JOURNAL OF MULTIVARIATE ANALYSIS 10, 603–610 (1980)

Incomplete Dirichlet Integrals with Applications to Ordered Uniform Spacings*

J. S. RAO AND MILTON SOBEL

Department of Mathematics, University of California at Santa Barbara, Santa Barbara, California 93106

Communicated by the Editors

Sobel, Uppuluri, and Frankowski, (*Selected Tables in Math. Statistics*, Vol. IV. Amer. Math. Soc., Providence, R. I., 1977) consider an incomplete Dirichlet integral of type I with several interesting applications connected with the multinomial distribution and provide tables of this integral along with other useful tables. Two incomplete Dirichlet integrals are discussed here along with some useful recurrence relations, providing simple methods of deriving the distribution theory of ordered uniform spacings.

1. INTRODUCTION

A collection of random variables $(Y_1,..., Y_b)$ is said to have a Dirichlet distribution with parameters $(\nu_1,..., \nu_b; \nu_{b+1})$, written as $D(\nu_1,..., \nu_b; \nu_{b+1})$ if they have the joint density

$$\frac{\Gamma(\nu_1 + \cdots + \nu_{b+1})}{\Gamma(\nu_1) \cdots \Gamma(\nu_{b+1})} \left(\prod_1^b y_i^{\nu_i - 1}\right) (1 - y_1 \cdots - y_b)^{\nu_{b+1} - 1} \qquad (1.1)$$

over the b-dimensional simplex $S_b = \{(y_1,..., y_b): y_i \geqslant 0, \quad i = 1,..., b, \sum_1^b y_i \leqslant 1\}$ and zero outside S_b. This is a b-variate generalization of the familiar beta density. For an elementary exposition on Dirichlet distributions and some properties, see Wilks [11] and for a more detailed discussion as well as tables, see Sobel *et al.* [9]. Incomplete Dirichlet integrals are strongly connected with the multinomial and negative multinomial probabilities and these relations are explored in Olkin and Sobel [7] and Khatri and Mitra [5].

Received July 16, 1979; revised December 4, 1979.

AMS Subject classifications: Primary 62E15; Secondary 62G30.

Key words and phrases: Dirichlet Distribution, Ordered Uniform Spacings, Test Statistics.

* Work supported by the National Science Foundation under Grant MCS78-01881.

0047-259X/80/040603–08$02.00/0

In Section 2 we consider two incomplete integrals based on (1.1) called the $\mathscr{I}$ and J functions and derive simple recurrence relations as well as closed series expansions for them. For the main application in Section 3, let $X_1,\dots,X_n$ denote independent $U(0,1)$ random variables. If $0 \leqslant X_1' \leqslant \cdots \leqslant X_n' \leqslant 1$ denote the order statistics, then the uniform spacings are defined by

$$Y_i = X_i' - X_{i-1}', \qquad i = 1,\dots, n+1, \tag{1.2}$$

with the definition $X_0' = 0$ and $X_{n+1}' = 1$. It is easy to check that the joint distribution of $(Y_1,\dots, Y_n)$ is an n-variate Dirichlet distribution $D(1,\dots, 1; 1)$ with density

$$\begin{aligned} f(y_1,\dots,y_n) &= n! \qquad \text{over the simplex } S_n = \left\{ y_i \geqslant 0, \sum_1^n y_i \leqslant 1 \right\} \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{1.3}$$

This is also the joint density of all the $(n+1)$ spacings which, of course, is degenerate because of the restriction $\sum_1^{n+1} y_i = 1$.

Statistics like the largest spacing, the kth smallest spacing, sum of the k largest spacings, etc., have been used in statistical literature to construct tests of goodness of fit and related hypotheses. See, for instance, Mauldon [6], Darling [3] or Barton and David [1]. Even earlier, Fisher [4] used the distribution of the largest spacing to construct a test of significance of the largest amplitude in harmonic analysis. Rao [8] defined (1-largest gap) as the "circular range" when the observations $(X_1,\dots, X_n)$ are on a circle of unit circumference and proposed a test of uniformity on the circle based on this circular range. Our objective here is to illustrate how one can derive the required distributional results for the ordered spacings using $\mathscr{I}$ and J functions and simple recurrence relations based on them.

2. Incomplete Dirichlet Integral and Recurrence Formulae

Let $(Y_1,\dots, Y_b)$ have Dirichlet distribution $D(1,\dots, 1; n-b+1)$ with $n \geqslant b$. For $0 < p < 1/b$ the probability of the event $\{Y_i > p,\ i = 1,\dots, b\}$ may be represented by the integral

$$J_p^{(b)}(1, n) = \frac{n!}{(n-b)!} \int_p^1 \underset{\{\sum_1^b y_i \leqslant 1\}}{\cdots} \int_p^1 \left(1 - \sum_1^b y_i\right)^{n-b} \prod_1^b dy_i, \tag{2.1}$$

where the expression in braces indicates that the integration is within the simplex $S_b = \{y_i \geqslant 0, \sum_1^b y_i \leqslant 1\}$ in R^b. It is clear that this integral is zero for $p > 1/b$. For $b = 0$, we define $J_p^{(0)}(1, n)$ to be equal to 1 if p is positive and zero otherwise. A recurrence formula for $J_p^{(b)}(1, n)$ is obtained by integrating y_b from p to $1-\sum_1^{b-1} y_i$ and transforming y_i to $(y_i/(1-p))$. Thus

$$J_p^{(b)}(1,n) = \frac{n!}{(n-b+1)!} \int_p^1 \underset{\{\sum_1^{b-1} y_i \leqslant 1-p\}}{\cdots} \int_p^1 \left(1-p-\sum_1^{b-1} y_i\right)^{n-b+1} \prod_1^{b-1} dy_i$$
$$= (1-p)^n J_{p/(1-p)}^{(b-1)}(1,n). \tag{2.2}$$

Note that the region $p/(1-p) > 1/(b-1)$ is the same as $p > 1/b$, where the J-function vanishes. Using the notation $\langle x \rangle = x$ if $x > 0$ and $= 0$ if $x \leqslant 0$, and repeatedly using the relation in (2.2) gives

$$J_p^{(b)}(1,n) = \langle 1 - bp \rangle^n \tag{2.3}$$

since for $0 < p < 1/b$, this iteration gives $(1-bp)^n$ and otherwise we have zero at the outset. It may be pointed out that this particular result (2.3) can also be obtained by combining the results in Theorems 2.1 and 2.2 of Khatri and Mitra [5] which express a somewhat more general incomplete Dirichlet integral in terms of sums of multinomial probabilities.

Similarly, when $(Y_1,..., Y_b)$ have a $D(1,..., 1; n-b+1)$ the probability of the event $\{Y_i < p, i = 1,..., b\}$ for $0 < p \leqslant 1$ is given for $b \leqslant n$ by the incomplete integral

$$\mathscr{I}_p^{(b)}(1,n) = \frac{n!}{(n-b)!} \int_0^p \underset{\{\sum_1^b z_i \leqslant 1\}}{\cdots} \int_0^p \left(1-\sum_1^b z_i\right)^{n-b} \prod_1^b dz_i. \tag{2.4}$$

This integral $\mathscr{I}_p^{(b)}(1, n)$ which we will refer to as the script $\mathscr{I}$-function generalizes the I-function defined in Sobel *et al.* [9, p. 3] in that the value of p here need not be bounded by $1/b$ as is the case in the definition of $I_p^{(b)}(1, n)$. This $\mathscr{I}$-function is related to the J-function defined in (2.1) through the inclusion–exclusion principle. If A_i denotes the event $\{Y_i > p\}$, $i = 1,..., b$, since $\mathscr{I}_p^{(b)}(1, n)$ represents the probability of the complement of $(\bigcup_1^b A_i)$, it is clear that

$$\mathscr{I}_p^{(b)}(1,n) = \sum_{j=0}^{b} (-1)^j \binom{b}{j} J_p^{(j)}(1,n)$$
$$= \sum_{j=0}^{b} (-1)^j \binom{b}{j} \langle 1 - jp \rangle^n \tag{2.5}$$

using (2.3). Note that (2.5) holds for all p although for $p > 1/b$, some of the J-integrals on the right side of (2.5) vanish.

The following recurrence formula on the $\mathscr{I}$-function which reduces a b-dimensional integral to $(b-1)$ dimensions, plays an important role in the numerical evaluations of these functions. Similar relations formed the basis for the numerical computations of the tables in Sobel *et al.* [9].

THEOREM 2.1. *For $b \leqslant n$ and $0 < p < 1$, the integral defined in* (2.4) *satisfies the recurrence formula*

$$\mathcal{J}_p^{(b)}(1, n) = \mathcal{J}_p^{(b-1)}(1, n) - (1-p)^n \mathcal{J}_{p/(1-p)}^{(b-1)}(1, n). \tag{2.6}$$

Proof. From Eq. (2.5) and the recurrence relation (2.2) on the J-function, we have

$$\begin{aligned}
\mathcal{J}_p^{(b)}(1, n) &= \sum_{j=0}^{b} (-1)^j \binom{b}{j} J_p^j(1, n) \\
&= J_p^0(1, n) + \sum_{j=1}^{b-1} (-1)^j \left[\binom{b-1}{j} + \binom{b-1}{j-1}\right] J_p^j(1, n) \\
&\quad + (-1)^b J_p^b(1, n) \\
&= \sum_{j=0}^{b-1} (-1)^j \binom{b-1}{j} J_p^j(1, n) + \sum_{j=1}^{b} (-1)^j \binom{b-1}{j-1} J_p^j(1, n) \\
&= \mathcal{J}_p^{(b-1)}(1, n) + \sum_{j=1}^{b} (-1)^j \binom{b-1}{j-1} (1-p)^n J_{(p/1-p)}^{j-1}(1, n) \\
&= \mathcal{J}_p^{(b-1)}(1, n) - (1-p)^n \mathcal{J}_{(p/1-p)}^{(b-1)}(1, n).
\end{aligned}$$

A more direct (but somewhat longer) proof can be obtained from the definition (2.4) by a careful partial integration with respect to z_b.

3. EXACT DISTRIBUTIONS RELATED TO ORDERED SPACINGS

Let Z_i denote the ith largest among $(Y_1, ..., Y_{n+1})$ for $i = 1, ..., k$ $(1 \leqslant k \leqslant n)$ and $Z_{k+1}, ..., Z_{n+1}$ denote the remaining unordered spacings, all of which are, of course, less than Z_k. From the exchangeability of $(Y_1, ..., Y_{n+1})$ and (1.3), the joint density of $(Z_1, ..., Z_k, Z_{k+1}, ..., Z_n)$ is

$$g(z_1, ..., z_n) = (n!)(n+1)^{(k)} \quad \text{for} \quad z_1 \geqslant z_2 \geqslant \cdots \geqslant z_k,$$
$$z_i \leqslant z_k \quad \text{for} \quad i > k, \quad \sum_1^n z_i \leqslant 1 \tag{3.1}$$

and zero otherwise, where for $r \leqslant n$, we use the descending factorial notation $n^{(r)} = n(n-1) \cdots (n-r+1)$. Let $0 \leqslant a_k \leqslant a_{k-1} \leqslant \cdots \leqslant a_2 \leqslant a_1 \leqslant 1$ denote positive ordered constants with $\sum_1^k a_i \leqslant 1$. It is clear that the density of $(Z_1, ..., Z_k)$ is non-zero only at such values $(a_1, ..., a_k)$. One can obtain the density of $(Z_1, ..., Z_k)$ at such a point $(a_1, ..., a_k)$ from (3.1) by integrating $(z_{k+1}, ..., z_n)$ over the domain $\{(z_{k+1}, ..., z_n)$: $z_i \leqslant a_k$, $i = k+1, ..., n$ and

$1-\sum_1^k a_i - a_k \leqslant \sum_{k+1}^n z_i \leqslant 1-\sum_1^k a_i\}$. This range of integration can be split as the difference of two regions giving the joint density function of $(Z_1,..., Z_k)$, namely,

$$f_{Z_1,...,Z_k}(a_1,..., a_k) = (n!)(n+1)^{(k)} \left\{ \int_0^{a_k} \underset{\{\sum_{k+1}^n z_i \leqslant (1-\sum_1^k a_i)\}}{\cdots} \int_0^{a_k} \prod_{k+1}^n dz_i \right.$$

$$\left. - \int_0^{a_k} \underset{\{\sum_{k+1}^n z_i \leqslant (1-a_k-\sum_1^k a_i)\}}{\cdots} \int_0^{a_k} \prod_{k+1}^n dz_i \right\}. \tag{3.2}$$

Writing $A_k = (1-\sum_1^k a_i)$ and $p_k = a_k/A_k$ and making the transformation $x_i = z_i/A_k$, $i = k+1,..., n$ in the first integral and the transformation $y_i = z_i/(A_k - a_k)$, $i = k+1,..., n$ in the second, (3.2) may be expressed as

$$f_{Z_1,...,Z_k}(a_1,..., a_k) = \frac{n!}{(n-k)!}(n+1)^{(k)}$$

$$\times \{A_k^{n-k} \mathscr{I}_{p_k}^{n-k}(1, n-k) - (A_k - a_k)^{n-k} \mathscr{I}_{p_k/(1-p_k)}^{(n-k)}(1, n-k)\}$$

in terms of the $\mathscr{I}$-function defined in (2.1). Using (2.6) and (2.5), we obtain for $k \leqslant n$

$$\begin{aligned} f_{Z_1,...,Z_k}&(a_1,..., a_k) \\ &= n^{(k)}(n+1)^{(k)} A_k^{n-k} \mathscr{I}_{p_k}^{(n-k+1)}(1, n-k) \\ &= n^{(k)}(n+1)^{(k)} \sum_{j=0}^{\infty} (-1)^j \binom{n-k+1}{j} \langle A_k - ja_k \rangle^{n-k}, \end{aligned} \tag{3.3}$$

whenever $0 < a_k < \cdots < a_1 < 1$, $\sum_1^k a_i \leqslant 1$. It is of interest to ask for the density of the kth largest spacing Z_k or that of the sum of the k largest spacings $S = \sum_1^k Z_i$. Transforming from $(Z_1, Z_2,..., Z_k)$ to $(S, Z_2,..., Z_k)$, we get

$$\begin{aligned} f_{S,Z_2,...,Z_k}&(s, a_2,..., a_k) \\ &= n^{(k)}(n+1)^{(k)} \sum_{j=0}^{n-k+1} (-1)^j \binom{n-k+1}{j} \langle 1-s-ja_k \rangle^{n-k}. \end{aligned}$$

Integrating $a_2,..., a_{k-1}$ in turn, with the limits $a_{i+1} < a_i < (s - \sum_{i+1}^k a_j)/i$ for $i = 2, 3,..., k-1$ (since $s = \sum_1^k z_j \geqslant \sum_{i+1}^k z_j + iz_i$), we get

$$f_{S,Z_k}(s, a_k) = n^{(k)}(n+1)^{(k)} \sum_{j=0}^{n-k+1} (-1)^j$$

$$\times \binom{n-k+1}{j} \langle 1-s-ja_k \rangle^{n-k} \frac{(s-ka_k)^{k-2}}{(k-2)!\,(k-1)!}. \tag{3.4}$$

Now integrating a_k over the limits 0 to s/k and after considerable analysis, one gets for $k \leqslant n$

$$f_S(s) = (n+1)!\, n \sum_{j=q}^{n+1} \frac{(-1)^{k-j+1}}{(j-k)!\,(j-k)^{k-1}\,k!\,k^{n-k}(n-j+1)!} \langle js-k \rangle^{n-1} \tag{3.5}$$

for $k/q < s \leqslant k/(q-1)$, $q = k+1,..., n+1$. Mauldon [6] and Barton and David [1] also obtain this result using different approaches.

On the other hand, integrating (3.4) with respect to s inside the summation over the range $(ka_k, 1-ja_k)$ for any fixed j, we obtain the density of the kth largest spacing Z_k, namely, (for $k \leqslant n$)

$$f_{Z_k}(x) = nk \binom{n+1}{k} \sum_{j=0}^{n-k+1} (-1)^j \binom{n-k+1}{j} \langle 1-(j+k)x \rangle^{n-1} \tag{3.6}$$

for $0 < x < 1/k$. In particular for $k = 1$, the density of the largest spacing is

$$f_{Z_1}(x) = n(n+1) \sum_{j=0}^{n} (-1)^j \binom{n}{j} \langle 1-(j+1)x \rangle^{n-1},$$

with cdf

$$F_{Z_1}(x) = \sum_{j=0}^{\infty} (-1)^j \binom{n+1}{j} \langle 1-jx \rangle^n. \tag{3.7}$$

It can be shown (e.g., through the finite difference operator) that for $k = 1$ the above density and cdf vanish for $x < 1/(n+1)$. These results have a long and interesting history. Stevens [10] considered the following problem in geometric probability. Suppose that $(n+1)$ arcs of equal length x are marked off at random on the circumference of a circle of unit perimeter. What is the probability that these $(n+1)$ arcs will cover the entire circumference and, more generally, that there will be at most $(k-1)$ breaks? The answer to the first question is given by $P(Z_1 < x) = F_{Z_1}(x)$ given in (3.7) while the answer to the latter question is given by the cdf corresponding to the density (3.6). This is because the midpoints of the $(n+1)$ arcs divide the circumference of the unit circle into $(n+1)$ arcs Y_i with joint density (1.3). Clearly there will be no breaks if $\{Z_1 < x\}$ and at most $(k-1)$ breaks if $\{Z_k < x\}$.

Corresponding results on the distribution of small spacings can be derived analogously. If V_i denotes the ith smallest spacing among $(Y_1,..., Y_{n+1})$ for $i = 1,..., k$ and $(V_{k+1},..., V_{n+1})$ the remaining unordered spacings, clearly from (1.3)

$$f_{V_1,...,V_n}(v_1,..., v_n) = n!\,(n+1)^{(k)} \quad \text{for } v_1 < v_2 < \cdots < v_k,$$
$$v_i > v_k \quad \text{for } i = k+1,..., n, \quad \sum^{n} v_i \leqslant 1, \tag{3.8}$$

and zero otherwise. For $k \leqslant n$, let $0 \leqslant v_1 \leqslant \cdots \leqslant v_k \leqslant 1$ denote ordered constants with $\sum_1^k v_i + (n-k+1)v_k \leqslant 1$. The joint density of $(V_1,..., V_k)$ at $(v_1,..., v_k)$ is obtained by integrating $(v_{k+1},..., v_n)$ over the domain $\{(v_{k+1},..., v_n)\colon\ v_i \geqslant v_k,\ i = k+1,..., n \text{ and } 0 \leqslant \sum_{k+1}^n v_i \leqslant 1 - v_k - \sum_1^k v_i\}$. Thus for $k \leqslant n$, using the definition (2.1) and the relation (2.3), this joint density of $(V_1,..., V_k)$ can be written as

$$\begin{aligned} f_{V_1,\ldots,V_k}&(v_1,..., v_k) \\ &= n^{(k)}(n+1)^{(k)} \left(1 - v_k - \sum_1^k v_i\right)^{n-k} J^{(n-k)}_{v_k/(1-v_k-\sum_1^k v_i)}(1, n-k) \\ &= n^{(k)}(n+1)^{(k)} \left(1 - \sum_1^k v_i - (n-k+1)v_k\right)^{n-k}, \end{aligned} \tag{3.9}$$

whenever $v_1 < v_2 < \cdots < v_k$, $\sum_1^k v_i + (n-k+1)v_k \leqslant 1$. The distribution of the sum of the k smallest spacings $S_k^* = \sum_1^k V_i$ as well as that of V_k, the kth smallest spacing can be obtained from (3.9) analogous to the results (3.5) and (3.6). On the other hand, denoting by S_k the sum of the k largest spacings, because of the obvious relations

$$S_k^* = 1 - S_{n+1-k}, \qquad V_k = Z_{n+2-k}, \tag{3.10}$$

they can be derived for $k \geqslant 2$ more directly from (3.5) and (3.6). For instance, from (3.6) and (3.10)

$$\begin{aligned} f_{V_K}(v) &= f_{Z_{n+2-k}}(v) \\ &= nk \binom{n+1}{k} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} \langle 1 - (n+2-k+j)x \rangle^{n-1} \end{aligned} \tag{3.11}$$

for $0 < x < 1/(n+2-k)$. For the special case $k = 1$ we obtain the common result for $S_1^* = V_1$

$$f_{V_1}(v) = n(n+1)[1-(n+1)v]^{n-1}, \qquad 0 < v < (n+1)^{-1}.$$

References

[1] BARTON, D. E., AND DAVID, F. N. (1955). Sums of ordered intervals and distances. *Mathematika* **2**, 150–159.

[2] BLUMENTHAL, S. (1970). Tests of fit based on partial sums of ordered spacings. *Ann. Inst. Statist. Math.* 261–276.

[3] DARLING, D. A. (1953). On a class of problems related to the random division of an interval. *Ann. Math. Statist.* **24**, 239–253.

[4] FISHER, R. A. (1929). Tests of significance in harmonic analysis. *Proc. Roy. Soc. Ser. A* **125**, 54–59.

[5] KHATRI, C. G., AND MITRA, S. K. (1969). Some identities and approximations concerning positive and negative multinomial distributions In *Multivariate Analysis-II* (P. R. Krishnaiah, Ed.), pp. 241–260. Academic Press, New York.

[6] MAULDON, J. A. (1951). Random division of an interval. *Proc. Cambridge Philos. Soc.* **47**, 331–336.

[7] OLKIN, I., AND SOBEL, M. (1966). Integral expressions for tail probabilities of the multinomial and negative multinomial distributions. *Biometrika* **52**, 167–179.

[8] RAO, J. S. (1969). *Some Contributions to the Analysis of Circular Data.* Unpublished Ph. D. Thesis, Indian Statistical Institute, Calcutta.

[9] SOBEL, M., UPPULURI, V. R. R., AND FRANKOWSKI, K. (1977). Dirichlet distributions Type 1, In *Selected Tables in Math. Statistics* Vol. IV Amer. Math. Soc., Providence, R. I.

[10] STEVENS, W. L. (1939). Solution to a geometrical problem in probability *Ann. Eugen.* **9**, 315–320.

[11] WILKS, S. S. (1962). *Mathematical Statistics.* pp. 177–182. Wiley, N. Y.