The Annals of Statistics
1973, Vol. 1, No. 2, 356-358

DISCRETENESS OF FERGUSON SELECTIONS[1]

BY DAVID BLACKWELL

University of California, Berkeley

In a fundamental paper on nonparametric Bayesian inference, Ferguson [1] associated with each probability measure α on a set S and each positive number c a way of selecting a probability measure on S at random. One of his interesting results is that his method selects a discrete distribution with probability 1. Ferguson's proof uses an explicit representation of the gamma process; we present here a quite different and perhaps simpler proof.

THEOREM 1 (Ferguson). *Let S be a nonempty Borel subset of a complete separable metric space and let $B_1, B_2, \cdots$ be Borel subsets of S that form a separating sequence,* i.e. *for any two distinct points s_1 and s_2 of S there is an n for which $\xi_n(s_1) \neq \xi_n(s_2)$, where ξ_n is the indicator of B_n. For any finite sequence $t = (\varepsilon_1, \cdots, \varepsilon_k)$ of 0's and 1's, denote by $B(t)$ the set of all s for which $(\xi_1, \cdots, \xi_k) = t$; for the empty sequence e, put $B(e) = S$. For any probability measure α on the Borel sets of S and any positive number c, if we select a function y from the set T of all finite sequences of 0's and 1's to the unit interval $[0, 1]$ so that the $y(t)$ are independent and $y(t)$ has a beta distribution with parameters $u(t)$ and $v(t)$, where*

$$u(t) = c\alpha(B(t1))$$
$$v(t) = c\alpha(B(t0))$$

then, with probability 1, there will be a unique probability distribution p on the Borel sets of S such that

$$p(\xi_{k+1} = 1 \mid (\xi_1, \cdots, \xi_k) = t) = y(t) \qquad \textit{for all} \quad t \in T. \tag{1}$$

Moreover, with probability 1, p will be discrete.

The beta distribution for $u > 0$, $v = 0$ is concentrated at 1 and for $u = 0$, $v > 0$ is concentrated at 0; its definition for $u = v = 0$ is irrelevant. Uniqueness of p is clear, since given y we can calculate $p(B(t))$ for all t and, since $\xi = (\xi_1, \xi_2, \cdots)$ is separating, any two p's that agree on all $B(t)$ are identical.

It will be seen that what forces discreteness is convergence of $\sum_t Ey(t)(1 - y(t))$. To get this convergence we shall use Theorem 2.

THEOREM 2. *Put $x(t) = \alpha(B(t))$. Then*

(a) $\sum_{|t| \leq n} x(t0)x(t1) = \frac{1}{2}(1 - D_n)$, *where $|t|$ denotes the length of t and* $D_n = \sum_{|w|=n+1} x^2(w)$.

(b) $\sum_t x(t0)x(t1) = \frac{1}{2}(1 - D)$, *where $D = \sum_s \alpha^2(s)$ is the sum of the squares of the probabilities of all points of S that have positive probability.*

Received October 27, 1969; revised September 27, 1971.

[1] Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, USAF, under Grant AFOSR-1312-67.