The Annals of Statistics
1973, Vol. 1, No. 3, 583-587

TWO CHARACTERIZATIONS OF THE DIRICHLET DISTRIBUTION

BY J. FABIUS

University of Leiden

Let $X = (X_1, \cdots, X_k)$ be a random vector with all $X_i \geqq 0$ and $\sum X_i \leqq 1$. Let $k \geqq 2$, and suppose that none of the X_i, nor $1 - \sum X_i$ vanishes almost surely. Without any further regularity assumptions, each of two conditions is shown to be necessary and sufficient for X to be distributed according to a Dirichlet distribution or a limit of such distributions. Either condition requires that certain proportions between components of X be independent of one or more other components of X.

1. Summary. Let $X = (X_1, \cdots, X_k)$ be a random vector with nonnegative components adding up to at most one. Let $k \geqq 2$, and suppose that none of the X_i, nor $1 - \sum X_i$ vanishes almost surely.

DEFINITION 1. X is $(CM)_i$-neutral for a given $i \in \{1, \cdots, k\}$ iff, for any integers $r_j \geqq 0$, $j \neq i$, there is a constant c, such that

$$E(\prod_{j \neq i} X_j^{r_j} \mid X_i) = c(1 - X_i)^{\sum_{j \neq i} r_j} \quad \text{a.s.}$$

DEFINITION 2. X is $(DR)_i$-neutral for a given $i \in \{1, \cdots, k\}$ iff, for any integer $r \geqq 0$, there is a constant c, such that

$$E(X_i^r \mid X_j, j \neq i) = c(1 - \sum_{j \neq i} X_j)^r \quad \text{a.s.}$$

THEOREM. *The following assertions are equivalent*:

(i) *X is $(CM)_i$-neutral for all i*;
(ii) *X is $(DR)_i$-neutral for all i*;
(iii) *The distribution of X is a Dirichlet distribution or a limit of Dirichlet distributions.*

2. Discussion. Both $(CM)_i$-neutrality and $(DR)_i$-neutrality are independence properties: the first means that the fractions $X_j/(1 - X_i)$ with $j \neq i$ are independent of X_i, while the latter requires that $X_i/(1 - \sum_{j \neq i} X_j)$ be independent of the X_j with $j \neq i$. Nevertheless we prefer the less intuitive formulation of the definitions in terms of conditional moments as given above, because this does not involve any possibly undefined quantities. These and similar independence properties arise naturally in certain statistical problems in biology, chemistry and geology. For this reason they have been studied under various names by a number of authors, including in particular Connor and Mosimann (1969), and Darroch and Ratcliff (1971).

Received April 1972; revised August 1972.
AMS 1970 *subject classification.* 62E10.
Key words and phrases. Dirichlet distribution, Dirichlet process, neutrality, tailfree random distribution function.

On the other hand, Fabius (1972) characterized $(CM)_i$-neutrality in terms of the posterior distributions one obtains when using the distribution of X as a prior distribution for the unknown probability vector of a multinomial distribution. $(DR)_i$-neutrality can be characterized in a similar manner. Moreover, Doksum (1971), (1972) and Fabius (1972) pointed out that there is an intimate connection between various notions of neutrality and various kinds of tailfree random distribution functions. Thus theorems to the effect that X must have a Dirichlet distribution if there is "enough" neutrality lead to theorems asserting that any random distribution function, which is tailfree in a sufficiently strong sense, is a Dirichlet process as defined by Ferguson (1972).

The present theorem is not at all new. In fact, Darroch and Ratcliff (1971) proved the equivalence of (ii) and (iii) under certain regularity conditions involving continuous densities. Moreover, Theorem 2 of Fabius (1972) asserts that (iii) is equivalent to a condition which is quite similar to (i), but slightly more restrictive. The difference stems from the fact that in the earlier paper the components of X were assumed to add up to exactly one. Thus, whereas (i) consists of k neutrality conditions, the earlier theorem, reformulated to fit the present terminology, imposes $k + 1$ such conditions. In consequence the proof (Lemma 3 below) has become somewhat more involved.

To conclude this section we note that (iii) is just a conveniently short way of saying that the distribution of X is either a Dirichlet distribution, or discrete and concentrated in the vertices of the simplex in which X takes its values, or degenerate.

3. Proof of the theorem. The implications (iii) $\Rightarrow$ (i) and (iii) $\Rightarrow$ (ii) are easy consequences of well-known properties of Dirichlet distributions as described by Wilks (1962). The converse implications follow from Lemmas 2 and 3 below. To simplify the notation we write

$$Y_i = \sum_{j \neq i} X_j\,, \qquad Y_I = \sum_{j \notin I} X_j$$

for any

$$i \in \{1, \cdots, k\}\,, \qquad I \subset \{1, \cdots, k\}\,.$$

LEMMA 1. (ii) *implies, for any proper subset I of $\{1, \cdots, k\}$, any $i \in I$, and any integer $r \geqq 0$, the existence of a constant c, such that*

$$E(X_i^r \mid X_j, j \notin I) = c(1 - Y_I)^r \quad \text{a.s.}$$

PROOF. We proceed by induction with respect to both $|I|$, i.e. the number of elements of I, and r. Let H_{ms} be the induction hypothesis that the assertion holds for any set I with $|I| \leqq m$ and any $r \leqq s$. Note that $H_{1,r}$ and $H_{m,0}$ are trivially true for any $r \geqq 0$ and any $m \in \{1, \cdots, k-1\}$. Thus we only need to show that $H_{m,r+1}$ and $H_{m+1,r}$ together imply $H_{m+1,r+1}$ for any $r \geqq 0$, $m \in \{1, \cdots, k-2\}$. To do this, we must fix a set I with $|I| = m + 1$ and $i \in I$. However, without loss of generality we may set $i = 1$, $I = \{1, \cdots, m + 1\}$. Because of

$H_{m,r+1}$ we know there is a constant c_1, such that

$$(1)\qquad E(X_1^{r+1} \mid X_j, j \notin I) = E\{E(X_1^{r+1} \mid X_j, j > m) \mid X_j, j \notin I\}$$
$$= c_1 E\{(1 - Y_I - X_{m+1})^{r+1} \mid X_j, j \notin I\} \quad \text{a.s.}$$

Moreover, taking expectations, we see that $0 < c_1 < 1$. Expanding the binomium $((1 - Y_I) - X_{m+1})^{r+1}$ in (1) and using $H_{m+1,r}$, we obtain

$$(2)\qquad E(X_1^{r+1} \mid X_j, j \notin I) = (-1)^{r+1} c_1 E(X_{m+1}^{r+1} \mid X_j, j \notin I) + c_2(1 - Y_I)^{r+1} \quad \text{a.s.}$$

where c_2 is a constant. In the same way we can show the existence of constants c_1' and c_2' with $0 < c_1' < 1$, such that

$$(3)\qquad E(X_{m+1}^{r+1} \mid X_j, j \notin I) = (-1)^{r+1} c_1' E(X_1^{r+1} \mid X_j, j \notin I) + c_2'(1 - Y_I)^{r+1} \quad \text{a.s.}$$

Substitution of (3) in (2) yields

$$E(X_1^{r+1} \mid X_j, j \notin I) = \frac{c_2 + (-1)^{r+1} c_1 c_2'}{1 - c_1 c_1'} (1 - Y_I)^{r+1} \quad \text{a.s.},$$

and the lemma is proved.

LEMMA 2. (ii) *implies, for any proper subset* I *of* $\{1, \cdots, k\}$ *and any integers* $r_i \geqq 0$, $i \in I$, *the existence of a constant* c, *such that*

$$E(\textstyle\prod_{i \in I} X_i^{r_i} \mid X_j, j \notin I) = c(1 - Y_I)^{\Sigma_{i \in I} r_i} \quad \text{a.s.}$$

Thus in particular (ii) *implies* (i).

PROOF. We again use induction on $|I|$, the assertion being trivially true for sets I with $|I| = 1$. Hence we start out from the assumption that the assertion holds for all sets I with $|I| \leqq m$ for a given $m \leqq k - 2$, and we fix a set I with $|I| = m + 1$. As before we put $I = \{1, \cdots, m + 1\}$ without loss of generality. Our assumption guarantees the existence, for any integers $r_i \geqq 0$, $i \in I$, of a constant c, such that

$$E(\textstyle\prod_{i \in I} X_i^{r_i} \mid X_j, j \notin I) = E\{X_{m+1}^{r_{m+1}} E(\prod_{i=1}^{m} X_i^{r_i} \mid X_j, j > m) \mid X_j, j \notin I\}$$
$$= cE\{X_{m+1}^{r_{m+1}} (1 - Y_I - X_{m+1})^{\Sigma_{i=1}^{m} r_i} \mid X_j j \notin I\} \quad \text{a.s.}$$

Expanding $((1 - Y_I) - X_{m+1})^{\Sigma r_i}$ and applying Lemma 1 we obtain the desired result.

LEMMA 3. (i) *implies* (iii).

PROOF. For arbitrary distinct $i, j \in \{1, \cdots, k\}$ and any integers r, $s \geqq 0$, the $(CM)_i$-and $(CM)_j$-neutrality implies

$$(4)\qquad \frac{EX_i^r X_j^s}{EX_i^r EX_j^s} = \frac{EX_i^r(1 - X_i)^s}{EX_i^r E(1 - X_i)^s} = \frac{EX_j^s(1 - X_j)^r}{EX_j^s E(1 - X_j)^r}.$$

In particular it follows that $q = EX_i(1 - X_i)/EX_i E(1 - X_i)$ does not depend on i. If either $q = 0$ or $q = 1$, one easily verifies that the distribution of X is a limit of Dirichlet distributions. In all other cases $0 < q < 1$. We then define

$a, a_1, \cdots, a_{k+1} > 0$, putting $q = a/(a+1)$, $a_i = aEX_i$ for $i = 1, \cdots, k$, and $\sum_1^{k+1} a_i = a$. It now turns out that the distribution of X is the Dirichlet distribution $D(a_1, \cdots, a_k; a_{k+1})$ in the notation of Wilks (1962). We prove this by showing that the moments of X coincide with the corresponding ones of $D(a_1, \cdots, a_k; a_{k+1})$.

We first show that all mixed moments can be expressed in terms of marginal moments, or, more precisely, that for any i any mixed moment of $X_1, \cdots, X_i$ can be expressed in terms of marginal moments of these same random variables. This follows by induction on i: It is trivially true for $i = 1$, and for any $i \leqq k - 1$ we have

$$E \prod_{j=1}^{i+1} X_j^{r_j} = \frac{E \prod_{j=i}^{i} X_j^{r_j}}{E(1 - X_{i+1})^{r_{i+1}}} EX_{i+1}^{r_{i+1}}(1 - X_{i+1})^{\sum_{j=1}^{i} r_j}$$

by the $(CM)_{i+1}$-neutrality.

It remains to be shown that the marginal moments of the X_i have the right values. This is in fact true for the first moments by the definition of the a_i. Supposing it to be true for all moments of order not exceeding a given $r \geqq 1$, we may use (4) to obtain

$$EX_i^r(1 - X_i) = \frac{EX_i^r E(1 - X_i)}{EX_j E(1 - X_j)^r} EX_j(1 - X_j)^r \tag{5}$$

and

$$EX_i^{r-1}(1 - X_i)^2 = \frac{EX_i^{r-1} E(1 - X_i)^2}{EX_j^2 E(1 - X_j)^{r-1}} EX_j^2(1 - X_j)^{r-1} \tag{6}$$

for arbitrary distinct i and j. Adding (5) and (6) we obtain an expression for $EX_i^{r-1}(1 - X_i)$ involving EX_j^{r+1} and moments of order not exceeding r. Up to its sign the coefficient of EX_j^{r+1} in this expression is given by

$$\begin{aligned} &\frac{EX_i^{r-1}E(1 - X_i)^2}{EX_j^2 E(1 - X_j)^{r-1}} - \frac{EX_i^r E(1 - X_i)}{EX_j E(1 - X_j)^r} \\ &\qquad = \frac{a_i(a_i + 1) \cdots (a_i + r - 2)(a - a_i)}{a_j(a - a_j)(a - a_j + 1) \cdots (a - a_j + r - 2)} \\ &\qquad\quad \times \left\{\frac{a - a_i + 1}{a_j + 1} - \frac{a_i + r - 1}{a - a_j + r - 1}\right\} > 0 , \end{aligned}$$

and hence does not vanish. Thus we can solve for EX_j^{r+1}, expressing it in terms of moments of lower order. Without further computation we may conclude that all moments of order $r + 1$ coincide with those of $D(a_1, \cdots, a_k; a_{k+1})$, and thus the proof is complete.

REFERENCES

[1] CONNOR, R. J. and MOSIMANN, J. E. (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. *J. Amer. Statist. Assoc.* **64** 194–206.

[2] DARROCH, J. N. and RATCLIFF, D. (1971). A characterization of the Dirichlet distribution. *J. Amer. Statist. Assoc.* **66** 641–643.

[3] DOKSUM, K. A. (1971). Tailfree and neutral processes and their posterior distributions. Report ORC 71-27, Operations Research Center, Univ. of California, Berkeley.

[4] DOKSUM, K. A. (1972). Decision theory for some nonparametric models. *Proc. Sixth Berkeley Symp. Math. Statist. Prob* **1** 331-343. Univ. of California Press.

[5] FABIUS, J. (1972). Neutrality and Dirichlet distributions. *Trans. Sixth Prague Conf. Inform. Theor., Statist. Dec. Functions, Random Processes.*

[6] FERGUSON, T. S. (1972). A Bayesian analysis of some nonparametric problems. To appear.

[7] WILKS, S. S. (1962). *Mathematical Statistics*. Wiley, New York.

CENTRAAL REKEN-INSTITUUT
DER RIJKSUNIVERSITEIT TE LEIDEN
WASSENAARSEWEG 80
POSTBUS 2060
LEIDEN, NETHERLANDS