The Annals of Mathematical Statistics
1968, Vol. 39, No. 5, 1615–1627

THREE MULTIDIMENSIONAL-INTEGRAL IDENTITIES WITH BAYESIAN APPLICATIONS[1]

BY JAMES M. DICKEY[2]

Yale University

1. Summary. The first identity (Section 2) expresses a moment of a product of multivariate t densities as an integral of dimension one less than the number of factors. This identity is applied to inference concerning the location parameters of a multivariate normal distribution.

The second identity (Section 3) expresses the density of a linear combination of independently distributed multivariate t vectors, a multivariate Behrens-Fisher density (Cornish (1965)), as an integral of dimension one less than the number of summands. The two-summand version of the second identity is essentially equivalent to the two-factor version of the first identity. A synthetic representation is given for the random vector, generalizing Ruben's (1960) representation in the univariate case. The second identity is applied to multivariate Behrens-Fisher problems.

The third identity (Section 4), due to Picard (Appell and Kampé de Fériet (1926)), expresses the moments of Savage's (1966) generalization of the Dirichlet distribution as a one-dimensional integral. A generalization of Picard's identity is given. Picard's identity is applied to inference about multinomial cell probabilities, to components-of-variance problems, and to inference from a likelihood function under a Student t distribution of errors.

2. The first identity. The identity is presented in Section 2.1 (Theorem 1). Applications are then briefly given in Section 2.2.

2.1. *Moments of products of multivariate t densities.* Consider the p-dimensional vector $\boldsymbol{\xi}$,

$$\boldsymbol{\xi} = \mathbf{x} + \boldsymbol{\zeta}/(\chi_\nu^2/\nu)^{\frac{1}{2}}, \tag{2.1}$$

where $\boldsymbol{\zeta}$ has the p-variate normal distribution with mean $\mathbf{0}$ and variance $\mathbf{M}^{-1}$ (precision $\mathbf{M}$), $\boldsymbol{\zeta} \sim N(\mathbf{0}, \mathbf{M}^{-1})$, independently of χ_ν^2. The vector $\boldsymbol{\xi}$ is said to have a p-variate t distribution with mean $\mathbf{x}$ and degrees-of-freedom ν [Cornish (1954), (1962), and Dunnett and Sobel (1954)]. The density of $\boldsymbol{\xi}$ is proportional to

$$[1 + \mathbf{G}((\boldsymbol{\xi} - \mathbf{x}))]^{-\frac{1}{2}n}, \tag{2.2}$$

Received 8 November 1967.

[1] Results presented here overlap with the author's Ph.D. dissertation at the University of Michigan, supported by a National Science Foundation grant. Further work at Yale University was supported by the Army, Navy, Airforce, and NASA under a contract administered by the Office of Naval Research; and at the University of Southern California, by a grant from the United States Public Health Service.

[2] Presently at the State University of New York at Buffalo.

where

$$\mathbf{G} = \nu^{-1}\mathbf{M}, \qquad n = \nu + p.$$

(Throughout the paper, $\mathbf{G}((\mathbf{y}))$, with double parentheses, will denote the quadratic form $\mathbf{y}'\mathbf{G}\mathbf{y}$; and $\mathbf{G} \geqq 0$ and $\mathbf{G} > 0$ will mean that $\mathbf{G}$ is symmetric and nonnegative definite and symmetric and positive definite, respectively.) The complete integral of (2.2) is finite if $\mathbf{G} > 0$ and $\nu = n - p > 0$, and it equals

$$(2.3) \qquad [C(\nu, p)\cdot|\mathbf{M}|^{\frac{1}{2}}]^{-1} = (\pi\nu)^{\frac{1}{2}p}\Gamma(\tfrac{1}{2}\nu)/\Gamma[\tfrac{1}{2}(\nu + p)]\cdot|\mathbf{M}|^{-\frac{1}{2}}.$$

Hence the density of $\boldsymbol{\xi}$ is the quotient of (2.2) and (2.3). The moments of $\boldsymbol{\xi}$ follow easily from the synthetic representation (2.1).

This Section 2.1 develops the moments of a product of t-like functions (2.2), each of which need not have a finite integral. Let

$$(2.4) \qquad g(\boldsymbol{\xi}) = \textstyle\prod_{k=1}^{K} [1 + \mathbf{G}_k((\boldsymbol{\xi} - \mathbf{x}_k))]^{-\frac{1}{2}n_K},$$

where each

$$\mathbf{G}_k \geqq 0, \qquad n_k > 0.$$

We seek, then, the complete p-dimensional integral of $f\cdot g$ where $f(\boldsymbol{\xi})$ is a polynomial in the coordinates of $\boldsymbol{\xi}$.

The limiting case, $\nu_1 \to \infty$, of one normal-density factor (more than one such factor can be combined by Lemma 1 below) will be left to the reader, except for brief mention of a special case at the end of this Section 2.1.

The basic tool is the representation of a t-like function as a gamma mixture of exponentials,

$$(2.5) \quad [1 + \mathbf{G}((\boldsymbol{\xi} - \mathbf{x}))]^{-\frac{1}{2}n} = [\Gamma(\tfrac{1}{2}n)2^{\frac{1}{2}n}]^{-1} \int_0^\infty u^{\frac{1}{2}n-1} \exp\{-\tfrac{1}{2}u[1 + \mathbf{G}((\boldsymbol{\xi} - \mathbf{x}))]\}\, du,$$

$n > 0$. We shall also need the following immediate generalization of a result of Raiffa and Schlaifer (1961), p. 312.

LEMMA 1. $\sum_k \mathbf{M}_k((\boldsymbol{\xi} - \mathbf{x}_k)) = \mathbf{D}((\boldsymbol{\xi} - \bar{\mathbf{x}})) + \sum \mathbf{M}_k((\mathbf{x}_k)) - \mathbf{D}((\bar{\mathbf{x}}))$, *where* $\mathbf{D} = \sum \mathbf{M}_k$ and $\bar{\mathbf{x}} = \mathbf{D}^{-1}(\sum \mathbf{M}_k\mathbf{x}_k)$. *The terms constant in* $\boldsymbol{\xi}$ *can also be written* $\sum \mathbf{M}_k((\mathbf{x}_k)) - \mathbf{D}((\bar{\mathbf{x}})) = \sum \mathbf{x}_k'\mathbf{M}_k(\mathbf{x}_k - \bar{\mathbf{x}}) = \sum \mathbf{M}_k((\mathbf{x}_k - \bar{\mathbf{x}}))$.

An application of (2.5) to each t-like factor of g and a completion of the square by Lemma 1 yield

$$g(\boldsymbol{\xi}) = [\textstyle\prod_k \Gamma(\tfrac{1}{2}n_k)2^{\frac{1}{2}n_k}]^{-1} \cdot \int_0^\infty du_1 \cdots \int_0^\infty du_K (\prod_k u_k^{\frac{1}{2}n_k-1}) \exp[-\tfrac{1}{2}\mathbf{D}_u((\boldsymbol{\xi} - \bar{\mathbf{x}}_u)) - \tfrac{1}{2}W_u],$$

where

$$\mathbf{D}_u = \textstyle\sum u_k\mathbf{G}_k, \qquad \bar{\mathbf{x}}_u = \mathbf{D}_u^{-1} \sum u_k\mathbf{G}_k\mathbf{x}_k,$$

$$W_u = \textstyle\sum u_k[1 + \mathbf{G}_k((\mathbf{x}_k))] - \mathbf{D}_u((\bar{\mathbf{x}}_u)).$$

After a formal change in the order of integration, the integral of $f \cdot g$ over Euclidean p-space is expressed as a K-dimensional integral,

$$\int_{R^p} f(\xi)\cdot g(\xi)\, d\xi = [\textstyle\prod_k \Gamma(\frac{1}{2}n_k)2^{\frac{1}{2}n_k}]^{-1}(2\pi)^{\frac{1}{2}p} \cdot \int_0^\infty du_1 \cdots \int_0^\infty du_K\, |\mathbf{D}_u|^{-\frac{1}{2}} N_{f|u}(\textstyle\prod_k u_k^{\frac{1}{2}n_k - 1}) \exp(-\frac{1}{2}W_u), \tag{2.6}$$

where

$$N_{f|u} = Ef(\zeta), \tag{2.7}$$

given

$$\zeta \sim N(\bar{\mathbf{x}}_u, \mathbf{D}_u^{-1}).$$

By Fubini's theorem, equation (2.6) is valid if either (thus each) member is finite when f is replaced by $|f|$. Hence, in particular, a necessary condition for absolute convergence of the p-dimensional integral of $f \cdot g$ is that $\sum \mathbf{G}_k > 0$. We note in passing that the right-hand sides of equation (2.6) and its descendants below are susceptible to the δ-method of asymptotic expansion (Wallace, 1958).

After a change of variables in the right-hand side of equation (2.6), we can perform one of the integrations to reduce the K-dimensional integral to a $(K - 1)$-dimensional integral. Given constants $c_k > 0$, $k = 1, \cdots, K$, let $u. = \sum c_k u_k$ and $u_k = v_k u.$. Then $\sum c_k v_k = 1$ and the Jacobian of partial derivatives of $u_1, \cdots, u_K$ with respect to $u., v_1, \cdots, v_{K-1}$ is $c_K^{-1} u.^{K-1}$. Expressed in these variables, $N_{f|u}$ is a polynomial in $u.^{-1}$, the coefficients of which are rational functions of the v_k. Denote

$$N_{f|u} = \textstyle\sum_r h_r(v_1, \cdots, v_K) u.^{-r}. \tag{2.8}$$

Then for each term of $N_{f|u}$, the integration with respect to $u.$ is in the form of the gamma function of argument $\frac{1}{2}(\sum n_k - p) - r$.

THEOREM 1. *Given $g(\xi)$ a product (2.4) of K-many t-like functions and $f(\xi)$ a polynomial in the coordinates of ξ. If the p-dimensional integral of $f \cdot g$ is absolutely convergent,*

$$\int_{R^p} f(\xi) g(\xi)\, d\xi = \mathrm{Const} \cdot \textstyle\sum_r \Gamma[\frac{1}{2}(n. - p) - r]2^{-r} \cdot c_K^{-1} \int_\sigma dv_1 \cdots dv_{K-1}\, |\mathbf{D}_v|^{-\frac{1}{2}} h_r(v_1, \cdots v_K)(\prod_k v_k^{\frac{1}{2}n_k - 1}) W_v^{-\frac{1}{2}(n. - p) + r},$$

where

$$\mathrm{Const} = [\textstyle\prod_k \Gamma(\frac{1}{2}n_k)]^{-1}\pi^{\frac{1}{2}p},$$
$$n. = \textstyle\sum n_k,$$
$$\mathbf{D}_v = \textstyle\sum v_k \mathbf{G}_k,$$
$$W_v = \textstyle\sum v_k[1 + \mathbf{G}_k((\mathbf{x}_k))] - \mathbf{D}_v^{-1}((\sum v_k \mathbf{G}_k \mathbf{x}_k)),$$

h_r is defined by (2.7) *and* (2.8), *and σ is the simplex,*

$$\sigma = \{(v_1, \cdots, v_K) : \textstyle\sum c_k v_k = 1 \quad \textit{and each} \quad v_k > 0\}.$$

Another necessary condition for absolute convergence of the integral of $f \cdot g$ is, thus, that $\sum n_k > p + 2 \max(r)$. Note possibilities for simplifying the integrand by choices of the c_k for which $\sum c_k v_k = 1$.

We turn now to specific integrands for Theorem 1.

I. A special circumstance in which $|\mathbf{D}_v|$, $\mathbf{D}_v^{-1}$, and $N_{f|u}$ are explicitely calculable functions of the v_k is when the $\mathbf{G}_k$ are simultaneously diagonalizable in the following sense. Write for each k

$$\mathbf{G}_k = \mathbf{A}'\mathbf{\Lambda}_k\mathbf{A},$$

where

$$|\mathbf{A}| = 1 \quad \text{and} \quad \mathbf{\Lambda}_k = \operatorname{diag}(\lambda_k^{(1)}, \cdots, \lambda_k^{(p)}).$$

Also,

$$\mathbf{z}_k = \mathbf{A}\mathbf{x}_k = (z_k^{(1)}, \cdots z_k^{(p)})'.$$

Then we have

$$|\mathbf{D}_v| = \prod_i (\textstyle\sum_k v_k \lambda_k^{(i)}),$$
$$W_v = \sum v_k + \sum_i [\textstyle\sum_{k<h} v_k v_h \lambda_k^{(i)} \lambda_h^{(i)} (z_k^{(i)} - z_h^{(i)})^2] / (\sum_k v_k \lambda_k^{(i)}).$$

The expression for W_v is obtained by noting that simultaneous diagonalizability of the $\mathbf{M}_k = v_k\mathbf{G}_k$ is equivalent to $\mathbf{M}_h\mathbf{D}_v^{-1}\mathbf{M}_k = \mathbf{M}_k\mathbf{D}_v^{-1}\mathbf{M}_h$ for all h, k, and hence the constant term of Lemma 1 becomes $\sum \mathbf{M}_k((\mathbf{x}_k)) - \mathbf{D}_v^{-1}((\sum \mathbf{M}_k\mathbf{x}_k)) = \sum_{k<h} [\mathbf{M}_k\mathbf{D}_v^{-1}\mathbf{M}_h]((\mathbf{x}_k - \mathbf{x}_h))$.

For the linear form, $f(\xi) = (\tilde{\mathbf{b}}'\mathbf{A})\xi$, $r = 0$ and

$$N_{f|u} = h_0 = \textstyle\sum_i \tilde{b}^{(i)} (\sum_k v_k \lambda_k^{(i)} z_k^{(i)}) / (\sum_k v_k \lambda_k^{(i)}).$$

For the uncentered quadratic form, $f(\xi) = [\mathbf{A}'\mathbf{L}\mathbf{A}]((\xi - \mathbf{x}_0))$, $r = 0, 1$, and

$$\begin{aligned} N_{f|u} &= u.^{-1}h_1 + h_0 \\ &= u.^{-1} \textstyle\sum_i \tilde{l}_{ii} / (\sum_k v_k \lambda_k^{(i)}) \\ &\quad + \textstyle\sum_{i,j} \tilde{l}_{ij} \prod_{q=i,j} [\sum_k v_k \lambda_k^{(q)} (z_k^{(q)} - z_0^{(q)})] / (\sum_k v_k \lambda_k^{(q)}), \end{aligned}$$

where $\tilde{\mathbf{L}} = (\tilde{l}_{ij})$ and $\mathbf{z}_0 = \mathbf{A}\mathbf{x}_0 = (z_0^{(1)}, \cdots, z_0^{(p)})'$.

I.A. The case of two t-like factors, $K = 2$, is especially important for applications, as discussed in the next Section 2.2. Fortunately, if $\mathbf{G}_1 + \mathbf{G}_2 > 0$, then $\mathbf{G}_1$ and $\mathbf{G}_2$ are simultaneously diagonalizable as defined above. In this case, $\sigma = (0, 1/c_1)$, and the integral of $f \cdot g$ has a representation as a one-dimensional integral.

The simultaneous diagonalization of $\mathbf{G}_1$ and $\mathbf{G}_2$ is achieved in practice by performing successive diagonalizations, by orthogonal matrices, of $\mathbf{G}_1 + \mathbf{G}_2$ and $\mathbf{H} = \mathbf{\Lambda}_3^{-\frac{1}{2}}\mathbf{O}'\mathbf{G}_1\mathbf{O}\mathbf{\Lambda}_3^{-\frac{1}{2}}$, where $\mathbf{\Lambda}_3$ and $\mathbf{O}$ are the diagonal matrix of roots and the orthogonal matrix of row eigenvectors of $\mathbf{G}_1 + \mathbf{G}_2$. Then $\mathbf{\Lambda}_1 = \mathbf{\Lambda}_3\mathbf{\Delta}$, $\mathbf{\Lambda}_2 = \mathbf{\Lambda}_3 - \mathbf{\Lambda}_1$, and $\mathbf{A} = \mathbf{\Lambda}_3^{-\frac{1}{2}}\mathbf{Q}\mathbf{\Lambda}_3^{\frac{1}{2}}\mathbf{O}$, where $\mathbf{\Delta}$ and $\mathbf{Q}$ are the matrices of roots and eigenvectors of $\mathbf{H}$. The one-dimensional quadrature is easily and accurately carried out by Simpson's rule.

A computer program in FORTRAN (listings available) has been developed to calculate the complete integral of $f \cdot g$ in the case $K = 2$ for $f(\xi) = 1$, $\mathbf{b}'\xi$, or $\mathbf{L}((\xi - \mathbf{x}_0))$. The program's performance compares favorably with that of Tiao and Zellner's (1964) asymptotic expansion of a (posterior) density function proportional to g. Tiao and Zellner consider a two-sample regression problem with parameters β_i and obtain approximations, .03726 and 9.6158×10^{-5}, to the posterior mean and variance of β_1. The values with the approximating normal density are .03730 and 9.0145×10^{-5}. The quadrature program yields .0372826 and 10.02397×10^{-5}.

I.B. Another subcase of interest occurs when the K quadratic forms are proportional: there exist a $\mathbf{G} > 0$ and $\gamma_1, \cdots, \gamma_K > 0$ such that each $\mathbf{G}_k = \gamma_k \mathbf{G}$. (This holds, of course, when $p = 1$.) Letting $c_k = \gamma_k$, we have $|\mathbf{D}_v| = |\mathbf{G}|$ and

$$W_v = \sum v_k + \sum_{k<h} v_k v_h \gamma_k \gamma_h \mathbf{G}((\mathbf{x}_k - \mathbf{x}_h)).$$

For the vector-valued function

$$f(\xi) = \xi, \qquad N_{f|u} = \sum v_k \gamma_k \mathbf{x}_k\,;$$

and for

$$f(\xi) = \mathbf{L}((\xi - \mathbf{x}_0)), \qquad N_{f|u} = u.^{-1} \operatorname{tr}(\mathbf{L}\mathbf{G}^{-1}) + \mathbf{L}((\textstyle\sum v_k \gamma_k \mathbf{x}_k - \mathbf{x}_0)).$$

I.AB. Consider the case of two t-like factors ($K = 2$) with their quadratic forms proportional, and let $\mathbf{G} = \mathbf{I}_p$. Then

$$g(\xi) = [1 + \gamma_1 \|\xi - \mathbf{x}_1\|^2]^{-\frac{1}{2}n_1} \cdot [1 + \gamma_2 \|\xi - \mathbf{x}_2\|^2]^{-\frac{1}{2}n_2},$$

and with $n. = n_1 + n_2$ and

$$C = [\Gamma(\tfrac{1}{2}n_1)\Gamma(\tfrac{1}{2}n_2)]^{-1} \pi^{\frac{1}{2}p} \Gamma[\tfrac{1}{2}(n. - p)] \gamma_1^{-\frac{1}{2}n_1} \gamma_2^{-\frac{1}{2}n_2},$$

we have for $f \equiv 1$,

$$\begin{aligned} &\int_{R^p} g(\xi)\, d\xi \\ (2.9) \qquad &= C \cdot \int_0^1 dt\; t^{\frac{1}{2}n_1 - 1}(1 - t)^{\frac{1}{2}n_2 - 1}[t\gamma_1^{-1} + (1 - t)\gamma_2^{-1} \\ &\qquad + t(1 - t)\|\mathbf{x}_1 - \mathbf{x}_2\|^2]^{-\frac{1}{2}(n. - p)} \\ &= C \cdot \gamma_2^{\frac{1}{2}(n. - p)} \cdot B(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2) \cdot F_1(\tfrac{1}{2}n_1\,; \tfrac{1}{2}(n. - p), \tfrac{1}{2}(n. - p); \tfrac{1}{2}n.\,; z_1, z_2), \end{aligned}$$

where z_1 and z_2 are the two (real) roots of

$$z^2 + (\gamma_2 \|\mathbf{x}_1 - \mathbf{x}_2\|^2 + \gamma_2 \gamma_1^{-1} - 1)z - \gamma_2 \|\mathbf{x}_1 - \mathbf{x}_2\|^2 = 0.$$

B is the complete beta integral, and F_1 is Appell's hypergeometric function of two variables,

$$F_1(\alpha; \beta, \beta'; \gamma; x, y) = [B(\alpha, \gamma - \alpha)]^{-1} \int_0^1 dt\; t^{\alpha - 1}(1 - t)^{\gamma - \alpha - 1}(1 - tx)^{-\beta}(1 - ty)^{-\beta'}$$

[Appell and Kampé de Fériet (1926), or Erdélyi *et al.* (1953)].

Double series in ascending or descending powers of x and y result from applying the binomial series to the integrand of F_1. Known transformations of F_1 lead to additional useful series.

Integrals of $f \cdot g$ for polynomial f can be similarly expressed in terms of Appell's F_1.

I.AB.O. As a final, limiting case, consider a function g_0, the product of an exponential function and a t-like function with proportional quadratic forms,

$$g_0(\boldsymbol{\xi}) = \exp\left(-\tfrac{1}{2}\lambda \|\boldsymbol{\xi} - \mathbf{x}_1\|^2\right) \cdot [1 + \gamma \|\boldsymbol{\xi} - \mathbf{x}_2\|^2]^{-\frac{1}{2}n}.$$

By setting in equation (2.9) $\gamma_1 = \lambda/n_1$, $\gamma_2 = \gamma$, and letting $n_1 \to \infty$, we obtain the representation,

$$(2.10) \quad \int_{R^p} g_0(\boldsymbol{\xi})\, d\boldsymbol{\xi} = [\Gamma(\tfrac{1}{2}n)2^{\frac{1}{2}n}]^{-1}(2\pi)^{\frac{1}{2}p}(\lambda\gamma^{-1})^{\frac{1}{2}n}\lambda^{-\frac{1}{2}p}$$
$$\cdot \exp\left(\tfrac{1}{2}\lambda\gamma^{-1} - \tfrac{1}{2}\lambda \|\mathbf{x}_1 - \mathbf{x}_2\|^2\right) \cdot \Upsilon(-\tfrac{1}{2}(n-p), \tfrac{1}{2}n; \tfrac{1}{2}\lambda\gamma^{-1}; \tfrac{1}{2}\lambda \|\mathbf{x}_1 - \mathbf{x}_2\|^2),$$

where we write for $b > 0$, $c \geqq 0$, and $a > 0$ if $c = 0$,

$$(2.11) \qquad \Upsilon(a, b; c; d) = \int_0^1 dt\, t^{a-1}(1-t)^{b-1} \exp\left(-c \cdot t^{-1} + d \cdot t\right).$$

The (arbitrarily symbolized) special function Υ, new to the author, merits study. For its boundary values and a general representation, see Erdélyi *et al.* (1953), p. 255, and Duff (1956), p. 242.

2.2. *Inference about multivariate normal location parameters.* Stein ((1962), e.g.), and Edwards, Lindman, and Savage (1963), p. 233, have objected to the use of the formal constant prior distribution for a high dimensional ($p \geqq 3$) multivariate normal location vector $\boldsymbol{\xi}$ (either the mean vector, or a vector of regression coefficients with the mean vector equal to $\mathbf{Z}\boldsymbol{\xi}$, $\mathbf{Z}$ known). Anscombe (1963) and Stein (1964) have criticized multivariate normal prior distributions for $\boldsymbol{\xi}$, the Raiffa-Schlaifer "conjugate class" for multivariate normal sampling with scale matrix parameter known.

Anscombe (1963) and Tiao and Zellner (1964) have recommended multivariate t prior distributions for $\boldsymbol{\xi}$. Such a prior distribution implies, in the case of known scale, a posterior density proportional to the product of a multivariate t density and a multivariate normal density, a limiting form of $g(\boldsymbol{\xi})$ (2.4).

The ability to handle prior independence of scale and location is clearly desirable. A normal likelihood function with an unknown scale integrated out by a gamma or Wishart kernel (including formal ignorance types) is proportional to a t-like function (Raiffa and Schlaifer (1961); Ando and Kaufman (1965)). Hence with a t-like prior for $\boldsymbol{\xi}$ and an independent gamma or Wishart-like prior for the unknown scale, we obtain a posterior density for $\boldsymbol{\xi}$ proportional to a product $g(\boldsymbol{\xi})$ (Anscombe (1963); Tiao and Zellner (1964)). Also, it is natural to combine information from K separate experiments (prior independence of the scales) into a likelihood function proportional to a product $g(\boldsymbol{\xi})$ (Tiao and Zellner (1964)).

The posterior distribution is the sole objective of much Bayesian work. But in cases of high dimensionality (p large), summarizing quantities are called for.

The role of low-order posterior moments in estimation with quadratic loss is well known (Savage, (1954)). See Dickey (1967c) for an hypothesis-testing procedure based on the posterior mean.

3. The second identity. It will be of advantage to first discuss applications in Section 3.1 and to then present the identity in Section 3.2 (Theorem 2).

3.1. *Multivariate Behrens-Fisher problems.* We treat Behrens-Fisher problems generalized in two ways, first to multivariate normal populations and secondly to any small number of such populations. Consider K many independent p_k-variate normal distributions π_k with unknown mean vectors $\mathfrak{u}_k$ and unknown variance matrices $\mathbf{H}_k^{-1}$. We discuss problems of inference about a linear combination $\mathfrak{n}$ of the means,

$$\mathfrak{n} = \sum_{k=1}^{K} \mathbf{A}_k \mathfrak{u}_k ,$$

with given $r \times p_k$ matrices $\mathbf{A}_k$, on the basis of a sample of independent vectors, $\mathbf{x}_{k1}, \mathbf{x}_{k2}, \cdots, \mathbf{x}_{kN_k}$, from each distribution π_k.

Solutions to traditional Behrens-Fisher problems ($p_k = 1$, $r = 1$, $K = 2$) based on the usual Behrens-Fisher distributions (variate $d = t_1 \cos\theta + t_2 \sin\theta$, the t_k being independent Student t variates with ν_k degrees of freedom) have been proposed by Behrens (1929), Fisher (1935), and Jeffreys (1940). In an unpublished manuscript, Savage (1961) has given a personalistic-Bayesian foundation for the use of Jeffreys' solutions as approximations.

We follow the precedent of former treatments of the traditional problem by working with independent prior opinions about the K experimental distributions, and hence independent posterior opinions. Suppose that each prior distribution of a pair $(\mathfrak{u}_k, \mathbf{H}_k)$ is independently as proposed in Section 2.2 (namely $\mathfrak{u}_1, \cdots, \mathfrak{u}_K$, $\mathbf{H}_1, \cdots, \mathbf{H}_K$ jointly independent). Then the posterior distributions of the means $\mathfrak{u}_k$ are independent with densities proportional to products $g_k(\mathfrak{u}_k)$ of t-like functions. Regression vectors $\boldsymbol{\beta}_k$ ($\mathfrak{u}_k = Z_k\boldsymbol{\beta}_k$) take posterior distributions of the same form, and hence could replace the means $\mathfrak{u}_k$ throughout this section.

It was suggested in Section 2.2 that inference can be based on the low-order moments of unknown parameters. In principle, the posterior moments of the $\mathfrak{u}_k$ can be computed according to the methods of Section 2.1. The low-order moments of $\mathfrak{n}$ are easily obtained from those of the $\mathfrak{u}_k$,

$$E\mathfrak{n} = \sum \mathbf{A}_k E\mathfrak{u}_k ,$$

$$E\mathbf{L}((\mathfrak{n} - \hat{\mathfrak{n}})) = \sum \operatorname{tr}(\mathbf{L}_k\mathbf{V}_k) + \mathbf{L}((E\mathfrak{n} - \hat{\mathfrak{n}})),$$

where $\mathbf{V}_k = E(\mathfrak{u}_k - E\mathfrak{u}_k)(\mathfrak{u}_k - E\mathfrak{u}_k)'$, and $\mathbf{L}_k = \mathbf{A}_k'\mathbf{L}\mathbf{A}_k$.

If each prior distribution of a pair $(\mathfrak{u}_k, \mathbf{H}_k)$ can be approximated by a Raiffa-Schlaifer or Ando-Kaufman "conjugate" distribution, in particular, if the dimensionality p_k is low and stable estimation applies, then the posterior distribution of $\mathfrak{u}_k$ is approximately multivariate t. Such distributions are of interest for summarizing opinion of the $\mathfrak{u}_k$ under the alternatives to a sharp null hypothesis in the Savage version of Bayesian tests of hypotheses (Edwards, Lindman, and Savage (1963)).

We describe briefly the Bayesian test for the hypothesis,

$$h\colon \boldsymbol{\eta} = \mathbf{0}.$$

We test h against its logical complement, $\bar{h}\colon \boldsymbol{\eta} \neq \mathbf{0}$. Define the prior and posterior odds for h, $\Omega_1 = P\{h\}/P\{\bar{h}\}$, $\Omega_2 = P\{h \mid x\text{'s}\}/P\{\bar{h} \mid x\text{'s}\}$, and their ratio, $\Lambda = \Omega_2/\Omega_1$. By Bayes' formula,

$$\Lambda = P'\{x\text{'s} \mid h\}/P'\{x\text{'s} \mid \bar{h}\}, \tag{3.1}$$

the likelihood ratio, in which P' denotes a probability density.

The denominator in (3.1) can be expressed as the integral, $\int P'\{x\text{'s} \mid \boldsymbol{\eta}, \bar{h}\} \cdot P'\{\boldsymbol{\eta} \mid \bar{h}\}\, d\boldsymbol{\eta}$. By approximating the numerator by the limit quantity $P'\{x\text{'s} \mid \boldsymbol{\eta} = \mathbf{0}, \bar{h}\}$, and then multiplying both the numerator and denominator by $P'\{\boldsymbol{\eta} = \mathbf{0} \mid \bar{h}\}$, we obtain, after another application of Bayes' formula, the approximation,

$$\Lambda = P'\{\boldsymbol{\eta} = \mathbf{0} \mid x\text{'s}, \bar{h}\}/P'\{\boldsymbol{\eta} = 0 \mid \bar{h}\}. \tag{3.2}$$

The posterior density of $\boldsymbol{\eta}$ in the numerator of (3.2) can conceivably be approximated by the density of the convolution of uncentered multivariate t distributions of the independently distributed summands $\mathbf{A}_k\boldsymbol{\mu}_k$. Specifically, if $\boldsymbol{\mu}_k = \mathbf{C}_k\boldsymbol{\mu}_k^*$ and each $\boldsymbol{\mu}_k^*$ is q_k-dimensional t with center $\boldsymbol{\zeta}_k^*$, matrix $\mathbf{G}_k = \nu_k^{-1}\mathbf{M}_k > 0$, and degrees of freedom $\nu_k > 0$, then $\boldsymbol{\eta} - \sum \mathbf{A}_k\mathbf{C}_k\boldsymbol{\zeta}_k$ is distributed like the r-vector $\boldsymbol{\delta}$,

$$\boldsymbol{\delta} = \sum \mathbf{B}_k\boldsymbol{\tau}_k, \tag{3.3}$$

where $\mathbf{B}_k = \mathbf{A}_k\mathbf{C}_k\mathbf{M}_k^{-\frac{1}{2}}$ and each $\boldsymbol{\tau}_k (= \mathbf{M}_k^{\frac{1}{2}}\boldsymbol{\mu}_k^*)$ independently has a standard q_k-dimensional t distribution on ν_k degrees of freedom (center $\mathbf{0}$, matrix $\mathbf{G}_k = \nu_k^{-1}\mathbf{I}_{q_k}$).

3.2. *Multivariate Behrens-Fisher distributions.* In the two-summand case ($K = 2$), Cornish (1965) termed the distribution of $\boldsymbol{\delta}$ (3.3) a "multiple Behrens-Fisher distribution". If $K = 2$ and $r = 1$, the distribution of $\boldsymbol{\delta} = \delta$ is a scaled version of a usual Behrens-Fisher distribution. A vector formed of two usual Behrens-Fisher variables which are different linear combinations of the same two Student variables is distributed like $\boldsymbol{\delta}$ for $r = 2$, $K = 2$. Indeed, from the definition (3.3), any linear transformation $\mathbf{D}\boldsymbol{\delta}$ has a multivariate Behrens-Fisher distribution. For example, marginal distributions are Behrens-Fisher. We develop the density and a synthetic representation for $\boldsymbol{\delta}$.

Returning to equations (2.2)–(2.4) and Theorem 1, note that we have already developed the density of $\boldsymbol{\delta}$ for the case $K = 2$. If $\mathbf{G}_1, \mathbf{G}_2 > 0$, and $n_1, n_2 > p$, the integral of g can be interpreted as proportional to a convolution of two multivariate t densities,

$$\psi(\mathbf{x}_1 - \mathbf{x}_2) = \int_{R^p} \prod_{k=1}^{K} C(\nu_k, p)|\mathbf{M}_k|^{\frac{1}{2}}[1 + \nu_k^{-1}\mathbf{M}_k((\boldsymbol{\xi} - \mathbf{x}_k))]^{-\frac{1}{2}(\nu_k+p)}\, d\boldsymbol{\xi}, \tag{3.4}$$

with $\psi(\boldsymbol{\delta})$ the density of the p-dimensional vector,

$$\boldsymbol{\delta} = \mathbf{B}_1\boldsymbol{\tau}_1 + \mathbf{B}_2\boldsymbol{\tau}_2, \qquad \mathbf{B}_k\mathbf{B}_k' = \mathbf{M}_k^{-1}.$$

Here, again, the $\boldsymbol{\tau}_k$ independently have standard q_k-dimensional ($q_k > p$) t distributions with $\nu_k > 0$ degrees of freedom. We have used the result of the synthetic representation of $\boldsymbol{\tau}_k$ (2.1) that $\mathbf{B}_k\boldsymbol{\tau}_k$ is distributed like $(\mathbf{B}_k\mathbf{B}_k')^{\frac{1}{2}}\mathbf{t}_k$ for a p-dimensional standard t vector, $\mathbf{t}_k$, with ν_k degrees of freedom.

If the first factor in the integrand of (3.4) ($k = 1$) is replaced by the normal $N(\mathbf{x}_1, \mathbf{M}_1^{-1})$ density, then $\boldsymbol{\tau}_1$ in the synthetic representation below (3.4) is replaced by a standard normal vector.

In the case I.AB. of $K = 2$ and proportional matrices (say $\mathbf{M}_k = \beta_k^{-2}\mathbf{I}_p$, $k = 1, 2$), $\psi(\boldsymbol{\delta})$ can now be read from (2.9) in terms of F_1. For $\boldsymbol{\delta} = \beta_1\boldsymbol{\tau}_1 + \beta_2\boldsymbol{\tau}_2$, with $\nu. = \nu_1 + \nu_2$,

$$\begin{aligned}\psi(\boldsymbol{\delta}) &= \Gamma[(\nu. + p)/2][\Gamma(\nu_1/2)\Gamma(\nu_2/2)\pi^{\frac{1}{2}p}]^{-1}(\nu_1\beta_1^2)^{\frac{1}{2}\nu_1}(\nu_2\beta_2^2)^{-\frac{1}{2}(\nu_1+p)}\\ &\quad\cdot B(\tfrac{1}{2}(\nu_1 + p), \tfrac{1}{2}(\nu_2 + p))\cdot F_1(\tfrac{1}{2}(\nu_1 + p);\\ &\quad\quad \tfrac{1}{2}(\nu. + p), \tfrac{1}{2}(\nu. + p); \tfrac{1}{2}\nu. + p; z_1; z_2),\end{aligned} \tag{3.5}$$

where z_1 and z_2 satisfy

$$z^2 + (\|\boldsymbol{\delta}\|^2/\nu_2\beta_2^2 + \nu_1\beta_1^2/\nu_2\beta_2^2 - 1)z - \|\boldsymbol{\delta}\|^2/\nu_2\beta_2^2 = 0.$$

This appears to be a new expression for the traditional Behrens-Fisher density, $p = 1$.

In the limiting case I.AB.O., from (2.10) and the limiting form of (3.4), we obtain the density $\psi_0(\boldsymbol{\delta}_0)$ of a simple linear combination, $\boldsymbol{\delta}_0 = \beta_1\mathbf{z} + \beta_2\boldsymbol{\tau}$, of a standard normal vector $\mathbf{z}$ and a standard multivariate t vector $\boldsymbol{\tau}$,

$$\begin{aligned}\psi_0(\boldsymbol{\delta}_0) &= (\tfrac{1}{2}\nu)^{\frac{1}{2}\nu}[\Gamma(\tfrac{1}{2}\nu)(2\pi)^{\frac{1}{2}p}]^{-1}\beta_1^{-(\nu+p)}\beta_2^{\nu}\\ &\quad\cdot\exp(\tfrac{1}{2}\nu\beta_2^2\beta_1^{-2} - \tfrac{1}{2}\beta_1^{-2}\|\boldsymbol{\delta}_0\|^2)\cdot\Upsilon(-\tfrac{1}{2}\nu, \tfrac{1}{2}(\nu + p);\\ &\quad\quad \tfrac{1}{2}\nu\beta_2^2\beta_1^{-2}; \tfrac{1}{2}\beta_1^{-2}\|\boldsymbol{\delta}_0\|^2),\end{aligned} \tag{3.6}$$

where equation (2.11) defines Υ. Equation (3.6) corrects an erroneous expression for ψ_0 in the one-dimensional case given by Ruben (1960).

With a general number K of summands, we have the following theorem.

THEOREM 2. *Let $\boldsymbol{\tau}_1, \cdots, \boldsymbol{\tau}_K$ have independent standard q_k-dimensional multivariate t distributions with $\nu_1, \cdots, \nu_K > 0$ degrees of freedom (centers $\mathbf{0}$, matrices $\nu_k^{-1}\mathbf{I}_{q_k}$). Then the random r-vector $\boldsymbol{\delta}$,*

$$\boldsymbol{\delta} = \sum \mathbf{B}_k\boldsymbol{\tau}_k,$$

has the representation,

$$\boldsymbol{\delta} = (\sum u_k^{-1}\nu_k\mathbf{B}_k\mathbf{B}_k')^{\frac{1}{2}}\mathbf{z}, \tag{3.7}$$

where the u_k are independently chi-squared distributed with ν_k degrees of freedom, and $\mathbf{z}$ is an independent r-dimensional standard normal vector. Consequently, $\boldsymbol{\delta}$ has the further representation,

$$\boldsymbol{\delta} = [\sum v_k^{-1}(\nu_k/\nu.)\mathbf{B}_k\mathbf{B}_k']^{\frac{1}{2}}\boldsymbol{\tau}, \tag{3.8}$$

where, with $\nu. = \sum \nu_k$, the $v_k = u_k/\sum u_j$ ($v_1, \cdots, v_K$ are jointly Dirichlet dis-

tributed: $v_k > 0$, $\sum v_k = 1$, *with density* $\Gamma(\frac{1}{2}\nu.) \prod_k [v_k^{\frac{1}{2}\nu_k - 1}/\Gamma(\frac{1}{2}\nu_k)]$ *in* $v_1, \cdots, v_{K-1}$, *Wilks* (1962), *pp.* 177–182), *and* $\boldsymbol{\tau}$ *has independently an r-dimensional standard multivariate t distribution* with $\nu.$ *degrees of freedom. If the matrix* $\sum \mathbf{B}_k\mathbf{B}_k'$ *is nonsingular, the distribution of* $\boldsymbol{\delta}$ *is nondegenerate with the density function,*

$$\psi(\boldsymbol{\delta}) = \Gamma[\tfrac{1}{2}(\nu. + r)]/[\pi^{\frac{1}{2}r} \textstyle\prod_k \Gamma(\tfrac{1}{2}\nu_k)]$$

$$\cdot \int_\sigma dv_1 \cdots dv_{K-1} (\textstyle\prod_k v_k^{\frac{1}{2}\nu_k - 1}) |\sum v_k^{-1}\nu_k \mathbf{B}_k\mathbf{B}_k'|^{-\frac{1}{2}} \tag{3.9}$$

$$\cdot [1 + (\textstyle\sum v_k^{-1}\nu_k\mathbf{B}_k\mathbf{B}_k')^{-1}((\boldsymbol{\delta}))]^{-\frac{1}{2}(\nu.+r)},$$

the range σ *of the* v_k *as above.*

Proof. Since each $\boldsymbol{\tau}_k = \mathbf{z}_k/(u_k/\nu_k)^{\frac{1}{2}}$, the $\mathbf{z}_k$ being independent standard normal vectors, then the conditional distribution of $\boldsymbol{\delta}$ given $u_1, \cdots, u_K$ is normal with mean $\mathbf{0}$ and variance $\sum u_k^{-1}\nu_k\mathbf{B}_k\mathbf{B}_k'$. Hence (3.7) holds. Let $u. = \sum u_k$ and write $u_k = v_k u.$. Then $u.$ is chi-squared distributed with $\nu.$ degrees of freedom, independently of the v_k (as in the discussion preceding equation (2.8)). By substituting for the u_k in (3.7), we obtain (3.8). The conditions $\sum \mathbf{B}_k\mathbf{B}_k' > 0$ and $\sum v_k^{-1}\nu_k\mathbf{B}_k\mathbf{B}_k' > 0$ (with probability one) are equivalent, by consideration of the corresponding quadratic forms.

Equation (3.9) generalizes Ruben's (1960) integral representation for the usual Behrens-Fisher densities.

4. The third identity.

4.1. *Mathematics.* Let $v_1, \cdots, v_K$ be Dirichlet distributed with parameters $b_1, \cdots, b_K$, each $\operatorname{Re}(b_k) > 0$; namely, $\sum v_k = 1$ and each $v_k > 0$, with density $B(b_1, \cdots, b_K)^{-1} \cdot \prod v_k^{b_k-1}$ in $v_1, \cdots, v_{K-1}$, where $B(b_1, \cdots, b_K) = [\prod \Gamma(b_k)]/\Gamma(\sum b_k)$ (Wilks (1962), pp. 177–82). Let u_1, $u_2 = 1 - u_1$ be Dirichlet distributed (beta distributed) with parameters a_1, $a_2 = \sum b_k - a_1$, each $\operatorname{Re}(a_i) > 0$. Then $(a_2/a_1)w = (a_2/a_1)(u_1/u_2)$ is F distributed with $\frac{1}{2}a_1$ and $\frac{1}{2}a_2$ degrees of freedom. Then given any linear form $\sum v_k z_k$,

$$\begin{aligned} E(\textstyle\sum v_k z_k)^{-a_1} &= E \textstyle\prod (u_1 z_k + u_2)^{-b_k} \\ &= E(w+1)^{\Sigma b_k} \textstyle\prod (w z_k + 1)^{-b_k}. \end{aligned} \tag{4.1}$$

The identity (4.1) in integral form is attributed to Picard by Appell and Kampé de Fériet (1926), p. 115. Each member is an integral representation of Lauricella's hypergeometric function F_D. This function has many interesting properties (Carlson (1963), and sequelae).

Theorem 3. *The following generalization of* (4.1) *holds. Given* $\operatorname{Re}(a_i) > 0$, $i = 1, \cdots, I$, $\operatorname{Re}(b_k) > 0, k = 1, \cdots, K$, *and* $\sum a_i = \sum b_k$. *Visualize a matrix* $\mathbf{Z}$ *of entries* z_{ik}. *Then,*

$$\begin{aligned} &\int_{\sigma_K} [\textstyle\prod_i (\sum_k v_k z_{ik})^{-a_i}](\prod_k v_k^{b_k-1})\, dv_1 \cdots dv_{K-1}/B(b_1, \cdots, b_K) \\ &\quad = \int_{\sigma_I} [\textstyle\prod_k (\sum_i u_i z_{ik})^{-b_k}](\prod_i u_i^{a_i-1})\, du_1 \cdots du_{I-1}/B(a_1, \cdots, a_I) \end{aligned} \tag{4.2}$$

where $\sigma_J = \{(w_1, \cdots, w_J): \sum w_j = 1$ *and each* $w_j > 0\}$. *In case* $\operatorname{Re}(\sum a_i) < \operatorname{Re}(\sum b_k)$, *define* $a_{I+1} = \sum b_k - \sum_1^I a_i$, *let* $z_{I+1,k} \equiv 1$, *and replace* I *by* $I + 1$ *in*

(4.2). *The expected-value form of* (4.2) *is an obvious analogue of* (4.1) *and can include multiple F, or "inverted Dirichlet", variables* (*Tiao and Guttman* (1965)).

PROOF. Assign the usual notation $(a, b) = \Gamma(a + b)/\Gamma(a)$. Let $z_{ik} = 1 - y_{ik}$. In a neighborhood of $z_{ik} \equiv 1$, the integrand of the left-hand member of (4.2) can be expanded, and the integration performed term-by-term,

$$\begin{aligned}\textstyle\sum_{\mathrm{each}\,m_{ik}=0}^{\infty} [\prod_i (a_i, \sum_k m_{ik})[\prod_k (1, m_{ik})]^{-1} \prod_k y_{ik}^{m_{ik}}]&\\ \cdot \textstyle\prod_k (b_k, \sum_i m_{ik})[(\sum b_k, \sum_{i,k} m_{ik})]^{-1}&\\ = \textstyle\sum_{\mathrm{each}\,m_{ik}=0}^{\infty} [\prod_k (b_k, \sum_i m_{ik})[\prod_i (1, m_{ik})]^{-1} \prod_i y_{ik}^{m_{ik}}]&\\ \cdot \textstyle\prod_i (a_i, \sum_k m_{ik})[(\sum b_k, \sum_{i,k} m_{ik})]^{-1},&\end{aligned}$$

yielding the right-hand member of (4.2), since $\sum b_k = \sum a_i$.

4.2. *Statistical applications.* The final member of equation (4.1), written as an integral, reads

$$\int_0^\infty w^{a_1-1} \prod (wz_k + 1)^{-b_k}\, dw/B(a_1, a_2), \tag{4.3}$$

which, when $b_i \equiv \frac{1}{2}(\nu + 1)$, $K = N$, and $z_k = (y_k - \hat{y}_k)^2$, we recognize as a likelihood function for errors independently Student t distributed with the unknown scale w integrated out. Known properties of Lauricella's F_D, for example differential relations, apply to inference from the likelihood function (4.3), for example maximum "likelihood" estimation of location parameters.

Savage (1966) has noted that if $v_1, \cdots, v_K$ are Dirichlet-distributed with parameters $b_1, \cdots, b_K$, then the transformed variables $\tilde{v}_k$,

$$\tilde{v}_k = v_k x_k / (\textstyle\sum_j v_j x_j), \qquad v_k = \tilde{v}_k x_k^{-1} / (\textstyle\sum_j \tilde{v}_j x_j^{-1}),$$

have density in $\tilde{v}_1, \cdots, \tilde{v}_{K-1}$,

$$(\textstyle\prod x_k^{-b_k})(\prod \tilde{v}_k^{b_k-1})(\sum \tilde{v}_j x_j^{-1})^{-\Sigma b_j}/B(b_1, \cdots, b_K).$$

Say, each $x_k > 0$; then $\tilde{v}_k > 0$ and $\sum \tilde{v}_k \equiv 1$. Then the moments of the $\tilde{v}_k$ are one-dimensional integrals, according to the integral form of the identity (4.1),

$$\begin{aligned}E \textstyle\prod \tilde{v}_k^{c_k} = B(b_1 + c_1, \cdots, b_K + c_K)[B(b_1, \cdots, b_K)B(\sum b_k, \sum c_k)]^{-1}&\\ \cdot (\textstyle\prod x_k^{-b_k}) \int_0^1 du\, u^{\Sigma b_k - 1}(1 - u)^{\Sigma c_k - 1} \prod [u(x_k^{-1} - 1) + 1]^{-(b_k+c_k)}.&\end{aligned} \tag{4.4}$$

Note the interpretation of the right-hand member as proportional to an expectation.

Note in passing that the relation between random vectors under a transformation of the above form is an equivalence relation. The transformation group is isomorphic to the quotient group of the direct sum of K copies of the multiplicative group of positive real numbers, divided by the subgroup of equicoordinate vectors $(v, v, \cdots, v)$.

The v_k are distributed like homogenized chi-squared variables,

$$v_k \sim \chi^2_{(2b_k)} / \textstyle\sum \chi^2_{(2b_j)},$$

with the usual degrees-of-freedom notation and the chi-squared variables independent (Wilks (1962)). Hence,

$$\tilde{v}_k \sim \chi^2_{(2b_k)} x_k / \sum \chi^2_{(2b_j)} x_j . \tag{4.5}$$

With this interpretation, the distribution of the $\tilde{v}_k$'s is important in some Bayesian approaches to components-of-variance problems (Hill (1965); Tiao and Box (1967)). I am grateful to George C. Tiao for pointing out this application.

Our final discussion applies to the quite general problem of inference from a multinomial observation, $n_1, n_2, \cdots, n_K$. Imagine a sequence $y_1, y_2, \cdots$ of observable independent random variables, each with unknown probability mass function $p(k)$ on $1, \cdots, K$. Then the predictive probability mass function of y_{N+1}, given $y_1, y_2, \cdots, y_N$ with cell counts $n_1, \cdots, n_K$, is, as is well known and easily derived,

$$E[p(k) \mid n_1, \cdots, n_K] = Ep(k) \prod p(j)^{n_j} / E \prod p(j)^{n_j},$$

where the expectations in the right-hand member are taken with respect to the prior distribution on the mass function p. More generally, jointly,

$$E[p(k_{N+1})p(k_{N+2}) \cdots p(k_{N+M}) \mid n_1, \cdots, n_K] \\ = Ep(k_{N+1}) \cdots p(k_{N+M}) \prod p(k)^{n_k} / E \prod p(k)^{n_k}.$$

For a prior Dirichlet distribution on p, say the $p(k)$'s distributed like the v_k's in this section (parameters $b_1, \cdots, b_K$), the needed high-order mixed moments are, of course, easily calculated in closed form,

$$\begin{aligned} E \prod p(k)^{c_k} &= B(b_1 + c_1, \cdots, b_K + c_K)/B(b_1, \cdots, b_K) \\ &= \Gamma(\sum b_k)[\Gamma(\sum b_k + \sum c_k)]^{-1} \prod \Gamma(b_k + c_k)[\Gamma(b_k)]^{-1}. \end{aligned}$$

For a prior Savage distribution on p, the needed moments are given by (4.4). Although this generalized Dirichlet distribution permits some more flexibility in the choice of a prior, still it cannot have special correlations between the probabilities of adjacent pairs of cells.

5. A grand extension. The multivariate normal, multivariate t, and Dirichlet distributions play important roles in the three theorems of this paper. It is an amusing exercise to try to replace these multivariate distributions by their matricvariate analogues: the distribution of a multivariate normal sample (James (1954)); the matricvariate t distribution (Kshirsagar (1960); Dickey (1967b)); and the multivariate beta distribution extended to any number K of matrices (Olkin and Rubin (1964)).

Acknowledgment. I am grateful to Leonard J. Savage and William A. Ericson for their patience and encouragement.

REFERENCES

ANDO, ALBERT, AND KAUFMAN, G. W. (1965). Bayesian analysis of the independent multinormal process—neither mean nor precision known. *J. Amer. Statist. Assoc.* **60** 347–58.

ANSCOMBE, F. J. (1963). Bayesian inference concerning many parameters with reference to supersaturated designs. *Bull. Inst. Internat. Statist.* **40** 721–33.

APPELL, P., AND KAMPÉ DE FÉRIET, J. (1926). *Fonctions hypergéométriques à plusieurs variables*. Gauthier-Villars, Paris.

BEHRENS, W. U. (1929). Ein Beitrag zur Fehlen-Berechnung bei wenigen Beobachtungen. *Landwirtschaftlich Jahrbuch* **68** 807–37.

CARLSON, B. C. (1963). Lauricella's hypergeometric function F_D. *J. Math. Anal. Appl.* **7** 452–70.

CORNISH, E. A. (1954). The multivariate t-distribution. *Austral. J. Phys.* **7** 531–42.

CORNISH, E. A. (1962). *The Multivariate t-distribution Associated with the General Multivariate Normal Distribution* (Technical Paper No. 13). Div. of Mathematical Statistics, Commonwealth Scientific and Industrial Organization, Melbourne.

CORNISH, E. A. (1965). A multiple Behrens-Fisher distribution. *Multivariate Analysis*, (ed. by P. R. Krishnaiah). Academic Press.

DICKEY, JAMES M. (1967a). Expansions of t densities and related complete integrals. *Ann. Math. Statist.* **38** 503–10.

DICKEY, JAMES M. (1967b). Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *Ann. Math. Statist.* **38** 511–18.

DICKEY, JAMES M. (1967c). A Bayesian hypothesis decision procedure. *Ann. Inst. Statist. Math.* **19** 367–69.

DUFF, G. F. D. (1956). *Partial Differential Equations*. Univ. Toronto Press, Toronto.

DUNNETT, CHARLES W. AND SOBEL, MILTON (1954). A bivariate generalization of Student's t-distribution, with tables for certain special cases. *Biometrika* **41** 153–67.

EDWARDS, WARD, LINDMAN, HAROLD, AND SAVAGE, LEONARD, J. (1963). Bayesian statistical inference for psychological research. *Psych. Rev.* **70** 193–242.

ERDÉLYI, ARTHUR, ET AL. (1953). *Higher Transcendental Functions*, **1** McGraw-Hill, New York.

FISHER, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugen., London.* **6** 391–8.

HILL, B. M. (1965). Inference about variance components in the one-way model. *J. Amer. Statist. Assoc.* **60** 806–25.

JAMES, ALAN T. (1954). Normal multivariate analysis and the orthogonal group. *Ann. Math. Statist.* **25** 40–75.

JEFFREYS, HAROLD (1940). Notes on the Behrens-Fisher formula. *Ann. Eugen., London* **10** 48–51.

KSHIRSAGAR, A. M. (1960). Some extensions of the multivariate t-distribution and the multivariate generalization of the distribution of the regression coefficient. *Proc. Cambridge Philos. Soc.* **57** 80–5.

OLKIN, INGRAM AND RUBIN, HERMAN (1964). Multivariate beta distributions and independence properties of the Wishart distribution. *Ann. Math. Statist.* **35** 261–69.

RAIFFA, HOWARD AND SCHLAIFER, ROBERT (1961). *Applied Statistical Decision Theory.* Harvard Univ. Press.

RUBEN, H. (1960). On the distribution of the weighted difference of two independent Student variates. *J. Roy. Statist. Soc. Ser. B.* **22** 188–94.

SAVAGE, LEONARD J. (1954). *The Foundations of Statistics*. Wiley, New York.

SAVAGE, LEONARD J. (1961). *The Subjective Basis of Statistical Practice* (unpublished notes). Univ. of Michigan.

SAVAGE, LEONARD J. (1966). Personal communication.

STEIN, C. M. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc. Ser. B.* **24** 265–85.

STEIN, C. M. (1964). Discussion on Bayesian and non-Bayesian decision theory. 109th annual meeting, Institute of Mathematical Statistics, Amherst.

TIAO, G. C. AND BOX, G. E. P. (1967). Bayesian analysis of a three-component hierarchical design model. *Biometrika* **54** 109–26.

TIAO, G. C. AND GUTTMAN, I. (1965). The inverted Dirichlet distribution with applications. *J. Amer. Statist. Assoc.* **60** 793–805.

TIAO, GEORGE C. AND ZELLNER, ARNOLD (1964). Bayes's theorem and the use of prior knowledge in regression analysis. *Biometrika* **51** 219–30.

WALLACE, DAVID L. (1958). Asymptotic approximations to distributions. *Ann. Math. Statist.* **29** 455–475.

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