Extending Families of Densities
By Special-Function Factors

By

Chong-Hong Chen and James M. Dickey*

Technical Report No. 511
School of Statistics
University of Minnesota
January 1988

*Research by Professor Dickey was supported by the U.S. National Science Foundation (Research Grant DMS-8614793).

ABSTRACT

A variety of parameterized distributions are generalized by multiplying the usual density by a generalized hypergeometric function or a multiple hypergeometric function with an appropriately transformed argument variable. A benefit of our choice of the new special-function factor is that the normalizing constant has a closed functional form. Also, from the series representation of the special function, the moments and characteristic functions of the generalized distributions are averages of the moments and characteristic functions of the usual distributions. The new random quantities, in the cases of the multiple gamma, Dirichlet, and inverted Dirichlet, have functional relations similar to the usual random quantities. The new generalized chi-squared distribution has the usual noncentral chi-squared distribution as a special case. The method, in general, preserves the property of being a Bayesian conjugate prior density.

1. Introduction

The multitude of ways to extend a family of distributions or, otherwise, develop a new parameterized family include the following:

(i). Develop the nonnull distribution of a test statistic [e.g. the noncentral chi-squared, noncentral t, noncentral F, double noncentral t, and double noncentral F distributions].

(ii) Directly introduce a further parameter [e.g. the generalized gamma distribution, with density function $P(x|\alpha,\beta,c) = cx^{c\alpha-1}e^{-x^c/\beta}/(\beta^x\Gamma(\alpha))$, $x \geq 0$].

(iii). Imbed a familiar distribution as a univariate margin of a new multivariate or matrix-variate distribution [e.g. the beta distribution, relative to the Dirichlet distribution].

(iv). Work with a function of the coordinates of a multivariate distribution [e.g. the Rayleigh distribution].

(v). Mix over a random parameter. That is, using a mixing distribution $H(\theta)$, one obtains a probability density function as $\int g(x;\theta)dH(\theta)$ or $\sum_1^c g(x;\theta_i)H(\theta_i)$, where $g(x;\theta)$ is the parameterized mixand density [e.g. the mixed normal, compound Poisson, Bayesian prior-predictive or posterior-predictive distributions such as the Dirichlet-Bernoulli distribution or Dirichlet-multinomial distribution].

(vi). Multiply by a further factor [e.g. truncated distributions, Bayesian posterior distributions].

In this article, we will extend several distributions by method (vi), using a generalized hypergeometric function or a multiple hypergeometric function as the new factor, following a suitable transformation in its argument. We will find that distributions obtained by other methods often can be rewritten in the form of method (vi). For example, the noncentral chi-squared distribution, a mixture of central chi-squared distributions with Poisson mixing, has a density expressible as the product of the central chi-squared density and a further factor - a generalized hypergeometric function. For another example, noncentarl F and double noncentral F, both defined as mixtures of central F distributions, have densities expressible as the product of a central F density and, respectively, a generalized hypergeometric function, and a multiple hypergeometric function. As a benefit of our general choice of factor and its argument, the normalizing constant will have a closed functional form, and the new moments and charasteristic functions will be averages of the familiar quantities. In particular cases, the new random variables will be seen to satisfy relations similar to familiar relations holding between the usual random variables.

We begin, Section 2, by providing a brief review of generalized hypergeometric functions and multiple hypergeometric functions. These special functions are well studied as solutions to differential equations. Their series representations and recursive integral representations will be useful for our purposes. After giving the definitions and relevant properties of the further factors, we develop the method of generalization In Section 3, we extend several families of univariate and multivariate distributions. Their properties and possible applications are given. We conclude, in Section 4, by discussing the effects of the further factors and suggesting an iterative procedure for estimating the parameters.

2. A Method of Generalization

2.1 The hypergeometric functions.

The special functions called generalized hypergeometric functions play an important role in statistics, as charateristic functions of distributions, moment-generating functions, probability-generating functions, cumulative distribution functions (Johnson & Kotz, 1970), and even probability density function themselves (Steyn, 1951). Their series representations can provide methods for calculating entities of statistical interest. This is increasingly so with the availability of computers.

The generalized hypergeometric function extends the gauss hypergeometric function by increasing the number of numerator and denominator parameters. Its series representation is

$$F[(a);(b);x] = \sum_{k=0}^{\infty} \frac{((a),k)x^k}{((b),k)k!}, \tag{2.1}$$

where the symbol (a) denotes the sequence of parameters $a_1,\ldots,a_A$, the multiple form $((a),k)$ is the product $(a_1,k)\ldots(a_A,k)$, and the Pochhammer symbol (a,r) is defined by the relations

$$\begin{aligned} (a,r) &= (a+r)/(a) \\ &= a(a+1)\ldots(a+r+1) \qquad (2.2) \\ (a,0) &= 1 \\ (a,-r) &= (-1)/(1-a,r). \end{aligned}$$

The B many denominator parameters $b_1,b_2,\ldots,b_B$ are required to take values other than negative integers. The series converges for all finite values of its argument if $A < B$, and for $|x| < 1$ if $A = B+1$.

The case $A > B+1$ is considered only if one or more of the numerator parameters is a negative integer, for which the series terminates and the question of convergence does not arise.

Many familiar functions are lower-order special cases of the generalized hypergeometric function. For example, the exponential function, ${}_0F_0(.;.;x) = e^x$, the binomial function, ${}_1F_0(a;.;x) = (1-x)^{-a}$, the Bessel function, ${}_0F_1(.;b;x) = \Gamma(b)(ix)^{(b-1)}J_{b-1}(2ix^{\frac{1}{2}})$, the incomplete gamma function, ${}_1F_1(a;a+1;-x) = ax^{-a}\gamma(a,x) = ax^{-a}\int_0^x y^{a-1}e^{-y}dy$ and the incomplete beta function, ${}_2F_1(a,1-b;x) = ax^{-a}B_x(a,b)$ $= ax^{-a}\int_0^x y^{a-1}(1-y)^{b-1}dy$.

The <u>multiple hypergeometric function</u> extends the generalized hypergeometric function by an increase in the number of argument variables. It has the following multiple series representation,

$$F^{A:B^1;\ldots;B^{(n)}}_{C:D^1;\ldots;D^{(n)}}\left[\begin{matrix}(a):(b^1);\ldots;(b^{(n)})\\(c):(d^1);\ldots;(d^{(n)})\end{matrix};\ x_1,\ldots,x_n\right]$$

$$= \sum\frac{((a),m_1+\ldots+m_n)((b^1),m_1)\ldots((b^{(n)}),m_n)x_1^{m_1}\ldots x_n^{m_n}}{((c),m_1+\ldots+m_n)((d^1),m_1)\ldots((d^{(n)}),m_n)m_1!\ldots m_n!} \qquad (2.3)$$

where the summation is over all integers m_i from 0 to infinity for each $i = 1,\ldots,n$. The most important hypergeometric functions of several variables are the Lauricella functions F_A, F_B, F_C, F_D which are the special cases defined by

(i) $$F_A(a;b_1,\ldots,b_n;c_1,\ldots,c_n;x_1,\ldots,x_n) = F^{1:1;\ldots;1}_{0:1;\ldots;1}\left[\begin{matrix}a:b_1;\ldots;b_n\\ \cdot:c_1;\ldots;c_n\end{matrix};x_1,\ldots,x_n\right] \qquad (2.4)$$

(ii) $$F_B(a_1,\ldots,a_n,b_1,\ldots,b_n;c;x_1,\ldots,x_n) = F^{0:2;\ldots 2}_{1:0;\ldots 0}\left[\begin{matrix}\cdot:(a_1,b_1);\ldots;(a_n,b_n)\\ c:\quad\cdot\quad;\ldots;\quad\cdot\end{matrix};x_1,\ldots,x_n\right]$$

(2.5)

(iii) $F_C(a,b;c_1,\dots,c_n;x_1,\dots,x_n) = F^{2:0;\dots;0}_{0:1;\dots;1}\left[\begin{matrix}(a,b): \dots\dots\\ \cdot\ :c_1;\dots;c_n\end{matrix};x_1,\dots,x_n\right]$ (2.6)

(iv) $F_D(a,b_1,\dots,b_n;c;x_1,\dots,x_n) = F^{1:1;\dots;1}_{1:0;\dots;0}\left[\begin{matrix}a:b_1,\dots;b_n\\ c:\cdot\ ,\dots,\cdot\end{matrix};x_1,\dots,x_n\right]$ (2.7)

If the number n of variables, is taken equal to two, these four functions reduce to the Appell functions F_2, F_3, F_4 and F_1, respectively; and if n = 1, all four functions become the Gauss function ${}_2F_1$. The convergence of multiple hypergeometric series was examined by Horn, 1889. In case $A+B^{(i)} < C+D^{(i)}+1$, the series converges for all finite value of x_i.

The Lauricella functions have a wide diversity of applications in statistics, mathematics, and theoretical physics, such as: an expectation associated with the multiple beta distribution; a Dirichlet distribution function associated with a transition density function in the field of genetics (for F_A); the distribution of the ratio of two Dirichlet variates and the expectation associated with the multiple gamma distribution (for F_D) (Exton, 1976).

Relevant properties of the multiple hypergeometric function include the following:

1. $F^{A:0;\dots;0}_{C:0;\dots;0}\left[\begin{matrix}(a);\cdot;\dots;\cdot\\ (c);\cdot;\dots;\cdot\end{matrix};x_1,\dots,x_n\right] = {}_AF_C((a);(c);x_1,\dots x_n)$ (2.8)

2. $F^{0:B^1;\dots;B^{(n)}}_{0:D^1;\dots;D^{(n)}}\left[\begin{matrix}\cdot:(b^1);\dots;(b^{(n)})\\ \cdot:(d^1);\dots;(d^{(n)})\end{matrix};x_1,\dots,x_n\right]$

$$= \prod_{j=1}^{n} {}_{B^{(j)}}F_{D^{(j)}}((b^{(j)});(d^{(j)});x_j) \quad (2.9)$$

3. $F^{A:1;\ldots;1}_{C:0;\ldots;0}\left[\begin{matrix}(a):b_1;\ldots,b^{(n)}\\(c):\cdot\,;\ldots;\cdot\end{matrix};x,\ldots,x\right]$

$$= {}_{A+1}F_C((a),b_1+\ldots+b^{(n)};(c);x) \tag{2.10}$$

4. $F_A(a,b_1,\ldots,b_n;c_1,\ldots,c_n;x_1,\ldots,x_n)$

$$= E(1-\sum_{j=1}^{n} U_i x_i)^{-a} \tag{2.11}$$

where U_i has the beta distribution $U_i \sim$ Beta (b_i,c_i-b_i) for $i = 1,2,\ldots,n$, independently.

5. $F_B(a_1,\ldots,a_n,b_1,\ldots,b_n;c;x_1,\ldots,x_n)$

$$= E\prod_{j=1}^{n}(1-x_jU_j)^{-b_j} \tag{2.12}$$

where $\underset{\sim}{U} = (U_1,\ldots,U_n)^T \sim$ Dirichlet $(a_1,\ldots,a_n,c-\sum_{j=1}^{n} a_j)$.

6. $F_D(a,b_1,\ldots,b_n;c;x_1,\ldots,x_n)$

$$= E(1-\sum_{1}^{n} U_j x_j)^{-a} \tag{2.13}$$

where $\underset{\sim}{U}$ has the Dirichlet distribution as in (2.12). Furthermore, the function F_D may also be represented by a single integral:

$$F_D(a,b_1,\ldots,b_n;c;x_1,\ldots,x_n)$$
$$= E\prod_{j=1}^{n}(1-x_iU)^{-b_j} \tag{2.14}$$

where U has beta distribution $U \sim$ Beta $(a,c-a)$.

2.2 The Method of generalization

By multiplying a familiar probability density function by a further factor and then renormalizing, the probability density function of a generalized distribution can be obtained. This is analogous to the operation performed in Bayes' Theorem. Let $p(x|\underset{\sim}{\theta})$ be such a generalized density function,

$$p(x|\underset{\sim}{\theta}) = p_0(x|\underset{\sim}{\theta}_0)\cdot g(z(x)|\underset{\sim}{\theta}_1)/c, \quad (2.15)$$

where $p_0(x|\underset{\sim}{\theta}_0)$ is a known probability density function having the parameter vector $\underset{\sim}{\theta}_0$ and $g(z(x)|\underset{\sim}{\theta}_1)$ is some function with argument $z(x)$ and parameter vector $\underset{\sim}{\theta}_1$. The parameter vector $\underset{\sim}{\theta}$ is a function $\underset{\sim}{\theta} = \underset{\sim}{\theta}(\underset{\sim}{\theta}_0,\underset{\sim}{\theta}_1)$ of $\underset{\sim}{\theta}_0$ and $\underset{\sim}{\theta}_1$ and c is the renormalizing constant. Here, x is a general notation for a single random variable, or a random vector. The argument $z(x)$ can be a real-valued or a vector-valued function of X.

2.3 Initial Examples:

1. Truncated distributions: the multiplier is an indicator function parameterized by the maximum and minimun possible values of the random variable, for example, the truncated normal distribution with density function,

$$p(x|\mu,\sigma,a,b) = p_0(x|\mu,\sigma)\cdot I_{[a,b]}(x)/[\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})] \quad (2.16)$$

where $p_0(x|\mu,\sigma)$ is the normal density with mean μ and standard deviation σ, and Φ is the standard normal c.d.f.

2. Rayleigh Distribution: Let $\underset{\sim}{X}$ be normally distributed in n dimensions with mean vector $\underset{\sim}{\mu}$ and covariance matrix σI_n, where σ is a positive constant and I_n is the identity matrix. Then $r = |\underset{\sim}{X}|$, the norm of $\underset{\sim}{x}$, follows the Rayleigh distribution with density function

$$p(r) = (\frac{\mu}{\sigma})(\frac{r}{\mu})^{n/2}e^{-(r^2+\mu^2)/2\sigma}\cdot I_{\frac{1}{2}(n-2)}(\frac{r\mu}{\sigma}), \quad r \geq 0 \quad (2.17)$$

where $\mu = |\underset{\sim}{\mu}|$ and $I_\nu(\cdot)$ is the modified Bessel function of the first kind of order ν.

3. Generalized beta and gamma distributions: Exton (1978) used a product of generalized hypergeometric functions of form ${}_AF_B(\ldots\ldots;z)$ as a multiplier to the density functions of a special beta distribution and the gamma distribution, respectively, to obtain new densities,

$$k^{-1}x^{d-1}\,{}_{A^1}F_{B^1}((a^1);(b^1);h_1(x)\cdots{}_{A^{(n)}}F_{B^{(n)}}((a^{(n)};(b^{(n)});h_nx),\quad 0\le x\le 1,$$

where

$$k = d^{-1}F^{1:A^1;\ldots;A^{(n)}}_{1:B^1;\ldots;B^{(n)}}\left[\begin{matrix}d:(a^1);\ldots;(a^{(n)})\\ d+1:(b^1);\ldots;(b^{(n)})\end{matrix};h_1,\ldots,h_n\right], \qquad (2.18)$$

and the density

$$k^{-1}e^{-px}x^{d-1}\,{}_{A^1}F_{B^1}((a^1);(b^1);h_1x)\cdots{}_{A^{(n)}}F_{B^{(n)}}((a^{(n)});(b^{(n)});h_nx) \qquad (2.19)$$

where

$$k = \frac{\Gamma(d)}{pd}\cdot F^{1:A^1;\ldots;A^{(n)}}_{0:B^1;\ldots;B^{(n)}}\left[\begin{matrix}d:(a^1);\ldots;(a^{(n)})\\ \cdot:(b^1);\ldots;(b^{(n)})\end{matrix};h_1/p,\ldots,h_n/p\right].$$

Properties and Requirements

1. The normal distribution, the beta distribution, and the gamma distribution are known as the conjugate prior distributions, respectively, for the normal sampling model, binomial sampling model and sampling Poisson process. There is a natural sense in which this relation extends to their generalized distributions as defined by this method. Generally, if $P_0(x|\underset{\sim}{\theta}_0)$ is a conjugate prior density for a sampling model, then the new generalized density $p(x|\underset{\sim}{\theta})$ is a family similiarly closed under sampling from the same sampling process.

2. If the further-factor function g can be expressed in series form or merely a polynomial by special choice of the numerator parameter in the generalized hypergeometric function, then the generalization can be considered as a mixture distribution. For a mixture, the moments and characteristic function are merely linear combinations of the moments or characteristic functions of the mixand distributions if such moments and characteristic functions exit. Examples include the mixed normal and the compound Poisson distributions.

3. In case the function g is a series having positive coefficients, and the argument z(x) is proportional to the kernel of the original distribution, then the generalized distribution is completely monotonic, that is, a scale-mixture of normal distributions, provided that the parent distribution is completely monotonic. This follows from the fact that the complete monotone class is closed under such summation.

4. Generally, it may be difficult to find the renormalizing constant, even though the factor function $g(z(x)|\underset{\sim}{\theta}_1)$ is a simple one. To avoid such difficulty, the choice of the function g and the argument z(x) should be made with care. In example 3, Exton used the generalized hypergeometric function as the function g to extend the beta and gamma distributions with the argument z(x) chosen proportional to the kernel of the respective beta or gamma density.

5. To form the new density (2.15), the further factor should be positive. Such a restriction can reduce the variety in the effect of the further factor. However, the numerous parameters of the special functions can allow important differences from the usual distribution. We shall consider the special functions to have positive coefficients throughout, except for cases of particular mention.

3. SPECIFIC PARAMETERIZED FAMILIES

3.1. Generalized Multiple Gamma Distribution:

We extend the multiple gamma distribution, a joint distribution of independent gamma variables with vectors of parameters $\underset{\sim}{\alpha}$, $\underset{\sim}{\beta}$, by multiplying the density by the generalized multiple hypergeometric function and then renormalizing, to yield

$$\left[\prod_{j=1}^{k} \frac{\beta_j^{\alpha_j} x_j^{\alpha_j-1} e^{-\beta_j x_j}}{\Gamma(\alpha_j)}\right] F^{A:B^1;\ldots;B^{(k)}}_{C:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1);\ldots;(b^{(k)})\\(c):(d^1);\ldots;(d^{(k)})\end{matrix};\rho_1 x_1,\ldots,\rho_k x_k\right]$$
$$\Big/ F^{A:B^1+1;\ldots;B^{(k)}+1}_{C:\,D^1\;;\ldots;\;D^{(k)}}\left[\begin{matrix}(a):(b^1),\alpha_1;\ldots;(b^{(k)}),\alpha_k\\(c):(d^1)\;;\ldots;(d^{(k)})\end{matrix};\rho_1/\beta_1,\ldots,\rho_k/\beta_k\right] \tag{3.1}$$

As a generalized multiple gamma distribution, this comprises a Bayesian conjugate prior family for the multiple Poisson sampling process, the joint distribution of K independent Poisson variables. With such a sampling model, if we denote the prior distribution by $\underset{\sim}{x} \sim MG(\underset{\sim}{\alpha},\underset{\sim}{\beta},F)$, where F denotes the new factor, then the posterior distribution takes the form,

$$\underset{\sim}{x}|\underset{\sim}{n} \sim MG(\underset{\sim}{\alpha}+\underset{\sim}{n},\underset{\sim}{\beta}+1;F) \tag{3.2}$$

The corresponding Bayesian prior predictive distribution, the marginal distribution of vector $\underset{\sim}{n}$, is obtained as a generalized multiple gamma mixture of the multiple Poisson distribution,

$$p(\underset{\sim}{n}) = E_{\underset{\sim}{x}|\underset{\sim}{\alpha},\underset{\sim}{\beta},F}(p(\underset{\sim}{n}|\underset{\sim}{x})) = \prod_{j=1}^{k} n_j^{-1} B(n_j,\alpha_j)^{-1}\left(\frac{\beta_j}{1+\beta_j}\right)^{\alpha_j}\left(\frac{1}{\beta_j+1}\right)^{n_j}$$
$$\cdot\frac{F^{A:B^1+1;\ldots;B^{(k)}+1}_{C:D^1\;;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1),\alpha_1+n_1;\ldots;(b^{(k)}),\alpha_k+n_k\\(c):(d^1)\;;\ldots;(d^{(k)})\end{matrix};\frac{\rho_1}{\beta_1+1},\ldots,\frac{\rho_k}{\beta_k+1}\right]}{F^{A:B^1+1;\ldots;B^{(k)}+1}_{C:D^1\;;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1),\alpha_1;\ldots;(b^{(k)}),\alpha_k\\(c):(d^1)\;;\ldots;(d^{(k)})\end{matrix};\frac{\rho_1}{\beta_1},\ldots,\frac{\rho_k}{\beta_k}\right]} \tag{3.3}$$

The vector $\underset{\sim}{\eta}$ no longer has independent coordinates, due to the dependent coordinates in the generalized multiple gamma distribution. Such a prior predictive distribution can be viewed as a generalized multiple negative binomial distribution if the α_i's are integers.

The characteristic function of (3.1) has the form,

$$E(e^{i\underset{\sim}{t}^T\underset{\sim}{x}}) = \left[\prod_{j=1}^{k}\left(\frac{\beta_j}{\beta_j - it_j}\right)^{\alpha_j}\right] \cdot \frac{F^{A:B^1+1;\ldots;B^{(k)}+1}_{C:D^1\;;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1),\alpha_1;\ldots;(b^{(k)}),\alpha_k \\ (c):(d^1)\;;\ldots;(d^{(k)})\end{matrix};\left(\frac{\rho_1}{\beta_1-it_1}\right),\ldots,\left(\frac{\rho_k}{\beta_k-it_k}\right)\right]}{F^{A:B^1+1;\ldots;B^{(k)}+1}_{C:D^1\;;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1),\alpha_1;\ldots;(b^{(k)}),\alpha_k \\ (c):(d^1)\;;\ldots;(d^{(k)})\end{matrix};\left(\frac{\rho_1}{\beta_1}\right);\ldots;\left(\frac{\rho_k}{\beta_k}\right)\right]} \qquad (3.4)$$

This is a product of the characteristic function of the multiple gamma distribution and a ratio of multiple hypergeometric functions. The moment generating function has a similar form.

<u>Univariate Case:</u>

For the univariate case, somewhat more generally than merely letting $k = 1$ in (3.1), we can allow the further factor to be a multiple hypergeometric function. Then the generalized gamma distribution will have the following density,

$$\frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)} \cdot F^{A:B^1;\ldots;B^{(k)}}_{C:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1);\ldots;(b^{(k)}) \\ (c):(d^1);\ldots;(d^{(k)})\end{matrix};\rho_1 x,\ldots,\rho_k x\right] \Big/ F^{A+1:B^1;\ldots;B^{(k)}}_{C\;\;:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a),\alpha;(b^1);\ldots;(b^{(k)}) \\ (c)\;\;:(d^1);\ldots;(d^{(k)})\end{matrix};\frac{\rho_1}{\beta_1},\ldots,\frac{\rho_k}{\beta_k}\right]. \qquad (3.5)$$

This yields, as special cases, Exton's generalized gamma distribution as in (2.19) for all B_i's and D_i's equal to zero, and also the noncentral chi-squared distribution, for which the density function can be expressed as the mixture of central chi-square density functions,

$$p_{\chi^2_{\delta(\lambda)}}(x) = \sum_{j=0}^{\infty} \frac{(\frac{\lambda}{2})^j e^{-\frac{1}{2}\lambda}}{j!}\, p_{\chi^2_{\delta+2j}}(x)$$

$$= \frac{(\frac{1}{2})^{\delta/2} x^{\frac{1}{2}\delta-1} e^{-\frac{1}{2}x}}{\Gamma(\frac{\delta}{2})} \cdot \sum_{j=0}^{\infty} \frac{((\frac{\lambda}{4})x)^j}{(\frac{\delta}{2},j)j!} \Big/ e^{\lambda/2}$$

$$= p_{\chi^2_{\delta}(x)} \cdot {}_0F_1(\cdot;\tfrac{\delta}{2};\tfrac{\lambda}{4}x) \Big/ {}_1F_1(\tfrac{\delta}{2};\tfrac{\delta}{2};\tfrac{\lambda}{2}). \quad (3.6)$$

We define the generalized chi-squared distribution having the density

$$\frac{x^{\frac{1}{2}\delta-1} e^{-\frac{1}{2}x}}{2^{\delta/2}\Gamma(\frac{\delta}{2})}\ {}_AF_B((a);(b);\rho x) \Big/ {}_{A+1}F_B((a),\tfrac{\delta}{2};(b);2\rho), \quad (3.7)$$

This contains (3.6) as a special case. For $\delta = 1$ in (3.7), we can define a generalized normal distribution by the transformation $y = \pm\sqrt{x}$,

$$P(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}y^2} \cdot {}_AF_B((a);(b);\tfrac{\mu}{\sigma^2}y) \Big/ {}_{2A}F_{2B}((\tfrac{a}{2}),(\tfrac{a+1}{2});(\tfrac{b}{2});(\tfrac{b+1}{2});2^{2A-2B-1}\tfrac{\mu^2}{\sigma^2}). \quad (3.8)$$

When A = B = 0, this reduces to the usual normal distribution with mean μ and variance σ^2. A sum of independent generalized chi-squared variables, each distributed as (3.7), will have a generalized gamma density, as in (3.1).

For application to Bayesian inference, in addition to the new conjugate prior family for the Poisson process, let us consider, as a sampling model, one of the generalized gamma distributions introduced by Stacy (1962). Such a conditional distribution has the density,

$$P(x|p) = \frac{cp^{w/c}}{\Gamma(\frac{w}{c})}\, x^{w-1}\exp\{-px^c\}, \quad \text{for } x > 0, \quad (3.9)$$

(w, c, and $p > 0$). This distribution includes as special cases the exponential ($p = c = 1$), the gamma ($c = 1$), and the Weibull ($w = c$).

The log-normal is also a limiting special case as $w \to \infty$. The distribution (3.9) is easily visualized in terms of $y = \log x$, the log survival time. We let the parameter p have the prior density $\pi(p)$ as follows,

$$\pi(p) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x} \prod_{j=1}^{k} {}_1F_1(b_j;d_j;\rho_j x)}{\Gamma(\alpha)\cdot F_A(\alpha;b_1,\ldots,b_k;d_1,\ldots,d_k;\frac{\rho_1}{\beta},\ldots,\frac{\rho_k}{\beta})}, \tag{3.10}$$

a special case of (3.1). The density of the posterior $\pi(p|x)$ then has the closed form,

$$\frac{(\beta+x^c)^{\frac{w}{c}+\alpha} p^{(\frac{w}{c}+\alpha)-1} e^{-(\beta+x^c)p} \prod_{j=1}^{k}\{{}_1F_1(b_j;d_j;\rho_j p)\}}{\Gamma(\frac{w}{c}+\alpha)\cdot F_A(\beta+x^c;b_1,\ldots,b_k;d_1,\ldots,d_k;\frac{\rho_1}{\beta+x^c},\ldots,\frac{\rho_k}{\beta+x^c})}. \tag{3.11}$$

The prior predictive density is given by

$$P(x) = \frac{cx^{w-1}(\beta+x^c)^{-(\frac{w}{c}+\alpha)} F_A(\beta+x^c;b_1,\ldots,b_k;d_1,\ldots,d_k;\frac{\rho_1}{\beta+x^c},\ldots,\frac{\rho_k}{\beta+x^c})}{B(\frac{w}{c},\alpha)\cdot F_A(\beta;b_1,\ldots,b_k;d_1,\ldots,d_k;\frac{\rho_1}{\beta},\ldots,\frac{\rho_k}{\beta})}. \tag{3.12}$$

3.2. Extended Dirichlet Distributions:

The moments of the Dirichlet distribution have the form of a ratio of two generalized beta functions. It is easy to derive from this the Dirichlet expectation of a multiple hypergeometric function.

THEOREM 3.1

If $\underset{\sim}{u} \sim D(\underset{\sim}{b})$, then

$$E^{(k-1)}_{\underset{\sim}{u}|\underset{\sim}{b}} F^{A:B^1;\ldots;B^{(k)}}_{C:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1);\ldots;(b^{(k)})\\(c);(d^1);\ldots;(d^{(k)})\end{matrix};\rho_1u_1,\ldots,\rho_ku_k\right]$$

$$= F^{A:B^1+1;\ldots;B^{(k)}+1}_{C+1:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1),b_1;\ldots;(b^{(k)}),b_k\\(c),\Sigma b_i:(d^1);\ldots;(d^{(k)})\end{matrix};\rho_1;\ldots;\rho_k\right] \tag{3.13}$$

again a multiple hypergeometric function.

We define the generalized Dirichlet distribution, to have the density,

$$P(\underset{\sim}{u}|\underset{\sim}{b},\Omega) = B(\underset{\sim}{b})^{-1}\prod_{i=1}^{k} u_i^{b_i-1}\, F^{A:B^1;\ldots;B^{(k)}}_{C:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1);\ldots;(b^{(k)})\\(c):(d^1);\ldots;(d^{(k)})\end{matrix};\rho_1u_1,\ldots,\rho_1u_k\right]$$

$$\Big/ F^{A:B^1+1;\ldots;B^{(k)}+1}_{C+1:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1),b_1;\ldots;(b^{(k)},b_k\\(c),b.:(d^1);\ldots;(d^{(k)}\end{matrix};\rho_1,\ldots,\rho_k\right] \qquad (3.14)$$

where $\Omega = \{(a),(c),(b^1),\ldots,(b^{(k)}),(d^1),\ldots,(d^{(k)}),\rho_1,\ldots,\rho_k\}$.

The density (3.14) can also be obtained from the generalized multiple gamma distribution (3.1), by the transformation, $u_i = x_i/\Sigma x_i$, $i = 1,2,\ldots,k$. This generalization gives Janardo (1973) and Dickey (1983) as special cases, as follows.

(i) Janardo's extended Dirichlet distribution has the density function,

$$B(\underset{\sim}{b})^{-1}\left(\prod_{i=1}^{k-1} u_i^{b_i-1}\right)\left(1-\sum_{i=1}^{k-1} u_i\right)^{b_k-1} e^{\sum_{i=1}^{k-1}\rho_i u_i}$$

$$\Big/ \Phi_2^{(k-1)}(b_1,\ldots,b_{k-1};b.;-\rho_1,\ldots,-\rho_{k-1}), \qquad (3.15)$$

where $\Phi_2^{(n)}(b_1,\ldots,b_n;c;x_1,\ldots,x_n) = \lim_{\varepsilon\to 0} F_D(\frac{1}{\varepsilon},b_1,\ldots,b_n;c;\varepsilon x_1,\ldots,\varepsilon x_n)$,

where $0 < u_i < 1$ for all i, $\sum_1^{k-1} u_i < 1$

$b_i > 0$ for all i, $b. = \sum_{i=1}^{k-1} b_i + b_k$

$\rho_i > 0$ for all i.

If $A = C = B^1 = \ldots = B^{(k)} = D^1 = \ldots = D^{(k)} = 0$, by properties (2.9), (2.2), ${}_0F_0(\cdot;\rho_1u_1,\ldots,\rho_{k-1}u_{k-1},0)$ reduces to $\exp(\sum_1^{k-1}\rho_i u_i)$.

(ii) Dickey's extended Dirichlet distribution has the density function,

$$B(\underset{\sim}{b})^{-1}\left(\prod_{i=1}^{k} u_i^{b_i-1}\right)(\underset{\sim}{u}\cdot\underset{\sim}{z})^{-\beta}\Big/R_{-\beta}(\underset{\sim}{b},\underset{\sim}{z}) \tag{3.16}$$

where $R_a(\underset{\sim}{b},\underset{\sim}{z})$ is B.C. Carlson's symmetrized multiple hypergeometric function, a special case of the Lauricella F_D function, $R_a(\underset{\sim}{b},z) = F_D^{(k)}(a,b_1,\ldots,b_k;b_1+\ldots+b_k;1-z_1,\ldots,1-z_k)$. For $A = 1$, $C = B^1 = \ldots = B^{(k)} = D^1 = \ldots = D^{(k)} = 0$, $a = \beta$ and property (2.8), $F^{1:0;\ldots;0}_{0:0;\ldots;0}\left[\begin{matrix}\beta:\cdot;\\ \cdot:\ ;\end{matrix}(1-z_1)u_1,\ldots,(1-z_k)u_k\right]$ reduces to $(\underset{\sim}{u}\ \underset{\sim}{z})^{-\beta}$. We note that Dickey's other extensions, involving several linear forms and a double average, are not special cases of (3.14).

From Theorem (3.1), we have the following expectation associated with the Lauricella functions.

<u>Corollary</u>

$$\underset{\underset{\sim}{u}|\underset{\sim}{b}}{E^{(k-1)}} F_A^{(k)}(c;d_1,\ldots,d_k;b_1,\ldots,b_k;\rho_1u_1,\ldots,\rho_ku_k)$$
$$= F_D^{(k)}(c,d_1,\ldots,d_k;b_1+\ldots+b_k;\rho_1,\ldots,\rho_k) \tag{3.17}$$

$$\underset{\underset{\sim}{u}|\underset{\sim}{b}}{E^{(k-1)}} F_D^{(k)}(b_1+\ldots+b_k;d_1,\ldots,d_k;e;\rho_1u_1,\ldots,\rho_ku_k)$$
$$= \underset{\underset{\sim}{u}|\underset{\sim}{b}}{E^{(k-1)}}\ \underset{\underset{\sim}{v}|(\underset{\sim}{d},e)}{E^{(k)}}\left(1-\sum_1^k\rho_iu_iv_i\right)^{-(b_1+\ldots+b_k)}$$
$$= F_B^{(k)}(d_1,\ldots,d_k;b_1,\ldots,b_k;e;\rho_1,\ldots,\rho_k) \tag{3.18}$$

<u>Corollary</u>

The moment of the generalized Dirichlet distribution is the product of a moment of the ordinary Dirichlet distribution and a ratio of multiple hypergeometric functions; that is,

$$\underset{\tilde{u}|\tilde{b},\Omega}{E}\left(\prod_{i=1}^{k} u_i^{n_i}\right)$$

$$= \frac{B(\tilde{b}+\tilde{n})}{B(\tilde{b})} \cdot \frac{F^{A:B^1+1;\ldots;B^{(k)}+1}_{c+1:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1),b_1+n_1;\ldots;(b^{(k)}),b_k+n_k \\ (c),b+n:(d^1);\ldots;(d^{(k)})\end{matrix};\rho_1,\ldots,\rho_k\right]}{F^{A:B^1+1;\ldots;B^{(k)}+1}_{c+1:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1),b_1;\ldots;(b^{(k)}),b_k \\ (c),b_.:(d^1);\ldots;(d^{(k)})\end{matrix};\rho_1,\ldots,\rho_k\right]} . \quad (3.19)$$

Marginal and Conditional Distributions

In general, the marginal density of a joint extended Dirichlet distribution with density function (3.14) can be derived as a series form, but not a closed (or named) functional form. However, if we restrict discussion to the special case $B^1 = \ldots = B^{(k)} = D^1 = \ldots = D^{(k)} = 0$, the density (3.14) reduces, by (2.8) to

$$B(\tilde{b})^{-1}\left[\prod_{i=1}^{k} u_i^{b_i-1}\right] {}_AF_C((a);(c);\rho_1u_1+\ldots+\rho_ku_k)$$

$$\Big/ F^{A:\ 1;\ldots;1}_{C+1:0;\ldots;0}\left[\begin{matrix}(a)\quad :b_1;\ldots;b_k \\ (c),b_.:\cdot\ ;\ldots;\cdot\end{matrix};\rho_1,\ldots,\rho_k\right]. \quad (3.20)$$

Conformably partition $\tilde{u} = (\tilde{u}^T_{(1)},\tilde{u}^T_{(2)})^T$, $\tilde{b} = (\tilde{b}^T_{(1)},\tilde{b}^T_{(2)})^T$ and $\tilde{\rho} = (\rho_1,\ldots,\rho_k)^T = (\tilde{\rho}^T_{(1)},\tilde{\rho}^T_{(2)})^T$, where $\tilde{u}_{(i)}:k_i\times 1$ $(i = 1,2)$ and $k_1+k_2 = k$. Then we have the following theorem concerning the marginal and conditional densities of (3.20).

Theorem 3.2

Suppose $\tilde{u}$ has the density function (3.20), and let $\tilde{v} = (\tilde{u}^T_{(1)},u^T_{(2).})$, $\tilde{w} = \tilde{u}_{(2)}/u_{(2).}$. Then the conditional distrubution of $\tilde{w}$ given $\tilde{v}$ has the density function,

$$P(\underset{\sim}{w}|\underset{\sim}{v}) = B(\underset{\sim}{b}_{(2)})^{-1}\left[\prod_{i=1}^{k_2} w_i^{b_{(2)i}-1}\right] \cdot {}_AF_C((a);(c);\tilde{\rho}_1 w_1,\ldots,\tilde{\rho}_{k_2} w_{k_2})$$

$$\Big/ F^{A\ :1;\ldots;1}_{C+1:0;\ldots;0}\left[\begin{matrix}(a):b_{(2)1};\ldots;b_{(2)k_2} \\ (c),b_{(2)\cdot}:\cdot;\ldots;\cdot\end{matrix};\tilde{\rho}_1,\ldots,\tilde{\rho}_{k_2}\right], \qquad (3.21)$$

and the marginal distribution of v has the density function

$$P(\underset{\sim}{v}) = B(\underset{\sim}{b}_{(1)},b_{(2)\cdot})^{-1}\left[\prod_{i=1}^{k_1} v_i^{b_{(1)i}-1}\right] v_{k_1+1}^{b_{(2)\cdot}-1}\, F^{A\ :1;\ldots;1}_{C+1:0;\ldots;0}\left[\begin{matrix}(a):b_{(2)1};\ldots;b_{(2)k_2} \\ (c),b_{(2)\cdot}:\ldots\ldots\end{matrix};\rho_1,\ldots,\rho_k\right]$$

$$\Big/ F^{A\ :1;\ldots;1}_{C+1:0;\ldots;0}\left[\begin{matrix}(a)\ :b_1;\ldots;b_k \\ (c),b\ :\ldots\ldots\end{matrix};\rho_1,\ldots,\rho_k\right]. \qquad (3.22)$$

where $\tilde{\rho}_i = \rho_1 u_1+\ldots+\rho_{k_1} u_{k_1} + u_{(2)\cdot}\rho_{k_1+1}$, $i = 1,2,\ldots,k_2$.

Proof: Indicate by (*) the different distribution for the same variables, u ~* $D(\underset{\sim}{b})$ with density $P^*(\underset{\sim}{u})$. Then the given distribution for (3.20) satisfies

$$P(\underset{\sim}{u}) = P^*(\underset{\sim}{u})\cdot {}_AF_C((a);(c);\underset{\sim}{u}^T_{(1)}\underset{\sim}{\rho}_{(1)}+\underset{\sim}{u}^T_{(2)}\underset{\sim}{\rho}_{(2)})/\text{constant}$$

where the constant denominator is

$$F^{A\ :1;\ldots;1}_{C+1:0;\ldots;0}\left[\begin{matrix}(a):b_1,\ldots,b_k \\ (c),b\cdot;\ldots\ldots\end{matrix};\rho_1,\ldots,\rho_k\right].$$

Change variables to $\underset{\sim}{w}$, $\underset{\sim}{v}$,

$$P(\underset{\sim}{w},\underset{\sim}{v}) = P^*(\underset{\sim}{w},\underset{\sim}{v})\,{}_AF_C((a);(c);\underset{\sim}{w}^T\underset{\sim}{1}_{k_2}\underset{\sim}{u}^T_{(1)}\underset{\sim}{\rho}_{(1)}+\underset{\sim}{w}^T u_{(2)\cdot}\underset{\sim}{\rho}_{(2)})/\text{const.}$$

Since $\underset{\sim}{w}|\underset{\sim}{v}$ ~* $D(\underset{\sim}{b}_{(2)})$ and $\underset{\sim}{v}$ ~* $D(\underset{\sim}{b}_{(1)},b_{(2)\cdot})$, then

$$P(\underset{\sim}{w}|\underset{\sim}{v})P(\underset{\sim}{v}) = \Bigg[P^*(\underset{\sim}{w}|\underset{\sim}{v})\,{}_AF_C((a);(c);\tilde{\rho}_1,\ldots,\tilde{\rho}_{k_2})$$

$$\Big/ F^{A\ :1;\ldots;1}_{C+1:0,\ldots,0}\left[\begin{matrix}(a):b_{(2)1};\ldots;b_{(2)k_2} \\ (c),(b_{(2)}\ :\ldots\ldots\ldots\end{matrix};\tilde{\rho}_1,\ldots,\tilde{\rho}_{k_2}\right]\Bigg]$$

$$\cdot P^*(\underset{\sim}{v})F^{A\ :1;\ldots;1}_{C+1:0;\ldots;0}\left[\begin{matrix}(a):b_{(2)1};\ldots;b_{(2)k_2} \\ (c),b_{(2)\cdot}:\ \ldots\ldots\end{matrix};\tilde{\rho}_1,\ldots,\tilde{\rho}_{k_2}\right]$$

$$\Big/ \text{constant.}$$

A special case of this theorem was obtained by Dickey for his generalizations of the Dirichlet family.

In addition to the convenient use of such a generalization for the prior and posterior distributions for multiple Bernoulli sampling, this class can also be used to model the variability of personal posterior probabilities over a population of Bayesian scientists (Dickey (1968) and Dickey and Freeman (1975)).

Univariate Case (Generalized Family of Beta Distributions)

The beta distribution is the univariate case of the Dirichlet (k=2). A generalization of the beta distribution can be obtained as the univariate case of our extended Dirichlet distribution. More generally, consider the following expectations of the beta distribution, $u \sim \text{Beta}(b_1, b_2)$,

$$E_{u|b_1,b_2}\, F^{A:B^1;\ldots;B^{(k)}}_{C:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1);\ldots;(b^{(k)})\\(c):(d^1);\ldots;(d^{(k)})\end{matrix};\rho_1 x^n,\ldots,\rho_k x^n\right]$$

$$= F^{A+n:B^1;\ldots;B^{(k)}}_{C+n:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a),\frac{b_1}{n},\ldots,\frac{b_1+n-1}{n}:(b^1);\ldots;(b^{(k)})\\(c),\frac{b_1+b_2}{n},\ldots,\frac{b_1+b_2+n-1}{n}:(d^1);\ldots;(d^{(k)})\end{matrix};\rho_1,\ldots,\rho_k\right] \quad (3.23)$$

and

$$E_{u|b_1,b_2}\, F^{A:B^1;\ldots;B^{(k)}}_{C:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a):(b^1);\ldots;(b^{(k)})\\(c):(d^1);\ldots;(d^{(k)})\end{matrix};\rho_1 x^n(1-x)^n,\ldots,\rho_k x^n(1-x)^n\right]$$

$$= F^{A+n:B^1;\ldots;B^{(k)}}_{C+n:D^1;\ldots;D^{(k)}}\left[\begin{matrix}(a),\frac{b_1}{n},\ldots,\frac{b_1+n-1}{n}:(b^1);\ldots;(b^{(k)})\\(c),\frac{b_1+b_2}{n},\ldots,\frac{b_1+b_2+n-1}{n}:(d^1);\ldots;(d^{(k)})\end{matrix};\frac{\rho_1}{4^n},\ldots,\frac{\rho_k}{4^n}\right]. \quad (3.24)$$

We can use the functions under these expectations as further factors to extend the beta distributions, and their expectations become the renormalizing constants. The generalization by (3.23) contains Exton's generalized beta distribution (2.18) as a special case, and also the noncentral beta distribution. The latter is defined as the distribution of the ratio

$$\chi'^2_{\nu_1}(\lambda)\Big/[\chi^2_{\nu_2} + \chi'^2_{\nu_1}(\lambda)], \tag{3.25}$$

and has the density function

$$e^{-\lambda/2}\sum_{j=0}^{\infty}\left(\frac{\lambda}{2}\right)^j (j!)^{-1} u^{\frac{1}{2}\nu_1+j-1}(1-u)^{\frac{\nu_2}{2}-1}\Big/B\left(\frac{\nu_1}{2}+j,\frac{2}{2}\right)$$

$$= B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)^{-1} u^{\frac{\nu_1}{2}-1}(1-u)^{\frac{\nu_2}{2}-1}\sum_{j=0}^{\infty}\frac{\left(\frac{\nu_1+\nu_2}{2},j\right)\left(\frac{\lambda u}{2}\right)}{\left(\frac{\nu_1}{2},j\right)j!}\Big/e^{\lambda/2}$$

$$= P_{u|b_1,b_2}(u)\cdot {}_1F_1\left(\frac{\nu_1+\nu_2}{2},\frac{\nu_1}{2},\frac{\lambda u}{2}\right)\Big/e^{\lambda/2}.$$

If we use the generalized chi-squared random variable instead of noncentral chi-squared in the representation (3.25), we obtain a generalized beta distribution which is a special case of (3.14) for k = 1 or (3.23) for k = 1, n = 1.

For interest, let us consider the very special case with density function

$$P(x) = \frac{x^{b_1-1}(1-x)^{b_2-1}\prod_{i=1}^{n}(1-\alpha_i x)^{-a_i}}{B(b_1,b_2)F_D^{(n)}(b_1,a_1,\ldots,a_n;b_1+b_2;\alpha_1,\ldots,\alpha_n)};\ 0 \le x \le 1. \tag{3.26}$$

The cumulative distribution function then has the following form,

$$F(x) = B(b_1,b_2)^{-1}\frac{x_1^{b_1}F_D^{(n+1)}(b_1,a_1,\ldots,a_n,1-b_2;1+b_1;\alpha_1,\ldots,\alpha_n,x)}{b_1\,F_D^{(n)}(b_1,a_1,\ldots,a_n,b_1+b_2;\alpha_1,\ldots,\alpha_n)} \tag{3.27}$$

Of course, when we let the a's equal zero, the well-known special case of the incomplete beta function is obtained,

$$B(b_1,b_2)^{-1}\frac{x^{b_1}}{b_1}\,{}_2F_1(b_1,1-b_2;1+b_1;x). \tag{3.28}$$

The moments of (3.26) are easy to derive as ratios of the Lauricella function F_D. We also have the expectation of a particular function in the same form as the further factor,

$$E\prod_{i=1}^{m}(1-\beta_i x)^{c_i} = \frac{F_D^{(n+m)}(b_1,a_1,\ldots,a_n,c_1,\ldots,c_m;b_1+b_2;\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_m)}{F_D^{(n)}(b_1,a_1,\ldots,a_n;b_1+b_2;\alpha_1,\ldots,\alpha_n)}. \tag{3.29}$$

For Bayesian inference, similarly to the Dirichlet family, the distributions of our generalized family are conjugate prior for Bernoulli sampling trials or the Binomial model. We can also consider a sampling model having the density function,

$$P(x|\beta) = e^{-\beta x}/\beta(1-\rho^{-\beta}), \quad 0 \le x \le 1, \tag{3.30}$$

the density of a truncated exponential distribution. Let the unknown parameter have the prior density $\pi(\beta)$ as in (3.26). Then the prior-predictive density is as given by Exton (1976)

$$P(x) = \sum_{m=0}^{\infty}\Phi_D^{(n+1)}(b_1-1,a_1,\ldots,a_n,\cdot;\ b_1+b_2-1;\ \alpha_1,\ldots,\alpha_n,-x-m), \tag{3.31}$$

where $\Phi_D^{(n+1)}$ is the confluent form of the Lauricella Function F_D.

3.3. The Extended Multiple F and Inverted Dirichlet Distrubutions

Let $\underset{\sim}{u} = (u_1,u_2,\ldots,u_k)^T$ have the extended Dirichlet distribution (3.14). Consider the usual transformation,

$$T: x_i = u_i/(1-u_1-\dots-u_{k-1}), \quad i = 1,2,\dots,k-1. \tag{3.32}$$

Then we have the extended inverted Dirichlet distribution with density function

$$P(\underset{\sim}{x}) = B(\underset{\sim}{b})^{-1}\left(\prod_{j=1}^{k-1} x_j^{b_j-1}\right)(1+\sum_{j=1}^{k-1} x_j)^{-b\circ}$$
$$\cdot F^{A:B^1;\dots,B^{(k)}}_{C:D^1;\dots;D^{(k)}}\left[\begin{matrix}(a):(b^1);\dots;(b^{(k)})\\(c):(d^1);\dots;(d^{(k)})\end{matrix};\frac{\rho x_1}{1+\Sigma x_i},\dots,\frac{\rho_{k-1}x_{k-1}}{1+\Sigma x_i},\frac{\rho_k}{1+\Sigma x_i}\right]$$
$$\Big/ F^{A:B^1+1;\dots;B^{(k)}+1}_{C+1:D^1;\dots;D^{(k)}}\left[\begin{matrix}(a):(b^1),b_1;\dots;(b^{(k)}),b_k\\(c),b\bullet:(d^1);\dots;(d^{(k)})\end{matrix};\rho_1,\dots,\rho_k\right]. \tag{3.33}$$

The density (3.33) can also be obtained by the transformation

$$X_i = Y_i/Y_k, \quad i = 1,2,\dots,k-1 \tag{3.34}$$

where $(Y_1,\dots,Y_k)$ has the generalized multiple gamma disrtibution (3.1). The generalized multiple F distribution $(z_1,z_2,\dots,z_k)$ can be obtained by the scale transformation of (3.33), $z_i = (b_k/b_1)x_i$, $i = 1,2,\dots,k$. The moments are ratios of multiple hypergeometric functions, similarly to (3.19).

Univariate Case:

The univariate case of the extended multiple F is of particular interest. The usual noncentral and double noncentral F distributions are both mixtures of central F distributions (Johnson & Kotz, 1970). These are both special cases of (3.33), as follows.

The extended noncentral F density is

$$P_{F'}(x) = P_F(x) \sum_{j=0}^{\infty} \frac{(\frac{\nu_1+\nu_2}{2}, j)(\frac{\frac{1}{2}\nu_1\lambda_1 x}{\nu_2+\nu_1 x})^j}{(\frac{\nu_1}{2}, j) j!} \Big/ e^{\lambda/2}$$

$$= P_F(x) \ {}_1F_1(\frac{\nu_1+\nu_2}{2}, \frac{\nu_1}{2}; \frac{\frac{1}{2}\lambda_1\nu_1 x}{\nu_2+\nu_1 x}) \Big/ e^{\lambda/2} \qquad (3.35)$$

The double noncentral F has density

$$P_{F''}(x) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} [\exp(\frac{-\lambda_1}{2})(\frac{\lambda_1}{2})^j/j!][\exp(\frac{-\lambda_2}{2})(\frac{\lambda_2}{2})^k/k!]$$

$$\cdot B(\frac{\nu_1}{2}+j, \frac{\nu_2}{2}+k)^{-1} \nu_1^{\frac{1}{2}\nu_1+j} \nu_2^{\frac{1}{2}\nu_2+k} x^{\frac{1}{2}\nu_1+j-1} (\nu_2+\nu_1 x)^{-\frac{1}{2}(\nu_1+\nu_2)-j-k}$$

$$= P_F(x) \cdot F^{1:0.0}_{0:1;1}\left[\begin{matrix} \frac{\nu_1+\nu_2}{2}; \cdot ; \cdot \\ \cdot \ ; \frac{\nu_1}{2}, \frac{\nu_2}{2} \end{matrix} ; \frac{\frac{1}{2}\nu_1\lambda_1 x}{\nu_2+\nu_1 x}, \frac{\frac{1}{2}\lambda_2\nu_2}{\nu_2+\nu_1 x}\right] \Big/ e^{\frac{1}{2}(\lambda_1+\nu_2)} . \qquad (3.36)$$

We began this section by defining the generalized multiple gamma distribution, which has dependent coordinates. Its univariate case contains as a special case the noncentral chi-squared distribution. By transforming independent chi-squared random variables, one obtains the ordinary Dichlet and inverted Dirichlet distributions. The generalized Dirichlet and generalized inverted Dirichlet distributions, as we defined them, can also be derived by the same transformations from the generalized multiple gamma distribution. Moreover, using the generalized chi-squared distribution instead of the noncentral chi-squared distribution, the transformed variables associated with noncentral beta, noncentral F, and double noncentral F turn out to be special cases of our generalized beta and generalized F distributions.

4. CONCLUSION

The method of section 2 successfully extends various distributions of multivariate form by introducing, as a further factor, a transformed special function, a generalized hypergeometric function, of multiple hypergeometric function. It is important that the further factor selected be positive. For cases when the argument variables are nonnegative, the parameters can be allowed to be negative. If the number of negative parameters is even, the coefficients of the hypergeometric function are positive, and so is the function itself. But if the number of negative parameters is odd, the signs of the coefficients of the series alternate, and then the positivity of the function is not guaranteed if the arguments can be any real value. Gasper (1975) studied the positivity of special functions by reducing them to a simpler function involving fewer parameters. The positivity of our further factor will imply the positivity of the renormalizing constant. This will allow a recurrence-relation method to prove the positivity of a special function having a few more parameters than a simple one.

In this article, we restrict our attention to the special functions having positive coefficients. Such special functions can still have a large effect on the ordinary distribution. Firstly, the effect can be a change in the parameters of the ordinary distribution. For example, in the beta distribution, Beta(α,β), the choice of ${}_1F_0(a;\cdot;x) = (1-x)^{-a}$ will change β to β-a, or the choice of ${}_1F_0(a;\cdot;(1-x)$ will change α to α-a; and in the gamma distribution, Gamma(α,β), the choice of ${}_0F_0(-;-;\rho x)$ will change β to $\beta-\rho$. Secondly, after choosing a particular special function, if we still need a further change, we can introduce two further parameters in the function, e.g. ${}_2F_1(a,b;c;\cdot)$ to ${}_3F_2(a,b,d_1;c,d_2;\cdot)$. The greater the difference is between d_1 and d_2, the greater the effect of the factor will be. To allow an even greater effect by a wider choice of factor functions, we have the following remedy: Use a linear combination of special functions as a further factor. This will allow one to choose a special function having a lower bound. For example, by working with the cosine function $\cos\sqrt{z} = {}_0F_1(\frac{1}{2},-\frac{1}{4}z) \geq -1$, or the Legendre polynomials $P_n(x) = {}_2F_1(-n,n+1;1;\frac{1-x}{2}) \geq \frac{1}{2}$, we may use $\cos\sqrt{z}+1$ and $P_n(x)-\frac{1}{2}$ as further factors. Other examples include the Hermite polynomials and Laguerre polynomials. The renormalizing constant will then be just a linear combination of the renormalizing constants for each individual term as a factor. In general, we may consider the following weighted average

$$P(x|\xi) = \sum_i P_0(x|\theta_i(\xi))\cdot\tilde{w}_i(x|\eta_i(\xi)) \tag{4.1}$$

where the ith weight is the ith linear combination,

$$\tilde{w}_i(x|\eta_i) = \sum_j c_{ij} f_j(x|\psi_j(\eta_i)) \geq 0, \quad \text{and } \sum\tilde{w}_i \equiv 1,$$

and the f_j's are particular special functions. The weights $\tilde{w}(x|\eta_i)$ would be independent of the variable x in the case of constant f_j's.

Perhaps, the most conspicuous problem when using a parameterized distribution to fit data is the estimation of the parameters. Many methods have being developed for estimation of the "ordinary" distributions, such as the frequency substitution principle, the method of moments, least squares, and maximum likelihood. The method of moments seems to offer a promising way of estimating the parameters of the "generalized" distributions. However, it turns out to be difficult to solve the resulting simultaneous equations for many parameters. A general iterative method might be suggested. One can use the usual distribution as an initial solution, check diagnostically after each fit by introducing one more parameter on a trial basis, and terminate by inspecting the successive values of a goodness-of-fit criterion. Such a method would provide a simultaneous fit and choice of model.

REFERENCES AND BIBLIOGRAPHY

Bennett, B.M. (1961). On a certain multivariate non-normal distribution. Proceedings of the Cambridge Philosophical Society, 57, 434-36.

Carlson, B.C. (1963). Lauricella's hypergeometric function F. J. Math. Analysis and Applications, Vol.7, 452-70.

Carlson, B.C. (1977). Special Functions of Applied Mathematics, Academic Press New York.

Dickey, James M. (1980). Beliefs about beliefs, a theory for stochastic assessments of subjective probabilities. Proceedings of the International Meeting on Bayesian Statistics, Valencia, Spain, May 28 - June 1, 1979, Valencia University press, 1980.

Dickey, James M. (1983). Multiple hypergeometric functions: Probabilistic interpretations and statistical uses. J. Amer. Statist. Assoc. Vol.78, 628-37.

Dickey J.M. and Chen, C.H. (1983). Direct subjective-probability modeling using ellipsoidal distributions. Bayesian Statistics 2: Proceedings of the Second International Meeting on Bayesian Statistics, (ed. by Bernardo, DeGroot, Lindley and Smith) Valencia, Spain, 6-10 Sept. 1983. North-Holland Publ. Co..

Dickey, J.M. and Freeman, P. (1975). Population-distributed personal probabilities. J. Amer. Statist. Assoc., Vol.70, No.350, 362-64.

Everitt, B.S. and Hand, D.J. (1981). Finite Mixture Distributions. Chapman and Hill, New York.

Exton, Harold (1976). Handbook of Hypergeometric Integrals: Theory, Applications, Computer Programs. John Wiley & Sons Inc., New York.

Exton, Harold (1978). Multiple Hypergeometric Functions and Applications. John Wiley & Sons Inc., New York.

Feller, William (1971). An Introduction to Probability Theory and its Applications, Vol.2, John Wiley & Sons, Inc., New York.

Gasper, George (1975). Positivity and special functions, in Richard A. Askey, ed., Theory and Application of Special Function. Academic Press, New York, 375-433.

Good, Irving John (1965). The Estimation of Probabilities. M.I.T. Press, Cambridge, Mass.

Jahnke, Eugene and Emde, Fritz (1945). Tables of Functions with Formulae and Curves. Dover Publications, New York.

Johnson, N.L. and Kotz, S. (1970). Distributions in Statistics: Continuous Univariate Distributions-2. John Wiley, New York.

Johnson, N.L. and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. John Wiley, New York.

Lebedev, N.N. (1965). Special Functions and Their Applications. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Miller, Kenneth S. (1964). Multidimensional Gaussion Distributions. John Wiley & Sons., New York.

Olkin, Ingram and Rubin, Herman (1964). Multivariate beta distributions and independence properties of the Wishart distribution. Annals of Mathematical Statistics, Vol.35, 261-69.

Stacy, E.W. (1962). A generalization of the gamma distribution. Annals of Mathematical Statistics, Vol.33, 1187-92.

Tiao, George G. and Guttman, Irwin (1965). The inverted Dirichlet distribution with applications J. Amer. Statist. Assoc., Vol.60, 793-805.