functions. *Bulletin of the American Mathematical Society*, **69** 548–551.
- [10] Dubins, L. and Freedman, D. (1964). Measurable sets of measures. *Pacific Journal of Mathematics*, **14**, 1211–1222.
- [11] Dubins, L. E. and Freedman, D. A. (1967). Random distribution functions. *Proc. Fifth Berkeley Symp. Math. Statist. Probl.*, **2**, 183–214.
- [12] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 209–230.
- [13] Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**(4), 615–629.
- [14] Gal, S. (1980). *Search Games*. Academic Press, Inc., Harcourt Brace Jovanovich, New York.
- [15] Graf, S., Mauldin, R. D. and Williams, S. C. (1986). Random homeomorphisms. *Advances in Math.*, **60**, 239–359.
- [16] Graf, S., Novak, E. and Papageorgiou, A. (1989). Bisection is not optimal on the average. *Numerische Mathematik*, **55**, 481–491.
- [17] Hengartner W. and Theodorescu, R. (1973). *Concentration Functions*. Academic Press, Inc., Harcourt Brace Jovanovich, New York.
- [18] Hill, T. (1996). A statistical derivation of the significant-digit law. *Statistical Science*, **10**, 354–363.
- [19] Hill, T. and Monticino, M. *Barycenter models for distribution of mass in 2 and 3 dimensions*. In preparation.
- [20] Hill, T. and Monticino, M. (1997). *Sequential barycenter arrays and random probability measures*. Technical Report, School of Mathematics, Georgia Institute of Technology.
- [21] Hill, T. and Monticino, M. (1998). Constructions of random distributions via sequential barycenters. *Ann. Statist.*, **26**(4), 1242–1253.
- [22] Kolmogorov, A. (1941). Über das logarithmisch normale Verteilungsgesetz der Dimension der Teilchen bei Zerstückelung. *Dokl. Acad. Nauk. SSSR*, **31**, 99–101.
- [23] Kraft, C. H. (1964). A class of distribution function processes which have derivatives. *Journal of Applied Probability*, **1**, 358–388.

- [24] Kraft, C. and van Eeden, C. (1964). Bayesian bio-assay. *Annals of Mathematical Statistics*, **35**, 886–890.
- [25] Lavine, M. (1992). Some aspects of Pólya tree distributions for statistical modeling. *Ann. Statist.*, **20**(3), 1222–1235.
- [26] Lavine, M. (1994). More aspects of Pólya tree distributions for statistical modeling. *Ann. Statist.*, **22**(3), 1161–1176.
- [27] Mattner, L. (1993). *Extremal Problems of Probability Distributions: A General Method and Some Examples* Vol. 22, IMS Lecture Notes Monograph, Hayward, CA, 274–283.
- [28] Mauldin, R. D., Sudderth, W. D. and Williams, S. C. (1992). Pólya trees and random distributions. *Ann. Statist.*, **20**(3), 1203–1221.
- [29] Mauldin, R. D. and Monticino, M. G. (1995). Randomly generated distributions. *Israel Journal of Mathematics*, **91**, 215–237.
- [30] Mauldin, R. D. and Williams, S. C. (1990). Reinforced random walks and random distributions. *Proc. Amer. Math. Soc.*, **110**(1), 251–258.
- [31] Monticino, M. (1995). A note on the moments of the mean for a Dubins-Freedman prior. University of North Texas, Department of Mathematics Technical Report.
- [32] Monticino, M. (1998). Constructing prior distributions with trees of exchangeable processes, *Journal of Statistical Planning and Inference*, **73**, 113–133.
- [33] Monticino, M. (2001). How to construct a random probability measure, *International Statistical Review* **69**, 153–167.
- [34] Novak, E. (1988). Stochastic properties of quadrature formulas. *Numerische Mathematik*, **53**, 609–620.
- [35] Ritter, K. (1992), *Average errors for zero finding: lower bounds for smooth or monotone functions*. University of Kentucky Technical Report No. 209-92.
- [36] Sitton, D. (2001). *Generating Random Absolutely Continuous Distributions*, PhD Thesis, Georgia Institute of Technology.
- [37] Skibinsky, M. (1968). Extreme n th moments for distributions on $[0, 1]$ and the inverse of a moment space map, *J. Appl. Probab.* **5**, 693–701.