A Characterization of Dirichlet Distributions

B. V. RAO AND BIKAS K. SINHA*

Indian Statistical Institute, Calcutta, India

Communicated by C. G. Khatri

J. N. Darroch and D. Ratcliff (*J. Amer. Statist. Assoc.* **66** (1971), 641–643) have given a characterization of the Dirichlet distributions based on the properties of independence of various functions of the random variables $(X_1, X_2, ..., X_k)$ having a joint continuous distribution over the k-dimensional simplex: $0 \leqslant x_i \leqslant \sum x_i \leqslant 1$. In this paper we provide a further characterization of this family of distributions essentially based on the properties of linear regression. Some extra conditions have been imposed and these are indeed indispensable.   © 1988 Academic Press, Inc.

1. INTRODUCTION

Characterizations in statistics form a fascinating area of study and research. There is a huge literature on characterizations of statistical distributions based on various properties of such distributions. This paper aims at incorporating a characterization of the Dirichlet family of distributions based on the properties of regression. The motivation came from a recent work [2] in which Neyman's problem of characterization of the family of multinomial distributions based on the properties of regression was properly studied and the present problem posed.

In [1], a characterization of the Dirichlet family of distributions is available. Note that a two-dimensional random variable (X, Y) is said to have Dirichlet distribution if the joint density $f(x, y)$ is of the form

$$f_{\alpha, \beta, \gamma}(x, y) = \text{Const.} \; x^{\alpha-1} y^{\beta-1} (1 - x - y)^{\gamma-1};$$

$$\alpha, \beta, \gamma > 0; 0 < x, y < x + y < 1. \quad (1.1)$$

For the two-dimensional case, the result in [1] may be stated as follows:

25

Suppose $(X, Y) \sim f(x, y)$ over the simplex $0 < x, y < x + y < 1$. Assume $f(x, y)$, $f(x, \cdot)$, and $f(\cdot, y)$ (the marginals) to be continuous. Then $(X, Y) \sim$ Dirichlet iff the following two conditions hold simultaneously:

(a) $(X/(1 - Y), Y)$ are independently distributed,

(b) $(Y/(1 - X), X)$ are independently distributed.

Clearly, for the Dirichlet distribution (1.1),

$$E(X \mid Y) = \frac{\alpha}{\alpha + \gamma}(1 - Y), \qquad E(Y \mid X) = \frac{\beta}{\beta + \gamma}(1 - X). \qquad (1.2)$$

The present investigation arose in trying to characterize the Dirichlet distribution through the regression equations alone. Obviously, the two linear equations by themselves are not strong enough to characterize the Dirichlet (see Remarks 1 and 2). The extra conditions needed are that the density be of the product form and one of these terms be a power function. These conditions seem too strong but in spite of these, the proof that the density must be Dirichlet turns out to be interesting. That these conditions are the best possible is shown by Remark 2.

2. Some Preliminary Results

LEMMA 1. *Let X, Y be jointly distributed with law P having full support on the simplex $0 < x, y, x + y < 1$. Assume that the two regressions $E(X \mid Y)$ and $E(Y \mid X)$ are linear. Then there exist constants $0 < c < 1$, $0 < d < 1$ such that*

$$E(X \mid Y) = c(1 - Y) \qquad a.e. \ [Y]$$
$$E(Y \mid X) = d(1 - X) \qquad a.e. \ [X].$$

Proof. Assume that $E(X \mid Y) = \alpha(1 - Y) + \beta$ so that $0 < \alpha(1 - Y) + \beta < 1 - y$ a.e. $[P]$. Since the support of P is the full simplex, take $\{y_n\} \to 1$ such that for each y_n this inequality holds. This gives $\beta = 0$. It is now immediate that $0 < \alpha < 1$. This completes the proof.

LEMMA 2. *Let ϕ be a positive integrable function on $(0, 1)$. Suppose that*

$$\int_0^y (y - t)^{\alpha + 1} \phi(t) \, dt = cy \int_0^y (y - t)^\alpha \phi(t) \, dt \qquad \textit{for all} \quad y \in (0, 1),$$

where $0 < c < 1$, $\alpha > -1$ and the integral on the right-hand side exists. Then $\phi(t) = c_0 t^{\gamma - 1}$ for all $t \in (0, 1)$, c_0 is a constant, and $c\gamma = (\alpha + 1)(1 - c) > 0$.

Proof. Let $\psi(y) = \int_0^y (y-t)^{\alpha+1} \phi(t) dt$. Then $\psi'(y) = (\alpha+1) \int_0^y (y-t)^\alpha \phi(t) dt$ and taking $c_1 = (\alpha+1)/c$, the given equation can be written as

$$\psi'(y)/\psi(y) = c_1/y \qquad \text{for all } y \in (0, 1).$$

Hence

$$\psi(y) = c_0 \, y^{c_1} \qquad \text{for all } y \in (0, 1),$$

where c_0 is some constant of integration. Then, we can write

$$\int_0^z (z-y)^{\delta-1} \left(\int_0^y (y-t)^\alpha \, \phi(t) \, dt \right) dy = (c_0/c) \int_0^z (z-y)^{\delta-1} \, y^{c_1-1} \, dy$$

for all $z \in (0, 1)$ and for all $\delta > 0$. After interchanging the integrations of t and y, we can write

$$\int_0^z \phi(t) \left(\int_t^z (z-y)^{\delta-1} (y-t)^\alpha \, dy \right) dt = (c_0/c) \, z^{\delta+c_1-1} B(\delta, c_1)$$

or

$$\int_0^1 (1-t)^{\alpha+\delta} \, \phi(tz) \, dt = (c_0/c) \, z^{\gamma-1} B(\alpha+\delta+1, \gamma)/B(\alpha+1, \gamma)$$

for all $z \in (0, 1)$ and $\gamma = c_1 - \alpha - 1$ (>0). This gives Mellin's transform of $t^\alpha \phi((1-t))$, namely,

$$\int_0^1 t^\delta \{ t^\alpha \phi((1-t) z) \} \, dt = (c_0/c) \, z^{\gamma-1} B(\alpha+\delta+1, \gamma)/B(\alpha+1, \gamma),$$

and then using the uniqueness of Mellin's transformation, we get the required result. This proves Lemma 2.

3. MAIN RESULTS: TWO-DIMENSIONAL CASE

THEOREM 1.   *Let (X, Y) be a random vector having joint distribution over the simplex $S = \{0 < x, y, x+y < 1\}$ with density function $f(x, y)$. Assume that*

   (i) $f(x, y) > 0$ *for all $(x, y) \in S$;*
   (ii) $f(x, y) = f_1(x) f_2(y) f_3(1-x-y)$ *for all $(x, y) \in S$;*
   (iii) f_1 *or f_3 is a power function; and*
   (iv) *regression of X on Y is linear.*

Then $f(x, y) = c_0 x^{\alpha - 1}(1 - x - y)^{\gamma - 1} f_2(y)$ for $(x, y) \in S$, where $\alpha > 0$, $\gamma > 0$ and c_0 is a positive constant.

Proof. (A) Let $f_1(x) = kx^{\alpha - 1}$ for some $\alpha > 0$ and for all $x \in (0, 1)$. By Lemma 1,

$$E(X \mid Y = y) = c(1 - y) \qquad \text{for all} \quad y \in (0, 1) \text{ and } 0 < c < 1.$$

This is equivalent to

$$\int_0^{1 - y} x^\alpha f_3(1 - x - y) \, dx = c(1 - y) \int_0^{1 - y} x^{\alpha - 1} f_3(1 - x - y) \, dx$$

$$\text{for all} \quad y \in (0, 1)$$

or

$$\int_0^y (y - t)^\alpha f_3(t) \, dt = cy \int_0^y (y - t)^{\alpha - 1} f_3(t) \, dt$$

$$\text{for all} \quad y \in (0, 1).$$

Now, from Lemma 2, $f_3(t) = c_0 t^{\gamma - 1}$ for all $t \in (0, 1)$ and $\gamma = \alpha(1 - c)/c > 0$. This proves the required result.

    (B) Let $f_3(t) = \delta t^{\gamma - 1}$ for all $t \in (0, 1)$ and $\gamma > 0$. Then, the regression of X on Y can be written as

$$\int_0^{1 - y} (1 - y - x)^\gamma f_1(x) \, dx = (1 - c)(1 - y) \int_0^{1 - y} (1 - x - y)^{\gamma - 1} f_1(x) \, dx$$

for all $y \in (0, 1)$. By Lemma 2, we get $f_1(x) = kx^{\alpha - 1}$ for all $x \in (0, 1)$ and $\alpha = \gamma c/(1 - c)$. This proves the other part. Thus Theorem 1 is established.

THEOREM 2. *Suppose (X, Y) have a joint distribution over the simplex $S = \{0 < x, y, x + y < 1\}$ with density $f(x, y)$. Assume*

   (i)   *and* (ii) *as in Theorem 1;*

  (iii)   *regressions $E(X \mid Y)$ and $E(Y \mid X)$ are both linear; and*

  (iv)   *at least one f_i is a power function.*

Then $f(x, y)$ must admit of the Dirichlet density.

Proof. If $f_1(f_2)$ is a power function, by linearity of regression of $X(Y)$ on $Y(X)$, we derive, using part (A) of Theorem 1, the fact that f_3 is a power function. This then suggests, by part (B) of Theorem 1, that $f_2(f_1)$ is a

power function since linearity of regression of $Y(X)$ on $X(Y)$ is also assumed here. This essentially proves the result.

Remark 1. Unless $f(x, y)$ is of the product form (ii), we do *not* have the characterization.This is seen through the following example. Recall that

$$f_{\alpha,\beta,\gamma}(x, y) = C_{\alpha,\beta,\gamma} x^{\alpha - 1} y^{\beta - 1} (1 - x - y)^{\gamma - 1},$$

where $C_{\alpha,\beta,\gamma}$ is a normalizing constant. Here, as always below, $\alpha > 0$, $\beta > 0$, $\gamma > 0$. Fix such an α, β, γ. Fix $k \neq 1$, $k > 0$. Fix $0 < p < 1$ and $q = 1 - p$. Put

$$f(x, y) = p f_{\alpha,\beta,\gamma}(x, y) + q f_{k\alpha, k\beta, k\gamma}(x, y).$$

Then $f(x, y)$ is a density for which conditions (i) and (iii) of Theorem 2 hold. But $f(x, y)$ is *not* of the product form and $f(x, y)$ is *not* Dirichlet. Actually, we can take any nondegenerate probability Π on $(0, \infty)$ and define

$$f(x, y) = \int_0^\infty f_{k\alpha, k\beta, k\gamma}(x, y) \, d\Pi(k).$$

Remark 2. Unless one of the f_i's is a power function, we may again come up with a counterexample as is shown below.

Take any positive continuous function δ on $[0, 1]$, e.g., $\delta(x) = x + x^2$. Put $f(x, y) = k\delta(x) \delta(y) \delta(1 - x - y)$, where k is a normalizing constant. Assuming that $\delta > 0$ a.e., we see that $f(x, y)$ satisfies conditions (i) and (ii) of either theorem. We shall verify that both regressions are indeed linear. If $\int_0^y x\delta(x) \delta(y - x) \, dx = \alpha$, then changing x to $y - x$ and adding the resulting equation to this, we get

$$\int_0^y x\delta(x) \delta(y - x) \, dx = (y/2) \int_0^y \delta(x) \delta(y - x) \, dx$$

or

$$\int_0^{1 - y} x\delta(x) \delta(1 - y - x) \, dx = \{(1 - y)/2\} \int_0^{1 - y} \delta(x) \delta(1 - y - x) \, dx.$$

That is, $E(X \mid Y = y) = (1 - y)/2$.

By symmetry, the other regression is also linear. However, $f(x, y)$ is *not* Dirichlet *unless* $\delta(x)$ is a power function. Of course, here, though $f(x, y)$ is of the product form, none of the f_i's is a power function.

Remark 3. We have not been able to give a *nice* statistical interpretation to condition (iv) of Theorem 2.

4. MULTIDIMENSIONAL CASE

The precise formulation of the result in the k-dimensional case $(k \geqslant 3)$ may be given as follows.

Suppose $(X_1, X_2, ..., X_k)$ is a random k-vector distributed over the k-dimensional simplex $S = \{0 < x_1, x_2, ..., x_k,\ x_1 + x_2 + \cdots + x_k < 1\}$, with the density $f(x_1, x_2, ..., x_k)$. Assume that

  (i) $f(x_1, x_2, ..., x_k) > 0$ a.e. $(x_1, x_2, ..., x_k)$;

  (ii) all the regressions are linear in the conditioning variables;

  (iii) $f(x_1, x_2, ..., x_k)$ is of the product form

$$f(x_1, x_2, ..., x_k) = f_1(x_1)\, f_2(x_2) \cdots f_k(x_k)\, f_{k+1}\left(1 - \sum_1^k x_i\right),$$

where $f_1, f_2, ..., f_{k+1}$ are integrable functions on $(0, 1)$; and

  (iv) at least one f_i is a power function.

Then $f(x_1, x_2, ..., x_k)$ is a Dirichlet density.

An outline of the proof is given here. By the nature of the density, the regression of X_i on $\{X_j, j \neq i\}$ is indeed a function of $\sum_{j \neq i} X_j$. The rest of the proof is clear.

Remark 4. It is natural to inquire if condition (ii) of the above theorem can be replaced by pairwise linearity of regression. The theorem still remains valid if at least one of $f_1, f_2, ..., f_k$ is a power function. What happens if instead f_{k+1} is a power function?

REFERENCES

[1] DARROCH, J. N., AND RATCLIFF, D. (1971). A characterization of the Dirichlet distribution. *J. Amer. Statist. Assoc.* **66**, 641–643.
[2] SINHA, B. K., AND GERIG, T. (1985). On Neyman's conjecture: A characterization of the multinomials. *J. Multivariate Anal.* **16**, 440–450.