PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 110, Number 1, September 1990

REINFORCED RANDOM WALKS AND RANDOM DISTRIBUTIONS

R. DANIEL MAULDIN AND S. C. WILLIAMS

(Communicated by William D. Sudderth)

ABSTRACT. Consider a classical Polya urn process on a complete binary tree. This process generates an exchangeable sequence of random variables Z_n, with values in [0,1]. It is shown that the empirical distribution $^{\#}\{i \le n: Z_i \le s\}/n$ converges weakly and the distribution of this limit is the same as a standard Dubins-Freedman random distribution. As an application, the variance of the first moment of these Dubins-Freedman distributions is calculated.

1. INTRODUCTION

Consider the following urn or reinforced random process. Let $\{0, 1\}^*$ be the set of all finite sequences of 0's and 1's: $\{0, 1\}^* = \bigcup_{k=0}^{\infty}\{0, 1\}^k$. This includes $\varnothing$, the empty sequence. For each $\langle e_1, \ldots, e_k\rangle$ in $\{0, 1\}^*$, let $\mathfrak{U}(\langle e_1, \ldots, e_k\rangle)$ be an urn containing a ball labelled 0 and a ball labelled 1. Play Polya's game to generate a number Z_1 in the interval [0,1] and a new complete binary tree of urns as follows. Draw a ball from the urn $\mathfrak{U}(\varnothing)$ and replace that ball with two identical balls. Let $e_{1,1}$ be the label of the drawn ball. From the urn $\mathfrak{U}(\langle e_{1,1}\rangle)$, draw a ball and return to this urn two identical balls. Let $e_{1,2}$ be the label of the second drawn ball. Go to urn $\mathfrak{U}(\langle e_{1,1}, e_{1,2}\rangle)$. Continue this process. Thus, we generate a number Z_1 in the unit interval with dyadic expansion

$$Z_1 = 0.e_{1,1}e_{1,2}e_{1,3}\cdots .$$

The second stage of this process generates a second number, Z_2, in [0,1] as in the first stage. However, Polya's game is played on the new tree. This tree is the same as the initial tree except for those urns labelled by the sequence of numbers drawn. Continue the reinforced random draws. Let (Ω, Σ, μ) be a complete probability space and $\{Z_n\}_{n=1}^{\infty}$ a sequence of random variables modelling this process. Thus, $Z_n = 0.e_{n,1}e_{n,2}e_{n,3}\cdots$. Our first theorem is that this is an

Received by the editors April 13, 1989 and, in revised form, August 28, 1989.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 60A15, 62A15.
Key words and phrases. Polya urn process, exchangable, empirical distribution, random distribution.
Research of the first author was supported by National Science Foundation Grant DMS 88-03361.

0002-9939/90 $1.00 + $.25 per page

exchangeable sequence of random variables. Thus, there is a unique probability measure Q on $M = \Pr[0, 1]$, the space of probability measures on [0,1] such that, for each Borel subset A of $[0, 1]^\omega$,

$$\text{(R)} \qquad \mu(\{\omega: \{Z_n(\omega)\}_{n=1}^\infty \in A\}) = \int_M \nu^*(A)\, dQ(\nu),$$

where ν^* is infinite product measure on $[0, 1]^\omega$ with each factor being ν, see Hewitt and Savage (1955) or Aldous (1985). We also think of Q as defined on the space of distribution functions on $[0, 1]$. From this viewpoint, our second theorem is that Q is actually P, one of the basic measures on distributions first described by Dubins and Freedman (1967).

The probability measure P is generated by producing a distribution function, h, on $[0, 1]$ as follows. First, set $h(0) = 0$, and $h(1) = 1$. Choose $h(1/2)$ according to the uniform distribution on $[0, 1]$. Next, choose $h(1/4)$ according to the uniform distribution on $[0, h(1/2)]$ and, independently, choose $h(3/4)$ from the uniform distribution on $[h(1/2), 1]$. Continue. The function h defined on the dyadic rationals in $[0, 1]$ extends to a strictly increasing and *continuous* map or distribution on $[0, 1]$. (We note that in the notation of Dubins and Freedman, P is denoted by P_μ where μ is the uniform distribution on the line segment $x = 1/2$, $0 < y < 1$.)

We can now state a direct connection between the asymptotic behaviour of the random variables Z_n and P as follows. For any sequence of exchangeable random variables Z_n, it is known that empirical distribution $\varphi_n = \#\{i \le n: Z_i(\cdot) \le s\}/n$ converges almost surely and that the distribution of this limit is the measure Q satisfying (R). (See Theorem 3.1 and Lemma 2.15 of Aldous (1985) for example.) The distribution of this limit could be termed the de Finetti measure or as Aldous calls it the "directing measure". The point here is that the distribution of this limit is P. The main tool used in verifying this connection is the amalgamation operator developed by Graf *et al.* (1986). We turn to the proofs.

2. Results

That the sequence or process $\{Z_n\}_{n=1}^\infty$ is exchangeable means the distribution is invariant under finite permutations of the indices. In other words, for each positive integer n, Borel set $A \subset [0, 1]^n$, and permutation π of $\{1, \dots, n\}$,

$$\text{(E)} \qquad \mu((Z_1, \dots, Z_n) \in A) = \mu((Z_{\pi^{-1}(1)}, \dots, Z_{\pi^{-1}(n)}) \in A).$$

Theorem 1. *The sequence* $\{Z_n\}_{n=1}^\infty$ *is exchangeable.*

In order to prove this theorem let us make an observation. Fix n and, for each $i \le n$, let A_i be a binary interval of the form

$$(1) \qquad A_i = [0.t_{i,1}t_{i,2}\cdots t_{i,k},\, 0.t_{i,1}t_{i,2}\cdots t_{i,k} + 1/2^k],$$

where we use the dyadic expansion system. Observe that, to prove Theorem 1, it suffices to prove (E) for A of the form $A = \prod_{i=1}^n A_i$, where k in expression (1) varies over the positive integers. We do this by obtaining an equivalent formula for (E) which is clearly invariant under permutation. First, some notation. For each $\sigma \in \{0, 1\}^m$ with $m < k$, let $I_\sigma = \{i \le n: \langle t_{i,1}, \dots, t_{i,m}\rangle = \sigma\}$. By convention, $I_\varnothing = \{1, \dots, n\}$. Also, set $s_{n,\sigma} = \sum_{i \in I_\sigma} t_{i,m+1}$; *i.e.*, $s_{n,\sigma}$ is the number of 1's so far drawn from the urn $\mathfrak{U}(\sigma)$. For convenience, set $s_n = s_{n,\varnothing}$.

Lemma 2. *For each positive integer* n *and for each positive integer* k, *if* A_i *is of the form given in* (1), *for* $1 \le i \le n$, *then*

$$\text{(F)} \qquad \mu(Z_i \in A_i;\ i = 1, \dots, n) = \prod_{m=0}^{k-1} \prod_{\sigma \in \{0,1\}^m} \int_0^1 p^{s_{n,\sigma}} (1-p)^{|I_\sigma| - s_{n,\sigma}}\, dp.$$

Proof. Let us first verify this formula in the case $k = 1$. Thus, each A_i is of the form $A_i = [0.t_i, 0.t_i + 1/2]$, where $t_i = 0$ or 1. Therefore,

$$\mu(\{\omega: Z_i(\omega) \in A_i;\ i = 1, \dots, n\}) = \mu(\{\omega: e_{i,1}(\omega) = t_i\}).$$

But, $\{e_{j,1}\}_{j=1}^\infty$ is a classical Polya urn scheme which is certainly exchangeable (Blackwell and Kendall (1964)) and is a mixture of independent, identically distributed Bernoulli choices with probability p of being 1, where p is uniform on $[0, 1]$. Formally,

$$\begin{aligned}
\mu(e_{i,1} = t_i;\ i = 1, \dots, n) &= \prod_{k=1}^{n} \mu(e_{k,1} = t_k | e_{1,1} = t_1, \dots, e_{k-1,1} = t_{k-1}) \\
&= \prod_{k=1}^{n} (1 + \#\{j < k: t_j = t_k\})/k + 1 \\
&= \prod_{k=1}^{n} 1 + (s_{k-1} \text{ or } k - 1 - s_{k-1})/(n+1)! \\
&= 1/(n+1) \binom{n}{s_n} \\
&= \int_0^1 p^{s_n} (1-p)^{n - s_n}\, dp.
\end{aligned}$$

Thus, formula (F) is verified for $k = 1$ and all positive integers n. Assume (F) holds for k and for all positive integers n. Now, for each i, $1 \le i \le n$, let

$$A_i = [0.t_{i,1} t_{i,2} \cdots t_{i,k+1}, 0.t_{i,1} t_{i,2} \cdots t_{i,k+1} + 1/2^{k+1}].$$

Set $\overline{A}_i = [0.t_{i,1} t_{i,2} \cdots t_{i,k}, 0.t_{i,1} t_{i,2} \cdots t_{i,k} + 1/2^k])$. Notice

$$\begin{aligned}
&\mu(Z_i \in A_i;\ 1 \le i \le n) \\
&\quad = \mu(e_{i,k+1} = t_{i,k+1};\ 1 \le i \le n | Z_i \in \overline{A}_i;\ 1 \le i \le n) \cdot \mu(Z_i \in \overline{A}_i;\ 1 \le i \le n).
\end{aligned}$$

But,

(2)

$$\mu(e_{i,k+1} = t_{i,k+1};\ 1 \le i \le n | Z_i \in [0.t_{i,1}t_{i,2}\cdots t_{i,k},\ 0.t_{i,1}t_{i,2}\cdots t_{i,k} + 1/2^k]) = \prod_{\sigma\in\{0,1\}^k} \int_0^1 p^{s_{n,\sigma}}(1-p)^{|I_\sigma|-s_{n,\sigma}}\,dp.$$

Substituting this last product for the conditional probability and applying the induction hypothesis shows that (F) holds for $k+1$ and all positive integers n. Thus, formula (F) and the lemma follow.

Since formula (F) is invariant under permutation of the indices. Theorem 1 follows. Let Q be the unique probability measure on $M = \Pr[0, 1]$ satisfying

$$\mu(\{\omega\colon \{Z_n(\omega)\}_{n=1}^{\infty} \in A\}) = \int_M \nu^*(A)\,dQ(\nu). \tag{R}$$

Let P be the probability measure induced on M according to the process described at the beginning. Actually, P is defined on the space of probability distribution functions on $[0, 1]$. If h is such a distribution function, then h^* denotes the infinite product measure on $[0, 1]^N$. Our second goal is to show that Q is P. To do this we construct a probability space $(\Omega, \Sigma, \bar{\mu})$ and a sequence of random variables $\{\overline{Z}_n\}_{n=1}^{\infty}$ such that formula (F) holds.

Consider the process $\overline{Z}_1, \overline{Z}_2, \overline{Z}_3, \ldots$ given by

$$\bar{\mu}(\overline{Z} \in A) = \int_M h^*(A)\,dP(h). \tag{3}$$

Fix n and for each $i \le n$, let A_i be of the form given by (1). We first obtain a reduction formula on k. To this end, consider the product set A whose first n factors are the sets A_i and whose other factors are $[0, 1]$. Substituting into (3) and using the amalgamation formula of Graf *et al.* (1986), §3, we have

$$\begin{aligned}
\bar{\mu}(\overline{Z}_i \in A_i;\ 1 \le i \le n) &= \int_M h^*(z_i \in A_i;\ 1 \le i \le n)\,dP(h) \\
&= \int_0^1 \iint_{M\times M} ([h_1, h_2]_x)^*(z_i \in A_i;\ 1 \le i \le n)\,dP(h_1)\,dP(h_2)\,d\lambda(x) \\
&= \int_0^1 \iint_{M\times M} \prod_{i=1}^{n} [h_1, h_2]_x(z_i \in A_i;\ 1 \le i \le n)\,dP(h_1)\,dP(h_2)\,d\lambda(x) \\
&= \int_0^1 \iint_{M\times M} \prod_{i\in I_0} x h_1(z_i \in 2A_i) \\
&\qquad \times \prod_{i\in I_1} (1-x) h_2(z_i \in 2A_i - 1))\,dP(h_1)\,dP(h_2)\,d\lambda(x),
\end{aligned}$$

where $I_0 = \{i \le n: t_{i,1} = 0\}$ and $I_1 = \{i \le n: t_{i,1} = 1\}$. Thus,

(4)
$$\bar{\mu}(\overline{Z}_i \in A_i;\ 1 \le i \le n) = \int_0^1 x^{|I_0|}(1-x)^{|I_1|}\,dx \int_M \prod_{i\in I_0} x h_1(z_i \in 2A_i)\,dP(h_1) \times \int_M \prod_{i\in I_1}(1-x)h_2(z_i \in 2A_i - 1)\,dP(h_2).$$

Also, notice that $|I_1| = s_{n,\varnothing}$ and $|I_0| = n - s_{n,\varnothing} = |I_\varnothing| - s_{n,\varnothing}$.

Now, for the case $k = 1$, we have $2A_i = [0, 1]$ if $i \in I_0$ and $2A_i - 1 = [0, 1]$ if $i \in I_1$. Thus,

$$\bar{\mu}(\overline{Z}_i \in A_i;\ 1 \le i \le n) = \int_0^1 x^{|I_0|}(1-x)^{|I_1|}\,dx$$

and formula (F) holds for $k = 1$ and for all n.

Now, suppose (F) holds for k and each

$$A_i = [0.t_{i,1}t_{i,2}\cdots t_{i,k+1},\ 0.t_{i,1}t_{i,2}\cdots t_{i,k+1} + 1/2^{k+1}].$$

If $i \in I_0$, then

$$2A_i = [0.t_{i,2}t_{i,3}\cdots t_{i,k},\ 0.t_{i,2}t_{i,3}\cdots t_{i,k} + 1/2^k],$$

and if $i \in I_1$, then $2A_i - 1 = [0.t_{i,2}t_{i,3}\cdots t_{i,k},\ 0.t_{i,2}t_{i,3}\cdots t_{i,k} + 1/2^k]$. So, applying the induction hypothesis,

$$\bar{\mu}(\overline{Z}_i \in A_i;\ 1 \le i \le n) = \left[\int_0^1 x^{|I_0|}(1-x)^{|I_1|}\,dx \cdot \prod_{i=0}^{k-1}\prod_{\sigma\in\{0,1\}^i}\int_0^1 p^{s_{n,0*\sigma}}(1-p^{|I_{0*\sigma}|-s_{n,0*\sigma}}\,dp\right] \times \left[\int_0^1 p^{s_{n,1*\sigma}}(1-p)^{|I_{1*\sigma}|-s_{n,1*\sigma}}\,dp\right].$$

Again, substituting $s_{n,\varnothing}$ for $|I_1|$ and $n - s_{n,\varnothing}$ for $|I_0|$, we have formula (F). □

We have shown that

$$\mu(\{\omega: \{Z_n(\omega) \in A\}) = \int_M h^*(A)\,dP(h), \tag{5}$$

for all A of the form $\prod_{i=1}^n A_i \times \prod_{i>n}[0, 1]$, where each A_i is of the form (1). It follows that (6) holds for all Borel sets A. Since Q is unique, we have proven

Theorem 3. $Q = P$.

In view of the mentioned known results, we can restate Theorem 3 as

Theorem 3a. *For μ-a.e. ω and $\forall\ s \in [0, 1]$, $\lim_{n\to\infty} \#\{i \le n: Z_i(\omega) \le s\}/n$ exists. Moreover, P is the distribution on*

$$h_\omega(s) = \lim_{n\to\infty} \#\{i \le n: Z_i(\omega) \le s\}/n.$$

3. Application

As an application of the preceding results, we calculate the variance of the first moment of a P-random distribution.

Theorem 4. *The variance of the first moment of a P-random distribution is* $1/40$.

Proof. Since $\int_0^1 x\,dh(x) = 1 - \int_0^1 h(x)\,dx$, we will calculate the variance of $\int_0^1 h(x)\,dx$. First, by Fubini's theorem, we have

$$E\left[\int_0^1 h(x)\,dx\right] = \int_M \int_0^1 h(x)\,dx\,dP(h)$$
$$= \int_0^1 \int_M h(x)\,dP(h)\,dx.$$

But, Graf *et al* showed that the expected P distribution is the identity: for each x, $\int_M h(x)\,dP(h) = x$. (Dubins and Freedman had stated this fact in their article, p. 186.) Thus, $E[\int_0^1 h(x)\,dx] = 1/2 = E[\int_0^1 x\,dh(x)$. Next, we calculate the second moment. By Fubini's theorem,

$$E\left[\left(\int_0^1 h(x)\,dx\right)^2\right] = E\left(\left(\int_0^1 h(x)\,dx\right)\left(\int_0^1 h(y)\,dy\right)\right]$$
$$= \int_0^1 \int_0^1 E[h(x)h(y)]\,dx\,dy.$$

Since Q is P, we have from the representation (R),

$$E\left[\left(\int_0^1 h(x)\,dx\right)^2\right] = \int_0^1 \int_0^1 \mu(Z_1 \le x\,,\, Z_2 \le y)\,dx\,dy$$
$$= \int_0^1 \int_0^1 \int_0^x \int_0^y \rho(s\,,\,t)\,dt\,ds\,dx\,dy,$$

where $\rho(s\,,\,t)$ is the joint density of Z_1 and Z_2. Again, by Fubini's theorem,

$$E\left[\left(\int_0^1 h(x)\,dx\right)^2\right] = \int_0^1 \int_0^1 \int_s^1 \int_t^1 \rho(s\,,\,t)\,dy\,dx\,dt\,ds$$
$$= \int_0^1 \int_0^1 (1-s)(1-t)\rho(s\,,\,t)\,dt\,ds.$$

Set the following notation: if $\sigma \in \{0\,,\,1\}^k$, let $W(\sigma) = [.\sigma\,,\,.\sigma + 1/2^k]$, where $.\sigma$ is the number with dyadic expansion σ, i.e., $.\sigma = \sum_{i=1}^k s_i/2^i$. For each

σ, set $J(\sigma) = W(\sigma^* 0) \times W(\sigma^* 1)$. By the symmetry or exchangability of the process,

$$(6) \quad E\left[\left(\int_0^1 h(x)\,dx\right)^2\right] = 2\sum_{k=0}^{\infty} \sum_{\sigma\in\{0,1\}^k} \iint_{J(\sigma)} (1-s)(1-t)\rho(s,t)\,dt\,ds.$$

Figure 1 illustrates the domains of integration, $J(\sigma)$.
In order to calculate ρ on the square $J(\sigma)$, note the general formula which follows from (F). If $\sigma \in \{0, 1\}^k$ and α, $\beta \in \{0, 1\}^m$, then

$$(7) \quad \mu(Z_1 \in W(\sigma * 0 * \alpha), Z_2 \in W(\sigma * 1 * \beta)) = (1/3)^k (1/6)(1/4)^m = (1/2)(1/3)^{k+1}(1/4)^{m+1}.$$

Thus, the density ρ on the square $J(\sigma)$ is $1/2(4/3)^{k+1}$. This means

(8)

$$E\left[\left(\int_0^1 h(x)\,dx\right)^2\right]$$

$$= \sum_{k=0}^{\infty} \sum_{\sigma\in\{0,1\}^k} (4/3)^{k+1} \int_{.\sigma^*}^{.\sigma*(0+1/2^{k+1}} \int_{.\sigma^* 1}^{.\sigma*1+1/2^{k+1}} (1-s)(1-t)\,dt\,ds$$

$$= \sum_{k=0}^{\infty} \sum_{\sigma\in\{0,1\}^k} (4/3)^{k+1}[-(1-s)^2/2]\Big|_{.\sigma*0}^{.\sigma*0+1/2^{k+1}} [-(1-t)^2/2]\Big|_{.\sigma*1}^{.\sigma*1+1/2^{k+1}}$$

$$= \sum_{k=0}^{\infty} \sum_{\sigma\in\{0,1\}^k} (1/3)^{k+1}[1 - .\sigma*0 - 1/2^{k+2}][1 - .\sigma*1 - 1/2^{k+2}]$$

$$= (1/4)\sum_{k=0}^{\infty} \sum_{\sigma\in\{0,1\}^k} (1/12)^{k+1}[2^{k+2}(1-.\sigma) - 1][2^{k+2}(1 - .\sigma - 1/2^{k+1}) - 1]$$

$$= (1/48)\sum_{k=0}^{\infty}(1/12)^k \sum_{i=1}^{2^k}(4i-1)(4i-3)$$

$$= (1/48)\sum_{k=0}^{\infty}(1/12)^k \left[16\sum_{i=1}^{2^k} i^2 - 16\sum_{i=1}^{2^k} i + 3 - 2^k\right]$$

$$= (1/48)\sum_{k=0}^{\infty}(1/12)^k[16\cdot 2^k(2^k+1)(2^{k+1}+1)/6 - 16\cdot 2^k((2^k+1)/2 + 3\cdot 2^k]$$

$$= (1/48)\sum_{k=0}^{\infty}(1/12)^k[(16/3)\cdot 8^k - (7/3)\cdot 2^k] = 11/40.$$

Thus, the variance of the expected value of a P-random distribution is $1/40$. □

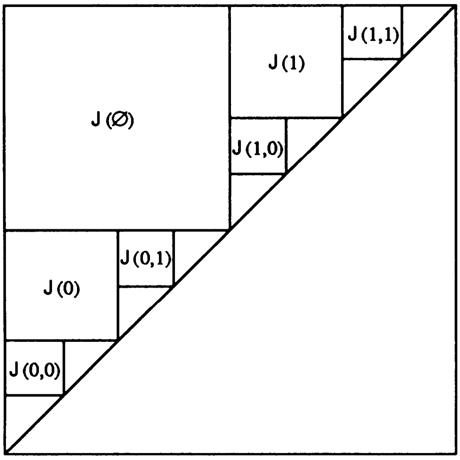

FIGURE 1.

REFERENCES

1. D. J. Aldous (1985), *Exchangeability and related topics*, Ecole d'ete de Probabilites de Saint-Flour XIII–1983, Lecture Notes in Math., vol. 1117, Springer-Verlag, New York, 1985, pp. 1–198.
2. D. Blackwell and D. Kendall, *The Martin boundary for Polya's urn scheme, and an application to stochastic population growth*, J. Appl. Probab. **1** (1964), 284–296.
3. P. Diaconis and D. A. Freedman, *de Finetti's theorem for Markov chains*, Ann. Probab. **8** (1980), 115–130.
4. L. E. Dubins and D. A. Freedman, *Random distribution functions*, Proceedings, Fifth Berkeley Symp. on Math. Statistics and Probability (L. M. LeCam and J. Neyman, eds.), Univ. of California Press, Berkeley/Los Angeles, 1967, pp. 183–214.
5. D. A. Freedman (1965), *Bernard Friedman's urn*, Ann. Math. Statist. **36**, 956–970.
6. S. Graf, R. D. Mauldin and S. C. Williams, *Random homeomorphisms*, Adv. in Math. **60** (1986), 239–359.
7. E. Hewitt and L. J. Savage, *Symmetric measures on Cartesian products*, Trans. Amer. Math. Soc. **80** (1955), 470–501.
8. B. M. Hill, D. Lane and W. Sudderth, *Exchangeable urn processes*, Ann. Probab. **15** (1987), 1586–1592.

MATHEMATICS DEPARTMENT, UNIVERSITY OF NORTH TEXAS, DENTON, TEXAS 76203–5116

MATHEMATICS DEPARTMENT, UTAH STATE UNIVERSITY, LOGAN, UTAH 84322