*International Statistical Review* (2001), **69**, 1, 153–167, Printed in Mexico


# How to Construct a Random Probability Measure

**Michael Monticino**

*Department of Mathematics, University of North Texas, Denton, USA. E-mail: monticno@unt.edu*

## Summary

**This article presents an overview of several constructive methods for generating random probability measures. Applications of random probability measures include Bayesian statistics, average optimal control problems, average error bounds for numerical equation solving methods, and models for random distributions of mass in space.**

*Key words:* Random probability measure; Nonparametric Bayesian inference; Prior distribution; Random homeomorphism; Distribution of mass.

## 1 Introduction

Suppose that you are observing the share price of your favorite stock trying to decide when to sell it so as to maximize the return on your investment. If you had a pretty good idea of the probability distribution of price fluctuations, then the problem of optimizing your return could be phrased in terms of a classical optimal stopping problem and under certain conditions an optimal selling strategy could be determined. More realistically, you may not be so sure about the price distribution. You might only have some partial information, like the average return on the stock. How would you select a good selling strategy in this setting? One way to approach this question would be to determine just how bad each of the available strategies could perform, and then select a strategy which performs best in the worst case. However, while the worst case scenario may be possible, it could be very unlikely. So instead of taking this minimax approach of comparing the worst case performance of strategies, it may be more reasonable to evaluate how well a strategy does on the average. This average performance can be determined by randomly generating probability distributions, determining the value of a strategy against each distribution, and then calculating the average. So how do you generate a probability distribution at random? (Or, how do you construct a homeomorphism or a convex function at random?) That is the topic of this article.

An overview of several methods for randomly generating probability measures is presented. The intention is not to give a rigorous measure-theoretic formulation of the methods, but rather an intuitive description of the generating schemes and a general discussion of the properties and applications of the constructions. The overview will hopefully encourage the reader to learn more about these constructions and their applications by going to the cited references. Moreover, all the constructions described here lend themselves to computer simulation and such simulations are a nice way to start investigating characteristics of the constructions.

Besides the optimal stopping problem mentioned above, random probability measures are used in a wide variety of problems in statistics, probability and analysis. An essential component of Bayesian treatments of statistical problems is the prior. The priors arising within the context of random probability measure constructions are priors on the space of probability measures. More specifically, a random probability measure construction is a technique for specifying a probability measure (prior) on the space of probability measures. The most familiar priors of this type are those defined on some parametric family of probability distributions. For example, a beta distribution on the success probability parameter is often used to define a prior on the class of binomial distributions. Perhaps less widely known are nonparametric priors like those discussed in this paper. An advantage that nonparametric constructions have over parameter-based priors is that the former have large support on the space of probability measures. Hence the general feeling that inferences based upon nonparametric priors may avoid biases potentially introduced by the selection of a particular parametric family. A class of random probability measures called Dirichlet processes is a common choice for nonparametric analysis. An excellent overview of Dirichlet processes is given by Ferguson (1974). Important considerations for applications of Dirichlet processes, and nonparametric random probability measure constructions in general, are given by Diaconis & Freedman (1986a,b). Good collections for delving into the wide variety of work in nonparametric Bayesian analysis, both theoretical and applied, are Bernardo, Berger, Dawid & Smith (1996) and Dey, Muller & Sinha (1998). An important recent article by Diaconis & Freedman (1999) attempts a unifying approach to random constructions, including Dirichlet processes and some of the techniques discussed here.

Dubins & Freedman (1967) note that any method for generating a probability measure at random is equivalent to selecting a non-decreasing right continuous function at random, and generating a continuous probability measure with full support is equivalent to selecting a homeomorphism[1] at random. A scheme for randomly generating homeomorphisms defines a probability measure on the space of homeomorphisms and such a measure allows quantitative statements about the behavior of homeomorphisms to be made. The importance of this is discussed by Ulam (1982) and Graf, Mauldin & Williams (1986). Furthermore, viewed from this latter perspective—as defining measures on function spaces—methods for randomly generating probability measures can be applied to obtain average case errors for numerical methods of equation solving. As mentioned in Novak (1988), worst case errors for numerical methods are generally much larger that those encountered with most functions. Therefore, average error bounds may provide a better way to compare different numerical algorithms. Such average case errors are investigated by Graf, Novak & Papageorgiou (1989) and Ritter (1992).

Random probability measures also occur naturally in models describing distributions of mass in space. These models come in varying degrees of complexity ranging from Kolmogorov's (1941) "rock-crushing" model, to multi-type branching random walks, to splitting diffusion processes, to embeddings of random graphs in continuous space studied by Aldous (1993).

Features shared by the constructions surveyed here are discussed in section 2. Section 3 describes the generation scheme presented by Dubins & Freedman (1967). A special case of the Dubins–Freedman approach is the random rescaling scheme explored by Graf, Mauldin & Williams (1986) and later generalized by Mauldin & Monticino (1995). Section 4 discusses generating techniques motivated by Polya urn schemes. A method given by Hill & Monticino (1998) for generating probability measures from sequential barycenter arrays is described in section 5. Some examples of using random probability measure constructions to evaluate average-case performance of stopping strategies are given in section 6.

[1] Recall that a homeomorphism from an interval $[a, b]$ onto an interval $[\alpha, \beta]$ is a one-to-one function, $f$, from $[a, b]$ onto $[\alpha, \beta]$ such that $f$ and its inverse function are continuous. So, for example, the distribution function of a continuous measure with full support on the interval $[0, 1]$ is a homeomorphism from $[0, 1]$ onto $[0, 1]$. In the other direction, a homeomorphism from $[0, 1]$ onto $[0, 1]$ which leaves the endpoints 0 and 1 fixed defines a distribution function for a continuous measure with full support on the interval.

## 2 Common Features, Different Motivations

The random probability measure (rpm) constructions discussed here share two important common features. First, each uses a base probability measure as the means of introducing randomness into the construction. For example, Graf, Mauldin & Williams (1986) use a fixed probability measure on the unit interval [0, 1] to randomly define the values of a distribution function on the dyadic rationals within [0, 1] (the function is then extended, in the natural way, to a distribution function on the entire interval). In the urn-based constructions studied by Mauldin, Sudderth & Williams (1992), the base measure lurks in the background as the directing or de Finetti measure of the exchangeable urn processes. Second, each method involves a recursive process which introduces a degree of "statistical self-similarity" into the construction.

Ferguson (1973) states that for a rpm construction to be useful it should be analytically manageable and the prior associated with the construction should have large support—that is, the construction should generate measures which are dense in the desired class of measures. Interpreting "analytically manageable" somewhat more broadly than Ferguson, the constructions discussed here are manageable in the sense that basic properties such as whether the measures generated are continuous, discrete, or singular can be verified, the average distribution can be calculated, and (as Ferguson intended) the posterior distributions are often tractable. This manageability stems largely from the similarity properties of the constructions. Conditions ensuring that the priors have large support are easily stated in terms of conditions on the base measure. Moreover, by choosing an appropriate base measure, all the techniques can give probability one to the set of continuous measures which have full support on, say, the interval [0, 1]. This contrasts with the Dirichlet priors given by Ferguson (1973, 1974) which have support on the set of discrete measures.

While sharing these common features, the constructions were often motivated by different applications. The random rescaling constructions developed by Graf, Mauldin & Williams (1986) and Mauldin & Monticino (1995) were motivated by questions about flows and homeomorphisms, while the study of Polya trees was initiated by problems in Bayesian statistics. The differing motivations are reflected in the properties investigated. For instance, Mauldin & Monticino (1995) determined the dimensions of the support of the generated distributions while Mauldin, Sudderth & Williams (1992) were more concerned with statistical ideas such as conjugate families, consistency, and the predictive distribution.

## 3 Dubins and Freedman Random Distribution Functions

Dubins' & Freedman's (1967) seminal work on random probability constructions generates a probability measure on the interval [0, 1] by randomly generating a distribution function, $h$, as follows. Let $\mu$ be a probability measure on the unit square $S = [0,1] \times [0,1]$—e.g., uniform (Lebesgue) measure on $S$. Set $h(0) = 0$ and $h(1) = 1$. Select a point $(x_0, y_0) \in S$ according to $\mu$. Let $S_1$ be the lower left rectangle of $S$ which has vertices $(0,0)$, $(0, y_0)$, $(x_0, 0)$, $(x_0, y_0)$. (See Figure 1.) Mapping $S$ onto $S_1$ scales $\mu$ into a measure $\mu_1$ on $S_1$. Analogously, let $S_2$ be the upper right rectangle with vertices $(1,1)$, $(1, y_0)$, $(x_0, 1)$, $(x_0, y_0)$. And mapping $S$ onto $S_2$ scales $\mu$ into a measure $\mu_2$ on $S_2$. Select a point $(x_1, y_1)$ at random in $S_1$ according to $\mu_1$ and independently select a point $(x_2, y_2)$ in $S_2$ according to $\mu_2$. These points now define upper right and lower left rectangles within $S_1$ and $S_2$ (four rectangles in all) and respective scaled versions of $\mu$. Continue in this manner so that at the $n$th stage there is a set of $2^n$ rectangles. Dubins & Freedman (1967, Section 2) show that if $\mu$ assigns measure 0 to the corners $(0,0)$ and $(1,1)$ of $S$, then in the limit the intersection of these sets of rectangles is a probability distribution function with $h(x_0) = y_0$, $h(x_1) = y_1$, $h(x_2) = y_2$, and so on.

Insight into the types of distribution functions which are generated by this method can be gained by considering a few straightforward (nonrandom) examples. For instance, suppose $\mu$ gives unit mass to the point $(\frac{1}{3}, \frac{1}{3})$. The first couple of stages of the construction yield $(x_0, y_0) = (\frac{1}{3}, \frac{1}{3})$, $(x_1, y_1) = (\frac{1}{9}, \frac{1}{9})$ and $(x_2, y_2) = (\frac{5}{9}, \frac{5}{9})$, which suggests that the distribution function generated is the identity function on [0, 1]—that is, the distribution function for the uniform probability measure over [0, 1]. Indeed this is the case. More generally, with a moments reflection on the scaling mappings occurring in the process, it is easy to see that if the support of $\mu$ lies on the main diagonal of $S$ then only (with probability one) the uniform distribution is generated.

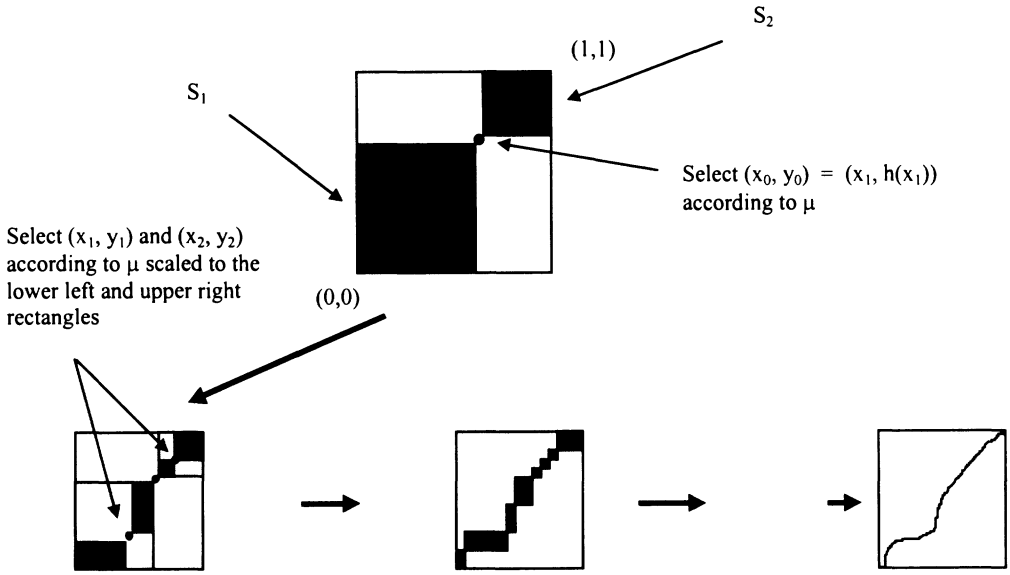



**Figure 1.** *Dubins–Freedman construction of a distribution function.*

On the other hand, suppose $\mu$ gives unit mass to the point $(\frac{1}{2}, \frac{1}{3})$. The distribution in this case is decidedly more interesting than in the first example. While the distribution function is continuous and strictly increasing, it does not have a finite positive derivative anywhere—for example, $h'(0) = \lim_{n\to\infty} \frac{(\frac{1}{3})^n}{(\frac{1}{2})^n} = 0$. Not having a finite positive derivative anywhere is termed *strictly singular*. Dubins & Freedman (1967) showed the following.

THEOREM.

*i) Almost all distributions generated are continuous if and only if $\mu$ assigns probability 0 to the vertical edges of S and $\mu$ assigns positive probability to the interior of S.*

*ii) Almost all distributions generated are purely discrete if either $\mu$ assigns probability 1 to the horizontal edges of S or $\mu$ assigns positive probability to the vertical edges of S.*

*iii) If $\mu$ does not give probability 1 to the main diagonal of S, then almost all the generated distribution functions are strictly singular.*

Let $P([0, 1])$ denote the set of probability measures on the interval [0, 1] and let $D_\mu$ be the prior on $P([0, 1])$ defined by the Dubins–Freedman scheme using base measure $\mu$. The *average distribution function* with respect to $D_\mu$ is defined to be

$$F_{\mu}(x) = \int_{\pi \in P([0,1])} \pi([0,x]) dD_{\mu}(\pi).$$

Dubins & Freedman (1967) demonstrate a number of properties of the average distribution function, including conditions under which it is continuous, singular or absolutely continuous (with respect to Lebesgue measure), and when it is the uniform distribution. Hill (1996) has recently shown a surprising connection between $F_{\mu}$ being the uniform distribution and Benford's Law—that is, the tendency of the first significant digit of numbers drawn at random from newspapers, almanacs, scientific tables, etc. to exhibit a logarithmic distribution.

**Random Rescaling RPMs.** A special case of the Dubins–Freedman construction is the random rescaling scheme investigated in depth by Graf, Mauldin & Williams (1986). Again, a probability measure on [0, 1] is produced by generating its associated distribution function. This time the base measure $\mu$ is a probability measure with support on the interval [0, 1]. Set $h(0) = 0$ and $h(1) = 1$.

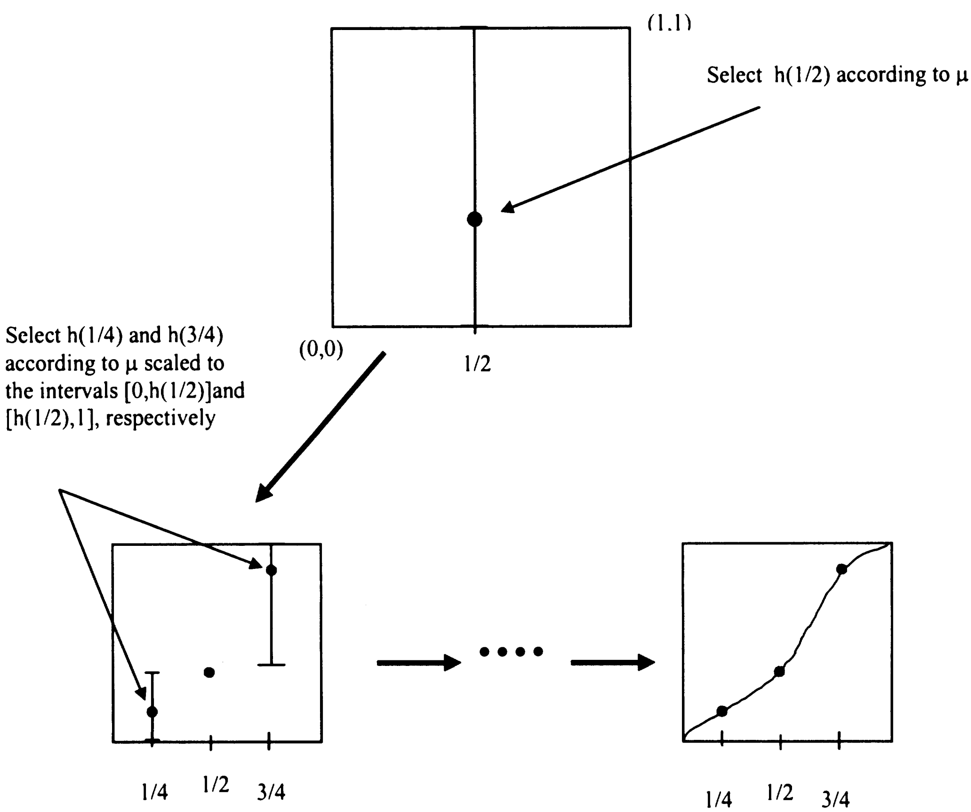



**Figure 2.** *The Graf–Mauldin–Williams rescaling scheme for generating a random probability measure.*

Now, as illustrated in Figure 2, select $h(1/2)$ according to $\mu$. Next select $h(1/4)$ according to $\mu$ rescaled to the interval $[0, h(1/2)]$ (according to the map of [0, 1] onto $[0, h(1/2)]$) and independently select $h(3/4)$ according to $\mu$ rescaled to the interval $[h(1/2), 1]$. Continue in this manner to define a function on the dyadic rationals. With probability one, $h$ extends to a probability distribution function on $[0, 1]$. This scheme is equivalent to the Dubins–Freedman method by taking a base measure on the square $S$ with support on the vertical line $\{1/2\} \times [0, 1]$. Graf, Mauldin & Williams' (1986) main focus is examining the properties of the generated distributions while viewing them as homeomorphisms on $[0, 1]$ (which they almost surely are if $\mu$ does not give mass to the endpoints 0 and 1). Their work was initially motivated by certain questions posed by Ulam about almost sure properties of homeomorphisms and flows. The three main classes of properties they investigate are

- invariance properties (e.g., time-reversal and inversion properties),
- derivative properties, and
- fixed point properties.

While rescaling $\mu$ at each step of the construction induces useful similarity properties, it is not the only way in which to "re-fit" $\mu$ into the subintervals at each stage of the construction. Graf, Mauldin & Williams (1986) also consider renormalizing $\mu$ at each stage and show, among other results, that if $\mu$ has full support on $[0, 1]$ then almost surely the re-normalizing scheme produces continuous distributions.

Mauldin & Monticino (1995) extend the Graf–Mauldin–Williams rescaling construction to allow possibly different base measures at each stage of the construction. The derivative structure as well as the dimension of the supports of the resulting distributions are examined. These results complement work by Kraft (1964) which shows how to choose base measures at each stage so that absolutely continuous (with respect to Lebesgue measure) distributions are generated almost surely.

A useful inferential tool in Bayesian statistics is the distribution of the mean with respect to a prior. That is the distribution of $\int x \, d\pi(x)$, where $\pi$ is a random probability measure chosen according to some prior. Determining this distribution, or even finding its expected value and variance, is not obvious for a general Dubins–Freedman or random rescaling prior. However, Mauldin & Williams (1990) used a construction introduced by Ferguson (1974) involving a tree of Polya urns for finding the mean and variance of the distribution of the mean for a specific (and important) Dubins–Freedman prior. Monticino (1995) extended their ideas to calculate all the moments of the distribution of the mean for this special prior. Besides providing some insight into the distribution of the mean, Mauldin & Williams (1990) initiated the development of two other classes of rpm constructions—the Polya tree constructions discussed in the next section and the sequential barycenter array constructions outlined in section 5.

## 4 Polya and Exchangeable Tree Constructions

Polya and exchangeable tree rpm constructions provide convenient conjugate families of priors for statistical models in which an experiment proceeds in several stages and each stage depends on the previous outcomes. Lavine (1992, 1994) discusses a variety of applications of these priors, including how to use them to model uncertainty in a model for pressurized vessel lifetimes.

To describe these priors, it is necessary to review some basic notions about exchangeable processes. A sequence $X = (X_1, X_2, ...)$ of $\mathcal{B}$-valued random variables is exchangeable if the distribution of $X$ is invariant under finite permutation of the indices. That is, for each $n$, measureable set $A \subseteq \mathcal{B}^n$ and permutation $\phi$ of $\{1, ..., n\}$,

$$P[(X_1, \ldots, X_n) \in A] = P[(X_{\phi^{-1}(1)}, \ldots, X_{\phi^{-1}(n)}) \in A].$$

Or equivalently, for each $n$, measurable function $f : \mathcal{B}^n \to \mathbb{R}$ and permutation $\phi$ of $\{1, ..., n\}$,

$$\int f(x_1, \ldots, x_n) d\rho^n = \int f(x_{\phi^{-1}(1)}, \ldots, x_{\phi^{-1}(n)}) d\rho^n,$$

where $\rho^n$ is the marginal distribution of $X$ on $\mathcal{B}^n$.

One way to generate a sequence of exchangeable $\{0, 1\}$-valued random variables is as follows. Using a fair die decide upon one of two possible sequences of coin tosses. Throw the die. If a 2 or 3 shows, flip a fair coin repeatedly (forever) recording a 1 for heads and a 0 for tails. Otherwise, flip a biased, say $P(\text{heads}) = .8$, coin forever. The resulting "mixture" of sequences of 0's and 1's is easily seen to be exchangeable. More generally, suppose $\Theta$ is a random probability measure. Select a distribution $\theta$ according to $\Theta$ and then generate $X_1, X_2, \ldots$ independent with common distribution $\theta$. The resulting process is exchangeable. De Finetti's theorem, given below, is essentially the converse to this observation—that every infinite exchangeable sequence is a random mixture of sequences of independent identically distributed random variables. In the die-coin example, given the outcome of the die, the resulting variables are independent with the same distribution.

Suppose $\mathcal{B}$ is a Borel space and let $P(\mathcal{B})$ denote the set of probability measures on $\mathcal{B}$.

THEOREM. *If $X = (X_1, X_2, \ldots)$ is an exchangeable sequence of $\mathcal{B}$-valued random variables, then there exists a unique probability measure $Q$ on $P(\mathcal{B})$ such that for each Borel subset A of $\mathcal{B}^\infty = \mathcal{B} \times \mathcal{B} \times \ldots$,*

$$P[X \in A] = \int_{\theta \in P(\mathcal{B})} \theta^\infty(A) dQ(\theta),$$

*where $\theta^\infty = \theta \times \theta \times \ldots$ is the infinite product measure on $\mathcal{B}^\infty$.*

The measure $Q$ is called the *de Finetti* or *directing measure* of $X$. Good resources for information on exchangeable sequences and their directing measures are Hewitt & Savage (1955) and Aldous (1983).

Another method for generating a sequence of exchangeable $\{0, 1\}$-valued random variables is the classic Polya urn scheme. Suppose an urn initially contains $n_0$ balls labeled "0" and $n_1$ balls labeled "1". On each play, a ball is selected at random from the urn and replaced by two balls of the same type. So if a 1 ball is drawn on the first play, there will be $n_1 + 1$ balls of type 1 and $n_0$ balls of type 0 in the urn just before the second play. Let $X_n$ be the type—0 or 1—of ball selected on the $n$th play. The $X_n$ are an exchangeable sequence of $\{0, 1\}$-random variables. A discussion of elementary properties and applications of Polya urn schemes is given by Feller (1968). If $X = (X_1, X_2, \ldots)$ is an exchangeable $\{0, 1\}$-valued sequence, then the empirical distribution converges to a Bernoulli distribution with parameter equal to the limit of the empirical fraction of 1's in $X$. The de Finetti measure of $X$ is a probability measure on the realizations of these limiting Bernoulli distributions, which in this setting allows a straightforward description. Identify the limiting Bernoulli distribution with its defining parameter. Then the de Finetti measure can be viewed as the distribution on the limiting fraction of 1's. For the case of a Polya urn consisting of $n_0$ 0 balls and $n_1$ 1 balls, the de Finetti measure has a beta distribution with parameters $n_0$ and $n_1$ (see Freedman (1965) and Blackwell & Kendall (1964)). So the Polya urn scheme provides a way to obtain a prior with full support on the set of probability measures on $\{0, 1\}$. Mauldin & Williams (1990) observed that by creating a tree of Polya urns they could extend this idea to randomly generate probability measures on the unit interval.

Let $F^*$ be the set of all finite sequences of elements of $F = \{0, 1\}$ including the empty sequence $\varnothing$. For each $p \in F^*$, let $U(p)$ be an urn containing balls labeled either 0 or 1. This can be pictured (see Figure 3) as a binary tree with root $\varnothing$ and an urn at each node. Motivated by the picture the construction is called a Polya tree. The tree of urns can be used to generate a sequence $X_1, X_2, \ldots$

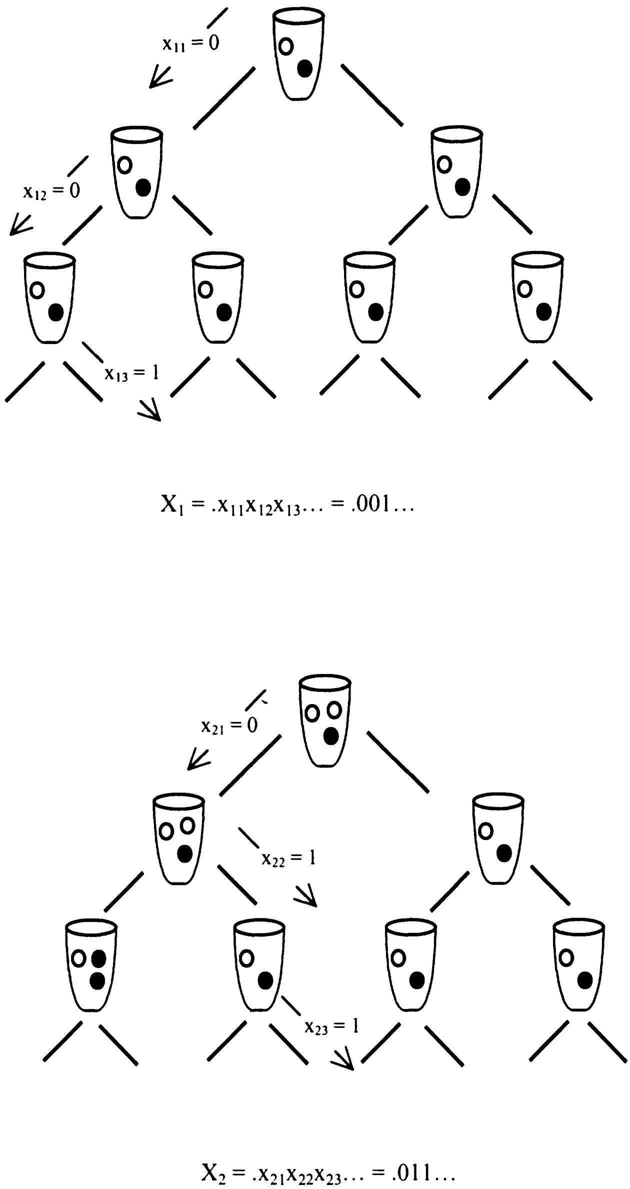



**Figure 3.** *Two walks down a Polya tree used to generate $X_1$ and $X_2$ (white balls represent "0" balls and black balls represent "1" balls). Note that on the second walk the urns are only modified along the path of the first walk.*

of $F^{\infty}$-valued random variables as follows. Draw a ball at random from urn $U(\varnothing)$ and replace it with two balls of the same type. Let $X_{1,1}$, the first coordinate of $X_1$, be the label of the selected ball. Now draw from urn $U(X_{1,1})$ and replace it by two of the same type. Let $X_{1,2}$ be the label of the ball selected from $U(X_{1,1})$. Next draw from urn $U(X_{1,1}, X_{1,2})$ to determine $X_{1,3}$, and so on. The random variable $X_2$ is obtained similarly to $X_1$, except $X_2$ is generated by the modified Polya tree. Note that this modified or reinforced tree is the same as the original tree except for the urns $U(\varnothing), U(X_{1,1}), U(X_{1,1}, X_{1,2}), \ldots$. Figure 3 illustrates two walks down the Polya tree. Continue the process to obtain $\{X_n\}_{n\geq 1}$. The $X_n$ can be viewed as $[0, 1]$-valued random variables by taking $X_n$ to be the number with dyadic expansion

$$X_n = 0.X_{n,1}X_{n,2}X_{n,3}\ldots$$

The $X_n$ are exchangeable. This was shown by Mauldin & Williams (1990) for the case in which each urn initially contains exactly one ball labeled 1 and one ball labeled 0. Mauldin, Sudderth & Williams (1992) generalized the construction to more than two types of balls and an arbitrary mix of balls in the urns. The directing measures of the generated sequences can be viewed as priors, *Polya tree priors*, which select probability measures on [0, 1] at random. (As discussed in Mauldin Sudderth & Williams (1992) and Lavine (1992), the sequences $X_{n,1}, X_{n,2}, \ldots$ can be mapped into spaces other than [0, 1]—for example, into a probability simplex of appropriate dimension.)

Mauldin, Sudderth & Williams (1992) establish a number of important properties of Polya tree priors. These include:

- showing that the set of Polya tree priors forms a conjugate class of priors,
- developing conditions under which a Polya tree prior has support on the set of continuous probability measures,
- establishing conditions under which a prior has full support on $P([0, 1])$, and
- specifying the form of the average distribution function.

The motivation behind the Polya tree scheme was to construct a useful class of conjugate priors for Bayesian applications and for which desirable properties, like having support on the set of continuous distributions, are easily verifiable. The construction does indeed meet these objectives. However, unlike the Dubins–Freedman or random rescaling constructions, the Polya tree scheme does not directly generate a probability distribution. Rather the walks down the tree of urns generate samples from the mixture of distributions defined by the Polya tree prior. Yet, it turns out that when viewed as measures on distribution functions, Polya trees produce the same rpms as those constructed through a random rescaling construction. For example, if each urn initially has exactly one ball labeled 0 and one labeled 1, then the resulting Polya tree prior is the same as that generated by the random rescaling scheme with the base measure $\mu$ equal to the uniform distribution on [0, 1]. Monticino (1998) establishes the generality of this connection between Polya tree and random rescaling rpms by connecting the directing measures associated with the urn processes at each node of the Polya tree to the base measures of the generalized random rescaling scheme given in Mauldin & Monticino (1995). So the distributions generated implicitly by the Polya urn scheme can be explicitly generated by translating the scheme to the corresponding random rescaling scheme. Moreover, Monticino (1998) shows that the tree of urn processes in the Polya scheme can be replaced by a tree of arbitrary exchangeable processes while retaining the fundamental properties of the Polya trees.

## 5 Sequential Barycenter Array Constructions

The motivation for the sequential barycenter array (SBA) construction was to develop a general and natural method for randomly generating probability measures with a prescribed mean or distribution on the mean. None of the other methods discussed here or elsewhere in the literature generate random measures with *a priori* specified means. Moreover, even the calculation of the distribution of the means for special cases of these constructions is difficult (see Cifarelli & Regazzini (1990) and Monticino (1995)).

Sequential barycenter arrays, though not called by that name, are used in standard proofs of Skorohod's embedding theorems (see, Billingsley (1986, Section 37)). It is the reversal of this standard procedure which forms the basis for the random probability measure construction. To specify the SBA construction, a couple of definitions are needed.

*Definition.* Let $F$ be the distribution function for a random variable $X$ with $E[|X|] < \infty$. The *F-barycenter* of $(a, c]$, $b_F(ac]$, is given by

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

In other words, the $F$-barycenter of $(a, c]$ is the conditional expectation of $X$ over the interval $(a, c]$. The sequential barycenter array of a random variable is an array of successive conditional expectations.

*Definition.* The *sequential barycenter array* of a random variable, $X$, with support on [0, 1] is the triangular array $\{m_{n,k}\}_{n=1,k=1}^{\infty,2^n-1}$ defined by

1. $m_{1,1} = E[X]$,
2. $m_{n,2j} = m_{n-1,j}$, for $n \geq 1$ and $j = 1, \ldots, 2^{n-1} - 1$,
3. $m_{n,2j-1} = b_F(m_{n-1,j-1}, m_{n-1,j}]$, for $j = 1, \ldots, 2^{n-1}$, with the convention that $m_{n,0} = 0$ and $m_{n,2^n} = 1$.

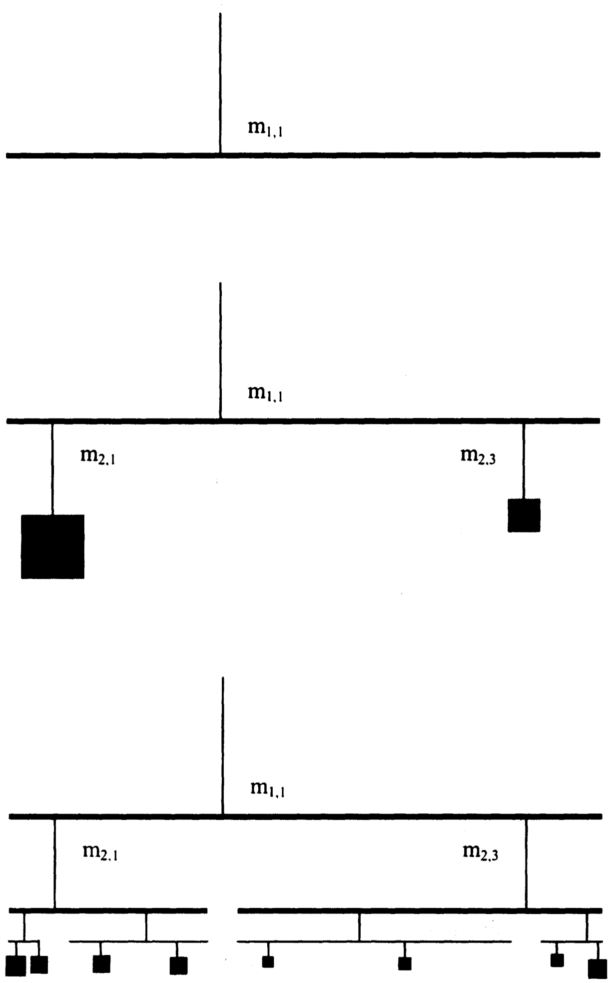
For example, if $X$ is the uniform distribution over [0, 1], then its sequential barycenter array is $\left\{\frac{k}{2^n}\right\}_{n=1,k=1}^{\infty,2^n-1}$.

Hill & Monticino (1998) show that a probability measure is completely determined by its sequential barycenter array, and they give an inversion formula for recovering the distribution function of the probability measure from its SBA. The basic idea of the SBA construction is as follows.

Let $\mu_0$ and $\mu$ be probability measures with support on [0, 1] and [0, 1), respectively. Select a point $m_{1,1}$ in [0, 1] according to $\mu_0$. This point will be the mean (expected value) of the rpm. Given $m_{1,1}$ next choose the distance to the two successive barycenters, $m_{2,1}$ and $m_{2,3}$, independently according to $\mu$ rescaled to the intervals $[0, m_{1,1}]$ and $[m_{1,1}, 1]$ respectively. Continue selecting each successive level of barycenters using rescaled versions of $\mu$. The resulting array is the SBA for a unique probability measure on [0, 1]. That is, a random probability measure is generated by generating its associated SBA. Figure 4 shows a picturesque way to view the scheme as a way of constructing a random mobile.

First select the balancing point, $m_{1,1}$, for the primary arm. Next select the balancing points, $m_{2,1}$ and $m_{2,3}$, for the secondary arms. Since we want the arms to balance like a mobile, once $m_{1,1}$, $m_{2,1}$ and $m_{2,3}$ are given, the mass assigned to each secondary arm (out of a total mass of 1) is completely determined—that is, the value of distribution function at $m_{1,1}$ is determined. Select the balancing points for the tertiary arms, and so on.

For example, consider the (non-random) example in which $\mu_0$ is unit mass at 1/3 and $\mu$ is unit mass at $\frac{1}{2}$. Then $m_{1,1} = 1/3$. The point $m_{2,1}$ is $\frac{1}{2}$ the distance from $m_{1,1}$ to $m_{1,0} = 0$, so $m_{2,1} = 1/6$. The point $m_{2,3}$ is $\frac{1}{2}$ the distance from $m_{1,1}$ to $m_{1,2} = 1$, so $m_{2,3} = 2/3$. Analogously, $m_{3,1} = 1/12$ is $\frac{1}{2}$ the distance from $m_{2,1}$ to $m_{2,0}$; $m_{3,3} = 1/4$ is $\frac{1}{2}$ the distance from $m_{2,1}$ to $m_{2,2}$; and so on. The corresponding distribution function for this SBA can be calculated by applying the definitions. For




First, select the balancing point for the primary arm of the mobile.

Then, select the balancing points for the secondary arms which determines the mass assigned to each arm.

Continue selecting the successive balancing points and the corresponding distribution of mass.

**Figure 4.** *The SBA scheme viewed as a mobile construction.*

instance,

$$m_{1,1} = m_{2,1}(F\left(\frac{1}{3}\right) - F(0)) + m_{2,3}(F(1) - F\left(\frac{1}{3}\right)) = m_{2,1}(F\left(\frac{1}{3}\right) - 0) + m_{2,3}(1 - F\left(\frac{1}{3}\right))$$

which yields

$$F\left(\frac{1}{3}\right) = \frac{m_{2,3} - m_{1,1}}{m_{2,3} - m_{2,1}} = \frac{2/3 - 1/3}{2/3 - 1/6} = \frac{2}{3}.$$

Moreover, in this case

$$F(x) = \begin{cases} 2x & 0 \le x \le \frac{1}{3} \\ \frac{2}{3} + \frac{1}{2}(x - \frac{1}{3}) & \frac{1}{3} \le x \le 1 \end{cases}$$

Note that $\mu_0$ is the distribution on the mean of the generated measures. So the SBA construction gives a relatively straightforward way to produce rpm's with any prescribed distribution on the mean. Hill & Monticino (1998) discuss several other properties of the SBA construction, including when the generated measures have full support on the interval [0, 1], conditions under which the measures are strictly singular, as well as the following result.

THEOREM. *1) If $\mu_0(\{0, 1\}) = 0 = \mu(\{0\})$, then almost all SBA measures are continuous.*
*2) If $\mu(\{0\}) > 0$, then almost all SBA measures are discrete.*
*3) If $\mu(\{0\}) \ge 1 - \frac{1}{\sqrt{2}}$, then almost all SBA measures have finite support.*

## 6 Applications and Extensions

As mentioned above, random probability measures have found applications in a variety of settings, from determining average-case errors for numerical equation solving methods to Bayesian models of pressurized vessel lifetimes. For another example of these applications, consider the following special cases of the optimal stopping problem discussed in the introduction.

A sequence of random variables, $X_1, X_2, X_3$ is to be observed. Suppose, initially, all that is known about the $\{X_i\}$ is that they are independent and take values in [0, 1]. The objective is to select a stop rule, $t$, which maximizes $E[X_t]$ on the average, over all such $\{X_i\}$.

Using a standard backwards induction argument (Chow, Robbins & Siegmund, 1971) it can be shown that there is an optimal stop rule, $t_c$, of the form

$$t_c = 1, \quad \text{if } X_1 > c$$

$$t_c = 2, \quad \text{if } t_c > 1 \text{ and } X_2 > m = \int \left[\int x dF(x)\right] dP(F)$$

$$t_c = 3, \quad \text{otherwise,}$$

where

$$c = \int \left[\int_{x>m} x dF(x) + mF(m)\right] dP(F)$$

and $P$ is the prior for $X_1, X_2, X_3$.

*Example.* Suppose it is additionally known that $X_1, X_2, X_3$ have identical means equal to $m$ (a constant). A natural prior in this case is the sequential barycenter array prior, $B_{\delta_m,\lambda}$, which takes $\mu_0$ equal to unit mass at $m$, $\delta_m$, and $\mu$ equal to Lebesgue measure, $\lambda$, on the unit interval [0, 1]. With this prior, $c$ can be explicitly calculated, yielding

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

$$= m + m(1-m) \int_0^1 \int_0^1 \frac{xy}{(1-m)y + mx} dxdy$$

$$= m + \frac{(1-m)^3 \ln(1-m) + m^3 \ln(m) + m(1-m)}{3m(1-m)}$$

For example, if $m = \frac{1}{2}$, then the optimal stop rule is given by

$$t_c = 1, \quad \text{if } X_1 > c \approx .6023$$

$$t_c = 2, \quad \text{if } t_c > 1 \text{ and } X_2 > m = \tfrac{1}{2}$$

$$t_c = 3, \quad \text{otherwise,}$$

*Example.* On the other hand suppose that no additional information about $X_1, X_2, X_3$ is known (other than they are independent and take values in [0, 1]). Then a reasonable "no information" prior is the Dubins–Freedman (random rescaling) prior, $D_\mu$, with base measure $\mu$ taken to be uniform on the horizontal line $\{(x, y) : 0 \leq y \leq 1, x = \frac{1}{2}\}$. This case can be shown to reduce to a standard *i.i.d.* optimal stopping problem in which the common distribution is the average distribution $F_\mu(x) = \int_{\pi \in P([0,1])} \pi([0, x]) dD_\mu(\pi)$. Dubins & Freedman (1967, Theorem 9.28) show that $F_\mu$ is the uniform distribution on the interval [0, 1] when $\mu$ is uniform on the vertical bisector. Which yields the following optimal stopping rule

$$t_c = 1, \quad \text{if } X_1 > .625$$

$$t_c = 2, \quad \text{if } t_c > 1 \text{ and } X_2 > 1/2 \; .$$

$$t_c = 3, \quad \text{otherwise,}$$

It is interesting to note that this is a slightly greedier (at time 1) stopping rule than in the case in which the means of $X_1, X_2, X_3$ are known to be $\frac{1}{2}$.

*Example.* Suppose now that the $X_1, X_2, X_3$ are known to have identical medians equal to $q$. It is straightforward to generalize the Dubins–Freedman construction so that the first point selected is $(q, 1/2)$. Then the construction proceeds as before with base measure $\mu$, say, equal to the uniform distribution on the horizontal line $\{(x, y) : 0 \leq x \leq 1, y = \frac{1}{2}\}$. This will generate distributions with medians equal to $q$ and will select the first and third quantiles (and successive "binary" percentiles) at random uniformly. Using Dubins & Freedman (1967, Theorem 9.21), the density of the average distribution can be derived as

$$f_q(x) = \begin{cases} \dfrac{1}{2\pi\sqrt{x(q-x)}} & x < q \\ \dfrac{1}{2\pi\sqrt{(x-q)(1-x)}} & x > q \end{cases} .$$

Giving the optimal strategy

$$t_c = 1, \text{ if } X_1 > c = \int_m^1 x f_q(x)dx + m \int_0^m f_q(x)dx$$

$$t_c = 2, \text{ if } t_c > 1 \text{ and } X_2 > m = \frac{1}{4} + \frac{q}{2}$$

$$t_c = 3, \text{ otherwise,}$$

If $q$ is assumed to be $\frac{1}{2}$, then the optimal strategy is the same as in the above no information case (although the values of the strategies are different under the different assumptions).

Lastly, an interesting area for further research is extending the constructions discussed in this paper to two, three and higher dimensional probability measures. It does not appear that any extensions to the Dubins–Freedman construction have been studied. Mauldin, Sudderth & Williams (1992) suggest a way to extend the Polya tree scheme to multi-dimensional simplexes, but little investigation has been done in this area. Hill & Monticino (in preparation) have devised several ways to generalize the SBA construction to higher dimensions. One method involves a rather direct extension to simplexes. Another is based on the "mobile" interpretation of the SBA construction, and can be formulated as a multitype branching random walk model. Aldous (1993) presents a related multi-dimensional mass distribution model based upon embedding random discrete graphs in continuous space.

## References


Aldous, D.J. (1993). Exchangeability and related topics, Ecole d'Ete de Probabilité de Saint-Flour XIII. *Lecture Notes in Math.*, **1117**, 1–197.

Bernardo, J.M., Berger, J.O., Dawid, A.P. & Smith, A.F.M. (eds). (1996). *Bayesian Statistics 5*: Proceedings of the Fifth Valencia International Meeting. Oxford: Clarendon Press.

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

Blackwell, D. & Kendall, D. (1964). The Martin boundary for Polya's urn scheme and an application to stochastic population growth. *J. Appl. Prob.*, **1**, 284–296.

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

Cifarelli, D.M. & Regazzini, E. (1990). Distribution functions and means of Dirichlet process. *Ann. Statist.*, **18**(1), 429–442.

Dey, D., Muller, P. & Sinha, D. (eds.) (1998). Practical Nonparametric and Semiparametric Bayesian Statistics. *Lecture Notes in Statistics*, **133**. New York: Springer-Verlag.

Diaconis, P. & Freedman, D. (1986a). On the consistency of Bayes estimates. *Ann. Statist.*, **14**(1), 1–67, (with discussion).

Diaconis, P. & Freedman, D. (1986b). On inconsistent Bayes estimates of location. *Ann. Statist.*, **14**(1), 68–87.

Diaconis, P. & Freedman, D. (1999). Iterated Random Functions. *SIAM Review*, **41**(1), 45–76.

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

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications* (3rd ed.). New York: John Wiley & Sons, Inc.

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

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

Freedman, D.A. (1965). Bernard Friedman's urn. *Ann. Math. Statistics*, **36**, 956–970.

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

Graf, S., Novak, E. & Papageorgiou, A. (1989). Bisection is not optimal on the average. *Numerische Mathematik*, **55**, 481–491.

Hewitt, E. & Savage, L.J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.*, **80**, 470–501.

Hill, T. (1996). A statistical derivation of the significant-digit law. *Statistical Science*, **10**, 354–363.

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

Hill, T. & Monticino, M. (1998). Constructions of random distributions via sequential barycenters. *Ann. Statist.*, **26**(4), 1242–1253.

Kolmogorov, A. (1941). Über das logarithmisch normale Verteilungsgesetz der Dimension der Teilchen bei Zerstuckelung. *Dokl. Acad. Nauk. SSSR*, **31**, 99–101.

Kraft, C.H. (1964). A class of distribution function processes which have derivatives. *Journal of Applied Probability*, **1**, 385–388.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modeling. *Ann. Statist.*, **20**(3), 1222–1235.

Lavine, M. (1994). More aspects of Polya tree distributions for statistical modeling. *Ann. Statist.*, **22**(3), 1161–1176.
Mauldin, R.D., Sudderth, W.D. & Williams, S.C. (1992). Polya trees and random distributions. *Ann. Statist.*, **20**(3), 1203–1221.
Mauldin, R.D. & Monticino, M.G. (1995). Randomly generated distributions. *Israel Journal of Mathematics*, **91**, 215–237.
Mauldin, R.D. & Williams, S.C. (1990). Reinforced random walks and random distributions. *Proc. Amer. Math. Soc.*, **110**(1), 251–258.
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.
Monticino, M. (1998). Constructing prior distributions with trees of exchangeable processes. *Journal of Statistical Planning and Inference*, **73**, 113–133.
Novak, E. (1988). Stochastic properties of quadrature formulas. *Numerische Mathematik*, **53**, 609–620.
Ritter, K. (1992). *Average errors for zero finding: lower bounds for smooth or monotone functions*. University of Kentucky Technical Report No. 209-92.
Ulam, S.M. (1982). Transformation, iterations, and mixing flows. In *Dynamical Systems II*, Eds. A.R. Bednarek & L. Cesari, pp. 419–416. New York: Academic Press.

## Résumé

Cet article présente un aperçu sur plusieurs méthodes constructives pour générer des mesures de probabilité aléatoires. Les applications des mesures de probabilité aléatoires incluent les statistiques Bayesiennes, des problèmes de contrôle optimal moyen, des bornes d'érreure moyennes pour les méthodes de résolution par des équations numériques et des modèles pour les distributions aléatoires de masse dans l'éspace.

*[Received July 1999, accepted March 2000]*