Multiple Hypergeometric Functions: Probabilistic Interpretations and Statistical Uses

JAMES M. DICKEY*

This article reviews and interprets recent mathematics of special functions, with emphasis on integral representations of multiple hypergeometric functions. B.C. Carlson's centrally important parameterized functions R and $\mathcal{R}$, initially defined as Dirichlet averages, are expressed as probability-generating functions of mixed multinomial distributions. Various nested families generalizing the Dirichlet distributions are developed for Bayesian inference in multinomial sampling and contingency tables. In the case of many-way tables, this motivates a new generalization of the function $\mathcal{R}$. These distributions are also useful for the modeling of populations of personal probabilities evolving under the process of inference from statistical data. A remarkable new integral identity is adapted from Carlson to represent the moments of quadratic forms under multivariate normal and, more generally, elliptically contoured distributions. This permits the computation of such moments by simple quadrature.

KEY WORDS: Generalized Dirichlet distributions; Multiple hypergeometric functions; Special functions; Carlson's R and $\mathcal{R}$; Multivariate distributions; Bayesian inference; Multinomial sampling; Contingency tables; Populations of personal probabilities; Generalized mean value; Moments of quadratic forms.

1. INTRODUCTION

Statisticians should keep informed of developments in the mathematics of special functions. This is true for applied mathematicians generally, but particularly so for statisticians, who make heavy use of parameterized families of probability distributions. Such families are badly needed in multivariate contexts, where joint distributions are required to model a wide variety of uncertainty relations. Many special functions give rise naturally to parameterized distributions under the following scheme.

A function $F(\mathbf{b}; \mathbf{z})$ may have an integral representation,

$$F(\mathbf{b}; \mathbf{z}) = \int g(\mathbf{b}, \mathbf{z}; \mathbf{u}) d\mathbf{u}, \tag{1.1}$$

where $g \geq 0$. Then one can simply define the probability density for a random vector $\mathbf{u}$ by

$$p(\mathbf{u}; \mathbf{b}, \mathbf{z}) = g(\mathbf{b}, \mathbf{z}; \mathbf{u})/F(\mathbf{b}; \mathbf{z}). \tag{1.2}$$

Note that $\mathbf{u}$ enters the representation (1.1) merely as the dummy variable of integration. The parameters and arguments of F have now become parameters of p. If the integrand g contains an arbitrary power of u as a factor, then the moments of p take a simple form as ratios of special functions F. For example, from $F(z) = \Gamma(z)$ we obtain the familiar gamma distribution $p(u; z) = u^{z-1}e^{-u}/\Gamma(z)$, $u, z > 0$, having moments $Eu^y = \Gamma(z + y)/\Gamma(z)$.

Integral and series representations of special functions can be important as characteristic functions of distributions, moment-generating functions, and probability-generating functions (Johnson and Kotz 1969,1970ab,1972). Relations satisfied by the special functions then provide iterative and other methods for calculating entities of statistical interest (e.g., $Eu^y = z(z + 1) \cdots (z + y - 1)$.) .

Efforts have been made over the years to unify the diverse field of special functions, as in early approaches based on hypergeometric series (Rainville 1960, Truesdell 1948), or the more modern matrix representations of Lie groups (Vilenkin 1968). A new analytic approach by B.C. Carlson (1977) treats special functions as averages of elementary functions. Carlson's work is based, to a large extent, on his homogeneous multiple hypergeometric function, denoted by R (1963). This is a reformulation of Appell's and Lauricella's functions, themselves generalizations of Gauss's well-known hypergeometric series.

We review relevant results on R and on its two-way generalization $\mathcal{R}$ and give probabilistic interpretations and statistical applications. We begin (Section 2) by setting up notation for the expectation operator of the Dirichlet distribution, relating it to the conjugate Bayesian inference for multiple-valued Bernoulli (multinomial) sampling, and providing a method for calculation of the associated Bayesian predictive-probability mass function as the Dirichlet mixed moment. We introduce the functions R and $\mathcal{R}$ as moments of linear forms in Dirichlet variables (Sections 3, 6, and 7) and then exhibit them as generating functions of the aforesaid predictive probabilities and corresponding characteristic function. Picard's identity and its generalization express R and $\mathcal{R}$, respectively, as lower-dimensional integrals. Carlson's hypergeometric mean value extends the more familiar Hardy-Littlewood-Polya family of homogeneous mean values, which includes the harmonic, geometric, and usual arithmetic mean. We give a remarkable new one-

* James M. Dickey is Professor, Department of Mathematics and Statistics, State University of New York at Albany, Albany, NY 12222. The author is grateful to B.C. Carlson, B.K. Shah, and an editor for references to the literature, and to J.-M. Jiang for informing him of errors in a previous version of this article.

© Journal of the American Statistical Association
September 1983, Volume 78, Number 383
Theory and Methods Section

dimensional integral representation for the moments of quadratic forms under multivariate normal distributions and more generally elliptically contoured distributions (Section 4).

The scheme of (1.1) and (1.2) is carried out for R and $\mathcal{R}$ to yield nested families of distributions generalizing the Dirichlet (Sections 5, 8, 9, and 10). The closure properties of these families are established in two contexts under operations of updating to account for statistical data. First, the distribution is considered as an ordinary Bayesian prior distribution; and second, it is used to model the variability of personal opinions over a population of persons. The theory for the first operation, of course, provides Bayesian inference tools for multiple-valued Bernoulli sampling.

We treat the case of Bayesian inference for contingency tables (Sections 7, 8, 9, and 10). A new family of prior distributions is proposed for the model of independence of sampling between row and column outcomes, providing prior dependence between the unknown row and column sampling probabilities. This family is extended to many-way tables, thereby motivating our conclusion (Section 10), in which we define a new generalization of $\mathcal{R}$. A many-way array generalizing the notion of a matrix serves as the argument variable for the new function.

2. DIRICHLET MOMENTS AND TWO FAMILIES OF PROBABILITY MASS FUNCTIONS

For a fixed integer $K \geq 2$, denote the Dirichlet family of probability distributions $D(\mathbf{b})$, parameterized by the vector $\mathbf{b} = (b_1, \ldots, b_K)$. This is the multivariate generalization of the beta family ($K = 2$) defined on the probability simplex of K coordinate variables $\mathbf{u} = (u_1, \ldots, u_K)$, where each $u_i \geq 0$ and $u. = 1$, where $u.$ denotes $\sum_1^K u_i$. There are only $K - 1$ free variables, and the density expressed in terms of any set consisting of all but one of the K variables takes the identical form,

$$p(\mathbf{u}) = B(\mathbf{b})^{-1} \prod_{i=1}^{K} u_i^{b_i - 1}, \tag{2.1}$$

in terms of Dirichlet's complete integral of the K-fold product in (2.1), $B(\mathbf{b}) = [\prod \Gamma(b_i)]/\Gamma(\sum b_j)$. (Product and summation signs are understood to extend over the full range of index variables unless otherwise indicated.) The random quantities u_i are nearly independent, with a slight negative association, from the constraint on their sum. This is intuitively clear from the familiar representation $u_i = x_i/x.$, $x. \equiv \sum x_i$, where the x_i's are independent chi-squared random variables on $2b_i$ degrees of freedom. For a leisurely development of the Dirichlet family and properties, see Wilks (1962).

The reader may recall that the Dirichlet distributions comprise the Bayesian conjugate prior family for the multiple-valued Bernoulli sampling model (Good 1965). If y_1, $y_2, \ldots$ are independently distributed each according to a common finite distribution with unknown atomic probabilities $u_1, \ldots, u_K$, then the vector formed from a length-n sequence $\mathbf{y} = (y_1, \ldots, y_n)$ (n chosen by a "noninformative" process) has sampling probability mass $pr(\mathbf{y} \mid \mathbf{u}) = \prod u_i^{m_i}$, in terms of the frequency counts $\mathbf{m} = (m_1, \ldots, m_K)$, where $m. = n$. We write $\mathbf{y} \mid \mathbf{u} \sim \text{Ber}(n, \mathbf{u})$ to mean that $\mathbf{y}$ given $\mathbf{u}$ has the above sampling distribution. Then the prior distribution $\mathbf{u} \sim D(\mathbf{b})$ implies the posterior distribution,

$$\mathbf{u} \mid \mathbf{y} \sim \mathbf{u} \mid \mathbf{m} \sim D(\mathbf{b} + \mathbf{m}). \tag{2.2}$$

We shall say the family $\{D\}$ is closed under sampling in the sense of personal updating, to distinguish this from the effect of data on distributions used later to model populations of opinion.

Of course, the frequency counts $\mathbf{m}$, themselves, have a multinomial sampling distribution, say $\mathbf{m} \mid \mathbf{u} \sim M(n, \mathbf{u})$, with mass $pr(\mathbf{m} \mid \mathbf{u}) = \binom{m.}{\mathbf{m}} \prod u_i^{m_i}$, and as sufficient statistics, they yield the same inference as $\mathbf{y}$. (The usual Bayesian notational practice is followed here whereby a generic notation, p or pr is used for a probability density or mass, and the argument variable indicates the distribution intended.)

The Bayesian prior-predictive distribution, the marginal distribution of vector $\mathbf{y}$, is obtained as the Dirichlet mixture of the Bernoulli sampling distribution, $pr(\mathbf{y}) = E_{\mathbf{u}|\mathbf{b}}[pr(\mathbf{y} \mid \mathbf{u})]$. The probability mass of this distribution is the same as the Dirichlet moment, one representation of which is immediate from the form of the Dirichlet density,

$$pr(\mathbf{y}) = E_{u|b}{}^{(K-1)} \prod u_i^{m_i} = B(\mathbf{b} + \mathbf{m})/B(\mathbf{b}). \tag{2.3}$$

(For emphasis we explicitly indicate the dimensionality of integral averages, here $K - 1$.) Call this the *Dirichlet-Bernoulli* distribution, parameterized by $\mathbf{b}$, and write $\mathbf{y} \sim DB(\mathbf{b})$. Under $DB(\mathbf{b})$, the coordinates of $\mathbf{y}$ are not independent, but they are exchangeable, that is, permutation symmetric.

The corresponding predictive distribution for the frequencies $\mathbf{m}$, the *Dirichlet-Multinomial* distribution, has probability mass $pr(\mathbf{m}) = \binom{m.}{\mathbf{m}} pr(\mathbf{y}) = \binom{m.}{\mathbf{m}} E_{\mathbf{u}|\mathbf{b}}{}^{(K-1)} \prod u_i^{m_i}$. In this case we write $\mathbf{m} \sim DM(n, \mathbf{b})$. Subsets of coordinates of $\mathbf{m}$ are exchangeable if the respective parameters b_i are equal. Raiffa and Schlaifer (1961) treat this family in the case $K = 2$, under the name Beta-binomial distribution. Of course, each of the predictive families $DB(\mathbf{b})$ and $DM(n, \mathbf{b})$ contains its associated sampling model as the limiting special case of prior certain knowledge regarding $\mathbf{u}$: $\mathbf{b} \to (\infty, \ldots, \infty)$, $\mathbf{b}/b.$ fixed, where $b.$ denotes $\sum b$.

A new expression for the Dirichlet moment yields an interpretation as a product of successive posterior means. We have

$$E_{\mathbf{u}|\mathbf{b}}{}^{(K-1)} \prod u_i^{m_i} = \frac{\prod (b_i, m_i)}{(\sum b_i, \Sigma m_i)} \tag{2.4}$$

in Appell's notation, $(b, m) = \Gamma(b + m)/\Gamma(b) = b(b + 1) \cdots (b + m - 1)$. Note that the new numerator and denominator have the same number $m.$ of factors, and thus we have a product of $m. = n$ ratios. Hence, obtain the following result, important for practical computations. (This result corresponds to the known equivalence of the Dirichlet-Multinomial and the Polya urn models, cf Feller 1966, p. 226. See also Hill 1974, p. 1024.)

Theorem 2.1. If the product $\prod u_i^{m_i}$ under the expectation operator in the Dirichlet mixed moment, (2.3), (2.4), is written in any order whatever as a product of $m.$ individual u_i's to first powers, then the mixed moment can be calculated by merely substituting for each factor its successive posterior mean based on the preceding factors as "data." Writing $\prod u_i^{m_i} = u_{y_1} u_{y_2} \cdots u_{y_n}$, for any fixed vector y with frequency counts $m(m. = n)$, yields

$$E_{\mathbf{u}|\mathbf{b}}^{(K-1)} \prod u_i^{m_i} = (E_{\mathbf{u}|\mathbf{b}}\, u_{y_1})(E_{\mathbf{u}|\mathbf{b},y_1}\, u_{y_2})(E_{\mathbf{u}|\mathbf{b},y_1,y_2}\, u_{y_3} \ldots (E_{\mathbf{u}|\mathbf{b},y_1,\ldots,y_{n-1}}\, u_{y_n}), \quad (2.5)$$

in which each $E_{\mathbf{u}|\mathbf{b}^*}\, u_y = b_y^*/b.^*$, where $\mathbf{b}^*$ denotes the Dirichlet parameter posterior (2.2) to the data given, so far, in the condition of the respective factor. Each new jth factor is both the posterior mean of the respective u_i and the posterior predictive probability for the event "observation" $y_j = i$.

For example, for $m. = 3$,

$$E_{\mathbf{u}|\mathbf{b}}^{(K-1)}\, u_1^2 u_2 = \frac{b_1}{b.}\, \frac{b_1+1}{b.+1}\, \frac{b_2}{b.+2} = \frac{b_1}{b.}\, \frac{b_2}{b.+1}\, \frac{b_1+1}{b.+2}. \quad (2.6)$$

This result motivates our new notation,

$$E_{\mathbf{u}|\mathbf{b}}^{(K-1)} \prod u_i^{m_i} = [\mathbf{b}:b.]^{\mathbf{m}}, \quad (2.7)$$

whereby

$$pr(\mathbf{y}) = [\mathbf{b}:b.]^{m} \quad \text{and} \quad pr(\mathbf{m}) = \binom{m.}{\mathbf{m}} [\mathbf{b}:b.]^{\mathbf{m}}.$$

Several new families that generalize the Dirichlet and are useful as prior distributions are given below. Theorem 2.1 will be applicable in two ways to the new distributions. First, expressions for the moments of the new distributions will involve the Dirichlet moment (2.7), which can be calculated by Theorem 2.1. Secondly, obvious modifications of the theorem apply directly to the new distributions in place of the Dirichlet.

3. THE HYPERGEOMETRIC FUNCTIONS R

Carlson (1977) defines the transform $F(\mathbf{z})$ of an arbitrary function of a single variable $f(x)$ by performing a Dirichlet average through the argument

$$F(\mathbf{b}, \mathbf{z}) = E_{\mathbf{u}|\mathbf{b}}^{(K-1)} f(\mathbf{u} \cdot \mathbf{z}), \quad (3.1)$$

where $\mathbf{z} = (z_1, \ldots, z_K)$, $\mathbf{u} \cdot \mathbf{z} = u_1 z_1 + \cdots + u_K z_K$, and $\mathbf{u} \sim D(\mathbf{b})$. The new function F of K variables with K parameters is an average of $f(x)$ over the interval [min z_i, max z_i]. See Figure 1 for $K = 2$. Carlson considers

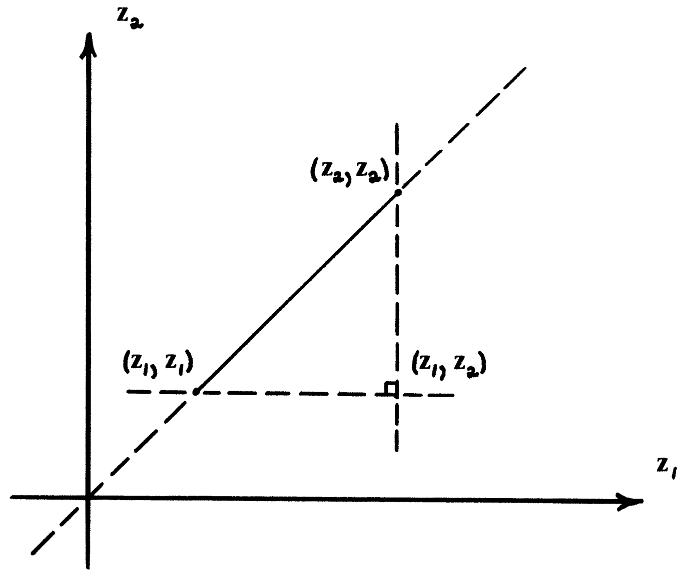

Figure 1. Carlson's F as an Average of f over a Segment of the Equiangular Line. Case K = 2.

complex-valued functions f and complex coordinates b_i and z_i, but here we treat, for the most part, merely real functions and variables. If each $b_i > 0$ and if the function f is C^n (nth derivative continuous) over an open interval I, then F will also be C^n over the cube I^K. Clearly, by the obvious marginalization property of the Dirichlet distribution, $(u_1, \ldots, u_{J-1}, \sum_J^K u_i) \sim D(b_1, \ldots, b_{J-1}, \sum_J^K b_i)$, and by the invariance to simultaneous reorderings of $\mathbf{u}$ and $\mathbf{b}$ coordinates, we have a lowering in dimensionality of F whenever a subset of equal coordinates of $\mathbf{z}$ occurs. For example, $F(\mathbf{b}; x, x, \ldots, x) = f(x)$.

The transform F and its confluent and other forms provide derivations for a wide variety of known special functions from familiar elementary functions. A central role is played by the particular case of the transform of a simple power $f(x) = x^a$. Define for $\mathbf{u} \sim D(\mathbf{b})$ and each $b_i > 0$,

$$R_a(\mathbf{b}, \mathbf{z}) = E_{\mathbf{u}|\mathbf{b}}^{(K-1)} (\mathbf{u} \cdot \mathbf{z})^a. \quad (3.2)$$

One reason for the interest in R has been that power series for f yield, by term-by-term integration, expansions of F as a series of R functions. Expansions of the integrand of R (3.2) and term-by-term integration yield the following representations for R.

Theorem 3.1 (Polynomial form for R). For nonnegative integer n, we obtain the *probability generating function* of the Dirichlet-Multinomial distribution,

$$R_n(\mathbf{b}, \mathbf{z}) = \sum_{\mathbf{m}|m.=n} \binom{m.}{\mathbf{m}} [\mathbf{b}:b.]^{\mathbf{m}} \prod z_i^{m_i} = E_{\mathbf{m}|n,\mathbf{b}} \prod z_i^{m_i}, \quad (3.3)$$

where $\mathbf{m} \sim DM(n, \mathbf{b})$.

Corollary 3.2. A complex argument provides the *characteristic function* of the Dirichlet-Multinomial, $R_n(\mathbf{b}, \mathbf{z})$

for $\mathbf{z} = (\exp(\mathbf{i}t_1), \ldots, \exp(\mathbf{i}t_K))$. Recall that the Multinomial characteristic function is the power

$$[u_1 \exp(\mathbf{i}t_1) + \cdots + u_K \exp(\mathbf{i}t_K)]^n.$$

Equation (3.3), which holds for restricted $a = n$, gives a polynomial homogeneous in the vector $\mathbf{z}$. It shows that the definition of R can be extended to nonpositive parameter values b_i, where the integral representation (3.2) fails to exist. Carlson gives a thorough treatment of the valid ranges for parameters and arguments of R. From the binomial-series expansion of $(1 - x)^a$, obtain the following representation exhibiting R as a multiple hypergeometric series (and thereby further motivating the function $f(x) = z^a$).

Corollary 3.3 (Series form). For each $|z_i| < 1$, with $\mathbf{1} = (1, \ldots, 1)$, we have the multiple series,

$$R_a(\mathbf{b}, \mathbf{1} - \mathbf{z}) = \sum_{n=0}^{\infty} \frac{(-a, n)}{n!} R_n(\mathbf{b}, \mathbf{z})$$

$$= \sum_{m_1=0}^{\infty} \cdots \sum_{m_K=0}^{\infty} (-a, m.)[\mathbf{b}:b.]^{\mathbf{m}} \prod \frac{z_i^{m_i}}{m_i!}. \tag{3.4}$$

Of course, the series terminates for nonnegative integer a.

Depending on the dimension K, the function R becomes a special form of Lauricella's F_D, Appell's F_1, or Gauss's familiar ${}_2F_1$. The following classical identity, attributed to Picard by Appell and Kampé de Fériet (1926), reduces the dimension of the integral representation, thus permitting the computation of R by simple quadrature. Although this refers only to a restricted range of the parameter a, a contour integral applies more generally (Carlson 1977, Theorem 6.8-2, p. 155). The author is presently developing computer algorithms for calculation of R and its generalizations in parameter ranges important for statistics.

Theorem 3.4 (Picard's identity). For $-\sum b_i < a < 0$, $\mathbf{v} = (v_1, v_2)$, $v_1 = w$, $v_2 = 1 - w$, and $\mathbf{c} = (a, \sum b_i + a)$,

$$R_a(\mathbf{b}, \mathbf{z}) = E_{v|c}{}^{(1)} \prod_{i=1}^{K} (v_1 z_i + v_2)^{-b_i}$$

$$= B(\mathbf{c})^{-1} \int_0^1 w^{-a-1} (1 - w)^{\Sigma b_j + a - 1} \times \left[\prod_1^K (wz_i + 1 - w)^{-b_i}\right] dw. \tag{3.5}$$

The distributions of linear combinations $\mathbf{u} \cdot \mathbf{z}$ of Dirichlet coordinates have been investigated by Bloch and Watson (1967). The functions R_n now give the moments of such a combination.

We finish this section by presenting Carlson's (1965) beautiful generalized mean value. For real parameters a, c, and a probability vector of "weights" $\mathbf{w}$, define the *hypergeometric mean value* of the positive quantities z_i, where $\mathbf{z} = (z_1, \ldots, z_K)$, as

$$M(a, c; \mathbf{z}, \mathbf{w}) = [R_a(c\mathbf{w}, \mathbf{z})]^{1/a}. \tag{3.6}$$

This extends, as follows, the more familiar Hardy, Littlewood, and Polya (1959) generalized mean value, $(w_1 z_1{}^a + \cdots + w_K z_K{}^a)^{1/a}$, of which the usual arithmetic, geometric, and harmonic means are special cases. Both generalized mean values are homogeneous in the vector $\mathbf{z}$. The familiar monotonicity and convexity properties are carried over to M. For example, M is increasing in a if c is positive.

Theorem 3.5.

$$\lim_{c \to 0} M = (w_1 z_1{}^a + \cdots + w_K z_K{}^a)^{1/a} \tag{3.7}$$

and

$$\lim_{c \to \infty} M = w_1 z_1 + \cdots + w_K z_K. \tag{3.8}$$

Proof. Note that as $c \to 0$, the Dirichlet distribution approaches the finite distribution with respective probabilities w_i over the set of vertices $\mathbf{V}_i = (\mathbf{u}: u_i = 1$ and $u_j = 0$ for $j \neq i)$, $i = 1, \ldots, K$. The effect of $b_i \to 0$ for a single value of i, only, would be to pull the probability away from $\mathbf{V}_i$, up against the boundary subsegment $\{\mathbf{u}: u_i = 0\}$; but since this must happen simultaneously in all $i = 1, \ldots, K$, the result is a sharing out of the limiting probability among the segment-intersection points $\mathbf{V}_i$. Also, as $c \to \infty$, all the Dirichlet probability accumulates to the single point $\mathbf{u} = \mathbf{w}$.

4. MOMENTS OF QUADRATIC FORMS

The following result regarding the distribution of centered quadratic forms was given in a nonprobabilistic form for multiple integrals by Carlson (1972). We state the result first for a multivariate normal vector.

Theorem 4.1. Consider a K-variate normal random vector $\mathbf{x}$, $\mathbf{x} \sim \text{Normal}^{(K)}(\mathbf{0}, V)$; consider the matrix A $(K \times K)$, symmetric and positive definite; and let $n > -K/2$. Then

$$E^{(K)}[(\mathbf{x}^T A \mathbf{x})^n] = E(Q^n) \cdot R_n(\tfrac{1}{2}, \ldots, \tfrac{1}{2}; \lambda_1, \ldots, \lambda_K), \tag{4.1}$$

where the λ_i's are the eigenvalues of the matrix AV, and where $Q \sim \chi_K{}^2$, so that

$$E(Q^n) = \Gamma(\tfrac{1}{2}K + n)/[\Gamma(\tfrac{1}{2}K)(\tfrac{1}{2})^n]. \tag{4.2}$$

This, in conjunction with Theorem 3.4, provides a new one-dimensional integral representation for the moments of a quadratic form. Previous work has concentrated on series expansions of the distribution and moment (see, e.g., Kotz, Johnson, and Boyd 1967 and references there). (This is also true of the more recent work on quadratic forms with matrix argument, e.g., Khatri 1966, Shah and Kahtri 1974.)

The proof of Theorem 4.1 is based on the following

Lemma from Carlson (1977), which yields also an alternative definition for R_n ($n = 0, 1, 2, \ldots$) applying with no restriction whatever on (complex valued) arguments $\mathbf{b} = (b_1, \ldots, b_K)$ and $\mathbf{z} = (z_1, \ldots, z_K)$.

Lemma 4.2 (Generating function for R_n). If each $| tz_i | < 1$ ($i = 1, \ldots, K$),

$$\prod_{i=1}^{K} (1 - tz_i)^{-b_i} = \sum_{n=0}^{\infty} [\Gamma(b. + n)/\Gamma(b.)] R_n(\mathbf{b}, \mathbf{z}) \frac{t^n}{n!}. \quad (4.3)$$

Proof of Theorem 4.1. The moment-generating function of $\mathbf{x}^T A \mathbf{x}$ is defined as

$$E \exp(\tfrac{1}{2}t(\mathbf{x}^T A \mathbf{x})) = \sum E[(\mathbf{x}^T A \mathbf{x})^n] 2^{-n} \frac{t^n}{n!}, \quad (4.4)$$

the left side of which is equal to

$$(2\pi)^{-(1/2)K} | V^{-1} |^{1/2} \int \exp(-\tfrac{1}{2}\mathbf{x}^T(V^{-1} - tA)\mathbf{x}) d\mathbf{x}$$

$$= | V^{-1}|^{1/2} / | V^{-1} - tA |^{1/2} = \prod_{1}^{K} (1 - t\lambda_i)^{-1/2}. \quad (4.5)$$

Comparison of the resulting equation to (4.3) yields the theorem.

Note that Q in Theorem 4.1 has the same distribution as $\mathbf{x}^T V^{-1} \mathbf{x}$, the maximal invariant of the symmetry group for the distribution of $\mathbf{x}$. This observation leads to the following generalization.

Theorem 4.3 (Arbitrary elliptically symmetric distribution). Let matrices A and B ($K \times K$) be positive-definite and symmetric. Assume that $\mathbf{x}$ has a distribution that depends on $\mathbf{x}$ only through

$$Q = \mathbf{x}^T B \mathbf{x}. \quad (4.6)$$

That is, the density of $\mathbf{x}$ has contours in the form of concentric ellipsoids defined by fixing values Q. Then if the nth absolute moment of Q exists, Equations (4.1) and (4.6) hold simultaneously, with the λ_i's defined as the eigenvalues of the matrix AB^{-1}.

This says that the nth moments of $\mathbf{x}^T A \mathbf{x}$ and $\mathbf{x}^T B \mathbf{x}$ have the ratio $R_n(\frac{1}{2}, \ldots, \frac{1}{2}, \lambda_1, \ldots, \lambda_K)$. To compute the moments of $Q = \mathbf{x}^T B \mathbf{x}$, it may be helpful to have its density. For $p(\mathbf{x}) = h(Q)$, the density of Q takes the form

$$[\pi^{(1/2)K}/\Gamma(\tfrac{1}{2}K)] \, | B |^{-\frac{1}{2}} Q^{(1/2)K-1} h(Q). \quad (4.7)$$

5. EXTENSIONS OF THE DIRICHLET DISTRIBUTIONS

The function R is associated with various extensions of the Dirichlet family of distributions. If the Dirichlet coordinate variables v_i, where $\mathbf{v} \sim D(\mathbf{b})$, are transformed by scaling and renormalizing to sum to unity,

$$u_i = v_i z_i^{-1} / \sum v_j z_j^{-1} \; (i = 1, \ldots, K), \quad (5.1)$$

where each $z_i > 0$, then the new vector $\mathbf{u}$ has the probability density for any subset of $K - 1$ coordinates,

$$B(\mathbf{b})^{-1} (\prod u_i^{b_i - 1})(\mathbf{u} \cdot \mathbf{z})^{-b.} \prod z_i^{b_i}, \quad (5.2)$$

where each $u_i \geq 0$ and $\sum u_i = 1$. (The Jacobian for this change of variable is $| \partial \mathbf{v}/\partial \mathbf{u} | = (\prod z_i)(\mathbf{u} \cdot \mathbf{z})^{-K}$.) In this case we write $\mathbf{u} \sim S(\mathbf{b}, \mathbf{z})$. We shall say that the distributions D transform to S under the operation of population updating.

These new density functions are homogeneous of degree zero in the parameter vector $\mathbf{z}$, $S(\mathbf{b}, c\mathbf{z}) \sim S(\mathbf{b}, \mathbf{z})$ for any $c > 0$; and the Dirichlet family is a special case, $D(\mathbf{b}) \sim S(\mathbf{b}, c\mathbf{1})$ where $\mathbf{1} = (1, \ldots, 1)$. These distributions were introduced by Dickey (1968a) and credited there to private conversations with L.J. Savage.

Geometrically, the transformation (5.1) says that the $\mathbf{v}$ simplex is first displaced and extended or contracted by the pure linear scaling operation in the numerator of (5.1), in which the vertices track along their respective axes. In this motion, all points in the simplex retain their proportional positions and their relative probability density ordinates. Then, the whole probability mass is returned onto the original simplex by a stereographic projection with the origin as reference point. This process is illustrated in Figure 2 for $K = 2$.

The class of transformations (5.1) is closed under composition. Indeed, it forms a group isomorphic to R^K/G where R denotes the group of positive real numbers under multiplication and G is the subgroup consisting of vectors having all K coordinates identical. The inverse of the transformation (5.1) is just $v_i = u_i z_i / \sum u_j z_j$. For the family of densities, also, we have closure: if $\mathbf{v} \sim S(\mathbf{b}, \mathbf{y})$ then for the change of variable (5.1), $\mathbf{u} \sim S(\mathbf{b}, \mathbf{z} \times \mathbf{y})$, where $\mathbf{z} \times \mathbf{y}$ denotes the vector $(z_1 y_1, \ldots, z_K y_K)$.

Equation (5.1) has an obvious resemblance to Bayes's formula. Dickey (1968a) and Dickey and Freeman (1975) studied the following model in which (5.1) is interpreted as the transformation of prior to posterior probability vec-

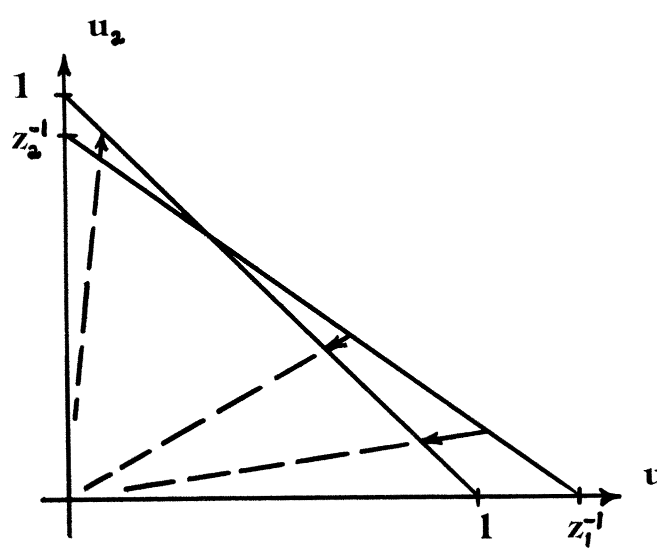

Figure 2. Savage's Transformation of the Probability Simplex. Case K = 2.

tors for a population of Bayesian scientists. If, for a specified scientist, the unknown state of nature θ ($\theta = 1, 2, \ldots, K$) has his personal prior probability $v_i = \Pr\{\theta = i\}$ ($i = 1, \ldots, K$), and if each scientist in the population observes the same experimental data with common likelihood function $l_i = pr(\text{data} \mid \theta = i)$ ($i = 1, \ldots, K$), then each scientist's coherent personal posterior probability vector $\mathbf{u}$, $u_i = \Pr\{\theta = i \mid \text{data}\}$, will satisfy equation (5.1) with $z_i^{-1} = l_i$ ($i = 1, \ldots, K$). Then to model the population of scientists' prior opinions by a Dirichlet distribution of probability vectors, $\mathbf{v} \sim D(\mathbf{b})$, will induce the population distribution of coherent posterior opinions, $\mathbf{u} \sim S(\mathbf{b}, \mathbf{l}^{-1})$, where $\mathbf{l}^{-1} = (l_1^{-1}, \ldots, l_K^{-1})$. More generally, we have the following statement.

Theorem 5.1. The family $\{S\}$ is closed under the operation of updating opinion populations. If before the data, $\mathbf{v} \sim S(\mathbf{b}, \mathbf{z})$, then after the data of likelihood $\mathbf{l}$, $\mathbf{u} \sim S(\mathbf{b}, \mathbf{z} \times \mathbf{l}^{-1})$ where $\mathbf{z} \times \mathbf{l}^{-1} = (z_1 l_1^{-1}, \ldots, z_K l_K^{-1})$.

The following consequence of the form of the density (5.2) will be important in the sequel.

Theorem 5.2 (Simple form for R). Under the parameter restriction $a = -b.$,

$$R_{-b.}(\mathbf{b}, \mathbf{z}) = \prod z_i^{-b_i}. \quad (5.3)$$

The moments of $\mathbf{u} \sim S(\mathbf{b}, \mathbf{z})$ are easily obtained as follows:

$$E_{u|b,z}^{(K-1)} (\prod u_i^{m_i})(\mathbf{u} \cdot \mathbf{z})^c = [\mathbf{b}:b.]^{\mathbf{m}} R_{-b.+c}(\mathbf{b} + \mathbf{m}, \mathbf{z}) \prod z_i^{b_i}. \quad (5.4)$$

For example, under the restriction $c = -m.$,

$$E(\prod u_i^{m_i})(\mathbf{u} \cdot \mathbf{z})^{-m.} = E \prod (v_i/z_i)^{m_i} = [\mathbf{b}:b.]^{\mathbf{m}} \prod z_i^{-m_i}. \quad (5.5)$$

In contrast to Theorem 5.1 and to previous property (2.2), the family $\{S\}$ is not closed under sampling in the personal-updating sense, that is, when considered as the source of a prior distribution modeling a single person's opinion. This motivates the following more general family, which *is* closed under personal-updating operations (but not, however, under population-updating operations).

Write $\mathbf{u} \sim D^1(\mathbf{b}, \mathbf{z}, \beta)$ when $\mathbf{u}$ has the density,

$$B(\mathbf{b})^{-1}(\prod u_i^{b_i - 1})(\mathbf{u} \cdot \mathbf{z})^{-\beta}/R_{-\beta}(\mathbf{b}, \mathbf{z}). \quad (5.6)$$

The moments are ratios of hypergeometric functions,

$$E_{u|b,z,\beta}^{(K-1)} (\prod u_i^{m_i})(\mathbf{u} \cdot \mathbf{z})^c = [\mathbf{b}:b.]^{\mathbf{m}} R_{-\beta+c}(\mathbf{b} + \mathbf{m}, \mathbf{z})/R_{-\beta}(\mathbf{b}, \mathbf{z}). \quad (5.7)$$

Theorem 5.3. The family $\{D^1\}$ is closed in the personal-updating sense. If $\mathbf{u} \sim D^1(\mathbf{b}, \mathbf{z}, \beta)$, then

$$\mathbf{u} \mid \mathbf{m} \sim D^1(\mathbf{b} + \mathbf{m}, \mathbf{z}, \beta). \quad (5.8)$$

Note that $S(\mathbf{b}, \mathbf{z}) \sim D^1(\mathbf{b}, \mathbf{z}, b.)$, thus motivating a new notation for S, $D_0^1(\mathbf{b}, \mathbf{z}) \sim D^1(\mathbf{b}, \mathbf{z}, b.)$. Property (5.8)

Table 1. Nesting and Closure Properties of the D^κ Distributions. Dimension K Arbitrarily Fixed ($\{D^0\} = \{D\}$, $\{D_0^1\} = \{S\}$)

	Family					
	Dirichlet	Generalizations Based on R		Generalizations Based on $\mathcal{R}$		
Nesting	$\{D^0\} \subset$	$\{D_0^1\} \subset$	$\{D^1\} \subset$	$\{D_0^2\} \subset$	$\{D^2\} \subset$	$\cdots$
Closure						
Personal-Updating	Y	N	Y	N	Y	$\cdots$
Population-Updating	N	Y	N	Y	N	$\cdots$

then exhibits the lack of personal-sense closure for $\{D_0^1\}$. The situation regarding subfamily relations and personal- and population-sense closure for these families of distributions is recorded in Table 1. The further families referred to there will be introduced in a later section, partially motivated by closure considerations.

Neither $D^0(\sim D)$, D_0^1 $(\sim S)$, D^1 nor any of the other distributions to be given here provides a solution to the outstanding Bayesian need for a convenient prior distribution having locally smooth realizations: that is, high positive correlation for adjacent pairs of multinomial probabilities, say u_i being nearly equal to u_{i-1} and u_{i+1}, but rather unlike u_j for farther values j. This need was discussed in Dickey (1968b).

6. THE DOUBLE AVERAGES $\mathcal{R}$

Generalizations of the function R will provide further generalizations of the Dirichlet family. The generalization introduced in this section, denoted by $\mathcal{R}$, was given by Carlson (1971), who used it for essentially the following elegant derivation of a generalized Picard's identity. The generalized identity had been derived earlier by Dickey (1968a), using multiple-series expansions.

Consider a matrix Z ($K \times \kappa$), having κ-dimensional row vectors $\mathbf{z}_{i*}$ ($i = 1, \ldots, K$) and K-dimensional column vectors $\mathbf{z}_{*j}$ ($j = 1, \ldots, \kappa$),

$$Z = \begin{bmatrix} \mathbf{z}_{1*} \\ \mathbf{z}_{2*} \\ \vdots \\ \mathbf{z}_{K*} \end{bmatrix} = (\mathbf{z}_{*1}, \ldots, \mathbf{z}_{*\kappa}). \quad (6.1)$$

Define the K-vectors $\mathbf{u}$, $\mathbf{b}$, and the κ-vectors $\mathbf{v}$, $\boldsymbol{\beta}$. The usual matrix product defines a bilinear form, $\mathbf{u}^T Z \mathbf{v}$, in which $\mathbf{u}$ and $\mathbf{v}$ are taken as vertical arrays. We define the double Dirichlet average as

$$\mathcal{R}_a(\mathbf{b}, Z, \boldsymbol{\beta}) = E_{u|b}^{(K-1)} E_{v|\boldsymbol{\beta}}^{(\kappa-1)} (\mathbf{u}^T Z \mathbf{v})^a = E_{u|b}^{(K-1)} R_a(\boldsymbol{\beta}; \mathbf{u} \cdot \mathbf{z}_{*1}, \ldots, \mathbf{u} \cdot \mathbf{z}_{*\kappa}). \quad (6.2)$$

Apparently, we have homogeneity of degree a in the matrix Z: $\mathcal{R}_a(\mathbf{b}, cZ, \boldsymbol{\beta}) = c^a \mathcal{R}_a(\mathbf{b}, Z, \boldsymbol{\beta})$ for all $c > 0$.

Lemma 6.1 (Transposed argument).

$$\mathcal{R}_a(\mathbf{b}, Z, \boldsymbol{\beta}) = \mathcal{R}_a(\boldsymbol{\beta}, Z^T, \mathbf{b})$$
$$= E_{\mathbf{v}|\boldsymbol{\beta}}^{(\kappa-1)} R_a(\mathbf{b}; \mathbf{z}_{1*} \cdot \mathbf{v}, \ldots, \mathbf{z}_{K*} \cdot \mathbf{v}). \quad (6.3)$$

Theorem 5.2 and Lemma 6.1 imply the following preliminary generalization of Picard's identity.

Lemma 6.2.

$$\mathcal{R}_{-\beta.}(\mathbf{b}, Z, \boldsymbol{\beta}) = E_{u|b}^{(K-1)} \prod_1^{\kappa} (\mathbf{u} \cdot \mathbf{z}_{*j})^{-\beta_j}. \quad (6.4)$$

This applies to $\mathcal{R}_a$ for which $a = -\beta.$ (the *First Restriction*). Furthermore, if $\beta. = b.$ (the *Second Restriction*), we also have the expression symmetric to (6.4),

$$\mathcal{R}_{-\beta.}(\mathbf{b}, Z, \boldsymbol{\beta}) = E_{v|\boldsymbol{\beta}}^{(\kappa-1)} \prod_1^{\kappa} (\mathbf{z}_{i*} \cdot \mathbf{v})^{-b_i}. \quad (6.5)$$

$\mathcal{R}$ satisfying both restrictions is called a *bare form*, $a = -\beta. = -b..$

Corollary 6.3. (Dickey 1968a), If $\beta._d = b.$,

$$E_{u|b}^{(K-1)} \prod_1^{\kappa} (\mathbf{u} \cdot \mathbf{z}_{*j})^{-\beta_j}$$
$$= E_{v|\boldsymbol{\beta}}^{(\kappa-1)} \prod_1^{K} (\mathbf{z}_{i*} \cdot \mathbf{v})^{-b_i}. \quad (6.6)$$

Let $Z' = (Z, \mathbf{1})$, where $\mathbf{1} = (1, \ldots, 1)^T$.

Lemma 6.4 (Unit-column property). Under the First Restriction, for arbitrary γ,

$$\mathcal{R}_{-\beta.}(\mathbf{b}, Z, \boldsymbol{\beta}) = \mathcal{R}_{-\gamma}(\mathbf{b}, Z', \boldsymbol{\beta}'), \quad (6.7)$$

where $\boldsymbol{\beta}' = (\boldsymbol{\beta}, \gamma - \beta.)$, and hence $\beta.' = \gamma$.

Corollary 6.5. The First Restriction yields a bare form,

$$\mathcal{R}_{-\beta.}(\mathbf{b}, Z, \boldsymbol{\beta}) = \mathcal{R}_{-\beta.}(\mathbf{b}, Z', \boldsymbol{\beta}'), \quad (6.8)$$

where $\boldsymbol{\beta}' = (\boldsymbol{\beta}, b. - \beta.)$, and hence $\beta.' = b..$

Theorem 6.6 (Generalized Picard's identity, Dickey 1968a).

$$\mathcal{R}_{-\beta.}(\mathbf{b}, Z, \boldsymbol{\beta}) = E_{u|b}^{(K-1)} \prod_1^{\kappa} (\mathbf{u} \cdot \mathbf{z}_{*j})^{-\beta_j}$$
$$= E_{v'|\boldsymbol{\beta}'}^{(\kappa)} \prod_1^{K} (\mathbf{z}_{i*} \cdot \mathbf{v} + v'_{\kappa+1})^{-b_i}, \quad (6.9)$$

where $\mathbf{v}' = (\mathbf{v}, v'_{\kappa+1})$ and $\boldsymbol{\beta}' = (\boldsymbol{\beta}, b. - \beta.)$.

Note that $v'_{\kappa+1}$ vanishes from (6.9) if $\beta. = b.$, by $\beta'_{\kappa+1} = 0$, thus producing (6.6). Picard's original identity (3.5) is the special case $\kappa = 1$. The generalized identity can be useful for computations when $\kappa \ll K$.

By Lemma 6.2, the First Restriction implies homogeneity for $\mathcal{R}$ in each column vector $\mathbf{z}_{*j}$:

$$\mathcal{R}_{-\beta.}(\mathbf{b}; c_1\mathbf{z}_{*1}, \ldots, c_\kappa\mathbf{z}_{*\kappa}; \boldsymbol{\beta})$$
$$= \left(\prod_1^{\kappa} c_j^{-\beta_j}\right) \mathcal{R}_{-\beta.}(\mathbf{b}; \mathbf{z}_{*1}, \ldots, \mathbf{z}_{*\kappa}; \boldsymbol{\beta}),$$

for all $c_j > 0$ ($j = 1, \ldots, \kappa$). A bare form, of course, will have homogeneity both in each column and in each row of Z.

7. CONTINGENCY TABLES AND A NEW REPRESENTATION FOR $\mathcal{R}$

The obvious sampling model for contingency tables with sampling independence between row and column outcomes is the multiple Bernoulli with restricted parameters,

$$pr(\mathbf{y} \mid \mathbf{u}, \mathbf{v}) = \prod_{i=1}^{K} \prod_{j=1}^{\kappa} (u_i v_j)^{m_{ij}}, \quad (7.1)$$

or, for the frequency counts array $M = (m_{ij}: i = 1, \ldots, K, j = 1, \ldots, \kappa)$, its multinomial version:

$$pr(M \mid \mathbf{u}, \mathbf{v}) = \binom{m..}{M} \Pi_{i,j} (u_i v_j)^{m_{ij}}. \quad (7.2)$$

Taking independent prior distributions $\mathbf{u} \sim D(\mathbf{b})$ and $\mathbf{v} \sim D(\boldsymbol{\beta})$ yields the independent posterior distributions, $\mathbf{u} \mid M \sim D(\mathbf{b} + \mathbf{m}_{*}.)$ and $v \mid M \sim D(\boldsymbol{\beta} + \mathbf{m}._{*})$, where $\mathbf{m}_{*}. = (\sum_j m_{ij}: i = 1, \ldots, K)$ and $\mathbf{m}._{*} = (\sum_i m_{ij}: j = 1, \ldots, \kappa)$. The corresponding prior predictive distribution has the probability mass function, parameterized by n, $\mathbf{b}$, $\boldsymbol{\beta}$,

$$pr(M) = \binom{m..}{M} [\mathbf{b}:b.]^{\mathbf{m}_*.} [\boldsymbol{\beta}:\beta.]^{\mathbf{m}._*}, \quad (7.3)$$

where each $m_{ij} = 0, \ldots, n$ and $m.. \equiv n$. (The independent-Dirichlets prior distribution used here will be extended in a later section.)

We obtain the following generalization of Theorem 3.1.

Theorem 7.1 (Polynomial form). For nonnegative integer n, we have the *probability generating function* of the predictive distribution (7.3):

$$\mathcal{R}_n(\mathbf{b}, Z, \boldsymbol{\beta}) = E_{M|n,b,\boldsymbol{\beta}} \Pi_{i,j} z_{ij}^{m_{ij}}. \quad (7.4)$$

This identity appears in a nonprobabilistic form in Carlson (1974, Eq. (3.4)). Again, the characteristic function follows by a substitution.

8. FURTHER GENERALIZED FAMILIES

The function $\mathcal{R}$ under the First Restriction yields an extension of the distribution families $\{D_0{}^1\}$ and $\{D^1\}$, as follows. The integral representation (6.4) of $\mathcal{R}$ under the First Restriction differs from the integral representation (3.2) of R in allowing multiple factors that are powers of linear forms. In this same way, we shall extend consideration to more general density functions. For K-dimensional vectors $\mathbf{u}$, $\mathbf{b}$, the matrix of positive entries Z ($K \times \kappa$), and the κ-dimensional vector $\boldsymbol{\beta}$, define the distribution

$$\mathbf{u} \sim D^\kappa(\mathbf{b}, Z, \boldsymbol{\beta}) \quad (8.1)$$

to mean that $\mathbf{u}$ has the density function on the probability simplex,

$$B(\mathbf{b})^{-1} \left(\prod_1^K u_i^{b_i - 1}\right) \left[\prod_1^{\kappa} (\mathbf{u} \cdot \mathbf{z}_{*j})^{-\beta_j}\right] \Big/ \mathcal{R}_{-\beta.}(\mathbf{b}, Z, \boldsymbol{\beta}). \quad (8.2)$$

Note that the normalizing constant satisfies the First Restriction. We write D_0^κ for the case when the normalizing constant $\mathfrak{R}$ is a bare form, that is, when also $\beta. = b.$

The distribution (8.2) has an important role as the posterior distribution of a variance proportion in Bayesian inference for normal sampling. This role and the consequent role of Appell's hypergeometric function of two variables F_1 was discovered by Dickey (1965, p. 49; 1968a, Eq. (2.9); 1974, Eq. (32)). The corresponding roles in Bayesian and non-Bayesian inference about variance components were discovered by Culver (1971) and Hill (1977, Eq. (2.4)), who contributed to the mathematical analysis of the Appell functions, obtaining asymptotic expansions. For a representation by F_1 of the Behrens-Fisher density, see Dickey (1968a, Eq. (3.5)); and for related asymptotic expansions, see Dickey (1967, Eq. (7)). For an extended statement on the relevance of the distribution (8.1) and the related "poly-t" posterior density of location parameters (proportional to a product of multivariate-t densities), see Drèze (1977). Such posterior distributions arise from prior independence of location and scale.

For a given distribution, define the *multiplicity* as the minimal nonnegative integer κ permitting such an expression (8.2) of the density. Thus, if any two $\mathbf{z}_{*j}$'s are proportional, or any $\mathbf{z}_{*j} = c\mathbf{1}$, or $\beta_j = 0$, then the multiplicity is diminished. The multiplicity of a family of such distributions is the maximum of the multiplicities of its member distributions.

Theorem 8.1 (Nesting). Writing $D^0(\mathbf{b})$ for the usual Dirichlet distributions, we have for fixed dimensionality K,

$$\{D^{\kappa-1}\} \subset \{D_0^\kappa\} \subset \{D^\kappa\}, \tag{8.3}$$

for the multiplicities $\kappa = 1, 2, \ldots$.

Proof. Since for $\beta. = b.$, $D_0^\kappa(\mathbf{b}, Z, \boldsymbol{\beta}) \sim D^\kappa(\mathbf{b}, Z, \boldsymbol{\beta})$, we have $\{D_0^\kappa\} \subset \{D^\kappa\}$. To see that $\{D^\kappa\} \subset \{D_0^{\kappa+1}\}$ for $\kappa = 0, 1, \ldots$, proceed as in the unit-column property for $\mathfrak{R}$ (Lemma 6.4). Including the formal further factor $1 = (\mathbf{u} \cdot \mathbf{1})^{-(b.-\beta.)}$ in the density (8.2), write

$$D^\kappa(\mathbf{b}, Z, \boldsymbol{\beta}) \sim D_0^{\kappa+1}(\mathbf{b}, Z', \boldsymbol{\beta}'), \tag{8.4}$$

where $Z' = (Z, \mathbf{1})$ and $\boldsymbol{\beta}' = (\boldsymbol{\beta}, b. - \beta.)$.

As in D^1, the density functions are obviously all homogeneous of degree zero in each column vector $\mathbf{z}_{*j}$ of the matrix Z,

$$D^\kappa(\mathbf{b}, Z, \boldsymbol{\beta}) \sim D^\kappa(\mathbf{b}; c_1\mathbf{z}_{*1}, \ldots, c_\kappa\mathbf{z}_{*\kappa}; \boldsymbol{\beta}) \tag{8.5}$$

for any $c_j > 0$ ($j = 1, \ldots, \kappa$).

The moments are proportional to ratios of $\mathfrak{R}$ functions. If $\mathbf{u} \sim D^\kappa(\mathbf{b}, Z, \boldsymbol{\beta})$, then

$$E_{u|b,Z,\boldsymbol{\beta}}^{(K-1)} \left(\prod_1^K u_i^{m_i}\right) \prod_1^\kappa (\mathbf{u} \cdot \mathbf{z}_{*j})^{-\mu_j}$$
$$= [\mathbf{b} : b.]^{\mathbf{m}}\, \mathfrak{R}_{-(\beta.+\mu.)}(\mathbf{b} + \mathbf{m}, Z, \boldsymbol{\beta} + \boldsymbol{\mu}) / \mathfrak{R}_{-\beta.}(\mathbf{b}, Z, \boldsymbol{\beta}). \tag{8.6}$$

Note that any linear forms under the expectation here involve the same coefficient vectors $\mathbf{z}_{*j}$ that appear in the density (8.2). This restriction is removed in the following remarkable extension of (8.6).

Theorem 8.2. Define the $K \times (\kappa + \Lambda)$ matrix,

$$Z' = (Z, Y), \quad \text{where} \quad Y = (\mathbf{y}_{*1}, \ldots, \mathbf{y}_{*\Lambda}) \tag{8.7}$$

and $\boldsymbol{\beta}' = (\boldsymbol{\beta} + \boldsymbol{\mu}, \boldsymbol{\nu})$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_\kappa)$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_\Lambda)$. Then for $\mathbf{u} \sim D^\kappa(\mathbf{b}, Z, \boldsymbol{\beta})$ and arbitrary $\mathbf{m}$, $\boldsymbol{\mu}$, and $(Y, \Lambda, \boldsymbol{\nu})$, the moment defined by including the further factors $\Pi_1^\Lambda(\mathbf{u} \cdot \mathbf{y}_{*j})^{-\nu_j}$ under the expectation operator in (8.6) has the expression on the right-hand side of (8.6), generalized by substituting into the numerator $(Z', \boldsymbol{\beta}'$, and $\beta. + \mu. + \nu.)$ for $(Z, \boldsymbol{\beta} + \boldsymbol{\mu}$, and $\beta. + \mu.)$, respectively.

Theorem 8.3. The operator (2.2) applied to $D^\kappa(\mathbf{b}, Z, \boldsymbol{\beta})$ yields $D^\kappa(\mathbf{b} + \mathbf{m}, Z, \boldsymbol{\beta})$. Hence, $\{D^\kappa\}$ is closed and $\{D_0^\kappa\}$ is not closed under sampling in the personal-updating sense.

Theorem 8.4. If $\mathbf{v} \sim D^\kappa(\mathbf{b}, Z, \boldsymbol{\beta})$, then the population-updating operation $u_i = v_i y_i^{-1} / \sum v_j y_j^{-1}$ ($i = 1, \ldots, K$) yields $\mathbf{u} \sim D_0^{\kappa+1}(\mathbf{b}, Z', \boldsymbol{\beta}')$, where

$$Z' = (\mathbf{z}_{*1} \times \mathbf{y}, \ldots, \mathbf{z}_{*\kappa} \times \mathbf{y}, \mathbf{y}),$$
$$\mathbf{y} = (y_1, \ldots, y_K)^T, \tag{8.8}$$

and $\boldsymbol{\beta}' = (\boldsymbol{\beta}, b. - \beta.)$. Of course, if $\beta. = b.$,

$$\mathbf{u} \sim D_0^\kappa(\mathbf{b}; \mathbf{z}_{*1} \times \mathbf{y}, \ldots, \mathbf{z}_{*\kappa} \times \mathbf{y}; \boldsymbol{\beta}).$$

Hence, $\{D^\kappa\}$ is not closed and $\{D_0^\kappa\}$ is closed under population-updating operations.

Theorem 8.4 reminds us that our density functions fail to be homogeneous in each *row* vector of the matrix parameter Z, in contrast to the *column*-vector homogeneity (8.5). Theorems 8.1, 8.3, and 8.4 complete and extend the subfamily and closure information in Table 1.

Further extended families suggest themselves. For example, dropping the First Restriction, define $\mathbf{u} \sim D_+^\kappa(a, \mathbf{b}, Z, \boldsymbol{\beta})$ for the density

$$B(\mathbf{b})^{-1}\left(\prod_1^K u_i^{b_i-1}\right) \times R_a(\boldsymbol{\beta}; \mathbf{u} \cdot \mathbf{z}_{*1}, \ldots, \mathbf{u} \cdot \mathbf{z}_{*\kappa}) / \mathfrak{R}_a(\mathbf{b}, Z, \boldsymbol{\beta}), \tag{8.9}$$

the moments of which have an obvious simple form in the case of $E\,\Pi u_i^{m_i}$. Of course

$$D^\kappa(\mathbf{b}, Z, \boldsymbol{\beta}) \sim D_+^\kappa(-\beta., \mathbf{b}, Z, \boldsymbol{\beta}),$$

whereby $\{D^\kappa\} \subset \{D_+^\kappa\}$.

Note, finally, that D_+^κ is actually a marginal distribution in the new joint distribution for two vectors, $(\mathbf{u}, \mathbf{v}) \sim D^\kappa D^K(a, \mathbf{b}, Z, \boldsymbol{\beta})$, which has density,

$$B(\mathbf{b})^{-1}B(\boldsymbol{\beta})^{-1}\left(\prod_1^K u_i^{b_i-1}\right)\left(\prod_1^\kappa v_j^{\beta_j-1}\right) \times (\mathbf{u}^T Z \mathbf{v})^a / \mathfrak{R}_a(\mathbf{b}, Z, \boldsymbol{\beta}). \tag{8.10}$$

Closed-form expressions can easily be written down for general moments of this joint distribution, in the spirit of Theorem 8.2.

9. A NEW FAMILY OF PRIOR DISTRIBUTIONS FOR CONTINGENCY TABLES HAVING INDEPENDENT ROW AND COLUMN SAMPLE OUTCOMES

The joint distribution (8.10) can be used as a prior distribution for a contingency table with sampling independence between row and column outcomes, thus extending the independent-prior theory in Section 7. If prior to M, $(\mathbf{u}, \mathbf{v}) \sim D^\kappa D^K(a, \mathbf{b}, Z, \boldsymbol{\beta})$, then posterior we have

$$(\mathbf{u}, \mathbf{v}) \mid M \sim D^\kappa D^K(a, \mathbf{b} + \mathbf{m}_{*\cdot}, Z, \boldsymbol{\beta} + \mathbf{m}_{\cdot *}). \quad (9.1)$$

The corresponding predictive distribution, $M \sim D^\kappa D^K M(n, a, \mathbf{b}, Z, \boldsymbol{\beta})$, has mass function,

$$pr(M)\, \mathcal{R}_a(\mathbf{b} + \mathbf{m}_{*\cdot}, Z, \boldsymbol{\beta} + \mathbf{m}_{\cdot *})/\mathcal{R}_a(\mathbf{b}, Z, \boldsymbol{\beta}), \quad (9.2)$$

where $pr(M)$ denotes the previous Dirichlet-prior version of the predictive distribution, as given in (7.3).

Note that the distribution (8.10) can express prior dependence between the (random) row probability vector and the column probability vector, but it is not so effective for modeling prior prejudice for smoothness within either vector. For example, if a is positive and $Z = \text{diag}(1, 1, \ldots, 1)$, then the joint prior density is high for near equality of a pair, $u_i \doteq v_i$.

10. A NEW FUNCTION

In addition to statistics benefiting from results in the field of special functions, it is clear also that statistics can motivate further developments in that field. We have seen that probability provides a tidy notation, as in the definition of R and its representation as a probability generating function. We were led, by parallel reasoning, to a polynomial representation for the more general $\mathcal{R}$. I would like to propose here the study of a new family of special functions. Many-way contingency tables and the corresponding Bayesian inference suggest an immediate extension of the function $\mathcal{R}$ to many-way-array arguments. (An even more general function of a three-way array appears in Carlson 1971, Eq. (7.1).)

Write the Dirichlet density

$$q(\mathbf{u}; \mathbf{b}) = B(\mathbf{b})^{-1} \textstyle\prod_1^K u_i^{b_i - 1},$$

and the multilinear form constructed from the H-way array $\mathfrak{X}$,

$$\mathfrak{X}(\mathbf{u}, \mathbf{v}, \ldots, \mathbf{w}) = \sum_{i=1}^{K^{(1)}} \sum_{j=1}^{K^{(2)}} \cdots \sum_{k=1}^{K^{(H)}} u_i v_j \cdots w_k z_{i,j,\ldots,k}. \quad (10.1)$$

Then we define the H-way function,

$$\mathcal{R}_a(\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(H)}; \mathfrak{X}) = \int d\mathbf{u} \cdots \int d\mathbf{w}\, q(\mathbf{u}; \mathbf{b}^{(1)}) \cdots q(\mathbf{w}; \mathbf{b}^{(H)}) \times [\mathfrak{X}(\mathbf{u}, \ldots, \mathbf{w})]^a. \quad (10.2)$$

This is, of course, an iterated average of lower-way functions,

$$\mathcal{R}_a(\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(H-1)}, \mathbf{b}^{(H)}; \mathfrak{X}) = \int d\mathbf{w}\, q(\mathbf{w}; \mathbf{b}^{(H)}) \times \mathcal{R}_a[\mathbf{b}^{(1)}, \ldots, b^{(H-1)}; \mathfrak{X}(*, \ldots, *, \mathbf{w})], \quad (10.3)$$

where $\mathfrak{X}(*, \ldots, *, \mathbf{w})$ is the $(H - 1)$-way array having $(i^{(1)}, \ldots, i^{(H-1)})$th entry

$$\sum_{k=1}^{K^{(H)}}{}' w_k z_{i^{(1)},\ldots,i^{(H-1)},k}.$$

The function is invariant to permutations of the coordinates of a parameter vector, say $\mathbf{b}^{(h)}$, simultaneously with the values of the hth index of $\mathfrak{X}$. It is also invariant to a generalization of the matrix transpose operation, whereby the order *between* index variables of $\mathfrak{X}$ is permuted simultaneously with the order between parameters $\mathbf{b}^{(h)}$ $(h = 1, \ldots, H)$. Picard's identity has an immediate extension.

Our H-way distribution is defined by

$$p(\mathbf{u}, \ldots, \mathbf{w}) = q(\mathbf{u}; \mathbf{b}^{(1)}) \cdots q(\mathbf{w}; \mathbf{b}^{(H)}) [\mathfrak{X}(\mathbf{u}, \ldots, \mathbf{w})]^a/\mathcal{R}_a(\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(H)}; \mathfrak{X}). \quad (10.4)$$

The Bayesian inference theory given for the two-way family (8.10) has an obvious generalization. For example, the predictive mass function for the H-way counts array $\mathcal{M}$ is

$$\left(\prod_{h=1}^{H} [\mathbf{b}^{(h)} : b_{\cdot}^{(h)}]^{\mathbf{m}^{(h)}}\right) \times \frac{\mathcal{R}_a(\mathbf{b}^{(1)} + \mathbf{m}^{(1)}, \ldots, \mathbf{b}^{(H)} + \mathbf{m}^{(H)}; \mathfrak{X})}{\mathcal{R}_a(\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(H)}; \mathfrak{X})}, \quad (10.5)$$

where $\mathbf{m}^{(h)} = \mathcal{M}(\mathbf{1}, \ldots, *, \mathbf{1}, \ldots, \mathbf{1})$, the hth marginal-sum vector, having ith coordinate $m_{\cdot,\ldots,i,\cdot,\ldots,\cdot}$ $(i = 1, \ldots, K^{(h)})$. This is, again, exhibited as the predictive mass for the prior independence case, multiplied by a ratio of $\mathcal{R}$ functions. Again, this will provide a polynomial form representing $\mathcal{R}$ as a predictive probability generating function, in an obvious generalization of Theorem 7.1.

[*Received October 1981. Revised November 1982.*]

REFERENCES

APPELL, P., and KAMPÉ DE FÉRIET, J. (1926), *Fonctions Hypergéométriques et Hypersphériques; Polynomes d'Hermite*, Paris: Gauthier-Villars.

BLOCH, DANIEL A., and WATSON, GEOFFREY S. (1967), "A Bayesian Study of the Multinomial Distribution," *Annals of Mathematical Statistics*, 38, 1423–1434.

CARLSON, B.C. (1963), "Lauricella's Hypergeometric Function F_D," *Journal of Mathematical Analysis and Applications*, 7, 452–470.

——— (1965), "A Hypergeometric Mean Value," *Proceedings of the American Mathematical Society*, 16, 759–766.

——— (1971), "Appell Functions and Multiple Averages," *SIAM Journal on Mathematical Analysis*, 2, 420–430.

——— (1972), "Fonctions Spéciales—Intégrandes à Deux Formes Quadratiques," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Série A*, 274, 1458–1461.

——— (1974), "Inequalities for Jacobi Polynomials and Dirichlet Averages," *SIAM Journal on Mathematical Analysis*, 5, 586–596.

——— (1977), *Special Functions of Applied Mathematics*, New York: Academic Press.

CULVER, D.H. (1971), "A Bayesian Analysis of the Balanced One-Way Variance Components Model," Ph.D. Dissertation, University of Michigan.

DICKEY, JAMES M. (1965), "Bayesian Statistical Inference for Multivariate Location Parameters," Ph.D. Dissertation, Department of Mathematics, University of Michigan.

——— (1967), "Expansions of t Densities and Related Complete Integrals," *The Annals of Mathematical Statistics*, 38, 503–510.

——— (1968a), "Three Multidimensional-Integral Identities With Bayesian Applications," *The Annals of Mathematical Statistics*, 39, 1615–1627.

——— (1968b), "Estimation of Disease Probabilities Conditioned on Symptom Variables," *Mathematical Biosciences*, 3, 249–265.

——— (1974), "Bayesian Alternatives to the F-Test and Least-Squares Estimate in the Normal Linear Model," in *Studies in Bayesian Econometrics and Statistics*, eds. S.E. Fienberg and A. Zellner, Amsterdam: North-Holland.

DICKEY, JAMES M., and FREEMAN, PETER (1975), "Population-Distributed Personal Probabilities," *Journal of the American Statistical Association*, 70, 362–364.

DRÈZE, J.H. (1977), "Bayesian Regression Analysis Using Poly-t Densities," *Journal of Econometrics*, 6, 329–354.

FELLER, WILLIAM (1966), *An Introduction to Probability Theory and Its Applications, Vol. II*, New York: John Wiley.

GOOD, IRVING JOHN (1965), *The Estimation of Probabilities*, Cambridge, Mass.: M.I.T. Press.

HARDY, G.H., LITTLEWOOD, J.E., and POLYA, G. (1959), *Inequalities* (2nd ed.), Cambridge, U.K.: Cambridge University Press.

HILL, BRUCE M. (1974), "The Rank-Frequency Form of Zipf's Law," *Journal of the American Statistical Association*, 69, 1017–1026.

——— (1977), "Exact and Approximate Bayesian Solutions for Inference About Variance Components and Multivariate Inadmissibility," *New Developments in the Applications of Bayesian Methods*, eds. Ahmet Aykac and Carlo Brumat, Amsterdam: North Holland.

JOHNSON, NORMAN L., and KOTZ, SAMUEL (1969), *Discrete Distributions*, in *Distributions in Statistics* (Vol. 1), New York: John Wiley.

——— (1970a), *Continuous Univariate Distributions—1*, in *Distributions in Statistics* (Vol. 2), New York: John Wiley.

——— (1970b), *Continuous Univariate Distribution—2*, in *Distributions in Statistics* (Vol. 3), New York: John Wiley.

——— (1972), *Continuous Multivariate Distributions*, in *Distributions in Statistics* (Vol. 4), New York: John Wiley.

KAHTRI, C.G. (1966), "On Certain Distribution Problems Based on Positive Definite Quadratic Functions in Normal Vectors," *Annals of Mathematical Statistics*, 37, 468–479.

KOTZ, SAMUEL, JOHNSON, N.L., and BOYD, D.W. (1967), "Series Representations of Distributions of Quadratic Forms in Normal Variables. I. Central Case," *Annals of Mathematical Statistics*, 38, 823–837.

RAIFFA, HOWARD, and SCHLAIFER, ROBERT (1961), *Applied Statistical Decision Theory*, Division of Research, Graduate School of Business Administration, Harvard University. Reprinted by M.I.T. Press.

RAINVILLE, EARL D. (1960), *Special Functions*, New York: Macmillan Co.

SHAH, B.K., and KHATRI, C.G. (1974), "Proof of Conjectures About the Expected Values of the Elementary Symmetric Functions of a Noncentral Wishart Matrix," *Annals of Statistics*, 2, 833–836.

TRUESDELL, C.A. (1948), *An Essay Toward a Unified Theory of Special Functions*, Princeton, N.J.: Princeton University Press.

VILENKIN, N. JA. (1968), *Special Functions and the Theory of Group Representations*, Transl. Math. Mono., 22, American Mathematical Society, Providence, Rhode Island.

WILKS, S.S. (1962), *Mathematical Statistics*, New York: John Wiley.