New discrete Appell and Humbert distributions with relevance to bivariate accident data

Adrienne W. Kemp

Mathematical Institute, University of St Andrews, Fife, KY16 9SS, Scotland, United Kingdom

Abstract

The paper categorizes discrete multivariate distributions into classes according to the forms of their probability generating functions, putting especial emphasis on those with pgfs involving Lauricella functions. The LPSDs are Lauricella power series distributions where the arguments of the function are proportional to the generating variables. Lauricella factorial moment distributions, LFMDs, have arguments of the form \( \lambda_i (s_i - 1) \), where \( s_i \) is a generating variable. New LFMDs are created; the differences between these and Xekalaki’s generalized Waring distribution are clarified using bivariate accident models.

© 2012 Published by Elsevier Inc.

1. Introduction

The extensive literature on bivariate discrete distributions is discussed in detail in [13] and in Johnson et al. [8] where it is shown that the joint pgf for the random variables \( X \) and \( Y \) is a key tool for deriving their properties. The simplest of these distributions are termed homogeneous; here the pgf has the form \( H(p_1 s + p_2 t)/H(p_1 + p_1) \), where \( s \) and \( t \) are the generating variables and \( p_1 \) and \( p_2 \) are parameters consistent with convergency. Homogeneity implies that the conditional distribution of \( X \), given \( X + Y = k \), is binomial with the index parameter \( k \) and probability \( p = p_1/(p_1 + p_2) \), Kemp [10]. Examples are the trinomial distribution with pgf \( G(s, t) = (1 + \alpha(s - 1) + \beta(t - 1))^k \) and the inverse sampling bivariate binomial distribution with pgf \( G(s, t) = (1 - \alpha - \beta)^k/(1 - as - bt)^k \).

Homogeneity is a restrictive property. The majority of bivariate distributions in the literature are inhomogeneous and have pgfs of the form \( G(s, t) = H(p_1 s + p_2 t + p_3 st)/H(p_1 + p_1 + p_3) \). In many cases, the pgf can be stated in terms of a generalized hypergeometric function with

\[
G(s, t) = _pF_q [a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda_1 (s - 1) + \lambda_2 (t - 1) + \lambda (st - 1)].
\]

Kumar [14] has studied their properties in depth; his table of these distributions includes the bivariate binomial, bivariate Poisson, bivariate negative binomial and bivariate logarithmic distributions. He has called these BGHFMDs (bivariate generalized hypergeometric factorial moment distributions). However, not all bivariate discrete distributions can have their pgfs expressed in this way.

This paper is concerned with distributions, whose pgfs involve Lauricella generalizations of the generalized hypergeometric function. Kocherlakota and Kocherlakota [13, p. 187] give tables of distributions where the pgfs are stated in terms of a particular Lauricella function called an \( F_1[\cdot] \) Appell function. For these distributions, the arguments of the function are...
the generating variables. Xekalaki’s [22–24] bivariate generalized Waring distribution belongs to this class. The generalized Waring distribution exists in the multivariate case with a probability generating function based on Lauricella’s hypergeometric function; see [25]. We will call multivariate distributions of these kinds of LPDS (Lauricella power series distributions). Jensen [6] gives a concise account of multivariate distributions. Johnson et al. [8] deal in depth with multivariate discrete distributions, including the multivariate hypergeometric and generalized Waring distributions.

Kemp and Kemp [11], see also [7], made a clear distinction between GHPDS (univariate generalized hypergeometric power series distributions) and GHFDs (univariate generalized hypergeometric factorial moment distributions). Some of the simplest distributions, such as the Poisson, belong to both classes, but others, such as the Poisson and beta, do not.

The new distributions in this paper are LFMDs (Lauricella factorial moment distributions). Their pgfs can be stated in terms of $F_1[\cdot]$; however, the arguments of the function are not $s_1, s_2, \ldots, s_k$, but instead are proportional to $(s_1 - 1), (s_2 - 1), \ldots, (s_k - 1)$, yielding power series for the factorial moments. Notation for the Lauricella and other higher functions is given in Section 2. LPDS’s are reviewed in Section 3, and Section 4 studies the new distributions (LFMDs).

The derivation and usefulness of the new distributions are prompted by a reexamination of Xekalaki’s [22] bivariate accident model for her bivariate generalized Waring distribution. Xekalaki’s [25] extension from a bivariate to a multivariate situation enables the accident experience of a group of individuals to be studied over a sequence of years; trends over time in accident proneness and liability may then become apparent. Similarly, LFMDs can be extended from the bivariate to the multivariate domain.

In Section 4, we assume that an individual’s proneness is different in the two periods under consideration. We note that if proneness does not alter, but the visibility bias is different in subsequent periods, then the multivariate outcome distribution is the same. Changes in methods of ascertaining data from period to period is particularly important in actuarial contexts where underreporting is a problem; see for example [12,26].

Changes over time in methods of ascertainment give rise to weighted distributions. These occur not only in actuarial studies but in many other contexts, such as environment, demography, meta-analysis, and health economics; see [16,21] for further examples.

2. Notation

The standard mathematical symbol for a rising (ascending) factorial is Pochhammer’s symbol

$$(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)} = (a+n-1)!/(a-1)!;$$

an older notation in the mathematical literature is $(a, n) = (a)_n$; it was used for example by Appell and Kemp de Fériet [1]. Notations in the statistical literature include $a(n) = (a)_n$, e.g. [13], and $a^{(n)} = (a)_n$, e.g. [8].

The notation $a^{(n)} = a(a-1) \cdots (a-n+1)$ is in general use for a falling (descending) factorial.

Lauricella functions are generalizations of the generalized hypergeometric function. For example

$$F_D[\alpha; \beta_1, \beta_2, \ldots, \beta_k; \gamma; \lambda_1 s_1, \lambda_2 s_2, \ldots, \lambda_k s_k] = \sum_{m_1, m_2, \ldots, m_k = 0}^{\infty} \frac{(\alpha)_{m_1+\cdots+m_k} (\beta_1)_{m_1} \cdots (\beta_k)_{m_k} (\lambda_1 s_1)^{m_1} \cdots (\lambda_k s_k)^{m_k}}{(\gamma)_{m_1+\cdots+m_k} m_1! \cdots m_k!}.$$ 

Appell functions are particular cases of Lauricella functions.

The product of two Gaussian hypergeometric functions is

$$2F_1[\alpha, \beta; \gamma; x] \times 2F_1[\alpha', \beta'; \gamma'; y] = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\alpha)_{m+n} (\beta)_{m+n} x^n y^m}{(\gamma)_{m+n} m! n!}.$$ 

Appell (1855–1930) replaced one or more of the products $(\eta)_{m+n} \eta_n$ by $(\eta)_{m+n}$, giving rise to four different Appell functions in two variables with seven confluent (Humbert) forms.

The Appell function of interest in this paper is

$$F_1[\alpha; \beta; \gamma; x, y] = \sum_{m=0}^{\infty} \frac{\Gamma(\gamma) \Gamma(\gamma-\alpha)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(\gamma)} \int_0^1 u^{\alpha-1} (1-u)^{\beta-1} (1-ux)^{-\beta} (1-uy)^{-\beta} du. \quad (Picard’s \ integral)$$

For particular values of the parameters are:

$$F_1[\alpha; \beta; \gamma; x, y] = (1-x)^{-\beta} (1-y)^{-\beta}$$

$$F_1[\alpha; \beta; \gamma; x, 0] = 2F_1[\alpha, \beta; \gamma; x]$$
The corresponding factorial moment generating function is a power series obtained by setting 
\( s \) of accidents \( X \) in (13).

\[ F_{1}[\alpha; \beta, \beta'; \gamma; x, 1] = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta')}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta')}F_{1}[\alpha, \beta; \gamma - \beta'; x] \]  
\[ F_{1}[\alpha; \beta, \beta'; \gamma; x, x] = SF_{1}[\alpha, \beta + \beta'; \gamma; x]; \tag{7} \]
see [1, Ch. II] and [2, Ch. 5]. The function \( F_{1}[\cdot, \cdot] \) converges for \( |x| < 1, |y| < 1 \).

A confluent (Humbert) form of it is

\[ \Phi_{1}[\alpha; \beta; \gamma; x, y] = \lim_{\epsilon \to 0} F_{1}[\alpha; \beta, 1/\epsilon; \gamma; x, ey] = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\alpha)_{m+n} (\beta)_{m} x^{m} y^{n}}{(\gamma)_{m+n} m! n!} \int_{0}^{1} u^{\alpha-1}(1-u)^{\gamma-\alpha-1}(1-ux)^{-\beta} e^{uy} du. \]  
\[ \tag{8} \]
Humbert functions are discussed in [1], but the readers are warned that there is a typo in the definition of \( \Phi_{1}[\cdot] \) in [2]. Other books that deal with them are Slater [17], Srivastava and Manocha [18], and Mathai [15].

### 3. Lauricella Power Series Distributions (LPSDs)

The joint distribution of two independent negative binomial variables is LPSD, since the pgf can be expressed as

\[ G(s_{1}, s_{2}) = (1 - q_{1}s_{1})^{-b}(1 - q_{2}s_{2})^{-b'}/(1 - q_{1})^{-b}(1 - q_{2})^{-b'} = F_{1}[\alpha; b, b'; a; q_{1}s_{1}, q_{2}s_{2}]/F_{1}[a, b, b'; a; q_{1}, q_{2}] \]  
\[ \tag{9} \]
where \( 0 < q_{1} < 1, 0 < q_{2} < 1 \) and \( F_{1}[\cdot] \) is the Appell Type 1 function. More interesting examples of LPSDs have been studied in depth by Janardan and Patil [5].

Chance mechanisms for various multivariate hypergeometric-type distributions are discussed in [4]. Early researchers, e.g. [19,20], were mainly concerned with the derivation of their properties via their probability mass functions. Janardan and Patil [5] advanced the subject considerably by introducing a unified multivariate hypergeometric distribution with \( k \) variables and pgf

\[ G(s_{1}, s_{2}, \ldots, s_{k}) = \frac{(a_{0})_{n} F_{0}_{\alpha}[-n; -a_{1}, -a_{2}, \ldots, -a_{k}; a_{0} - n + 1; s_{1}, s_{2}, \ldots, s_{k}]}{(a_{1} + a_{2} + \ldots + a_{k})_{n} F_{1}[a, b, b'; a; q_{1}, q_{2}]}. \]  
\[ \tag{10} \]
where \( F_{0}_{\alpha}[-] \) is a Lauricella function. Their bivariate unified hypergeometric distribution has the pgf

\[ G(s, t) = \frac{(a_{3})_{n} F_{1}[a, b, c; a_{3} - n + 1; s, t]}{(a_{1} + a_{2} + a_{3})_{n} F_{2}[a, b, c; a_{1}, a_{2}, a_{3}]} \]  
\[ \tag{11} \]
they showed that the bivariate hypergeometric, bivariate inverse hypergeometric, bivariate negative hypergeometric, bivariate inverse negative hypergeometric, bivariate Polya and bivariate inverse Polya distributions all have pgfs that can be obtained from (11) by a suitable choice of parameters for \( a_{1}, a_{2} \) and \( a_{3} \). The marginal distributions of \( X, Y, X + Y \) are univariate hypergeometric type distributions; further properties of the bivariate unified hypergeometric distribution are discussed by Kocherlakota and Kocherlakota [13] and Johnson et al. [8]. Xekalaki’s [22–24] bivariate generalized Waring distribution belongs to this family.

A more general type of LPSD can be obtained by subjecting observations to visibility bias. Suppose that larger values are more difficult to observe and that the joint observation \( \{x_{1}, x_{2}, \ldots, x_{k}\} \) is seen with probability \( p_{1}^{x_{1}}p_{2}^{x_{2}} \ldots p_{k}^{x_{k}} \). Then the distribution (10) becomes a weighted distribution with pgf

\[ G(s_{1}, s_{2}, \ldots, s_{k}) = \frac{F_{0}_{\alpha}[-n; -a_{1}, -a_{2}, \ldots, -a_{k}; a_{0} - n + 1; p_{1}s_{1}, p_{2}s_{2}, \ldots, p_{k}s_{k}]}{F_{0}_{\alpha}[-n; -a_{1}, -a_{2}, \ldots, -a_{k}; a_{0} - n + 1; p_{1}, p_{2}, \ldots, p_{k}]}. \]  
\[ \tag{12} \]

### 4. Lauricella Factorial Moment Distributions (LFMDs)

These are defined to have pgfs of the form

\[ G(s_{1}, s_{2}, \ldots, s_{k}) = \frac{F_{0}_{1}[a; b_{1}, b_{2}, \ldots, b_{k}; p; \lambda_{1}(s_{1} - 1), \lambda_{2}(s_{2} - 1), \ldots, \lambda_{k}(s_{k} - 1)]}{F_{0}_{1}[a; b_{1}, b_{2}, \ldots, b_{k}; p; \lambda_{1}, \lambda_{2}, \ldots, \lambda_{k}]}. \]  
\[ \tag{13} \]
The corresponding factorial moment generating function is a power series obtained by setting \( s_{i} - 1 = t_{i}, i = 1, \ldots, k \) in (13).

Xekalaki’s [22] bivariate generalized Waring distribution has an underlying proneness-liability type model for numbers of accidents \( X \) and \( Y \) in two time periods. The joint distribution for \( X \) and \( Y \) is assumed to have a double negative binomial pgf

\[ G(s_{1}, s_{2}|v) = (1 + v - vs_{1})^{-\beta}(1 + v - vs_{2})^{-\kappa} \]
where \( v \) is a proneness parameter. She assumed that individuals have the same proneness \( v \) in the two periods, but that \( v \) varies over individuals according to a type-two beta distribution. The pgf for the outcome joint distribution is then

\[
G(s_1, s_2) = \int_0^\infty (1 + v - vs_1)^{-\beta}(1 + v - vs_2)^{-\xi} \times \frac{\Gamma(\rho + \alpha)\nu^{a-1}dv}{\Gamma(\rho)\Gamma(\alpha)\nu^{a+\rho}} = \frac{(\rho)^{\beta+\xi}}{(a + \rho)^{\beta+\xi}}F_1[\alpha; \beta, \xi; a + \rho + \beta + \xi; s_1, s_2].
\]

Further models for the bivariate generalized Waring distribution are given in [23]; factorial moment estimation is discussed in [24]. The properties of the distributions are described in [13,8].

Suppose that in the first time period, \( \nu = 1 \), \( \theta = \xi = \gamma \), and that \( \phi = \xi = \gamma \) are both known. The factorial moment generating function is

\[
G(s_1, s_2) = \sum_{m=0}^\infty \sum_{n=0}^\infty \frac{(\alpha)s_n(\beta)m(\xi)n\theta_1^m(s_1 - 1)^m\theta_2^n(s_2 - 1)^n}{(\gamma)^{m+n}m!n!} = F_1[\alpha; \beta, \xi; \gamma; \theta_1(s_1 - 1), \theta_2(s_2 - 1)].
\]

The factorial moment generating function is

\[
G(t_1, t_2) = F_1[\alpha; \beta, \xi; \gamma; \theta_1t_1, \theta_2t_2],
\]

whence \( \mu'_{[i,j]} = (\alpha)_{i+j}(\beta)_{i}(\xi)\theta_1^i\theta_2^j(\gamma)^{i+j} \). The marginal distribution of \( X_1 \) has the pgf

\[
G(s_1, 1) = 2F_1[\alpha; \beta, \gamma; \theta_1(s_1 - 1)]
\]

with \( \text{mgf}(t) = 2F_1[\alpha; \beta, \gamma; \theta_1t] \). This is the type \( H_2 \) distribution of Gurland [3] and Katti [9]. It is GHFMD; see Section 2.4.2 in [7]. Similarly, the marginal distribution of \( X_2 \) has the pgf \( G(1, s_2) = 2F_1[\alpha, \xi; \gamma; \theta_2(s_2 - 1)] \). Expressions for the factorial moments of these marginal distributions are therefore straightforward and can be used for moment estimation.

The distribution of \( X_1 + X_2 \) has the pgf

\[
G(s_1, s_2) = F_1[\alpha; \beta, \xi; \gamma; \theta_1(s_1 - 1), \theta_2(s_2 - 1)]
\]

with \( \text{mgf}(t) = 2F_1[\alpha; \beta, \gamma; \theta_1t] \). The type \( H_2 \) distribution of Gurland [3] and Katti [9]. It is GHFMD; see Section 2.4.2 in [7]. Similarly, the marginal distribution of \( X_2 \) has the pgf \( G(1, s_2) = 2F_1[\alpha, \xi; \gamma; \theta_2(s_2 - 1)] \). Expressions for the factorial moments of these marginal distributions are therefore straightforward and can be used for moment estimation.

The distribution of \( X_1 + X_2 \) has the pgf

\[
G(s_1, s_2) = F_1[\alpha; \beta, \xi; \gamma; \theta_1(s_1 - 1), \theta_2(s_2 - 1)]
\]
if \(v\) again has a type-one beta distribution then the joint outcome pgf is

\[
G^*(s_1, s_2) = \int_0^1 \left(1 + \theta_1 v - \theta_1 v s_1^{-1}\right)^{-\beta} \exp\left[\xi v (s_2 - 1)\right] \times \frac{v^{\alpha-1}(1-v)^{\gamma-1}}{B(\alpha, \gamma)} \, dv
\]

\[
= \Phi_1(\alpha; \beta; \gamma; \theta_1 (s_1 - 1), \zeta (s_2 - 1))
\]

\[
= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left(\begin{array}{c}
\frac{(\alpha)_{m+n}\beta m\theta_1^m (s_1 - 1)^m \zeta^n (s_2 - 1)^n}{(\gamma)_{m+n} m! n!}
\end{array}\right).
\]

(20)

The factorial moment generating function is now \(G^*(t_1, t_2) = \Phi_1(\alpha; \beta; \gamma; \theta_1, t_1, t_2), \) and \(\mu'_{(i,j)} = (\alpha)_i (\beta)_j (\gamma)_i (\gamma)_j.\) The marginal distributions of \(X\) and \(Y\) are type \(H_2\) and Poisson \(\land\) beta, respectively.

If the independent negative binomial distributions for \(X\) and \(Y\) tend to independent Poisson distributions with parameters \(\eta\) and \(\zeta\), where \(v\) has a type-one beta distribution, then the joint outcome distribution is a homogeneous bivariate Poisson \(\land\) beta distribution with pgf

\[
G^{**}(s_1, s_2) = \int_0^1 \exp[\eta v (s_1 - 1)] \exp[\xi v (s_2 - 1)] \times \frac{v^{\alpha-1}(1-v)^{\gamma-1}}{B(\alpha, \gamma)} \, dv
\]

\[
= \Phi_1(\alpha; \gamma; \eta (s_1 - 1) + \zeta (s_2 - 1)).
\]

(21)

In memoriam.

The paper is written in memory of Sam Kotz and also Paul Appell who died in 1930, the year in which Sam and I were born.

References