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.

PROOF OF THEOREM 2. Select two points σ_1 and σ_2 independently in S with distribution α and denote by F_k the event $\xi_i(\sigma_1) = \xi_i(\sigma_2)$ for $i < k$, $\xi_k(\sigma_1) = 0$, $\xi_k(\sigma_2) = 1$. The left and right sides of (a) are easily seen to be the probability of $\bigcup_1^{n+1} F_k$, and the left and right sides of (b) are easily seen to be the probability of $\bigcup_1^{\infty} F_k$.

To prove Theorem 1 we first check that, with probability 1, there will be a p related to y by (1). Any y determines a (unique) probability measure q on the space Ω of infinite sequences of 0's and 1's such that

$$q(\omega \text{ begins with } t1 \mid \omega \text{ begins with } t) = y(t) \qquad \text{for all } t\,.$$

Any p that makes $\xi = (\xi_1, \xi_2, \cdots)$ have distribution q will satisfy (1) and there will be such a p if (and only if) $q(\xi S) = 1$ (ξS is Borel, being the $1-1$ Borel measurable image of S.) As noted by Ferguson, if y is selected as in Theorem 1,

$$Eq(\omega \text{ begins with } t) = \alpha(B(t)) \qquad \text{for all } t\,. \tag{2}$$

Now $Eq(A)$ and $\alpha(\xi^{-1}A)$ are probability measures on Ω and (2) says they agree on sets of the form "ω begins with t." So they agree for all Borel sets. In particular for $A = \xi S$ we get $Eq(\xi S) = 1$, so that $q(\xi S) = 1$ with probability 1.

To see that p is discrete with probability 1, for any probability distribution p on S and any $s \in S$, say that S conforms to p at stage $k+1$ ($k \geqq 0$) if

$$\begin{aligned} &\xi_{k+1}(s) = 1 \quad \text{and} \quad y(\xi_1(s), \cdots, \xi_k(s)) \geqq \tfrac{1}{2} \qquad \text{or} \\ &\xi_{k+1}(s) = 0 \quad \text{and} \quad y(\xi_1(s), \cdots, \xi_k(s)) < \tfrac{1}{2}\,, \end{aligned}$$

i.e. if ξ_{k+1} has its more probable value given the previous ξ_j, with equality resolved (arbitrarily) in favor of 1. Say that s ultimately conforms to p if it conforms to p at all but a finite number of stages. For any p, there are only countably many ultimately conforming s. We show that if p is selected as in Theorem 1 and then s is selected according to p, the probability that s ultimately conforms to p is 1.

The probability that s fails to conform to p at stage $k+1$, given $\xi_1, \cdots, \xi_k$ and $y(t)$ for $|t| \leqq k$ is

$$w_k = \min y(\xi_1, \cdots, \xi_k)\,, \qquad 1 - y(\xi_1, \cdots, \xi_k)\,,$$

so that the probability that s fails to conform to p at step $k+1$ is Ew_k, and s will ultimately conform to p with probability 1 if

$$\textstyle\sum_k Ew_k \quad \text{converges.}$$

Now $E(w_k \mid (\xi_1, \cdots, \xi_k) = t) = m(t)$, where $m(t) = E \min(\beta, 1-\beta)$ and β has a beta $u(t)$, $v(t)$ distribution, so that

$$E(w_k) = \textstyle\sum_{|t|=k} P(B(t))m(t)\,.$$

But, from (2), $P(B(t)) = \alpha[B(t)]$, so that

$$E(w_k) = \textstyle\sum_{|t|=k} [u(t) + v(t)]m(t)/c\,.$$

To complete the proof, use $\min(\beta, 1-\beta) \leqq 2\beta(1-\beta)$ for $0 \leqq \beta \leqq 1$ to obtain (suppressing t for the moment)

$$m \leqq 2\left[\frac{u}{u+v} - \frac{u(u+1)}{(u+v)(u+v+1)}\right] \leqq \frac{2uv}{u+v},$$

so that

$$E(w_k) \leqq 2 \sum_{|t|=k} u(t)v(t)/c .$$

Since $u(t) = cx(t1)$ and $v(t) = cx(t0)$, we obtain, from Theorem 2,

$$\sum E(w_k) \leqq c(1-D),$$

where D is the sum of the squares of the probabilities that α assigns to points.

REFERENCE

[1] FERGUSON, THOMAS S. (1972). A Bayesian analysis of some non-parametric problems. *Ann. Statist.* **1** 209–230.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720