Polya Trees and Random Distributions

by

R. Daniel Mauldin*, University of North Texas

William D. Sudderth*, University of Minnesota

S.C. Williams, Utah State University

Technical Report No. 550
May 1990

*Research supported by National Science Foundation Grant DMS-8803361 (for Mauldin) and DMS-8911548 (for Sudderth).

Abstract

Trees of Polya urns are used to generate sequences of exchangeable random variables. By a theorem of de Finetti each such sequence is a mixture of independent, identically distributed variables and the mixing measure can be viewed as a prior on distribution functions. The collection of these Polya tree priors form a convenient conjugate family analogous to the Dirichlet processes of Ferguson (1973). Unlike Dirichlet processes, Polya tree priors can assign probability one to the class of continuous distributions.

American Mathematical Society 1980 subject classification. Primary 62A15. Secondary 62G99, 60G09, 60G57.

1. Introduction

The Polya urn scheme is, perhaps, the simplest and most concrete way to generate a sequence $X_1, X_2, \ldots$ of exchangeable random variables having values in a finite set $E = \{0,1,\ldots,k\}$. Suppose that the urn u has initially u_i balls of color i for $i \in E$ and, that, at each stage, a ball is drawn at random and replaced by two of the same color. Let $X_n = i$ if the nth ball selected is of color i. It is well-known that the X_n are exchangeable and that the sample distribution of $X_1, \ldots, X_n$ converges almost surely to a random probability vector $\Theta = (\Theta_o, \ldots, \Theta_k)$ which has a Dirichlet distribution with parameters $(u_o, \ldots, u_k)$. Furthermore, given $\Theta = \theta$, the variables $X_1, X_2, \ldots$ are independent with $P[X_n = i] = \theta_i$ for all n and i. (These facts are reviewed in the next section.)

All of these results were generalized by Blackwell and MacQueen (1973) who showed that the random distributions constructed by Ferguson (1973) can be viewed as the limit of the sample distributions of variables which are obtained from a Polya urn scheme based on a continuum of colors. The urn scheme makes many properties of Ferguson distributions intuitively clear. For example, the Ferguson distributions form a conjugate family of prior distributions for nonparametric problems just as the Dirichlet distributions form a conjugate family for multinomial sampling. (By the way, Ferguson (1973) called his distributions "Dirichlet processes." To avoid confusion with the processes constructed here, we will continue to call them Ferguson distributions.)

This paper presents another conjugate family of prior distributions which are constructed from trees of Polya urns. Mauldin and Williams (1990) gave the first such construction and the following is a natural generalization. Let E^* be the set of all finite sequences of elements of $E = \{0,1,\ldots,k\}$ including the empty sequence ϕ. Think of E^* as an infinite tree beginning at ϕ and suppose that u is a function which assigns to every p in E^* an urn u(p) which contains $u(p)_i$ balls of color i for each $i \in E$. Use this <u>Polya tree</u> u to generate a sequence of random variables $X_{11}, X_{12}, \ldots$ and a new tree $u^{(1)}$ as follows: Draw a ball at random from $u(\phi)$ and replace it by two of the same color. Set $X_{11} = i_1$ if the ball is of color i_1. Next draw a ball from $u(i_1)$, replace it by two of the same color, and set $X_{12} = i_2$ if the ball is of color i_2. Go on to $u(i_1, i_2)$ and continue in this fashion. Set $X_1 = (X_{11}, X_{12}, \ldots)$ and let $u^{(1)}$ be the Polya tree which was obtained in the construction. Iterate the entire process to obtain $X_1, X_2, \ldots$ and $u^{(1)}, u^{(2)}, \ldots$. It is shown below (Theorem 4.1) that the X_n are exchangeable. So, by a theorem of de Finetti, there is a measure $Q = Q_u$ defined on the space of probability measures on $E^N = E \times E \times \ldots$ such that the distribution of $X_1, X_2, \ldots$ can be obtained by first choosing Θ with distribution Q and then choosing $X_1, X_2, \ldots$ to be independent with distribution θ given $\Theta = \theta$. This de Finetti measure Q will be calculated explicitly (Theorem 4.2).

The X_n can also be regarded as ordinary random variables having values in the unit interval $I = [0,1]$. Just set

$$Y_n = \sum_{i=1}^{\infty} X_{ni}/(k+1)^i. \tag{1.1}$$

Again $Y_1, Y_2, \ldots$ are exchangeable and the form of the de Finetti measure $\bar{Q}$ is immediate from that of Q. The measure $\bar{Q} = \bar{Q}_u$ is now defined on the space of probability measures on [0,1] and can be regarded as picking a random distribution function.

It is natural to compare this family of priors based on Polya trees with the family of Ferguson distributions. Both families are relatively easy to understand and have easily described posterior distributions. (As will be seen, if Θ is a random probability with distribution Q_u (or $\bar{Q}_u$) and $X_1, \ldots, X_n$ is a random sample from Θ, then the posterior distribution of Θ is $Q_{u^{(n)}}$ $\left(\text{or } \bar{Q}_{u^{(n)}}\right)$ where $u^{(n)}$ is the Polya tree after the nth stage of the construction described above.) Both the Polya tree priors and the Ferguson distributions typically have large support. Both typically have posterior distributions which are consistent in the sense that, if data are generated from a "true" distribution θ_o, then the posterior distributions will converge weakly, with probability one, to θ_o. (For the Polya tree priors, this will be shown to be a consequence of results of Fabius (1964) on tailfree priors.) There is one interesting difference between the two families. As Ferguson (1973) himself showed, Ferguson distributions give probability one to the set of discrete probability measures. The priors constructed here often assign probability one to the class $C = \{\theta: \theta(\{x\}) = 0 \text{ for all } x\}$ of continuous measures and simple conditions are given in Theorem 5.2 and its corollary to guarantee that $Q_u(C) = 1$.

The priors $\bar{Q}_u$ defined on probability measures on [0,1] are closely related to the distributions constructed by Dubins and Freedman (1966) in a different fashion. In fact, if $k = 1$ and if $u(p)$ contains exactly one ball of color 0 and one ball of color 1 for every p, then $\bar{Q}_u$ was shown by Mauldin and Williams (1990) to be one of the Dubins and Freedman measures. The family of all $\bar{Q}_u$ is the natural conjugate family of priors which contains this basic measure. (See Graf, Mauldin, and Williams (1986) for another generalization of the Dubins and Freedman measures.)

The next section is a review of classical Polya urn schemes, Dirichlet distributions and their connection by way of de Finetti's Theorem. After these preliminaries a study is made in section 3 of Polya tree processes which end after a finite number of stages. This is in preparation for the study of infinite trees and also provides a new conjugate family of priors for finite, sequential sampling schemes. Section 4 treats infinite Polya trees u and characterizes the de Finetti measures Q_u. Section 5 gives conditions under which Q_u concentrates on continuous distributions and section 6 gives conditions under which the support of Q_u is large. A functional equation is given for Q_u in section 7. Section 8 is about using Polya tree priors to estimate distribution functions. The final section has a few remarks.

2. Polya urns and Dirichlet distributions

The initial data needed for the classical Polya urn scheme described in the introduction is the vector $u = (u_o,\dots,u_k)$ where each u_i represents the initial number of balls of color i. Such a vector will be called an <u>urn vector</u> if u_i is nonnegative for every $i \in E$ and the quantity $|u| = u_o + \dots + u_k$ is strictly positive. The probability of drawing a ball of color i from u is, by definition, $u_i/|u|$ for each i. After drawing a ball of color i, the urn at the next stage has $u_i + 1$ balls of color i. With these conventions, the Polya urn scheme, as described in the introduction, generates a sequence

$$X = (X_1,X_2,\dots)$$

of random variables with values in E. (As before, $X_n = i$ if the nth ball drawn has color i.) The sequence X is said to have a <u>Polya distribution with</u>

parameter u, or, more briefly, X is $\mathcal{P}(u)$.

In order to calculate the joint distribution of $X_1,\dots,X_n$, let $j = (j_1,\dots,j_n)$ be a sequence of n elements of E and set

$$P(j;u) = P[X_1=j_1,\dots,X_n=j_n]. \tag{2.1}$$

For each $i \in E$, let $c(i)$ be the number of i's occurring in the sequence j and define $s(i) = u_i + c(i)$. Notice that, if $X_1=j_1,\dots,X_n=j_n$, then $s(i)$ corresponds to the number of balls of color i in the urn after the nth stage.

A simple counting argument shows that

$$P(j;u) = \frac{(s(o)-1)_{c(o)}\cdots(s(k)-1)_{c(k)}}{(|u|+n-1)_n} \tag{2.2}$$

where, as usual,

$$(a)_b = a(a-1)\dots(a-b+1),\ b=1,2,\dots,\ (a)_o = 1.$$

It is clear from (2.2) that $P(j;u)$ is unchanged when the coordinates of j are permuted. We record this in a lemma.

Lemma 2.1. If $(X_1,X_2,\dots)$ has a Polya distribution, then $X_1,X_2,\dots$ are exchangeable.

The next result is a version of de Finetti's theorem on exchangeable variables. It is a special case of results in Meyer (1966) or in Aldous (1985).

Theorem 2.1. Let $X_1,X_2,\dots$ be exchangeable variables with values in $E = \{0,\dots,k\}$. For $n=1,2,\dots$ and $i \in E$, let $C_n(i) = \#\{j\le n: X_j=i\}$. Then

(a) $(C_n(0),\dots,C_n(k))/n$ converges almost surely to a random probability vector $\theta = (\theta_o,\dots,\theta_k)$, and
(b) given θ, the variables $X_1,X_2,\dots$ are independent and each has distribution θ.

The distribution of the random probability vector Θ in Theorem 1.1 is called the <u>de Finetti measure</u> for the sequence $X = (X_1, X_2, \ldots)$. (It is called the <u>directing measure</u> by Aldous (1985)).

Let

$$S_k = \{(\theta_o, \ldots, \theta_k): \theta_o \geq 0, \ldots, \theta_k \geq 0, \sum_{i=0}^{k} \theta_i = 1\}.$$

Here is a simple characterization of the de Finetti measure.

<u>Theorem</u> 2.2. A probability measure μ defined on the Borel subsets of S_k is the de Finetti measure for the exchangeable sequence X if and only if, for every finite sequence $i_1, \ldots, i_m$ of elements of E

$$(2.3) \qquad P[X_1 = i_1, \ldots, X_m = i_m] = \int \theta_o^{c_o} \ldots \theta_k^{c_k} \, d\mu(\theta_o, \ldots, \theta_k)$$

where c_j is the number of elements in $i_1, \ldots, i_m$ which are equal to j for each j.

<u>Proof</u>: If μ is the de Finetti measure for X, formula (2.3) follows from Theorem 2.1(b). Conversely, μ is uniquely determined by its moments since it has compact support. □

Notice that (2.3) is equivalent to

$$(2.4) \qquad P[X_1 = i_1, \ldots, X_m = i_m] = E\left(\prod_{a=1}^{m} \Theta_{i_a}\right)$$

where Θ is a random probability vector with distribution μ.

Our next task is to identify the de Finetti measure when X is a Polya sequence. First we recall the definition of a Dirichlet distribution.

The Dirichlet distribution defined here will be slightly more general than usual in that some of the variables can be degenerate at zero. Our definition is consistent with that of Ferguson (1973). Let $\Theta = (\Theta_o, \ldots, \Theta_k)$ be a random vector with values in S_k and let $u = (u_o,\ldots,u_k)$ be an urn vector. Suppose first that $u_i > 0$ for all i. Then we say Θ is Dirichlet with parameter u, written $\underset{\sim}{D}(u)$, if $k = 0$ and $\Theta_o = 1$ almost surely or if $k > 0$ and $(\Theta_o,\ldots,\Theta_{k-1})$ has the density function

$$f(\theta_o,\ldots,\theta_{k-1}) = \frac{\Gamma(|u|)}{\Gamma(u_o)\ldots\Gamma(u_k)} \theta_o^{u_o}\ldots\theta_k^{u_k}$$

for $(\theta_o,\ldots,\theta_k) \in S_k$. (Notice $\theta_k = 1 - (\theta_o+\ldots+\theta_{k-1})$). In the general case, take F to be that subset of $\{0,1,\ldots,k\}$ consisting of those i for which $u_i > 0$. Say $F = \{i_o,\ldots,i_r\}$ where $0 \le r \le k$. Now define Θ to be $\underset{\sim}{D}(u)$ if $(\Theta_{i_o},\ldots,\Theta_{i_r})$ is $\underset{\sim}{D}((u_{i_o},\ldots,u_{i_r}))$ and $\Theta_j = 0$ almost surely for $j \notin F$.

The next result gives the nice connection between Polya urn schemes and the Dirichlet distribution. It is a special case of the theorem in Blackwell and MacQueen (1973). Much of it can also be found in Blackwell and Kendall (1964). An elementary proof can be based on Theorem 2.2.

Theorem 2.3. If X is $\mathcal{P}(u)$, then the de Finetti measure for X is $\underset{\sim}{D}(u)$.

The Dirichlet family of distributions is useful for Bayesian analysis largely because it is the natural family conjugate to the multinomial. Here is a statement of this well-known fact.

Theorem 2.4. Suppose Θ is $\underset{\sim}{D}(u)$ and, given Θ, X_1 has distribution Θ. Then, given $X_1 = i$, Θ is $\underset{\sim}{D}(u+\delta(i))$ where $\delta(i)$ is the probability vector which has 1 in the ith coordinate.

An elementary proof can be given using Bayes formula. A more interesting proof in the present context is to embed X_1 in a sequence $X = (X_1,X_2,\ldots)$ of variables which are independent with distribution Θ given Θ. Then X is $\mathcal{P}(u)$ by Theorems 2.1 and 2.3. So, clearly, $(X_2,X_3,\ldots)$ is $\mathcal{P}(u+\delta(i))$ given $X_1 = i$. Use

Theorems 2.1 and 2.3 again to see that Θ is $\underset{\sim}{D}(u+\delta(i))$ given $X_1 = i$.

3. Finite Polya trees

Let s be a positive integer and let E_{s-1} be the set of all finite paths or sequences $p = (i_1,\ldots,i_m)$ of elements of E whose length m is less than or equal to s-1, including the empty sequence ϕ. A <u>Polya tree of height</u> s is a mapping u which assigns to each $p \in E_{s-1}$ an urn vector $u(p) = (u(p)_o,\ldots,u(p)_k)$. Given such a u, the procedure explained in the introduction generates a random vector

$$X_1 = (X_{11}, X_{12}, \ldots, X_{1s})$$

and a new tree $u^{(1)}$. (As before $X_{11} = i$ if the ball drawn from $u(\phi)$ is of color i, etc. The only difference is that the procedure ends after s stages. The conventions about urns with nonintegral numbers of balls are the same as in the previous section.) The function $u^{(1)}$ is given by

$$\begin{aligned} u^{(1)}(i_1,\ldots,i_m)_j &= u(i_1,\ldots,i_m)_j + 1 \\ &\qquad \text{if } X_{11}=i_1,\ldots,X_{1m}=i_m, X_{1,m+1}=j, \\ &= u(i_1,\ldots,i_m)_j \text{ if not} \end{aligned}$$

for each $p = (i_1,\ldots,i_m) \in E_{s-1}$ and $j \in E$.

Iterate the procedure using $u^{(1)}$ to get $X_2 = (X_{21},\ldots,X_{2s})$ and $u^{(2)}$, and so on.

The sequence $X = (X_1, X_2, \ldots)$ is called <u>s-stage Polya with parameter</u> u or $\mathcal{P}_s(u)$.

To calculate the joint distribution of the first n variables, let

$$j = (j^{(1)},\ldots,j^{(n)})$$

be a sequence of n vectors in E^s

$$j^{(a)} = (j_{a1},\ldots,j_{as}),\ a=1,\ldots,n.$$

For each such j and each path p of length r-1 in E_{s-1}, let j(p) be the vector corresponding to the colors of those balls drawn from u(p) if $(X_1,\ldots,X_n) = j$. That is,

$$(3.0) \qquad j(p) = \left(j_{a_1 r},\ldots,j_{a_i r}\right)$$

where $j^{(a_1)},\ldots,j^{(a_i)}$ are those vectors occurring in j (taken in order) whose first r-1 coordinates coincide with p. Set $j(p) = \phi$ if the path p is not traversed. Now the probability of drawing j(p) from u(p) is just P(j(p);u(p)) as in (2.2). Furthermore the outcomes of draws from urns associated with different paths are independent. Hence,

$$(3.1) \qquad P[(X_1,\ldots,X_n) = j] = \prod_{p \in E_{s-1}} P(j(p);u(p))$$

where the convention is made that $P(\phi;u(p)) = 1$.

<u>Lemma</u> 3.1. If $(X_1,X_2,\ldots)$ is $\mathcal{O}_s(u)$, then the variables $X_1,X_2,\ldots$ are exchangeable.

<u>Proof</u>: The probability P(j(p);u(p)) is invariant under permutations of the coordinates of j(p) and, hence, is invariant under permutations of $j^{(1)},\ldots,j^{(n)}$. □

Now Theorem 1.1 applies to the sequence $X_1,X_2,\ldots$ of exchangeable variables taking values in E^s, and we would like to identify the de Finetti measure. First some additional notation and terminology are needed.

It is convenient to view probability measures on E^s as being "strategies" in the sense of Dubins and Savage (1965).

<u>Definition</u>. An s-<u>day strategy</u> is a mapping θ which assigns to every $p \in E_{s-1}$ a probability measure $\theta(p)$ on E. The notation θ_o will often be used for $\theta(\phi)$.

An s-day strategy θ naturally determines a probability measure $\mu = \mu(\theta)$ on

E^s as follows: The marginal μ-distribution on the first coordinate is θ_o and, given that the first r coordinates are p where $p \in E^r$, $1 \le r < s$, the conditional μ-distribution of the r+1st coordinate is $\theta(p)$.

To simplify notation, the measure $\mu(\theta)$ associated with θ will be written as θ below. Notice that a random probability Θ on E^s is induced by putting a distribution on the collection $\{\Theta(p): p \in E_{s-1}\}$.

Let u be a Polya tree of height s.

Definition. A random probability measure Θ on E^s is a Dirichlet strategy with parameter u (written $D_s(u)$) if the random probability measures $\{\Theta(p): p \in E_{s-1}\}$ are independent and $\Theta(p)$ is $\underset{\sim}{D}(u(p))$ for every $p \in E_{s-1}$.

Theorem 3.1. If $X = (X_1, X_2, \ldots)$ is $P_s(u)$, then the de Finetti measure of X is $\underset{\sim}{D}_s(u)$.

Proof: The proof is an application of Theorem 2.2. For the application, E is replaced in that theorem by E^s and

$$j = (j^{(1)}, \ldots, j^{(n)})$$

is a finite sequence of elements of E^s. We must verify (2.4) which here becomes

$$P[(X_1, \ldots, X_n) = j] = E\left[\prod_{a=1}^{n} \Theta(\{j^{(a)}\})\right] \tag{3.2}$$

under the assumption that Θ is $\underset{\sim}{D}_s(u)$. The left-hand-side of (3.2) is given by (3.1). So it remains to calculate the right-hand-side.

First notice that, for

$$j^{(a)} = (j_{a1}, \ldots, j_{as})$$

and θ a strategy,

$$\theta(\{j^{(a)}\}) = \theta_o\{j_{a1}\}\theta(j_{a1})\{j_{a2}\}\ldots\theta(j_{a1}, \ldots, j_{a,s-1})\{j_{as}\}.$$

(This is just the ordinary product rule for calculating the probability of an intersection.) Now substitute into the right-hand-side of (3.2), collect terms, and use the independence of the $\Theta(p)$ to obtain

$$E\left(\prod_{a=1}^{n} \Theta(\{j^{(a)}\})\right) = \prod_{p\in E_{s-1}} E\left(\prod_{i\in j(p)} \Theta(p)(\{i\})\right). \tag{3.3}$$

Here $j(p)$ is the same as in (3.0) and i varies over the coordinates of $j(p)$ taken with their multiplicities. It follows from Theorems 2.2 and 2.3, and our assumption that $\Theta(p)$ is $\underset{\sim}{D}(u(p))$ that

$$E\left(\prod_{i\in j(p)} \Theta(p)(\{i\})\right) = P(j(p);\alpha(p)).$$

By (3.1) and (3.3), the proof is complete. □

Just as the Dirichlet family is conjugate to ordinary multinomial sampling, the family of Dirichlet strategies is conjugate for "strategic sampling" in which an experiment takes place in several stages each of which depends on the preceding outcomes. Even if the experiment is terminated (censored) before the last stage, the Dirichlet strategies remain conjugate as was pointed out by Jim Dickey.

<u>Theorem</u> 3.2. Suppose Θ is $\underset{\sim}{D}_s(u)$ and, given Θ, $X_1 = (X_{11},\ldots,X_{1s})$ has distribution Θ. Then, given $(X_{11},\ldots,X_{1r}) = (i_1,\ldots,i_r)$ where $1 \le r \le s$, Θ is $\underset{\sim}{D}_s(u')$ where

$$\begin{aligned} u'(\phi) &= u(\phi) + \delta(i_1), \\ u'(i_1,\ldots,i_a) &= u(i_1,\ldots,i_a) + \delta(i_{a+1}) \\ &\qquad \text{for } a=1,\ldots,r-1, \\ u'(p) &= u(p) \text{ for all other } p \in E_{s-1}. \end{aligned}$$

<u>Proof</u>: Do an induction on r using Theorem 2.4 and the independence of the

$\Theta(p)$'s. (Or use Bayes formula and the fact that the density for Θ is the product of the densities for the $\Theta(p)$.) □

4. Infinite Polya trees

An <u>infinite Polya tree</u> is a mapping u which assigns to every $p \in E^*$ an urn vector $u(p)$. Given such a u, the scheme described in the introduction generates sequences $X_1, X_2, \ldots$ and $u^{(1)}, u^{(2)}, \ldots$ where, for each n,

$$X_n = (X_{n1}, X_{n2}, \ldots)$$

is a random element of E^N and $u^{(n)}$ is an infinite Polya tree. The sequence $X = (X_1, X_2, \ldots)$ is <u>infinite stage Polya with parameter</u> u or $\mathcal{P}_\infty(u)$. We can also code each X_n using a k+1-ary expansion as in (1.1) to get a random variable Y_n with values in the unit interval I. We will also call the sequence $Y = (Y_1, Y_2, \ldots)$ <u>infinite stage Polya on</u> I or $\mathcal{P}_I(u)$.

The basic properties of the infinite stage Polya sequences are easily derived from the finite case. To do this, let $X = (X_1, X_2, \ldots)$ be $\mathcal{P}_\infty(u)$ and, for positive integers s and n, let

$$X_n^{(s)} = (X_{n1}, \ldots, X_{ns})$$

be the first s coordinates of X_n; let $X^{(s)}$ be the sequence $(X_1^{(s)}, X_2^{(s)}, \ldots)$; and let $u^{(s)}$ be the restriction of u to E_{s-1}. Here is an obvious but useful fact.

<u>Lemma</u> 4.1. If X is $\mathcal{P}_\infty(u)$, then, for every s, $X^{(s)}$ is $\mathcal{P}_s(u^{(s)})$.

<u>Theorem</u> 4.1. If $(X_1, X_2, \ldots)$ is $\mathcal{P}_\infty(u)$ ($(Y_1, Y_2, \ldots)$ is $\mathcal{P}_I(u)$), then $X_1, X_2, \ldots$ $(Y_1, Y_2, \ldots)$ are exchangeable.

<u>Proof</u>: Let n be a positive integer and let A be a Borel subset of $(E^N)^n$. To prove exchangeability of the X_i's, we need to check that $P[(X_1, \ldots, X_n) \in A]$ is invariant under permutations of the indices. It suffices to do this for sets A of the form

$$A = A_1 \times \ldots \times A_n$$

where each A_i is a cylinder set in E^N of the form

$$A_i = B_i \times E^N, \; B_i \subset E^{r_i}.$$

for some positive integer r_i. Take s to be the maximum of the r_i's so that each A_i depends on only the first s coordinates and is of the form

$$A_i = C_i \times E^N, \; C_i \subset E^s.$$

Thus

$$P[X_1 \in A_1, \ldots, X_n \in A_n] = P[X_1^{(s)} \in C_1, \ldots, X_n^{(s)} \in C_n].$$

Exchangeability of the X_i's now follows from that of the $X_i^{(s)}$'s. The exchangeability of the Y_i's is an easy consequence of that of the X_i's. □

A more general form of Theorem 2.1 (cf. Hewitt and Savage (1955) or Aldous (1985)) can now be used to see that there is a de Finetti measure for a sequence which is $\mathcal{P}_\infty(u)$ or $\mathcal{P}_I(u)$. To describe these measures, let $M(E^N)$ $(M(I))$ be the space of probability measures defined on the Borel subsets of $E^N(I)$ and give this space its usual topology of weak convergence. If X is $\mathcal{P}_\infty(u)$, then the de Finetti measure $Q = Q_u$ for X is a probability measure defined on the Borel subsets of $M(E^N)$ and satisfying

$$\text{(4.1)} \qquad P[X_1 \in A_1, \ldots, X_n \in A_n] = \int \left(\prod_{i=1}^{n} \theta(A_i) \right) Q(d\theta)$$

for all n and all Borel subsets $A_1, \ldots, A_n$ of E^N. The de Finetti measure $\bar{Q} = \bar{Q}_u$ for a Y which is $\mathcal{P}_I(u)$ can be defined similarly; just replace the X_i's by Y_i's and take the sets A_i to be Borel subsets of the unit interval. The existence and uniqueness of Q and $\bar{Q}$ are well-known (cf. Hewitt and Savage (1955) or Aldous (1985)). There is, of course, a simple relationship between Q and $\bar{Q}$.

Let ψ: $E^N \to I$ be the mapping defined by

$$\psi(x_1, x_2, \ldots) = \sum_{n=1}^{\infty} x_n/(k+1)^n. \tag{4.2}$$

Clearly, ψ is continuous and induces a continuous mapping $\theta \to \theta\psi^{-1}$ from $M(E^N)$ into $M(I)$ where $(\theta\psi^{-1})(B) = \theta(\psi^{-1}(B))$ for B a Borel subset of I. Now $\tilde{Q}$ can be thought of as the distribution of $\Theta\psi^{-1}$ when Θ has distribution Q; that is,

$$\tilde{Q}(F) = Q\{\theta: \theta\psi^{-1} \in F\} \tag{4.3}$$

for Borel subsets F of $M(I)$.

To characterize Q, it is again useful to use the notion of a strategy.

Definition. A strategy θ is a mapping from E^* to the collection of probability measures on E.

Each strategy θ naturally determines a measure $\mu = \mu(\theta) \in M(E^N)$ in the same fashion that an s-day strategy determines a measure on E^s. As in the s-day case, we will write θ for $\mu(\theta)$ and we can obtain a random measure Θ with values in $M(E^N)$ by putting a joint distribution on $\{\Theta(p): p \in E^*\}$.

Let u be an infinite Polya tree.

Definition. A random probability measure Θ with values in $M(E^N)$ is a Dirichlet strategy with parameter u (written $\underset{\sim}{D}_\infty(u)$) if the random probability measures $\{\Theta(p): p \in E^*\}$ are independent and $\Theta(p)$ is $\underset{\sim}{D}(u(p))$ for every $p \in E^*$. The measure corresponding to the distribution of such a Θ is also said to be $\underset{\sim}{D}_\infty(u)$.

Theorem 4.2. If $X = (X_1, X_2, \ldots)$ is $\mathcal{P}_\infty(u)$, then the de Finetti measure Q_u of X is $\underset{\sim}{D}_\infty(u)$.

Proof: If Θ is $\underset{\sim}{D}_\infty(u)$, then, for every positive integer s, Θ restricted to E_{s-1} is $\underset{\sim}{D}_s(u)$. So, by Theorem 3.1, equality (4.1) holds when every A_i depends only on the first s coordinates. □

Again the Dirichlet strategies are conjugate for strategic sampling even if the experiment is terminated at some finite stage.

<u>Theorem</u> 4.3. Suppose Θ is $\underset{\sim}{D}_\infty(u)$ and, given Θ, $X_1 = (X_{11}, X_{12}, \ldots)$ has distribution Θ. Then, given X_1 (or given $(X_{11}, \ldots, X_{1r})$), Θ is $\underset{\sim}{D}_\infty(u^{(1)})$ where

$$\begin{aligned} u^{(1)}(\phi) &= u(\phi) + \delta(X_{11}), \\ u^{(1)}(X_{11}, \ldots, X_{1r}) &= u(X_{11}, \ldots, X_{1r}) + \delta(X_{1,r+1}) \\ &\qquad (\text{for } r \le s-1), \\ u^{(1)}(p) &= u(p) \text{ for other } p \in E^*. \end{aligned}$$

<u>Proof</u>: In checking the properties of a conditional distribution, one can restrict attention to Θ restricted to the finite sets E_s and to finitely many of the X_{1n}'s. But then the desired properties follow from Theorem 3.2. □

Another way of expressing the result of Theorem 4.3 is that if Θ has the prior distribution Q_u on $M(E^N)$, then, given X_1, the posterior distribution of Θ is $Q_{u^{(1)}}$ where $u^{(1)}$ is the tree obtained in the construction of section 1. It follows that the posterior of Θ given $X_1, \ldots, X_n$ is $Q_{u^{(n)}}$. (Notice that $u^{(n)}$ is a random tree depending on the values of $X_1, \ldots, X_n$.)

Suppose next that Θ is a random element of $M(I)$ with distribution $\bar{Q}_u$ and we wish to calculate the posterior distribution of Θ given Y_1 where $Y_1 = \psi(X_1)$. It follows from (4.2) and (4.3) that this posterior distribution will be $\bar{Q}_{u^{(1)}}$ if we make the convention that $u^{(1)}$ should be the urn associated with an X_1 such that $\psi(X_1) = Y_1$ and having only finitely many 0's (say) and if the probability under $\bar{Q}_u$ that Y_1 is a k+1-ary rational is zero. This last condition is certainly satisfied if Y_1 has a continuous distribution in the sense of the next section.

5. Continuity of predictive distributions and random measures.

A <u>prior</u> $Q(\bar{Q})$ on $M(E^N)$ $(M(I))$ is a probability measure defined on the Borel subsets of $M(E^N)$ $(M(I))$. Suppose the random measure θ has distribution $Q(\bar{Q})$ and, given θ, the variables $X_1, X_2, \ldots$ are independent each having distribution θ. The marginal distribution of X_1 is called the <u>predictive distribution</u> for the prior $Q(\bar{Q})$.

The collection of <u>continuous measures</u> on $E^N(I)$ is given by

$$C = \{\theta \in M(E^N): \theta\{x\} = 0 \text{ for all } x \in E^N\}.$$
$$(\bar{C} = \{\theta \in M(I): \theta\{x\} = 0 \text{ for all } x \in I\}.)$$

We are interested in conditions on an infinite Polya tree u for the prior $Q_u(\bar{Q}_u)$ to have a continuous predictive distribution and also for $Q_u(C)$ $(\bar{Q}_u(\bar{C}))$ to be 1. For convenience, we will work on E^N, but the conditions given apply as well on I for the prior $\bar{Q}_u$.

Now, under Q_u, the probability that X_1 equals $x \in E^N$ is just the probability, in the Polya tree construction, of traversing the path $x = (x_1, x_2, \ldots)$ and this probability is given by the infinite product

$$(5.1) \qquad \Pi(x;u) = \frac{u(\phi)x_1}{|u(\phi)|} \cdot \frac{u\left(x_1\right)x_2}{|u(x_1)|} \cdot \frac{u\left(x_1,x_2\right)x_3}{|u(x_1,x_2)|} \cdots .$$

Here is an immediate consequence and an obvious corollary.

<u>Theorem</u> 5.1. The predictive distribution of Q_u $(\bar{Q}_u)$ is continuous if and only if $\Pi(x;u) = 0$ for every $x \in E^N$.

<u>Corollary</u> 5.1. If the proportion of balls of each color is bounded away from 1 in the sense that

$$(5.2) \qquad \sup\left\{\frac{u(p)_i}{|u(p)|}: p \in E^*, \ i \in E\right\} < 1,$$

then the predictive distribution of Q_u ($\bar{Q}_u$) is continuous.

In the special case when every $u(p)_i$ is an integer, condition (5.2) is satisfied if the total number of balls in each urn is uniformly bounded and if there are balls of different colors in every urn.

The next two lemmas hold for a general prior $Q(\bar{Q})$ on $M(E^N)$ ($M(I)$).

Lemma 5.1. If $Q(C)$ ($\bar{Q}(\bar{C})$) equals one, then the predictive distribution of $Q(\bar{Q})$) in continuous.

Proof: If $Q(C) = 1$ and $x \in E^N$, then

$$\begin{aligned} P[X_1=x] &= \int P[X_1=x|\Theta=\theta]\ dQ(\theta) \\ &= \int_C \theta\{x\}\ dQ(\theta) \\ &= 0. \quad \square \end{aligned}$$

Lemma 5.2. A necessary and sufficient condition for $Q(C)$ ($\bar{Q}(\bar{C})$) to be one is that $P[X_1=X_2]$ be zero.

Proof: Now X_1 and X_2 are independent with distribution θ given $\Theta = \theta$. Hence, $P[X_1=X_2|\Theta=\theta]$ is zero if and only if θ is continuous. Consequently,

$$P[X_1=X_2] = \int_{C^c} P[X_1=X_2|\Theta=\theta]\ dQ(\theta) > 0$$

if and only if $Q(C^c) > 0$. $\square$

To apply Lemma 5.2 to the case where $Q = Q_u$ for a Polya tree u, notice that, given $X_1 = x$, the distribution of X_2 is the predictive distribution for $Q_{u(1)}$ by Theorem 4.3. Thus

$$(5.3) \qquad P[X_2=x|X_1=x] = \Pi(x;u^{(1)}).$$

Also, if m is the distribution of X_1, then

$$(5.4) \qquad P[X_1=X_2] = \int P[X_2=x|X_1=x]\ dm(x).$$

Now we are ready to give conditions for $Q_u(C)$ to be one.

Theorem 5.2. If $Q_u(C)$ $(\bar{Q}_u(\bar{C}))$ equals one, then $\Pi(x;u) = 0$ for every $x \in E^N$. If $\Pi(x;u^{(1)}) = 0$ for every $x \in E^N$, then $Q_u(C)$ $(\bar{Q}_u(\bar{C}))$ equals one.

Proof: The first assertion is immediate from Theorem 5.1 and Lemma 5.1. The second follows from (5.3), (5.4), and Lemma 5.2. □

The reader should be aware that the tree $u^{(1)}$ in the expression $\Pi(x;u^{(1)})$ depends on x. Thus the condition, $\Pi(x;u^{(1)}) = 0$ for all x, is not a condition on a single tree. However, every $u^{(1)}(p)_i$ is either $u(p)_i+1$ or $u(p)_i$ depending on whether $p = (x_1,...,x_n)$ and $x_{n+1} = i$ for some n or not. Similarly every $|u^{(1)}(p)|$ is either $|u(p)| + 1$ or $|u(p)|$ according to whether $p = (x_1,...,x_n)$ for some n or not. It follows that the terms occurring in the infinite product expression (5.1) for $\Pi(x;u^{(1)})$ will be bounded away from 1 if (5.2) holds and the $|u(p)|$ are bounded away from zero. This observation yields a corollary to the second assertion of Theorem 5.2.

Corollary 5.2. If (5.2) holds and

$$(5.5) \qquad \inf\{|u(p)|: p \in E^*\} > 0,$$

then $Q_u(C)$ $(\bar{Q}_u(\bar{C}))$ is one.

Condition (5.5) obviously holds when every $u(p)_i$ is an integer because in that case $|u(p)| \geq 1$. (We are not allowing empty urns.)

6. Support and Consistency

As Ferguson (1973) remarked, it is often desirable for a prior to have large support. The <u>support</u> (or <u>topological carrier</u>) of a probability measure μ defined on the Borel subsets of a compact Hausdorff space M is the least compact set $S(\mu)$ which has μ-measure one. Notice that μ has <u>full support</u> in the sense that $S(\mu) = M$ if and only if every nonempty, open subset of M has positive μ-measure.

Here is a characterization of the Polya tree priors which have full support.

<u>Theorem</u> 6.1. The following are equivalent conditions on an infinite Polya tree u:

(a) The prior Q_u has full support.
(b) The prior $\bar{Q}_u$ has full support.
(c) For every $p \in E^*$, the Dirichlet measure $D(u(p))$ has full support.
(d) For every $p \in E^*$ and $i \in E$, $u(p)_i > 0$.

<u>Proof</u>: (a) $\Rightarrow$ (b). Assume (a) and let F be a nonempty, open subset of M(I). Then the set $G = \{\theta: \theta\psi^{-1} \in F\}$ is a nonempty, open subset of $M(E^N)$ and, by (4.3),

$$\bar{Q}(F) = Q(G) > 0.$$

(Here and below $\bar{Q} = \bar{Q}_u$ and $Q = Q_u$.)

(b) $\Rightarrow$ (d). Suppose (d) is false. So there exist $p = (i_1,\ldots,i_m) \in E^*$ and $i \in E$ such that $u(p)_i = 0$. Thus, if Θ is a random measure with distribution Q_u, then $\Theta(p)\{i\} = 0$ with probability one.

Let l and r be the numbers whose expansions to base k+1 are

$$l = .i_1 \ldots i_m i00\ldots,$$
$$r = .i_1 \ldots i_m ikk\ldots,$$

and let J be the closed interval [l,r]. Let $g: I \to [0,\infty)$ be a continuous, nonzero function which equals zero on the complement of J. Consider the nonempty, open set

$$U = \{\mu \in M(I): \int g d\mu > 0\}.$$

To prove (b) is false, we need only show $\bar{Q}(U) = 0$.

Notice that U is a subset of the set

$$F = \{\mu \in M(I): \mu(J) > 0\}.$$

So it suffices to show $\bar{Q}(F) = 0$. By (4.3),

$$\bar{Q}(F) = Q\{\theta: \theta(\psi^{-1}(J)) > 0\}.$$

Also,

$$\psi^{-1}(J) = \{x \in E^N: x_1=i_1, \ldots, x_m=i_m, x_{m+1}=i\}$$

and

$$\begin{aligned}\theta(\psi^{-1}(J)) &= \theta_o\{i_1\}\theta(i_1)\{i_2\}\ldots\theta(i_1,\ldots,i_m)\{i\} \\ &\leq \theta(i_1,\ldots,i_m)\{i\}. \\ &= \theta(p)\{i\}.\end{aligned}$$

Hence,

$$\bar{Q}(F) \leq Q\{\theta: \theta(p)\{i\} > 0\} = 0.$$

(c) $\Leftrightarrow$ (d). This is a trivial consequence of our conventions about the Dirichlet distribution. (Notice that to say $\underset{\sim}{D}(u(p))$ has full support means that the support of $\underset{\sim}{D}(u(p))$ is S_k.)

(d) $\Rightarrow$ (a). Assume (d). It suffices to show that each set in a base for the topology of $M(E^N)$ has positive Q - measure. The usual base for the weak topology consists of sets of the form

$$\{\theta \in M(E^N): |\int g_i d\theta - \int g_i d\theta_o| < \epsilon_i,\ i=1,\ldots,n\}$$

where the g_i are continuous, real-valued functions on E^N, θ_o is a fixed element of $M(E^N)$, and the ϵ_i are positive numbers (cf. section II.6 of Parthasarathy (1967)). However, every continuous g on E^N can be uniformly approximated by a finite, linear combination of indicator functions of clopen sets of the form

(6.1) $$C = \{x \in E^N: x_1=i_1, \ldots, x_n=i_n\}.$$

(This follows from the Stone-Weierstrass Theorem, for example.) Hence, another base for the topology of $M(E^N)$ consists of sets of the form

$$(6.2) \qquad B = \{\theta \in M(E^N): |\theta(C_i) - \theta_o(C_i)| < \epsilon_i, \ i=1,\ldots,n\}$$

where each C_i is a set of the form (6.1), $\theta_o \in M(E^N)$, and each ϵ_i is positive.

So we need only show Q(B) is positive for B as in (6.2). To do this, we begin by borrowing a trick from Ferguson (1973, Proposition 3.3). Call n the dimension of the set C in (6.1) and let b be the maximum of the dimensions of the C_i occurring in (6.2). Then each C_i is a disjoint union of at most 2^b clopen sets of the form

$$(6.3) \qquad D = \{x \in E^N: x_1=j_1,\ldots,x_b=j_b\}.$$

Let ϵ be the minimum of the ϵ_i in (6.2). Then B contains the set

$$(6.4) \quad F = \bigcap_D \{\theta \in M(E^N): |\theta(D) - \theta_o(D)| < \epsilon/2^b\}$$

where the intersection is over all D of the form (6.3). So it suffices to show that $Q(F) > 0$.

Now each D occurring in (6.4) is of dimension b. So a strategy θ belongs to F if and only if its restriction $\theta^{(b)}$ to E_{b-1} belongs to

$$F_b = \bigcap_{D_b} \{\theta \in M(E^b): |\theta(D_b) - \theta_o^{(b)}(D_b)| < \epsilon/2^b\}$$

where D_b is that subset of E^b such that

$$D = D_b \times E^N.$$

The set F_b is a nonempty, open subset of $M(E^b)$, and if Θ has distribution Q, then

$$\begin{aligned} Q(F) &= P[\Theta \in F] \\ &= P[\Theta^{(b)} \in F_b]. \end{aligned}$$

The final probability is positive because $\theta^{(b)}$ is $\underset{\sim}{D}_b(u^{(b)})$ and has, under condition (d), a density which is positive on all of $M(E^b)$. (The space $M(E^b)$, can be identified with the set S_n where $n = k^b-1$ and the density of $\theta^{(b)}$ is taken with respect to Lebesgue measure.) □

Let $\lambda \in M(E^N)$ and let u be an infinite Polya tree. Suppose data variables $X_1, X_2, \ldots$ are independent with distribution λ and consider the posterior distributions $Q_{u^{(n)}}$ calculated from these data variables. Following Freedman (1963) and Fabius (1964), call the pair (λ, Q_u) <u>consistent</u> if, with probability one, $Q_{u^{(n)}}$ converges weakly to $\delta(\lambda)$, the measure concentrated on $\{\lambda\}$. For $\lambda \in M(I)$, the consistency of the pair $(\lambda, \bar{Q}_u)$ is defined similarly under the assumption that λ gives mass zero to the collection of k+1-ary rationals so that $\bar{Q}_{u^{(n)}}$ is the correct posterior given data $Y_1, \ldots, Y_n$ which are independent with distribution λ.

<u>Theorem</u> 6.2. If λ belongs to the support of Q_u ($\bar{Q}_u$ and assigns mass zero to the k+1-ary rationals), then the pair (λ, Q_u) $((\lambda, \bar{Q}_u))$ is consistent.

The proof of this theorem is similar to the proof of Theorem 2.2 in Fabius (1964), the main idea being a reduction to the finite, discrete case of Freedman 1963). Indeed, the measures $\bar{Q}_u$ are "tail-free" in the sense of Fabius so that this result is almost immediate from his.

7. Conditional trees, conditional strategies, and an amalgamation formula.

Let θ be a strategy on E^N and recall that we regard θ also as being that measure in $M(E^N)$ such that, for each clopen set of the form,

$$C = \{x \in E^N: x_1 = i_1, \ldots, x_n = i_n\},$$

$$\theta(C) = \theta_o\{i_1\}\theta(i_1)\{i_2\}\ldots\theta(i_1, \ldots, i_{n-1})\{i_n\}.$$

For each $i \in E$, the <u>conditional strategy</u> $\theta[i]$ is defined by

$$\theta[i](p) = \theta(ip)$$

for each $p \in E^*$, where $ip \in E^*$ is the concatenation of i with p (cf. Dubins and Savage (1976, section 2.5)). Now the measure $\theta[i]$ corresponds to the conditional distribution under θ of $(x_2,x_3,\ldots)$ given $x_1=i$. So, for any Borel set $B \subset E^N$,

$$(7.1) \qquad \theta(B) = \sum_{i=0}^{k} \theta[i](Bi)\theta_o\{i\}$$

where $Bi = \{(x_1,x_2,\ldots): (i,x_1,x_2,\ldots) \in B\}$.

Notice that, conversely, given probability measures $\lambda \in M(E)$ and $\theta^{(o)},\ldots,\theta^{(k)}$ in $M(E^N)$, we can always construct a measure $A(\lambda,\theta^{(o)},\ldots\theta^{(k)})$ in $M(E^N)$ by the formula

$$(7.2) \qquad A(\lambda,\theta^{(o)},\ldots,\theta^{(k)})(B) = \sum_{i=0}^{k} \theta^{(i)}(Bi)\lambda\{i\}$$

for Borel $B \subset E^N$. The function A is clearly Borel measurable.

Next consider an infinite Polya tree u and, for each $i \in E$, define the <u>conditional tree</u> u[i] by

$$u[i](p) = u(ip)$$

for all $p \in E^*$. (Pictorially, u[i] is just that part of the tree u which grows out of the initial branch i.)

<u>Theorem</u> 7.1. Let u be an infinite Polya tree and let Θ be a random meaure with distribution Q_u. Then the following are true.

(a) The distribution R_u of Θ_o is $\underset{\sim}{D}(u(\phi))$.
(b) The distribution of $\Theta[i]$ is $Q_{u[i]}$ for each $i \in E$.
(c) $\Theta_o,\Theta[o],\ldots,\Theta[k]$ are independent.
(d) If g is a bounded, Borel measurable function from $M(E^N)$ to the real line, then

(7.3) $\int g(\theta)Q_u(d\theta)=\int\ldots\int g(A(\lambda,\theta^{(o)},\ldots,\theta^{(k)}))R_u(d\lambda)Q_{u[o]}(d\theta^{(o)})\ldots Q_{u[k]}(d\theta^{(k)})$.

<u>Proof</u>: By Theorem 4.2, Q_u is $\underset{\sim}{D}_\infty(u)$. So $\Theta_o = \Theta(\phi)$ is $\underset{\sim}{D}(u(\phi))$ and (a) is true. Similarly, for every $p \in E^*$ and $i \in E$, $\Theta[i](p) = \Theta(ip)$ is $\underset{\sim}{D}(u(ip)) = \underset{\sim}{D}(u[i](p))$. Also, the measures $\{\Theta[i](p): p \in E^*\}$ are independent because the measures $\{\Theta(p): p \in E^*\}$ are independent. Thus $\Theta[i]$ is $\underset{\sim}{D}_\infty(u[i])$ which is $Q_{u[i]}$ and b is proved. Property (c) follows from the fact that $\{\Theta_o\},\{\Theta[o](p): p \in E^*\},\ldots,\{\Theta[k](p): p \in E^*\}$, are disjoint subfamilies of $\{\Theta(p): p \in E^*\}$, which is a family of independent random measures. The final property is an immediate consequence of (a), (b), and (c). It can also be proved directly by considering special g's of the form

$$g(\theta) = \theta\{x:\ x_1=i_1,\ldots,x_n=i_n\}$$

and inducting on n. □

There is a straightforward translation of our results for E^N into results for I. First identify each $y \in I$ with its expansion to base k+1

$$y = .y_1y_2\ldots$$

making the convention that only finitely many y_i's can be equal to k. Think of a measure $\mu \in M(I)$ as being the distribution of y and, for each $i \in E$, let $\mu[i]$ be the conditional distribution under μ of $.y_2y_3\ldots$ given $y_1=i$. Also, let μ_o be the μ-distribution of y_1. Then, for any Borel subset C of I,

$$\mu(C) = \sum_{i=1}^{k} \mu[i](C(i))\mu_o\{i\}, \tag{7.4}$$

where $C(i) = \{.y_2y_3\ldots:\ .iy_2y_3\ldots \in C\}$ for $i \in E$.

Also, given $\lambda \in M(E)$ and $\mu^{(o)},\ldots,\mu^{(k)}$ in M(I), we can construct a measure $\bar{A}(\lambda,\mu^{(o)},\ldots,\mu^{(k)})$ in M(I) by the rule

$$(7.5)\quad \bar{A}(\lambda,\mu^{(o)},\ldots,\mu^{(k)})(C) = \sum_{i=0}^{k} \mu^{(i)}(C(i))\lambda\{i\}.$$

Here now is the translation of (7.3) to the interval.

Theorem 7.2. If u is an infinite Polya tree and g is a bounded, Borel measurable function from M(I) to the real line, then

$$(7.6)\ \int g(\mu)\bar{Q}_u(d\mu) = \int\ldots\int g(\bar{A}(\lambda,\mu^{(o)},\ldots,\mu^{(k)}))R_u(d\lambda)\bar{Q}_{u[o]}(d\mu^{(o)})\ldots\bar{Q}_{u[k]}(d\mu^{(k)}).$$

Proof: By (4.3), the left side of (7.6) is equal to

$$\int g(\theta\psi^{-1})\ Q_u(d\theta)$$

which, by (7.3) is the same as,

$$\int\ldots\int g(A(\lambda,\theta^{(o)},\ldots,\theta^{(k)})\psi^{-1})R_u(d\lambda)Q_{u[o]}(d\theta^{(o)})\ldots Q_{u[h]}(d\theta^{(k)}).$$

However, it is easily checked that

$$A(\lambda,\theta^{(o)},\ldots,\theta^{(k)})\ \psi^{-1} = \bar{A}(\lambda,\theta^{(o)}\psi^{-1},\ldots,\theta^{(k)}\psi^{-1})$$

and, by (4.3), for each $i \in E$, the $Q_{u[i]}$ distribution of $\theta^{(i)}\psi^{-1}$ is $\bar{Q}_{u[i]}$. □

Consider a Polya tree u for which u(p) is the same for all $p \in E^*$. Then $u[i] = u$ for every $i \in E$ and the formula of (7.6) is an equation for the measure $\bar{Q}_u$. This formula has predecessors in Dubins and Freedman (1967) and Graf, Mauldin, and Williams (1986).

8. Estimation of a distribution function.

Suppose $Y_1,\ldots,Y_n$ is a sample from a distribution $\Theta \in M(I)$ and we wish to estimate the distribution function F for Θ given by

$$F(y) = \Theta[0,y],\ 0 \le y \le 1.$$

If Θ has prior distribution $\bar{Q}_u$ then, as Ferguson (1973, section 5(a)) explains, a natural Bayes estimator is the distribution function $\hat{F}_n(y|Y_1,\ldots,Y_n)$ which corresponds to the expected value of F(y) under the posterior $\bar{Q}_{u^{(n)}}$ and can be written as

$$\hat{F}_n(y|Y_1,\ldots,Y_n) = \int\theta[o,y]\ \bar{Q}_{u^{(n)}}(d\theta).$$

Since we know the form of the Polya tree $u^{(n)}$, the problem reduces to the no data case where $\hat{F}$ is the expected (or predictive) distribution function as in

$$\hat{F}(y) = \hat{F}(y;u) = \int\theta[o,y]\ \bar{Q}_u(d\theta).$$

For the calculation, introduce the notation $\Pi((y_1,\ldots,y_n);u)$ for the probability that the Polya tree process $X_{11},\ldots,X_{1n}$ traverses the finite path $(y_1,\ldots,y_n)$; that is,

$$\Pi((y_1,\ldots,y_n);u) = \frac{u(\phi)_{y_1}}{|u(\phi)|}\cdots\frac{u(y_1,\ldots,y_{n-1})_{y_n}}{|u(y_1,\ldots,y_n)|}.$$

The distribution function $\hat{F}(y;u)$ is, of course, the distribution function for the predictive distribution discussed in section 5. That is, $\hat{F}(y;u)$ is the distribution function for the random variable

$$Y = .X_{11}X_{12}\ldots \text{ (to base k+1)} \tag{8.1}$$

where $X_1 = (X_{11},X_{12},\ldots)$ is the random sequence constructed in the infinite Polya urn scheme of the introduction.

<u>Theorem</u> 8.1. The expected distribution function under the Polya tree prior $\bar{Q}_u$ is

$$\hat{F}(y;u) = \sum_{n=0}^{\infty}\left[\Pi((y_1,\ldots,y_n);u)\sum_{i=0}^{y_{n+1}-1} u(y_1,\ldots,y_n)_i/|u(y_1,\ldots,y_n)|\right] + \Pi((y_1,y_2,\ldots);u). \tag{8.2}$$

for every $y = .y_1y_2\ldots$ (to base k+1) in I. (Here $\Pi(\phi;u) = 1$ and $\Pi((y_1,y_2,\ldots);u)$ is as in (5.1).)

Proof: The proof is by repeated applications of the amalgamation formula (7.6) in which we set $g(\theta) = \theta[o,y]$. Notice that, if $C = [o,y]$ and $i \in E$, then

$$C(i) = \{.y_2\ldots: .iy_2\ldots \le .y_1y_2\ldots\}$$

$$= \begin{cases} I \text{ if } i < y_1, \\ [0,.y_2\ldots] \text{ if } i = y_1, \\ \phi \qquad \text{ if } i > y_1. \end{cases}$$

Now use (7.6) to get

$$(8.3)\quad \hat{F}(y;u) = \int \theta[0,y]\bar{Q}_u(dy)$$

$$= \int\ldots\int\left\{\sum_{i=0}^{y_1-1}\lambda\{i\} + \theta^{(y_1)}[0,.y_2\ldots]\lambda\{y_1\}\right\}R_u(d\lambda)Q_{u[o]}(d\theta^{(o)})\ldots Q_{u[k]}(d\theta^{(k)})$$

$$= \frac{\sum_{i=0}^{y_1-1} u(\phi)_i}{|u(\phi)|} + \frac{u(\phi)_{y_1}}{|u(\phi)|}\hat{F}\left(.y_2y_3\ldots;u[y_1]\right).$$

Now use (7.6) (or the above formula) to get an expression for $\hat{F}(.y_2y_3\ldots;u[y_1])$ and so forth. After n steps, we obtain an expression for $\hat{F}(y;u)$ consisting of the first n terms of (8.2) and a remainder term equal to

$$r_n = \Pi((y_1,\ldots,y_n);u)\hat{F}(.y_{n+1}\ldots;u[y_1][y_2]\ldots[y_n]).$$

Clearly,

$$0 \le \lim_n r_n \le \lim_n \Pi((y_1,\ldots,y_n);u)$$

$$= \Pi((y_1,y_2,\ldots);u).$$

So, if $\Pi((y_1,y_2,\ldots);u) = 0$, we are finished. On the other hand, if $\Pi((y_1,y_2,\ldots);u) > 0$, then

$$
\begin{aligned}
1 &\geq \hat{F}(.y_{n+1}\ldots;u[y_1]\ldots[y_n]) \\
&\geq \Pi((y_{n+1},y_{n+2},\ldots);u[y_1]\ldots[y_n]) \\
&= \frac{u(y_1,\ldots,y_n)_{y_{n+1}}}{|u(y_1,\ldots,y_n)|} \cdot \frac{u(y_1,\ldots,y_{n+1})_{y_{n+2}}}{|u(y_1,\ldots,y_{n+1})|} \cdots \\
&\to 1 \text{ as } n \to \infty. \quad \square
\end{aligned}
$$

Formula (8.1) is annoyingly complex when compared with Ferguson's formula [8,(5.2)]. However, it does simplify in interesting special cases.

Say that an infinite Polya tree u has <u>constant proportions</u> λ if there is a fixed probability vector $\lambda = (\lambda_o,\ldots,\lambda_k)$ such that, for every $p \in E^*$ and $i \in E$,

$$\frac{u(p)_i}{|u(p)|} = \lambda_i.$$

Let $\lambda^* = (k+1)^{-1}(1,1,\ldots,1)$ be the probability vector all of whose coordinates are equal to $(k+1)^{-1}$.

<u>Theorem</u> 8.2. Suppose u has constant proportions λ.

(a) If $\lambda = \lambda^*$, then $\hat{F}(\dot{y};u) = y$ for all $y \in I$.
(b) If $\lambda \neq \lambda^*$, then $\hat{F}(y;u)$ is singular with respect to Lebesgue measure.

<u>Proof</u>: If $\lambda = \lambda^*$, then, by (8.2),

$$\hat{F}(y;u) = \sum_{n=0}^{\infty} (k+1)^{-n}(y_{n+1}/(k+1))$$

$$= \sum_{n=1}^{\infty} y_n (k+1)^{-n}$$

$$= y.$$

Suppose now that $\lambda \neq \lambda^*$. So, for some $i \in E$, $\lambda_i \neq (k+1)^{-1}$. Now the variables $X_{11}, X_{12}, \ldots$ of (8.1) are clearly independent with distribution λ. By the strong law of large numbers $\hat{F}(\cdot;u)$ assigns probability one to the set of all $y = .y_1y_2\ldots$ such that

$$\#\{j \le n\colon y_j = i\}/n \to \lambda_i \text{ as } n \to \infty.$$

But this set has Lebesgue measure zero. □

Consider again the problem of calculating $\hat{F}_n(y|Y_1,\ldots,Y_n) = \hat{F}(y;u^{(n)})$. In principle, the problem is solved since we know the form of $u^{(n)}$ and can apply (8.2). However, it may be suggestive to rewrite (8.3) for $u^{(n)}$ as

$$\hat{F}(y;u^{(n)}) =$$

$$\frac{\left[\sum_{i=0}^{y_1-1} u(\phi)_i + \sum_{j=1}^{n} \delta_{Y_{j1}}\{0,\ldots,y_1-1\}\right] + \left[u(\phi)_{y_1} + \sum_{j=1}^{n} \delta_{Y_{j1}}\{y_1\}\right]\hat{F}(.y_2\ldots;u^{[n]}[y_1])}{|u(\phi)| + n},$$

where Y_{j1} is the first coordinate of Y_j to base $k+1$ and δ_Y is point mass at Y_{j1} for each j. In the interesting special case where $u(p) = (1,1,\ldots,1)$ for all $p \in E^*$, the formula becomes

$$\hat{F}(y;u^{(n)}) = \frac{y_1 + \sum_{j=1}^{n} \delta_{Y_{j1}}\{0,\ldots,y_1-1\} + \left[1 + \sum_{j=1}^{n} \delta_{Y_{j1}}\{y_1\}\right] \hat{F}(.y_2\ldots;u^{(n)}[y_1])}{k+n}.$$

This formula can be iterated as in the proof of Theorem 8.1. Notice that, if the observations are censored after the first stage (i.e. only $Y_{11},\ldots,Y_{n1}$ are observed), then $u^{(n)}[y_1] = u[y_1]$ and $\hat{F}(.y_2\ldots,u^{(n)}[y_1]) = .y_2\ldots$. So, in this case, the calculation is complete. Similarly, if observations are censored after n stages, then n iterations suffice for the calculation.

9. Remarks.

Although it was convenient for us to develop the theory of the Polya tree priors $\tilde{Q}_u$ using the unit interval and k+1-ary rationals, there is no serious difficulty in developing the theory for the real line or any standard Borel space and using a more general tree of partitions as in Fabius (1964) and Blackwell (1973). A nice way to choose partition intervals on the line is to use percentiles of the predictive distribution. As Michael Lavine showed, there is then a simple way to construct a Polya tree having any specified predictive distribution.

Michael Lavine and Robert Wolpert also noticed that the class of Polya tree priors includes that of Ferguson distributions. In fact, Theorem 1 of Blackwell (1973) implies that any Ferguson distribution arises from a Dirichlet strategy. Suppose, for example, that θ is a random element of $M(\{0,1\}^N)$ which has a Ferguson distribution with parameter λ. Then θ is $\mathcal{P}_\infty(u)$ where $u(i_1,\ldots,i_n)_i = \lambda\{x \in \{0,1\}^N\colon x_1=i_1,\ldots,x_n=i_n,x_{n+1}=i\}$ for each $(i_1,\ldots,i_n,i) \in \{0,1\}^*$.

If one sought to use Polya trees or mixtures of them to approximate an actual opinion, it would be natural to specify the first two or three stages of the tree carefully and then to use a reference prior such as the tree with one ball of each color in every urn. Thus one could be relatively precise about the first few coordinates of X_1 and vague about the rest.

References

[1] Aldous, David J. (1983). Exchangeability and related topics. École d'Été de Probabilité de Saint-Flour XIII. Lecture Notes in Math. 1117 1-197.

[2] Blackwell, D. (1973). Discreteness of Ferguson selections. Ann. Statist. 1 356-358.

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

[4] Blackwell, D. and MacQueen, J.B. (1973). Ferguson distributions via Polya urn schemes. Ann. Statist. 1 353-355.

[5] Dubins, L.E. and Freedman, D.A. (1967). Random distribution functions. Proc. Fifth Berkeley Symp. Math. Statist. Prob. 2 183-214. Univ. of California Press.

[6] Dubins, L.E. and Savage, L.J. (1965,1976). Inequalities for Stochastic Processes. Dover, New York.

[7] Fabius, J. (1964). Asymptotic behavior of Bayes estimates. Ann. Math. Statist. 35 846-856.

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

[9] Freedman, D.A. (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. Ann. Math. Statist. 34 1386-1403.

[10] Graf, S., Mauldin, R.D. and Williams, S.C. (1986). Random homeomorphisms. Advances in Math. 60 239-359.

[11] Hewitt, E. and Savage, L.J. (1955). Symmetric measures on Cartesian products. Trans. Amer. Math. Soc. 80 470-501.

[12] Mauldin, R.D. and Williams, S.D. (1990). Reinforced random walks and random distributions. Proc. Amer. Math. Soc. (to appear).

[13] Meyer, P.A. (1966). Probability and Potentials. Blaisdell, Waltham, Massachusetts.