Asian Journal of Statistical Sciences
Vol. 1, No. 1, (2021), pp. 01-07
© ARF India, Gurgaon, India

ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

On bivariate and multivariate gamma difference distributions

Barry C. Arnold

Department of Statistics, University of California, Riverside, USA.

ARTICLE HISTORY
Compiled June 20, 2021

Rceived 28 December 2020; Accepted 28 May 2021

ABSTRACT
A multivariate model involving differences of independent gamma distributed components was introduced by Arnold (2020) The present paper provides a more detailed discussion of this model.

1. Introduction

Arnold and Ng (2011) introduced a bivariate second kind beta or beta(2) distribution involving 8 independent random variables with gamma distributions (subsequently such random variables will be referred to as gamma components). It was identified as the most general bivariate model whose marginals are ratios of sums of independent gamma variables. The model involves 8 independent components $U_1, U_2, ..., U_8$ with $U_j \sim \Gamma(\delta_j, 1), \quad j = 1, 2, ..., 8$. The two-dimensional random vector (X, Y) is then defined by

$$\begin{aligned} X &= \frac{U_1 + U_5 + U_7}{U_3 + U_6 + U_8}, \\ Y &= \frac{U_2 + U_6 + U_7}{U_4 + U_5 + U_8}. \end{aligned} \tag{1}$$

This defines an 8-parameter family of bivariate distributions with beta(2) marginal distributions. If (X, Y) is defined as in (1) then we write: $(X, Y) \sim BB(2)(\underline{\delta})$, to be read as (X, Y) has a bivariate second kind beta distribution with parameter vector $\underline{\delta}$.

The flexible Arnold-Ng model subsumed and extended several previously available bivariate beta(2) models. The construction of this bivariate model relies on two well known facts. (1) A sum of independent gamma variables with a common scale parameter has again a gamma distribution. (2) Ratios of independent gamma variables with a common scale parameter have beta distributions of the second kind.

Why there are 8 U_j's and where they are located in the model, may require some explanation. There are four locations where a particular U_j may be placed. (1) In the

CONTACT Barry C. Arnold. Email: barnold@ucr.edu

numerator of X. (2) In the denominator of X. (3) In the numerator of Y and (4) In the denominator of Y. The variables U_1, U_2, U_3 and U_4 appear only once and each one of them appears in only one of the four possible locations. A variable U_j cannot appear in both the numerator and denominator of X, nor of Y, since otherwise the independence of numerators and denominators, required for beta(2) marginals, would be destroyed. U_5 appears in the numerator of X and in the denominator of Y. U_6 appears in the denominator of X and the numerator of Y. U_7 appears in both numerators, while U_8 appears in both denominators. No U_j can appear in 3 or in 4 of the possible locations, since that would destroy the required independence of at least one numerator and its corresponding denominator. If an additional independent gamma variable is introduced in one or two permissible locations in (1) then it can be combined with one of the existing 8 U_j's and no enrichment of the model will result. Thus for example, if U_9 is added to both numerators, then $U_7 + U_9$ will continue to play the role of U_7 with an adjusted shape parameter $\delta_7 + \delta_9$.

We adopt the convention that a random variable with a $\Gamma(\delta, 1)$ distribution with $\delta = 0$ will be defined to be a random variable that is degenerate at 0, By setting some of the δ_j's in the Arnold-Ng model (1) equal to zero, simplified submodels (some of which have been discussed in the literature) will be obtained. Note that after setting certain δ_j's equal to zero, we must retain $\delta_1+\delta_5+\delta_7 > 0$, $\delta_3+\delta_6+\delta_8 > 0$, $\delta_2+\delta_6+\delta_7 > 0$, and $\delta_4 + \delta_5 + \delta_8 > 0$, in order to continue to have beta(2) marginal distributions.

2. The bivariate gamma-difference model

If W_1 and W_2 are independent gamma distributed random variables with $W_j \sim \Gamma(\delta_j, 1)$ $j = 1, 2$, then the difference $Z = W_1 - W_2$ has what is known as a gamma-difference distribution with parameters δ_1 and δ_2, and we write $Z \sim GD(\delta_1, \delta_2)$. Klar (2015) provides details, possible applications and historical perspective on this model. Many early appearances of the model dealt with the symmetric case in which $\delta_1 = \delta_2$.

If $Z \sim GD(\delta_1, \delta_2)$ then its moment generating function is of the form

$$M_Z(t) = (1-t)^{-\delta_1}(1+t)^{-\delta_2}, \quad |t| < 1. \tag{2}$$

The moments of Z could be obtained from this moment generating function,or they may more easily be obtained by expanding $(W_1 - W_2)^k$, where k is a positive integer, and using available expressions for gamma moments. The mean, variance and skewness of Z are thus, respectively:

$$\delta_1 - \delta_2 \qquad \delta_1 + \delta_2 \qquad \frac{2(\delta_1 - \delta_2)}{(\delta_1 + \delta_2)^{3/2}} \tag{3}$$

The distribution of Z is asymmetric unless $\delta_1 = \delta_2$. Klar provided an expression for the density of Z that involves Whittaker-W functions.

Of course, if δ_1 and δ_2 are positive integers, successive integration by parts can be used to evaluate the density.

The representation $Z = X_1 - X_2$ not only allows for ready computation of the moments of Z, but also permits straightforward simulation of realizations from its distribution.

A simple bivariate versions of the GD distribution can be developed using the "variables in common" methodology. For it, we can begin with 3 independent gamma variables U_1, U_2 and U_3 with $U_j \sim \Gamma(\delta_j, 1), \quad j = 1, 2, 3$. Then define $X = U_1 - U_3$ and $Y = U_2 - U_3$.

Instead, we may develop a more flexible bivariate model in a manner analogous to that used in the development of the Arnold-Ng bivariate beta(2) model. We thus begin with 8 independent gamma variables $U_1, U_2, ..., U_8$ with $U_j \sim \Gamma(\delta_j, 1), \quad j = 1, 2, ..., 8$.. We then define (X, Y) by

$$\begin{aligned} X &= (U_1 + U_5 + U_7) - (U_3 + U_6 + U_8), \\ Y &= (U_2 + U_6 + U_7) - (U_4 + U_5 + U_8), \end{aligned} \tag{4}$$

analogous to (1). If (X, Y) is as defined by (4) then we write $(X, Y) \sim BGD(\underline{\delta})$, to be read as a bivariate gamma-difference distribution with parameter vector $\underline{\delta}$. Clearly (X, Y) has gamma-difference marginals. The BGD distribution is identifiable since each U_j plays a different role in the construction. Moreover, the BGD model represents the most general model in which both X and Y are linear combinations of independent gamma variables with all coefficients equal to 1 or -1. Of course submodels of (4), in which certain δ_j's are set equal to 0, with the usual convention that the corresponding U_j's are equal to 0 with probability 1, may frequently be found adequate to model particular data sets. For example, the three parameter "variables in common" model described in the previous paragraph is identifiable as a special case obtained by setting 5 of the parameters equal to 0. It exhibits a limited range of correlation values (only non-negative ones). Alternatively, it is sometimes appropriate to impose linear constraints on the δ_j's to arrive at a model with a parameter space of reduced dimension. An example, in which linear constraints are imposed is one which can be identified as having asymmetric Laplace marginal distributions.

Moments of the BGD distribution are readily evaluated, since they are functions of available gamma moments. For example we have:

$$E(X) = \delta_1 + \delta_5 + \delta_7 - \delta_3 - \delta_6 - \delta_8, \tag{5}$$
$$E(Y) = \delta_2 + \delta_6 + \delta_7 - \delta_4 - \delta_5 - \delta_8, \tag{6}$$
$$var(X) = \delta_1 + \delta_5 + \delta_7 + \delta_3 + \delta_6 + \delta_8, \tag{7}$$
$$var(Y) = \delta_2 + \delta_6 + \delta_7 + \delta_4 + \delta_5 + \delta_8, \tag{8}$$

and

$$cov(X, Y) = -\delta_5 - \delta_6 + \delta_7 + \delta_8. \tag{9}$$

An attractive feature of this bivariate model is that a simple expression is available for the covariance and correlation. It is clearly possible to have a full range of correlations in the BGD model. Zero correlation can occur even though X and Y are dependent. It only requires that $-\delta_5 - \delta_6 + \delta_7 + \delta_8 = 0$. Independence will be the case if $\delta_5 = \delta_6 = \delta_7 = \delta_8 = 0$.

3. Possible extensions involving dependent component variables

The flexible bivariate gamma difference model has basic components consisting of 8 independent variables having gamma distributions. It is intriguing to speculate regarding the consequences of allowing some degree of dependence among the 8 gamma components. In an extreme case, one might assume that the vector $\underline{U}$ has one of the many available multivariate gamma distributions. The dimension of the parameter spaces of such models would typically be unacceptably large. Moreover, in most such models we would no longer have marginal distributions of the gamma difference form.

One possible model which incorporates a mild degree of dependence among the U_j's while retaining gamma difference marginals may be described as follows. Assume that (X, Y) is defined by

$$\begin{aligned} X &= (U_1 + U_5 + U_7) - (U_3 + U_6 + U_8), \\ Y &= (U_2 + U_6 + U_7) - (U_4 + U_5 + U_8), \end{aligned} \tag{10}$$

where now $(U_1, U_2), (U_3, U_4), U_5, U_6, U_7, U_8$ are independent variables with (U_1, U_2) and (U_3, U_4) having bivariate gamma distributions (with gamma marginals and unit scale parameters), while $U_j \sim \Gamma(\delta_j, 1), \quad j = 5, 6, 7, 8$.

In order for (10) to constitute a genuine extenssion of the BGD model, it is necessary that at least one of the bivariate gamma distributions involved in its definition is not itself of the BGD form. If both are of the BGD form then the resulting model will simplify to become again of the BGD form.

4. On multivariate versions of the gamma difference model

k-dimensional versions of the Arnold-Ng beta(2) distribution were mentioned in Arnold and Ghosh (2014), in a context of copula models. We will describe the analogous approach to develop the k-variate gamma difference distribution. First we consider the three dimensional case. It will then be evident how to deal with higher dimensions.

A three dimensional gamma difference distribution will be one whose structure is of a form which involves 26 independent gamma distributed U_j's. This is the appropriate number of gamma distributed components since a trivariate model (X, Y, Z) expressed as differences of two independent sums of independent gamma variables (with unit scale parameter), will involve 6 places where a particular U can appear, three places in the first sums and three places in the subtracted second sums. But a particular U cannot appear in both the first sum and the subtracted second sum of any of the three variables X, Y and Z, There will be 6 U's which appear in just one of the 6 possible places. These will be denoted by $U_1, U_2, ..., U_6$. There will be 12 U's that appear in exactly two of the 6 possible positions, denoted by $U_7, U_8, ..., U_{18}$. Finally there are 8 U's that appear in 3 places, namely $U_{19}, U_{20}, ..., U_{26}$. No U can appear in more than 3 places without violating the requirement that first sums must be independent of their corresponding subtracted second sums.

Thus, there are a total of 26 parameters in the model where $U_j, \ j = 1, 2, ..., 26$ are independent variables with $U_j \sim \Gamma(\delta_j, 1)$ for each j. The model can then be expressed in the following form.

$$\begin{aligned} X &= (U_1+U_7+U_8+U_9+U_{10}+U_{19}+U_{20}+U_{21}+U_{22}) \\ &\quad -(U_4+U_{11}+U_{12}+U_{13}+U_{14}+U_{23}+U_{24}+U_{25}+U_{26}), \end{aligned} \tag{11}$$

$$\begin{aligned} Y &= (U_2+U_7+U_{11}+U_{15}+U_{16}+U_{19}+U_{20}+U_{23}+U_{24}) \\ &\quad -(U_5+U_9+U_{13}+U_{17}+U_{18}+U_{21}+U_{22}+U_{25}+U_{26}), \end{aligned} \tag{12}$$

and

$$\begin{aligned} Z &= (U_3+U_8+U_{12}+U_{15}+U_{17}+U_{19}+U_{21}+U_{23}+U_{25}) \\ &\quad -(U_6+U_{10}+U_{14}+U_{16}+U_{18}+U_{20}+U_{22}+U_{24}+U_{26}). \end{aligned} \tag{13}$$

The pattern for the dimensions of the parameter spaces of the multivariate models can now be recognized. The univariate model involves 2 U's, i.e., 3^1-1. The bivariate model involves 8 U's, i.e., 3^2-1. The trivariate case involves 26 U's, i.e., 3^3-1, and, in general, the k-dimensional model involves 3^k-1 U's.

Use of the fully parameterized k-dimensional model would almost never be recommended. Instead simplified sub-models, obtained by setting many of the δ's equal to zero, can be expected to be adequate for many data sets.

5. Parameter estimation

If a sample is available from the bivariate gamma difference distribution (BGD) with a full array of 8 parameters, the absence of a density function will rule out using a maximum likelihood approach for parameter estimation. What we do have available are relativly simple expressions for moments and mixed moments of the coordinate random variables. In principle then we could choose 8 sample moments and/or mixed moments, equate them to their expectations and solve the resulting 8 equations for the δ_i parameters. This, in many cases, will prove to be a non-trivial exercise. However, it will typically be the case that simplified sub-models involving only a few of the δ_i's will be utilized. With fewer equations to deal with, the method of moments approach is often not difficult to implement. We can illustrate this with two examples. As usual, even in more complicated cases, the estimates obtained will be consistent and jointly asymptotically normal.

Example 5.1. As a first example, consider the very simple model in which $\delta_3=\delta_4=\delta_5=\delta_6=\delta_7=0$. so that the three parameter model is of the form

$$X=U_1-U_8,$$

$$Y=U_2-U_8.$$

In this case we can set up the following moment equations:

$$M_X = (1/n)\sum_{i=1}^{n} X_i = \delta_1 - \delta_8, \qquad M_Y = (1/n)\sum_{i=1}^{n} Y_i = \delta_2 - \delta_8,$$

$$M_{XY} = (1/n)\sum_{i=1}^{n} X_i Y_i = \delta_1\delta_2 - \delta_1\delta_8 - \delta_2\delta_8 + 2\delta_8^2.$$

These can be readily solved to yield the following method of moments estimates for the three parameters in the model, i.e.,

$$\widetilde{\delta}_8 = \sqrt{M_{XY} - M_X M_Y},$$

$$\widetilde{\delta}_1 = M_X + \widetilde{\delta}_8, \qquad \widetilde{\delta}_2 = M_Y + \widetilde{\delta}_8.$$

Example 5.2. As a second example consider a 4-parameter model in which $\delta_1 = \delta_2 = \delta_3 = \delta_4 = 0$. The model is thus of the form

$$X = (U_5 + U_7) - (U_6 + U_8)$$

$$Y = (U_6 + U_7) - (U_5 + U_8).$$

In this case, observe that if we define $W_i = (X_i + Y_i)/2$ and $Z_i = (X_i - Y_i)/2$, then $E(W_i) = \delta_7 - \delta_8$ and $var(W_i) = \delta_7^2 + \delta_8^2$. Similarly $E(Z_I) = \delta_5 - \delta_6$ and $var(Z_i) = \delta_5^2 + \delta_6^2$. We may then set up the equations:

$$M_W = \delta_7 - \delta_8, \qquad S_W^2 = \delta_7^2 + \delta_8^2,$$

$$M_Z = \delta_5 - \delta_6, \qquad S_Z^2 = \delta_5^2 + \delta_6^2,$$

and solve to obtain the following estimates for the parameters:

$$\widetilde{\delta}_7 = M_W + \sqrt{2S_W^2 - M_W^2}, \qquad \widetilde{\delta}_8 = \widetilde{\delta}_7 - M_W,$$

$$\widetilde{\delta}_5 = M_Z + \sqrt{2S_Z^2 - M_Z^2}, \qquad \widetilde{\delta}_6 = \widetilde{\delta}_5 - M_Z.$$

Examp;es in which 5 or more of the δ's are non-zero may require iterative solution of the moment equations, but except for that, they can be expected to yield reasonable estimates of the parameters.

Klar (2015), observing the simplicity of the expression for the characteristic function of a gamma-difference variable, suggested a parameter estimation strategy using the empirical characteristic function. In higher dimensional cases the joint characteristic function could be utilized for estimation in parallel fashion.

6. Remarks

The present paper has provided an introduction to a broad spectrum of bivariate gamma-difference models and sub-models which can potentially be useful additions to the modeler's tool kit. These new flexible models can be expected to find application in cases in which the simpler well-known models prove to be inadequate to adapt to particular data sets. It will be unlikely that the full 8 parameter model will be frequently deemed appropriate.

References

[1] Arnold, B. C. (2020). Some bivariate and multivariate models involving independent gamma distributed components. To appear.

[2] Arnold, B.C., Ghosh, I. (2014). Bivariate Kumaraswamy models involving use of Arnold-Ng copulas. Journal of Applied Statistical Sciences, 22, 227-240.

[3] Arnold, B.C., Ng, H.K.T. (2011). Flexible bivariate beta distributions. Journal of Multivariate Analysis, 102, 1194-1202.

[4] Klar, B. (2015). A note on gamma difference distributions. Journal of Statistical Computation and Simulation, 85, 3708-3715.