Ferguson Distributions Via Polya Urn Schemes

David Blackwell; James B. MacQueen

The Annals of Statistics, Vol. 1, No. 2 (Mar., 1973), 353-355.

Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28197303%291%3A2%3C353%3AFDVPUS%3E2.0.CO%3B2-5

The Annals of Statistics
1973, Vol. 1, No. 2, 353-355

FERGUSON DISTRIBUTIONS VIA PÓLYA URN SCHEMES

BY DAVID BLACKWELL[1] AND JAMES B. MACQUEEN[2]

University of California, Berkeley

The Polya urn scheme is extended by allowing a continuum of colors. For the extended scheme, the distribution of colors after n draws is shown to converge as $n \to \infty$ to a limiting discrete distribution μ^*. The distribution of μ^* is shown to be one introduced by Ferguson and, given μ^*, the colors drawn from the urn are shown to be independent with distribution μ^*.

Let μ be any finite positive measure on (the Borel sets of) a complete separable metric space X. We shall say that a random probability measure μ^* on X has a *Ferguson distribution with parameter* μ if for every finite partition $(B_1, \cdots, B_r)$ of X the vector $\mu^*(B_1), \cdots, \mu^*(B_r)$ has a Dirichlet distribution with parameter $\mu(B_1), \cdots, \mu(B_r)$ (when $\mu(B_i) = 0$, this means $\mu^*(B_i) = 0$ with probability 1). Ferguson [3] has shown that, for any μ, Ferguson μ^* exist and when used as prior distributions yield Bayesian counterparts to well-known classical nonparametric tests. He also shows that μ^* is a.s. discrete. His approach involves a rather deep study of the gamma process.

One of us [1] has given a different and perhaps simpler proof that Ferguson priors concentrate on discrete distributions. In this note we give still a third approach to Ferguson distributions, exploiting their connection with generalized Pólya urn schemes.

We shall say that a sequence $\{X_n, n \geqq 1\}$ of random variables with values in X is a *Pólya sequence with parameter* μ if for every $B \subset X$

$$P(X_1 \in B) = \mu(B)/\mu(X) \tag{1}$$

and

$$P\{X_{n+1} \in B \mid X_1, \cdots, X_n\} = \mu_n(B)/\mu_n(X) , \tag{2}$$

where $\mu_n = \mu + \sum_1^n \delta(X_i)$ and $\delta(x)$ denotes the unit measure concentrating at x. Note that, for finite X, the sequence $\{X_n\}$ represents the results of successive draws from an urn where initially the urn has $\mu(x)$ balls of color x and, after each draw, the ball drawn is replaced and another ball of its same color is added to the urn. Note also that, without the restriction to finite X, for any (Borel measurable) function ϕ on X, the sequence $\{\phi(X_n)\}$ is a Pólya sequence with parameter $\phi\mu$, where $\phi\mu(A) = \mu\{\phi \in A\}$.

We now describe the connections between Pólya sequences and Ferguson distributions.

Received August 13, 1971.

[1] Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, USAF, under Grant AFOSR-71-2100.

[2] Visiting professor of Psychology and Statistics.

THEOREM. *Let $\{X_n\}$ be a Pólya sequence with parameter μ. Then*

(a) *$m_n = \mu_n/\mu_n(X)$ converges with probability 1 as $n \to \infty$ to a limiting discrete measure μ^*,*

(b) *μ^* has a Ferguson distribution with parameter μ and*

(c) *given μ^*, the variables $X_1, X_2, \cdots$ are independent with distribution μ^*.*

PROOF. Suppose first that X is finite, say $X = \{1, 2, \cdots, r\}$. Let μ^*, $\{X_n\}$ be variables whose joint distribution is defined by (b) and (c). If π_n is the empirical distribution of $X_1, \cdots, X_n$, it follows from the strong law of large numbers that $\pi_n \to \mu^*$ with probability 1 as $n \to \infty$. Since

$$m_n = (\mu + n\pi_n)/(\mu(X) + n)\,,$$

(a) follows. It remains to show that $\{X_n\}$ is a Pólya sequence with parameter μ, which is equivalent to

$$P(A) = \textstyle\prod_x \mu(x)^{[n(x)]}/\mu(X)^{[n]} \tag{3}$$

where $A = \{X_1 = x_1, \cdots, X_n = x_n\}$, $n(x)$ denotes the number of i with $x_i = x$ and $a^{[k]} = a(a+1)\cdots(a+k-1)$.

Since $P(A \mid \mu^*) = \prod_x \mu^*(x)^{n(x)}$, we get

$$P(A) = E \textstyle\prod_x \mu^*(x)^{n(x)}\,. \tag{4}$$

That the right sides of (3) and (4) are equal is a standard formula [2] for the moments of Dirichlet distributions.

For general X, let $\{X_n\}$ be a Pólya sequence with parameter μ, let I_j be the indicator of the event that X_j is different from all X_i with $i < j$ and define

$$f_{nj} = I_j m_n(X_j) \qquad \text{for} \quad 1 \leqq j \leqq n\,,$$
$$f_{nj} = 0 \qquad \text{for} \quad j > n\,.$$

We show that

$$\text{with probability 1,} \quad f_{nj} \quad \text{converges as} \quad n \to \infty\,, \tag{5}$$

say to f_j^* and

$$\textstyle\sum_j f_j^* = 1 \quad \text{with probability 1}\,. \tag{6}$$

Part (a) of the Theorem, with μ^* defined by

$$\mu^*(B) = \textstyle\sum_{x_j \in B} f_j^*$$

is an easy consequence of (5) and (6) since, for any B, we have, writing

$$s_{nr} = \textstyle\sum_{1 \leq j \leq r;\, x_j \in B} f_{nj}\,, \qquad t_{nr} = \sum_{1 \leq j \leq r} f_{nj}$$

$$s_{nr} \leqq m_r(B) \leqq s_{nr} + (1 - t_{nr}) \qquad \text{for} \quad 1 \leqq r \leqq n\,, \tag{7}$$

so that, letting first $n \to \infty$, then $r \to \infty$ we obtain (a).

To get (5) and (6), fix r and define

$$U_n = j \qquad \text{if} \quad 1 \leqq j \leqq r \quad \text{and} \quad I_j = 1 \quad \text{and} \quad X_{r+n} = X_j$$
$$= 0 \qquad \text{otherwise.}$$

Given $X_1, \cdots, X_r$, the sequence $\{U_n\}$ is a Pólya sequence on $\{0, 1, \cdots, r\}$ with parameter μ' defined by

$$\mu'(j) = \mu_r(X) f_{rj} \qquad \text{for } 1 \leqq j \leqq r,$$
$$\mu'(0) = \mu_r(X) - \textstyle\sum_{j=1}^r \mu'(j),$$

and the sequence m_n' associated with $\{U_n\}$ satisfies

$$m_n'(j) = f_{r+n,j} \qquad \text{for } 1 \leqq j \leqq r \tag{8}$$

and

$$m_n'(0) = 1 - \textstyle\sum_{j=1}^r f_{r+n,j}. \tag{9}$$

We apply the finite case of our Theorem to $\{U_n\}$. From (8) and part (a) of the Theorem we get (5), and from (9) and part (b) of the Theorem we conclude

$$E(1 - \textstyle\sum_1^r f_j^* \mid X_1, \cdots, X_r) = \frac{\mu'(0)}{\mu_r(X)} \leqq \frac{\mu(X)}{\mu(X) + r}. \tag{10}$$

Taking expectation in (10) and letting $r \to \infty$ gives $E(1 - \sum_{j=1}^\infty f_j^*) = 0$, and (6) follows.

Parts (b) and (c) are now easy consequences of the finite case. For any finite partition $B_1, \cdots, B_r$ of X, define ϕ on X by $\phi = i$ on B_i, so that $\{\phi(X_n)\}$ is a Pólya sequence with parameter $\phi\mu$. We conclude that the limit of $(m_n(B_1), \cdots, m_n(B_r))$, already identified as $(\mu^*(B_1), \cdots, \mu^*(B_r))$, has a Dirichlet distribution with parameter $\mu(B_1), \cdots, \mu(B_r)$, establishing (b). For (c), let $\{\phi_i\}$ be a sequence of functions on X, each with finitely many values, such that, if $\mathscr{F}_i$ is the (finite) field of X-sets determined by ϕ_i, we have $\mathscr{F}_{i+1} \supset \mathscr{F}_i$ and the Borel field determined by $\mathscr{F} = \bigcup \mathscr{F}_i$ consists of all Borel sets. Part (c) of the finite case of our Theorem, applied to $\{\phi_j(X_n)\}$, yields

(c′) given $\phi_j \mu^*$, the sequence $\{\phi_i(X_n)\}$ is independent with distribution $\phi_i \mu^*$ for $i \leqq j$.

Letting $j \to \infty$, we get

(c″) given μ^*, the sequence $\{\phi_i(X_n)\}$ is independent with distribution $\phi_i \mu^*$ for all i.

Since $\{\phi_i(X_n)\}$ is independent with distribution $\phi_i \mu^*$ for all i implies $\{X_n\}$ is independent with distribution μ^*, part (c) follows from (c″), completing the proof.

REFERENCES

[1] Blackwell, David (1973). Discreteness of Ferguson selections. *Ann. Statist.* **1** 356–358.
[2] DeGroot, Morris H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.
[3] Ferguson, Thomas S. (1972). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

Department of Statistics
University of California
Berkeley, California 94720