SORT 32 (1) January-June 2008, 3-36

# Construction of multivariate distributions: a review of some recent results

José María Sarabia[1] and Emilio Gómez-Déniz[2]

**Abstract**

The construction of multivariate distributions is an active field of research in theoretical and applied statistics. In this paper some recent developments in this field are reviewed. Specifically, we study and review the following set of methods: (a) Construction of multivariate distributions based on order statistics, (b) Methods based on mixtures, (c) Conditionally specified distributions, (d) Multivariate skew distributions, (e) Distributions based on the method of the variables in common and (f) Other methods, which include multivariate weighted distributions, vines and multivariate Zipf distributions.

*MSC:* 60E05, 62E15, 62H10

*Keywords:* Order statistics, Rosenblatt transformation, mixture distributions, conditionally specified distributions, skew distributions, variables in common, multivariate weighted distributions, vines, multivariate Zipf distributions, associated random variables

## 1 Introduction

The construction, study and applications of multivariate distributions is one of the classical fields of research in statistics, and it continues to be an active field of research.

In recent years several books containing theory about multivariate nonnormal distributions have been published: Hutchinson and Lai (1990), Joe (1997), Arnold, Castillo and Sarabia (1999), Kotz, Balakrishnan and Johnson (2000), Kotz and Nadarajah (2004), Nelsen (2006). In the discrete case specifically, we cannot ignore the books of Kocherlakota and Kocherlakota (1992) and Johnson, Kotz and Balakrishnan (1997) and the review papers by Balakrishnan (2004, 2005).

---

[1] Department of Economics, University of Cantabria, Santander, Spain.
[2] Department of Quantitative Methods, University of Las Palmas de Gran Canaria, Gran Canaria, Spain.
Received: November 2007

In this paper some recent methods for constructing multivariate distributions are reviewed. Reviews on constructions of discrete and continuous bivariate distributions are given by Lai (2004 and 2006). One of the problems of this work is the impossibility of producing a standard set of criteria that can always be applied to produce a unique distribution which could unequivocally be called the multivariate version (Kemp and Papageorgiou, 1982). In this sense, there is no satisfactory unified scheme of classifying these methods. In the bivariate continuous case Lai (2004 and 2006) has considered the following clusters of methods,

- Marginal transformation method
- Methods of construction of copulas
- Mixing and compounding
- Variables in common and trivariate reduction techniques
- Conditionally specified distributions
- Marginal replacement
- Geometric approach
- Constructions of extreme-value models
- Limits of discrete distributions
- Some classical methods
- Distributions with a given variance-covariance matrix
- Transformations

Some of these methods have merited considerable attention in the recent literature and they will not be revised here. For instance, a detailed study on the construction of copulas is provided by Nelsen (2006) and also in the review paper by Mikosch (2006).

The choice of the methods revised in this paper responds to a general interest and our own research experience. Therefore, and as is obvious, this revision cannot be considered as exhaustive regarding multivariate distributions.

The contents of this paper are as follows. In Section 2 we study multivariate distributions based on order statistics. Section 3 reviews methods based on mixtures. Conditionally specified distributions are studied in Section 4. Section 5 reviews multivariate skew distributions. Some recent distributions based on the method of the variables in common are studied in Section 6. Finally, other methods of construction (multivariate weighted distributions, vines and multivariate Zipf distributions.) are briefly commented in Section 7.

## 2 Multivariate Distributions based on Order Statistics

Order statistics and related topics (especially extreme value theory) have received a lot of attention recently, see Arnold *et al.* (1992), Castillo *et al.* (2005), David and Nagaraja

(2003) and Ahsanullah and Nevzorov (2005). In this section we review multivariate distributions beginning with the idea of order statistics.

### 2.1 An extension of the multivariate distribution of subsets of order statistics

Let $X_1, \dots, X_n$ be a sample of size $n$ drawn from a common probability density function (pdf) $f(x)$ and cumulative distribution function (cdf) $F(x)$, and let $X_{1:n} \le \cdots \le X_{n:n}$ denote the corresponding order statistics. Now, let $X_{n_1:n}, \dots, X_{n_p:n}$ be a subset of $p$ order statistics, where $1 \le n_1 < \cdots < n_p \le n$, $p = 1, 2, \dots, n$. The joint pdf of $X_{n_1:n}, \dots, X_{n_p:n}$ is

$$\frac{n!}{\prod_{j=1}^{p+1}(n_j - n_{j-1} - 1)!} \left\{\prod_{j=1}^{p} f(x_j)\right\} \prod_{j=1}^{p+1} \{F(x_j) - F(x_{j-1})\}^{n_j - n_{j-1} - 1}, \tag{1}$$

for $x_1 \le \cdots \le x_p$, where $x_0 = -\infty$, $x_{p+1} = +\infty$, $n_0 = 0$ and $n_{p+1} = n + 1$.

Beginning with the idea of fractional order statistics (Stigler (1977) and Papadatos (1995)) Jones and Larsen (2004) proposed generalizing (1) by considering real numbers $a_1, \dots, a_{p+1} > 0$ instead of integers $n_1, \dots, n_{p+1}$, to obtain the joint pdf,

$$g_F(x_1, \dots, x_p) = \frac{\Gamma(a_1 + \cdots + a_{p+1})}{\prod_{j=1}^{p+1} \Gamma(a_j)} \left\{\prod_{j=1}^{p} f(x_j)\right\} \prod_{j=1}^{p+1} \{F(x_j) - F(x_{j-1})\}^{a_j - 1}, \tag{2}$$

on $-\infty = x_0 \le x_1 \le \cdots \le x_p \le x_{p+1} = \infty$. Two particular cases merit our attention. If $F \sim \mathcal{U}[0, 1]$ is a uniform distribution on $[0, 1]$, (2) becomes

$$g_U(u_1, \dots, u_p) = \frac{\Gamma(a_1 + \cdots + a_{p+1})}{\prod_{j=1}^{p+1} \Gamma(a_j)} \prod_{j=1}^{p+1} (u_j - u_{j-1})^{a_j - 1}, \tag{3}$$

defined on $0 = u_0 \le u_1 \le \cdots \le u_p \le u_{p+1} = 1$, which is the generalization of uniform order statistics. If $\{U_j\}$, $j = 1, 2, \dots, p$ is distributed as (3), then $\{X_j = F^{-1}(U_j)\}$, $j = 1, 2, \dots, p$ is distributed as (1). Another important relation is obtained from the Dirichlet distribution. Let $(V_1, \dots, V_k)$ a Dirichlet distribution with joint pdf

$$\frac{\Gamma(a_1 + \cdots + a_{p+1})}{\prod_{j=1}^{p+1} \Gamma(a_j)} \prod_{j=1}^{p} v_p^{a_k - 1} \left(1 - \sum_{j=1}^{p} v_j\right)^{a_{p+1} - 1}, \tag{4}$$

defined on the simplex $v_j \ge 0$, $j = 1, 2, \dots, p$, $v_1 + \cdots + v_p \le 1$. In this case $U_i = V_1 + \cdots + V_i$, $i = 1, 2, \dots, p$ and $X_i = F^{-1}(V_1 + \cdots + V_i)$, $i = 1, 2, \dots, p$.

In the univariate case, family (2) becomes

$$g_F(x) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} f(x) F^{a_1-1}(x)[1 - F(x)]^{a_2-1}, \tag{5}$$

which was also proposed by Jones (2004) and it is a generalization of the $r$-order statistics. The idea of this author is to begin with a symmetric distribution $f$ ($a_1 = a_2 = 1$ in (5)) and enlarge this family with parameters $a_1$ and $a_2$, controlling skewness and tail weight. If $B \sim \mathcal{B}e(a_1, a_2)$ is a beta distribution with parameters $a_1$ and $a_2$, family (5) can be obtained by the simple transformation $X = F^{-1}(B)$.

As a last comment in this section, we mention the concept of generalized order statistics introduced by Kamps (1995), as a unified model for ordered random variables, which includes among others the usual order statistics, record values and $k$-record values as special cases.

### 2.1.1 An example with the normal distribution

In this section we include an example with the normal distribution. If $p = 2$, $F = \Phi$, $f = \phi$, where $\Phi$ and $\phi$ are the cdf and the pdf of the standard normal distribution, respectively, general expression (2) becomes

$$g_\Phi(x, y; \underline{a}) = \frac{\Gamma(a_1 + a_2 + a_3)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)} \phi(x)\phi(y)[\Phi(x)]^{a_1-1}[\Phi(y) - \Phi(x)]^{a_2-1}[1 - \Phi(y)]^{a_3-1}, \tag{6}$$

on $x < y$, and $a_1, a_2, a_3 > 0$. Both marginals distributions $X$ and $Y$ are like (5) with parameters $(a_1, a_2 + a_3)$ and $(a_1 + a_2, a_3)$, respectively. The local dependence function is given by

$$\gamma(x, y) = \frac{\partial^2 \log g_\Phi(x, y, \underline{a})}{\partial x \partial y} = \frac{(a_2 - 1)\phi(x)\phi(y)}{[\Phi(y) - \Phi_(x)]^2},$$

if $x < y$.

Figure 1 shows two examples of the bivariate distribution (6).

## 2.2 Multivariate distribution involving the Rosenblatt construction

As a multivariate version of Jones' (2004) univariate construction defined in equation (5), Arnold, Castillo and Sarabia (2006) have proposed multivariate distributions based on an enriching process using a representation of a $p$-dimensional random vector with a given distribution due to Rosenblatt (1952). Consider an initial family of $p$-dimensional joint distribution functions $F(x_1, \ldots, x_p)$. We assume that these

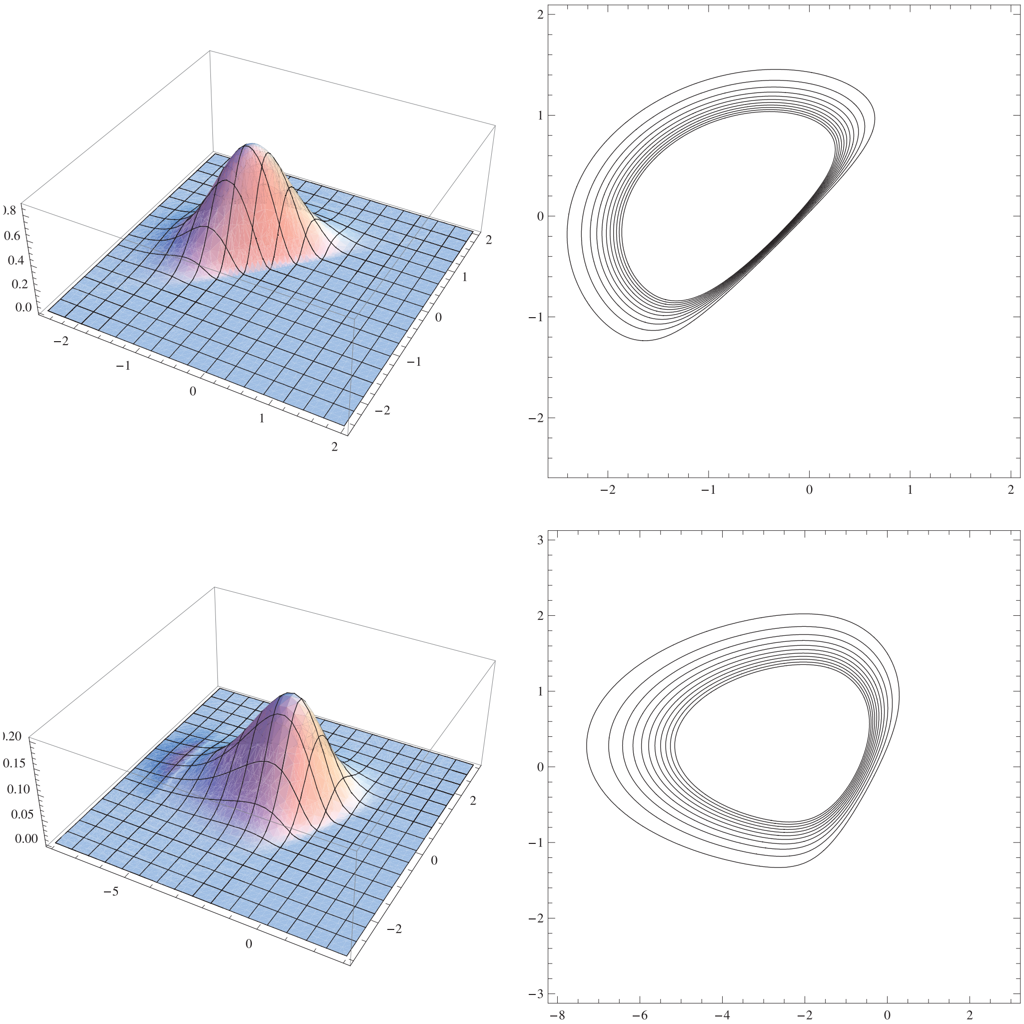


***Figure 1:*** *Joint pdf and contour plots of bivariate ordered normal distribution (6) with* $a_1 = 2$, $a_2 = 3$ *and* $a_3 = 4$ *(above) and* $a_1 = 0.2$, $a_2 = 3$ *and* $a_3 = 2$ *(below).*

distributions are absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^p$ and we will be denoted by $f(x_1, \ldots, x_p)$, the corresponding joint density function. We next introduce the notation

$$
\begin{aligned}
F_1(x_1) &= \Pr(X_1 \leq x_1), \\
F_2(x_2|x_1) &= \Pr(X_2 \leq x_2|X_1 = x_1), \\
&\vdots \\
F_p(x_p|x_1, \ldots, x_{p-1}) &= \Pr(X_p \leq x_p|X_1 = x_1, \ldots, X_{p-1} = x_{p-1}),
\end{aligned}
$$

and associated conditional densities

$$f_1(x_1),\ f_2(x_2|x_1), \dots, f_p(x_p|x_1, \dots, x_{p-1}). \tag{7}$$

In the spirit of the Rosenblatt transformation, these authors proposed the multivariate distribution of $\underline{X} = (X_1, \dots, X_p)$ defined by

$$\begin{aligned} X_1 &= F_1^{-1}(V_1), \\ X_2 &= F_2^{-1}(V_1|X_1), \\ &\vdots \\ X_p &= F_p^{-1}(V_p|X_1, \dots, X_{p-1}), \end{aligned}$$

where $V_1, \dots, V_p$ represent independent beta distributions $V_i \sim \mathcal{B}e(a_i, b_i)$. The resulting joint density for $\underline{X}$ is that given by

$$g(x_1, \dots, x_p; \underline{a}, \underline{b}) = f(x_1, \dots, x_p) \prod_{i=1}^{p} f_{\mathcal{B}e(a_i,b_i)}(F_i(x_i|x_1, \dots, x_{i-1})), \tag{8}$$

where $f_{\mathcal{B}e(a_i,b_i)}$, $i = 1, \dots, p$ denotes the density of a beta random variable with parameters $(a_i, b_i)$. It is clear from (8) that the initial joint density $f(x_1, \dots, x_p)$ is included in (8) as a special case setting $a_i = b_i = 1$, $i = 1, \dots, p$. All the conditional densities (7) are of the form (5). The proposed method is quite general, and several new $p$-dimensional parametric families have been proposed, including: Frank-beta distribution, the Farlie-Gumbel-Morgenstern-beta family, the normal-beta family, the Dirichlet-beta family and the Pareto-beta family. The families of distributions obtained in this way are very flexible and easy to estimate. Details can be found in Arnold, Castillo and Sarabia (2006).

## 3 Methods Based on Mixtures

The use of mixtures to obtain flexible families of densities has a long history, especially in the univariate case. The advantages of the mixtures mechanism are diverse. The new classes of distributions obtained by mixing are more flexible than the original, overdispersed with tails larger than the original distribution and often providing better fits.

The extension of a mixture to the multivariate case is usually simple, and the marginal distributions belong to the same family. On the other hand, simulation and Bayesian estimation of mixtures are quite direct. Since the introduction of simulation-based methods for inference (particularly the Gibbs sampler in a Bayesian framework), complicated densities such as those having mixture representation have been satisfactorily handled.

### 3.1 Common mixture models

In this model we assume conditional independence among components and a common parameter shared by all components. The joint cdf is given by

$$F_{\underline{X}}(\underline{x}) = \int_{-\infty}^{\infty} \left\{ \prod_{k=1}^{p} F_k(x_k|\theta) \right\} dF_{\Theta}(\theta), \quad \underline{x} = (x_1, \dots, x_p), \tag{9}$$

or in terms of joint densities,

$$f_{\underline{X}}(\underline{x}) = \int_{-\infty}^{\infty} \left\{ \prod_{k=1}^{p} f_k(x_k|\theta) \right\} dF_{\Theta}(\theta), \quad \underline{x} = (x_1, \dots, x_p). \tag{10}$$

In these models $\theta$ acts as a frailty parameter. In the joint cdf (9), if each component $F_k(x_k|\theta)$ is stochastically increasing in $\theta$, then $X_1, \dots, X_p$ are associated and

$$\operatorname{cov}(u(X_1, \dots, X_p), v(X_1, \dots, X_p)) \geq 0, \tag{11}$$

for all increasing functions $u$, $v$ for which the covariance exists.

### 3.2 A more general model

In this situation we have the general models,

$$F_{\underline{X}}(\underline{x}) = \int_{\mathbb{R}^p} \left\{ \prod_{k=1}^{p} F_k(x_k|\theta_k) \right\} dF_{\underline{\Theta}}(\underline{\theta}), \quad \underline{x} = (x_1, \dots, x_p), \tag{12}$$

or in terms of joint densities,

$$f_{\underline{X}}(\underline{x}) = \int_{\mathbb{R}^p} \left\{ \prod_{k=1}^{p} f_k(x_k|\theta_k) \right\} dF_{\underline{\Theta}}(\underline{\theta}), \quad \underline{x} = (x_1, \dots, x_p), \tag{13}$$

where $\underline{\theta} = (\theta_1, \dots, \theta_p)$.

In the following sections we include some recent multivariate distributions proposed in the literature obtained by using previous formulations.

### 3.3 Multivariate discrete distributions

The study of the variability of multivariate counts arises in many practical situations. In ecology the counts may be the different species of animals in different geographical areas whilst in insurance, the number of claims of different policyholders in the portfolio.

### 3.3.1 Multivariate Poisson-lognormal distribution

The multivariate Poisson-lognormal distribution (Aitchison and Ho, 1989) is one of the most relevant models. This distribution is the mixture of conditional independent Poisson distributions, where the mean vector $\underline{\theta} = (\theta_1, \ldots, \theta_p)$ follows a $p$-dimensional lognormal distribution with probability density function,

$$g(\underline{\theta}; \underline{\mu}, \Sigma) = (2\pi)^{-p/2}(\theta_1 \cdots \theta_p)^{-1}|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\log \underline{\theta} - \underline{\mu})^\top \Sigma^{-1}(\log \underline{\theta} - \underline{\mu})\right].$$

The probability mass function is given by (formula (13)):

$$\Pr(X_1 = x_1, \ldots, X_p = x_p) = \int_{\mathbb{R}_+^p} \prod_{i=1}^{p} f(x_i|\theta_i) g(\underline{\theta}; \underline{\mu}, \Sigma) d\underline{\theta}, \quad x_1, \ldots, x_p = 0, 1, \ldots, \tag{14}$$

where $\mathbb{R}_+^p$ denotes the positive orthant of $p$-dimensional real space $\mathbb{R}^p$. Although it is not possible to obtain a closed expression for the probability mass function in (14), its moments can be easily obtained by conditioning

$$E(X_i) = \exp\left(\mu_i + \frac{1}{2}\sigma_{ii}\right) = \alpha_i, \tag{15}$$

$$\mathrm{Var}(X_i) = \alpha_i + \alpha_i^2 \left[\exp(\sigma_{ii}) - 1\right], \tag{16}$$

$$\mathrm{cov}(X_i, X_j) = \alpha_i\alpha_j \left[\exp(\sigma_{ii}) - 1\right], \quad i, j = 1, \ldots, p, \; i \neq j. \tag{17}$$

From (15) and (16) it is obvious than the marginal distributions are overdispersed and from (17) the model admits both negative and positive correlations. Other versions of this model can be viewed in Tonda (2005).

### 3.3.2 Multivariate Poisson-generalized inverse Gaussian distribution

Departing from the Sichel distribution (Poisson-generalized inverse Gaussian distribution) in Sichel (1971) investigated bivariate extensions of that distribution. In Stein *et al.* (1987) one of them is studied in order to obtain the estimation of the parameters via the likelihood method.

Conditionally given $\lambda$, the $X_i$ are independently distributed as Poisson random variables with parameter $\lambda\xi_i$, where $\xi_i$ is a scale factor ($i = 1, \ldots, p$) and the parameter

$\lambda$ follows a generalized inverse Gaussian distribution with parameters 1, $w > 0$ and $\gamma \in \mathbb{R}$. The multivariate mixture is:

$$\begin{aligned}\Pr(X_1 = x_1, \ldots, X_p = x_p) &= \frac{K_{\sum x_i + y}\left(w^{1/2}(w + \sum \xi_i)^{1/2}\right)}{K_\gamma(w)} \\ &\times \left(\frac{w}{w + 2\sum \xi_i}\right)^{(\sum x_i + y)/2} \prod_{i=1}^{p} \frac{\xi_i^{x_i}}{x_i!}, \\ &\quad x_1, \ldots, x_p = 0, 1, \ldots, \ \alpha, \\ &\quad \xi_1, \ldots \xi_p > 0, -\infty < \gamma < \infty,\end{aligned}$$

where $K_\nu(z)$ denotes the modified Bessel function of the second kind of order $\nu$ and argument $z$. The basic moments and the correlation matrix are

$$\begin{aligned}E(X_i) &= \xi_i R_\gamma(w), \\ \mathrm{Var}(X_i) &= \xi_i^2 K_{\gamma+2}(w)/K_\gamma(w) + E(X_i)\left[1 - E(X_i)\right], \\ \mathrm{corr}(X_i, X_j) &= \left[(1 + 1/\xi_i g(w,\gamma))\left((1 + 1/\xi_j g(w,\gamma)\right)\right]^{-1/2}, \\ &\qquad i, j = 1, \ldots, p, \ i \neq j,\end{aligned}$$

where $R_\nu(z) = K_{\nu+1}(z)/K_\nu(z)$ and $g(w,\gamma) = R_{\gamma+1}(w) - R_\gamma(w)$. The correlations between marginals are positive.

### *3.3.3 Multivariate negative binomial-inverse Gaussian distribution*

Gómez-Déniz *et al*. (2008) have considered a new distribution by mixing a negative binomial distribution with an inverse Gaussian distribution, where the parameterization $\hat{p} = \exp(-\lambda)$ was considered. This new formulation provides a tractable model with attractive properties, which makes it suitable for application in disciplines where overdispersion is observed.

The multivariate negative binomial-inverse Gaussian distribution can be considered as the mixture of independent $\mathcal{NB}(r_i, \hat{p} = e^{-\lambda})$, $i = 1, 2, \ldots, p$ combined with an inverse Gaussian distribution for $\lambda$. The joint probability mass function is given by (formula (10)),

$$\begin{aligned}&\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p) \\ &= \prod_{i=1}^{p} \binom{r_i + x_i - 1}{x_i} \sum_{j=0}^{\tilde{x}} (-1)^j \binom{\tilde{x}}{j} \exp\left\{\frac{\psi}{\mu}\left[1 - \sqrt{1 + \frac{2(\tilde{r} + j)\mu^2}{\psi}}\right]\right\},\end{aligned}$$

where $x_1, x_2, \ldots, x_p = 0, 1, 2, \ldots; \mu, \psi, r_1, \ldots, r_p > 0$ and $r = r_1 + \cdots + r_p$, $\tilde{x} = x_1 + \cdots + x_p$ and the moments,

$$\begin{aligned} E(X_i) &= r_i\left[M_\lambda(1)-1\right],\ i=1,2,\ldots,r\\ \mathrm{Var}(X_i) &= (r_i+r_i^2)M_\lambda(2)-r_iM_\lambda(1)-r_i^2M_\lambda^2(1),\ i=1,2,\ldots,p\\ \mathrm{cov}(X_i,X_j) &= r_ir_j\left[M_\lambda(2)-M_\lambda^2(1)\right],\quad i,j=1,\ldots,p,\ i\neq j. \end{aligned}$$

where $M_\lambda(t)$ is the moment generating function of the inverse Gaussian distribution. Since $M_\lambda(2)-M_\lambda^2(1)=\mathrm{Var}(e^\lambda)$, the correlation is always positive. Applications of this model in insurance can be found in Gómez-Déniz *et al.* (2008).

### 3.3.4 Multivariate Poisson-beta distribution

Sarabia and Gómez-Déniz (2008) have proposed multivariate versions of the beta mixture of Poisson distribution considered by Gurland (1957) and Katti (1966). The new class of distributions can be used for modelling multivariate dependent count data when marginal overdispersion is also observed. The basic multivariate distribution Poisson-Beta $(X_1,\ldots,X_p)^\top$ is defined through $p$ independent Poisson distributions with parameters $\phi_i\theta,\ \phi_i>0,\ i=1,2,\ldots,p$, where $\theta\sim\mathcal{B}e(a,b)$ with $a,b>0$. The probability mass function is given by:

$$\begin{aligned} &\Pr(X_1=x_1,\ldots,X_p=x_p)\\ &= \prod_{i=1}^p\frac{\phi_i^{x_i}}{x_i!}\cdot\frac{\Gamma(a+b)\Gamma(a+\sum_{i=1}^p x_i)}{\Gamma(a)\Gamma(a+b+\sum_{i=1}^p x_i)}\,{}_1F_1\left(a+\sum_{i=1}^p x_i;a+b+\sum_{i=1}^p x_i;-\sum_{i=1}^p\phi_i\right), \end{aligned}$$

where ${}_1F_1(a;c;x)$ represents the confluent hypergeometric function. The previous model has the advantage of its simplicity but presents two shortcomings. On the one hand, the parameters in the marginal distributions are not free, in the sense that all marginal distributions share parameters $a$ and $b$. On the other hand, the model is not valid for representing multivariate count data with negative correlation between pairs of variables. This can be overcome by defining a multivariate Poisson-Beta distribution by the stochastic representation

$$\begin{aligned} X_i|\theta_i &\sim \mathcal{P}o(\phi_i\theta_i),\quad i=1,2,\ldots,p\ \text{ independent},\\ (\theta_1,\ldots,\theta_p) &\sim f(\theta_1,\ldots,\theta_p),\\ \theta_i &\sim \mathcal{B}e(a_i,b_i),\quad i=1,2,\ldots,p, \end{aligned}$$

where $f(\cdot)$ represents a multivariate distribution with beta marginals $\mathcal{B}e(a_i,b_i)$ and $\phi_i>0$, $i=1,2,\ldots,p$. Although there is not a closed form for the joint probability mass function, the means, variances and covariance vector are computed straightforwardly by conditioning. By choosing the Sarmanov-Lee distribution described in Sarmanov (1966), Lee (1996) and Kotz *et al.* (2000) and which has been used by Sarabia and

Castillo (2006), Sarabia and Gómez-Déniz (2008) built a bivariate distribution that admits non-limited correlations of any sign.

### 3.4 Continuous distributions

Walker and Stephens (1999) have observed a simple representation of the Weibull distribution as a mixture in such a way that,

$$\begin{aligned} X|\theta &\sim f(x|\theta) = \frac{ax^{a-1}}{\theta} I(0 < x^a < \theta), \\ \theta &\sim \mathcal{G}(2, c), \end{aligned}$$

where $I(\cdot)$ is the indicator function and $\mathcal{G}$ represents the gamma distribution. However, a Weibull distribution has only two parameters and the skewness is defined once the mean and variance are defined. The idea of these authors is to replace the gamma mixing with a two-parameter lognormal distribution

$$X|\theta \sim f(x|\theta) = \frac{ax^{a-1}}{e^{a\theta}} I(0 < x < e^{\theta}) \tag{18}$$

$$\theta \sim \mathcal{N}(\mu, \sigma^2), \tag{19}$$

which we will represent as $X \sim \mathcal{PLN}(a, \mu, \sigma)$. If we use formula (10) we obtain the joint density,

$$f(x_1, \ldots, x_p) = \prod_{i=1}^{p} a_i x_i^{a_i - 1} \exp(-\tilde{a}\mu + \tilde{a}^2\sigma^2/2) \left\{1 - \Phi(\log \tilde{z} + \tilde{a}\sigma)\right\}, \tag{20}$$

where $\log \tilde{z} = (\log(\max\{x_i\}) - \mu)/\sigma$ and $\tilde{a} = a_1 + \cdots + a_p$.

The multivariate version of Walker and Stephens (1999) based on (18)-(19) and general models (12)-(13) is quite direct. The joint $p$-dimensional density function can be written as

$$\begin{aligned} f(x_1, \ldots, x_p|\theta_1, \ldots, \theta_p) &= \prod_{k=1}^{p} \frac{a_k x_k^{a_k - 1}}{e^{a_k \theta_k}} I(0 < x_k < e^{\theta_k}) \\ (\theta_1, \ldots, \theta_p) &\sim \mathcal{N}_p(\mu, \Sigma). \end{aligned}$$

Marginally $X_j \sim \mathcal{PLN}(a_j, \mu_j, \sigma_j)$, $j = 1, \ldots, p$ and

$$\operatorname{cov}(X_j, X_k) = \frac{a_j a_k [\exp(\sigma_{jk}) - 1]}{(1 + a_j)(1 + a_k)} \exp[\mu_j + \mu_k + (\sigma_j^2 + \sigma_k^2)/2], \quad j \neq k,$$

where both positive and negative correlations are possible.

### 3.4.1 The multivariate normal-inverse Gaussian distribution

A multivariate version of the normal inverse Gaussian distribution introduced by Barndorff-Nielsen (1997) has been developed by Protassov (2004) and Øigard and Hanssen (2002). The model is a mean-variance mixture of a $p$-dimensional normal random variable with a univariate inverse Gaussian distribution. The probability density function is

$$f(\underline{x};\alpha,\underline{\beta},\underline{\mu},\delta,\Gamma) = \frac{\delta}{2^{\frac{p-1}{2}}}\left[\frac{\alpha}{q(\underline{x})}\right]^{\frac{p+1}{2}} \exp\left[p(\underline{x})\right] K_{\frac{p+1}{2}}\left[\alpha q(\underline{x})\right], \tag{21}$$

where

$$\begin{aligned} p(\underline{x}) &= \delta\sqrt{\alpha^2 - \underline{\beta}^\top\Gamma\underline{\beta}} + \underline{\beta}^\top(\underline{x}-\underline{\mu}), \\ q(\underline{x}) &= \sqrt{\delta^2 + \left[(\underline{x}-\underline{\mu})^\top\Gamma^{-1}(\underline{x}-\underline{\mu})\right]}, \end{aligned}$$

$K_\nu(z)$ denotes the modified Bessel function of the second kind of order $\nu$ and argument $z$, $\alpha > 0$, $\underline{\beta} \in \mathbb{R}^p$, $\delta > 0$, $\underline{\mu} \in \mathbb{R}^p$ and $\Gamma$ is a $p \times p$ matrix. The distribution is symmetric if and only if $\Gamma = \boldsymbol{I}$ and $\underline{\beta} = \underline{0}$. This multivariate distribution has been shown to be useful in risk theory and the framework of physics.

## 4 Conditionally Specified Distributions

A bivariate random variable can be written as the product of a marginal distribution and the corresponding conditional distribution,

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|x).$$

This is a simple method for generating bivariate distributions, and has been used in the practical literature as a common approach for obtaining dependent models, especially when $Y$ can be thought of as caused by $X$.

Now, the conditional distribution of $X$ given $Y$ together with the other conditional distribution $Y$ given $X = x_0$ determines the joint pdf from

$$f_{X,Y}(x,y) \propto \frac{f_{X|Y}(x|y) f_{Y|X}(y|x_0)}{f_{X|Y}(x_0|y)},$$

uniquely for each $x_0$. If we consider all possible values of $x_0$, we obviously obtain a richer model. In this sense, a bivariate random variable can be specified through its conditional distributions. If we assume that both conditional distributions belong to certain parametric classes of distributions, it is possible to obtain the joint distribution using the methodology proposed in Arnold, Castillo and Sarabia (1992, 1999) (see also Arnold, Castillo and Sarabia (2001)). To obtain the joint pdf it is necessary to solve certain functional equations. This methodology provides highly flexible multiparametric distributions, with some “unexpected” and interesting properties.

### 4.1 Compatible conditional densities

The existence of a bivariate distribution with given conditional distributions is a previous question. Let $(X, Y)$ be a random vector with joint density with respect to some product measure $\mu_1 \times \mu_2$ on $S(X) \times S(Y)$, where $S(X)$ denotes the set of possible values of $X$ and $S(Y)$ the set of possible values of $Y$ (note that one variable could be discrete and the other absolutely continuous with respect to the Lebesgue measure). The marginal, conditional and joint densities are denoted by $f_X(x)$, $f_Y(y)$, $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$, $f_{X,Y}(x, y)$ and the sets of possible values $S(X)$ and $S(Y)$ can be finite, countable or uncountable. Consider two possible families of conditional densities $a(x, y)$ and $b(x, y)$. We ask when it is true that there will exist a joint density for $(X, Y)$ such that

$$f_{X|Y}(x|y) = a(x, y), \quad x \in S(X), \; y \in S(Y)$$

and

$$f_{Y|X}(y|x) = b(x, y), \quad x \in S(X), \; y \in S(Y).$$

If such a density exists we will say that $a$ and $b$ are compatible families of conditional densities. We define

$$N_a = \{(x, y) : a(x, y) > 0\}$$

and

$$N_b = \{(x, y) : b(x, y) > 0\}.$$

The following compatibility theorem was stated by Arnold and Press (1989).

**Theorem 1 (Compatible conditionals)** *A joint density $f(x, y)$, with $a(x, y)$ and $b(x, y)$ as its conditional densities, will exist iff*

*(i) $N_a = N_b = N$, say*

*(ii) there exist functions $u(x)$ and $v(y)$ such that for every $(x, y) \in N$ we have*

$$\frac{a(x, y)}{b(x, y)} = u(x)v(y) \tag{22}$$

*in which $u(x)$ is integrable, i.e. $\int_{S(X)} u(x) d\mu_1(x) < \infty$.*

### 4.2 Results in exponential families

One of the most important results in conditional specification is a Theorem provided by Arnold and Strauss (1991), dealing with bivariate distributions with conditionals in prescribed exponential families. Then, we consider two different exponential families of densities $\{f_1(x;\underline{\theta}) : \underline{\theta} \in \Theta \subset \mathbb{R}^{\ell_1}\}$ and $\{f_2(y;\underline{\tau}) : \underline{\tau} \in T \subset \mathbb{R}^{\ell_2}\}$ where:

$$f_1(x;\underline{\theta}) = r_1(x)\beta_2(\underline{\theta}) \exp\left[\sum_{i=1}^{\ell_1} \theta_i q_{1i}(x)\right] \tag{23}$$

and

$$f_2(y;\underline{\tau}) = r_2(y)\beta_2(\underline{\tau}) \exp\left[\sum_{j=1}^{\ell_2} \tau_j q_{2j}(y)\right]. \tag{24}$$

We are interested in the identification of the class of bivariate densities $f(x,y)$ with respect to $\mu_1 \times \mu_2$ on $S_x \times S_y$ for which conditional densities are well defined and satisfy the following:

- for every $y$ for which $f(x|y)$ is defined, this conditional density belongs to family (23) for some $\underline{\theta}$ which may depend on $y$ and
- for every $x$ for which $f(y|x)$ is defined, this conditional density belongs to family (24) for some $\underline{\tau}$ which may depend on $x$.

The class of all bivariate pdf $f(x,y)$ with conditionals in these prescribed exponential families, can be obtained as follows.

**Theorem 2** *Let $f(x,y)$ be a bivariate density whose conditional densities satisfy:*

$$f(x|y) = f_1(x;\underline{\theta}(y))$$

*and*

$$f(y|x) = f_2(y;\underline{\tau}(x))$$

*for every $x$ and $y$ for some functions $\underline{\theta}(y)$ and $\underline{\tau}(x)$ where $f_1$ and $f_2$ are as defined in (23) and (24). It follows that $f(x,y)$ is of the form:*

$$f(x,y) = r_1(x)r_2(y) \exp\left\{\underline{q}^{(1)}(x)^\top M \underline{q}^{(2)}(y)\right\} \tag{25}$$

*in which*

$$\underline{q}^{(1)}(x) = (1, q_{11}(x), \ldots, q_{1\ell_1}(x))^\top$$

*and*

$$\underline{q}^{(2)}(y) = (1, q_{21}(y), \ldots, q_{2\ell_2}(y))^{\top}$$

*and $M$ is a matrix of parameters of dimension $(\ell_1+1)\times(\ell_2+1)$ subject to the requirement that:*

$$\int_{S_x}\int_{S_y} f(x,y)d\mu_1(x)d\mu_2(y) = 1. \tag{26}$$

*The term $e^{m_{00}}$ is the normalizing constant that is a function of the other $m_{ij}$'s determined by the constraint (26).*

Note that the class of densities with conditionals in the prescribed family is itself an exponential family with $(\ell_1+1)\times(\ell_2+1)-1$ parameters. Upon partitioning the matrix $M$ in (25) in the following manner:

$$M = \begin{pmatrix} m_{00} & | & m_{01} & \cdots & m_{0\ell_2} \\ -- & + & -- & -- & -- \\ m_{10} & | & & & \\ \cdots & | & & \tilde{M} & \\ m_{\ell_1 0} & | & & & \end{pmatrix}, \tag{27}$$

it can be verified that independent marginals will be encountered iff the matrix $\tilde{M} \equiv 0$. The elements of $\tilde{M}$ determine the dependence structure in $f(x,y)$.

### 4.3 Two examples

In this section we include two examples (discrete and continuous) of bivariate distributions with conditional specifications.

Using Theorem 2, and after a convenient parameterization, the most general bivariate distribution with Poisson conditionals has the following joint probability mass function,

$$\Pr(X=x, Y=y) = k(\lambda_1,\lambda_2,\lambda_3)\frac{\lambda_1^x}{x!}\frac{\lambda_2^y}{y!}\lambda_3^{xy}, \quad x,y = 0,1,2,\ldots \tag{28}$$

with $\lambda_1, \lambda_2 > 0$ and $0 < \lambda_3 \le 1$ and where $k$ is the normalizing constant. The conditional distribution of $X$ given $y$ is $\mathcal{P}o(\lambda_1\lambda_3^y)$ and $Y$ given $x$ is $\mathcal{P}o(\lambda_2\lambda_3^x)$. If $\lambda_3 = 1$, $X$ and $Y$ are independent and if $0 < \lambda_3 < 1$, $X$ and $Y$ are negatively correlated with correlation coefficient range $\rho(X, Y) \in (-1, 0)$. The marginal distributions of (28) are

$$\begin{aligned} \Pr(X = x) &= k\frac{\lambda_1^x}{x!}\exp(\lambda_2\lambda_3^x), \quad x = 0, 1, 2, \dots \\ \Pr(Y = y) &= k\frac{\lambda_2^y}{y!}\exp(\lambda_1\lambda_3^y), \quad y = 0, 1, 2, \dots, \end{aligned}$$

which are not Poisson except in the independence case. Wesolowski (1996) has characterized this distribution using a conditional distribution and the other conditional expectation.

The second example corresponds to the normal case. Again, using Theorem 2, the most general bivariate distribution with normal conditionals is given by

$$f_{X,Y}(x, y; \underline{m}) = \exp\left\{(1, x, x^2)\begin{pmatrix} m_{00} & m_{01} & m_{02} \\ m_{10} & m_{11} & m_{12} \\ m_{20} & m_{21} & m_{22} \end{pmatrix}\begin{pmatrix} 1 \\ y \\ y^2 \end{pmatrix}\right\}. \tag{29}$$

Distributions with densities of form (29) are called normal conditional distributions. Note that (29) is an eight parameter family of densities, and $m_{00}$ is the normalizing constant. The conditional expectations and variances are:

$$E(Y|X = x) = -\frac{m_{01} + m_{11}x + m_{21}x^2}{2(m_{02} + m_{12}x + m_{22}x^2)}, \tag{30}$$

$$\mathrm{Var}(Y|X = x) = -\frac{1}{2(m_{02} + m_{12}x + m_{22}x^2)}, \tag{31}$$

$$E(X|Y = y) = -\frac{m_{10} + m_{11}y + m_{12}y^2}{2(m_{20} + m_{21}y + m_{22}y^2)}, \tag{32}$$

$$\mathrm{Var}(X|Y = y) = -\frac{1}{2(m_{20} + m_{21}y + m_{22}y^2)}. \tag{33}$$

The normal conditional distributions give rise to models where the $m_{ij}$ constants satisfy one of the two sets of conditions

(a) $m_{22} = m_{12} = m_{21} = 0$; $m_{20} < 0$; $m_{02} < 0$; $m_{11}^2 < 4m_{02}m_{20}$.

(b) $m_{22} < 0$; $4m_{22}m_{02} > m_{12}^2$; $4m_{20}m_{22} > m_{21}^2$.

Models satisfying conditions (a) are the classical bivariate normal models with normal marginals and conditionals, linear regressions and constant conditional variances. More interesting are the models satisfying conditions (b). These models have normal conditional distributions, non-normal marginals, and the regression functions are either constant or non-linear given by (30) and (32). Each regression function is

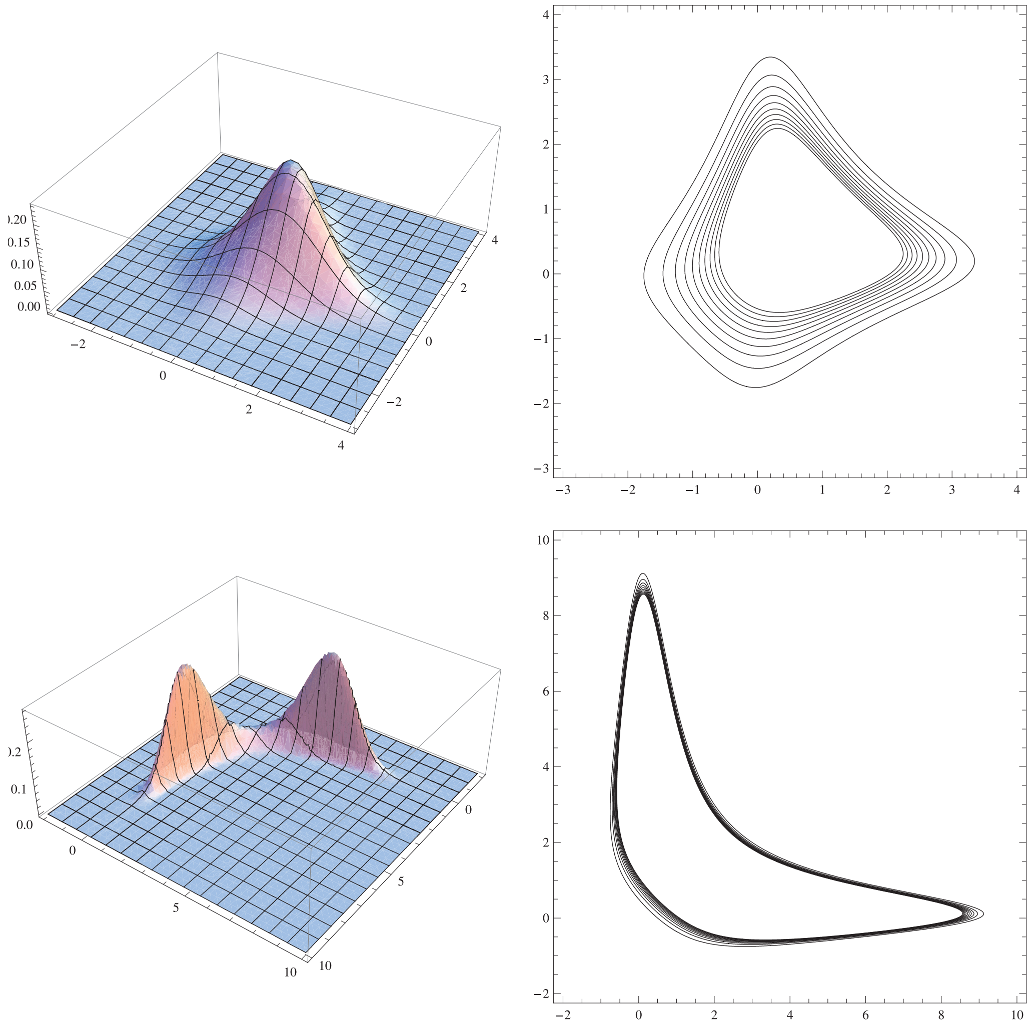

***Figure 2:*** *Joint pdf and contour plots of two bivariate distributions with normal conditionals.*

bounded (in contrast with the bivariate normal model) and the conditional variance functions are also bounded and non constant. They are given by (31) and (33). Another unexpected property of (29) is the multimodality, where one, two and three modes are possible (see Arnold *et al.*, 2000). Figure 2 presents two models of kind (b), one unimodal and the other one bimodal.

### *4.4 Multivariate extensions*

Previous results can be extended to higher dimensions. Technical details appear in Arnold, Castillo and Sarabia (1999 and 2001). As an important model, we consider the case in which the families of conditional densities $X_i|\underline{X}_{(i)} = \underline{x}_{(i)}$, $i = 1, \ldots, p$ are exponential families, where $X_{(i)}$ denotes the $p$-dimensional vector $\underline{X}$ with the $i$th coordinate deleted. In this situation, the most general joint density with exponential conditionals must be of the form

$$f_{\underline{X}}(\underline{x}) = \left[\prod_{i=1}^{p} r_i(x_i)\right] \exp\left\{\sum_{i_1=0}^{\ell_1}\sum_{i_2=0}^{\ell_2}\cdots\sum_{i_p=0}^{\ell_p} m_{i_1,i_2,\ldots,i_p}\left[\prod_{i=1}^{p} q_{ii_j}(x_j)\right]\right\}.$$

For example, the $p$-dimensional distribution with normal conditionals is of the form

$$f_{\underline{X}}(\underline{x}) = \exp\left\{\sum_{\underline{i}\in T_p} m_{\underline{i}}\left[\prod_{i=1}^{p} x_i^{ij}\right]\right\}, \tag{34}$$

where $T_p$ is the set of all vectors of 0's, 1's and 2's of dimension $p$. Densities of the form (34) have normal conditional densities for $X_i$ given $\underline{X}_{(i)} = \underline{x}_{(i)}$ for every $\underline{x}_{(i)}$, $i = 1, \ldots, p$. The classical $p$-variate normal density is a special case of (34).

### *4.5 Applications of the conditionally specified models*

Applications of these conditional models are contained in the book by Arnold, Castillo and Sarabia (1999). These applications include modelling of bivariate extremes, conditional survival models, multivariate binary response models with covariates (Joe and Liu, 1996) and Bayesian analysis using conditionally specified models. The use of this kind of distribution in risk analysis and economics in general is quite recent. Some applications have been provided by Sarabia, Gómez and Vázquez (2004) and Sarabia *et al.* (2005). The class of bivariate income distribution with lognormal conditionals has been studied by Sarabia *et al.* (2007). In the risk theory context, Sarabia and Guillén (2008) have proposed flexible bivariate joint distributions for modelling the couple $(S, N)$, where $N$ is a count variable and $S = X_1 + \cdots + X_N$ is the total claim amount.

## 5 Multivariate Skew Distributions

The skew-normal (SN) distribution, its different variants and their corresponding multivariate versions, have received considerable attention over the last few years. Two recent reviews of these classes appear in the book edited by Genton (2004) and the paper by Azzalini (2005). To introduce the multivariate version it is necessary to know the univariate case and its properties, which the multivariate version is based on. A random variable $X$ is said to have a skew-normal distribution with parameter $\lambda$, if the probability density function is given by

$$f(x;\lambda) = 2\phi(x)\Phi(\lambda x), \quad -\infty < x < \infty. \tag{35}$$

A random variable with pdf (35) will be denoted as $X \sim \mathcal{SN}(\lambda)$. Parameter $\lambda$ controls the skewness of the distribution and varies in $(-\infty, \infty)$. The linearly skewed version of this distribution is given by,

$$f(x;\lambda_0,\lambda_1) \propto \phi(x)\Phi(\lambda_0 + \lambda_1 x), \quad -\infty < x < \infty, \tag{36}$$

and $\lambda_0, \lambda_1 \in \mathbb{R}$, which we will denote by $X \sim \mathcal{SN}(\lambda_0, \lambda_1)$.

The next two properties hold for distribution (35) and allow us to understand multivariate extensions:

- Hidden truncation mechanism. Let $(X_0, X_1)$ be a bivariate normal distribution with standardized marginals and correlation coefficient $\delta$. Then, the variable,

$$\{X_1 | X_0 > 0\} \tag{37}$$

  is distributed as a $\mathcal{SN}(\lambda(\delta))$ distribution, where $\lambda(\delta) = \frac{\delta}{\sqrt{1-\delta^2}}$.

- Convolution representation. If $X_0$ and $X_1$ are independent $\mathcal{N}(0,1)$ random variables, and $-1 < \delta < 1$, then

$$Z = \delta|X_0| + (1-\delta^2)^{1/2} X_1 \tag{38}$$

  is a $\mathcal{SN}(\lambda(\delta))$.

A general treatment of the hidden truncation mechanism (37) is found in Arnold and Beaver (2002).

A multivariate version of the basic model (35) has been considered by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999). This multivariate version of the SN distribution is defined as

$$f(x) = 2\phi_p(x-\mu;\Sigma)\Phi(\alpha^\top w^{-1}(x-\mu)), \tag{39}$$

where $\phi_p(x-\mu;\Sigma)$ is the joint pdf of a multivariate normal distribution $\mathcal{N}_p(\mu,\Sigma)$, $\mu \in \mathbb{R}^p$ is a location parameter, $\Sigma$ is a positive definite covariance matrix, $\alpha \in \mathbb{R}^p$ is a parameter which controls skewness and $w$ is a diagonal matrix composed by the standard deviations of $\Sigma$. If we set $\alpha = 0$ in (39), we obtain a classical $\mathcal{N}_p(\mu,\Sigma)$ distribution. Similar to the univariate case, we can obtain (39) using a hidden truncation mechanism (37) and convolution representation (38).

Let $X_0$ and $X_1$ be random variables of dimensions 1 and $p$ such that

$$\begin{pmatrix} X_0 \\ X_1 \end{pmatrix} \sim \mathcal{N}_{1+p}(0, \Sigma^*), \quad \Sigma^* = \begin{pmatrix} 1 & \delta^\top \\ \delta & \tilde{\Sigma} \end{pmatrix},$$

where $\tilde{\Sigma}$ is a correlation matrix and

$$\delta = (1 + \alpha^\top \tilde{\Sigma} \alpha)^{-1/2} \tilde{\Sigma} \alpha.$$

Then, the $p$-dimensional random variable

$$Z = \{X_1 | X_0 > 0\},$$

has the joint pdf

$$f(z) = 2\phi_p(z; \tilde{\Sigma}) \Phi(\alpha^\top z), \tag{40}$$

which is an affine transformation of (39).

For the convolution representation let $X_0 \sim \mathcal{N}(0, 1)$ and $X_1 \sim \mathcal{N}_p(0, R)$ be independent random variables, where $R$ is a correlation matrix. Let $\Delta = diag\{\delta_1, \ldots, \delta_p\}$, $-1 < \delta_j < 1$, $j = 1, 2, \ldots, p$ and $I_p$ the identity matrix of order $p$ and $1_p$ the $p$-dimensional vector of all 1s. Then,

$$Z = \Delta 1_p |X_0| + (I_p - \Delta^2)^{1/2} X_1,$$

is distributed in the form (40). The relationship between $(R, \Delta)$ and $(\tilde{\Sigma}, \alpha)$ can be found in Azzalini and Capitanio (1999).

### 5.1 An alternative multivariate skew normal distribution

An alternative class of multivariate-normal distributions was considered by Gupta, González-Farías and Domínguez-Molina (2004). Previous multivariate versions (40) were obtained by conditioning that one random sample be positive; these authors condition that the same number of random variables be positive and then, in the univariate case both families are the same. A random vector of dimension $p$ is said to have a multivariate skew normal distribution (according to Gupta *et al.*, 2004) if its pdf is given by

$$f_p(x; \mu, \Sigma, D) = \frac{\phi_p(x; \mu, \Sigma) \Phi_p(D(x - \mu); 0, I)}{\Phi_p(0; 0, I + D\Sigma D^\top)}, \tag{41}$$

where $\mu \in \mathbb{R}^p$, $\Sigma > 0$, $D(p \times p)$, and $\phi_p(\cdot;\xi,\Omega)$ and $\Phi_p(\cdot;\xi,\Omega)$ denote the pdf and the cdf, respectively of a $\mathcal{N}_p(\xi,\Omega)$ distribution.

As an extension to (41), González-Farías *et al.* (2003, 2004) introduced the closed skew-normal family of distributions. This family is closed under conditioning, linear transformations and convolutions. It is defined as

$$f_p(x;\mu,\Sigma,D,\nu,\Delta) = \frac{\phi_p(x;\mu,\Sigma)\Phi_q(D(x-\mu);\nu,\Delta)}{\Phi_q(0_q;\nu;\Delta + D\Sigma D^\top)}, \tag{42}$$

where $x,\mu,\nu \in \mathbb{R}^p$, $\Sigma \in \mathbb{R}^p \times \mathbb{R}^p$, $D \in \mathbb{R}^q \times \mathbb{R}^p$, $\Delta \in \mathbb{R}^q \times \mathbb{R}^q$ and $\Sigma$ and $\Delta$ are positive definite matrices and $0_q = (0,\ldots,0) \in \mathbb{R}^q$.

The closed skew-normal distributions can be generated by conditioning the first components of a normal random vector in the event that the remaining components are greater than certain given values.

### *5.2 Conditional specification*

Arnold, Castillo and Sarabia (2002) have discussed the problem of identifying $p$-dimensional densities with skew-normal conditionals. They address the question of identifying joint densities for a $p$-dimensional random vector $X$ that has the property that for each $x_{(i)} \in \mathbb{R}^{p-1}$ we have

$$X_i|X_{(i)} = x_{(i)} \sim \mathcal{SN}(\lambda_0^{(i)}(x_{(i)}), \lambda_1^{(i)}(x_{(i)})), \quad i = 1,2,\ldots,p. \tag{43}$$

An important parametric family of densities takes the form

$$f(x_1,\ldots,x_p;\underline{\lambda}) \propto \prod_{i=1}^{p} \phi(x_i)\Phi\left(\sum_{\underline{s}\in S_p} \lambda_{\underline{s}} \prod_{i=1}^{p} x_i^{s_i}\right), \tag{44}$$

where $S_p$ denotes the set of all vectors of 0's and 1's of dimension $p$. In the bivariate case, we obtain the following bivariate distribution with linearly skewed-normal conditionals,

$$f(x,y;\underline{\lambda}) \propto \phi(x)\phi(y)\Phi(\lambda_{00} + \lambda_{10}x + \lambda_{01}y + \lambda_{11}xy). \tag{45}$$

Note (45) does not belong to class (39) except when $\lambda_{00} = \lambda_{11} = 0$. The normalizing constant is complicated in general, except when $\lambda_{00} = \lambda_{10} = \lambda_{10} = 0$, in which is equals 2 and the density is explicitly given by

$$f(x,y;\lambda) = 2\phi(x)\phi(y)\Phi(\lambda xy). \tag{46}$$

This model has normal marginals and skew-normal conditionals of type (35), and bimodality is possible.

Model (44) can be viewed as a generalized hidden truncation model, defining $\underline{\tilde{X}} = (X_0, X_1, \ldots, X_p)$, with $X_i$'s i.i.d. $\mathcal{N}(0,1)$, in which we retain only those $\underline{\tilde{X}}$ for which

$$X_0 \leq \sum_{\underline{s} \in S_k} \lambda_{\underline{s}} \prod_{i=1}^{p} X_i^{s_i},$$

and the resulting conditional density of $(X_1, \ldots, X_p)$ will then be given by (44). More about skew conditionals models can be found in Sarabia (2002) and Arnold, Castillo and Sarabia (2007a, 2007b).

### 5.3 Balakrishnan skew-normal distribution

Balakrishnan (2002) as a discussant of Arnold and Beaver (2002) generalized the SN distribution as

$$f_n(x;\lambda) = \frac{\phi(x)[\Phi(\lambda x)]^n}{c_n(\lambda)}, \quad x \in \mathbb{R}, \tag{47}$$

where $n$ is an integer and $c_n(\lambda) = \int_{-\infty}^{\infty} \phi(x)[\Phi(\lambda x)]^n dx$. This distribution is known as Balakrishnan skew-normal distribution. If we set $n = 0$ and $n = 1$ in (47) the above density reduces to the $\mathcal{N}(0,1)$ distribution and the SN distribution, respectively. Gupta and Gupta (2004) have studied some properties of (47).

Several multivariate versions are possible. If we think of an extension by conditionals, in the simpler bivariate case, we obtain the joint pdf,

$$f_n(x,y;\lambda) = \tilde{c}_n(\lambda)\phi(x)\phi(y)[\Phi(\lambda xy)]^n, \quad (x,y) \in \mathbb{R}^2. \tag{48}$$

For this distribution, both conditionals are like (47), but the marginal distributions are not.

Yadegari *et al.* (2008) have considered the extension of (47) given by

$$f_{n,m}(x;\lambda) = \frac{1}{c_{n,m}(\lambda)}[\Phi(\lambda x)]^n[1-\Phi(\lambda x)]^m\phi(x), \quad x \in \mathbb{R}, \tag{49}$$

where $c_{n,m}(\lambda) = \sum_{i=0}^{m} \binom{m}{i}(-1)^i c_{n+i}(\lambda)$. A natural extension of (49) to the multivariate case is

$$f_{n,m}(x;\lambda) = \frac{1}{c_{n,m}(\lambda)}[\Phi(\lambda^\top x)]^n[1-\Phi(\lambda^\top x)]^m\phi_p(x), \quad x \in \mathbb{R}^p. \tag{50}$$

For $m = 0$ and $n = 1$ this distribution reduces to the multivariate SN distribution.

### 5.4 Extensions and applications

The initial formulation (35) gives rise to an important number of extensions and variants. One of these variants appears replacing the normality assumption with alternative symmetric distribution. An interesting class of skewed densities is provided by the following elementary, but useful, result (Azzalini, 2005).

**Lemma 1** *If $f_0$ is a p-dimensional pdf such that $f_0(x) = f_0(-x)$ for $x \in \mathbb{R}^p$, $G$ is a one-dimensional differentiable cdf such that $G'$ is a density symmetric about zero, and $w$ is real-valued function such that $w(-x) = -w(x)$ for all $x \in \mathbb{R}^p$, then*

$$f(x) = 2 f_0(x) G\{w(x)\}, \quad x \in \mathbb{R}^p, \tag{51}$$

*is a genuine pdf on $\mathbb{R}^p$.*

Different choices for $f_0$, $G$ and $w$ in (51) give rise to a huge number of variants of skewed densities. In a more general setting, Wang *et al.* (2004) have shown that any $p$-dimensional multivariate pdf $g(x)$ admits for any fixed location parameter $\lambda \in \mathbb{R}^p$ a unique skew-symmetric representation

$$g(x) = 2 f(x-\lambda)\pi(x-\lambda), \quad x \in \mathbb{R}^p, \tag{52}$$

where $f : \mathbb{R}^p \to \mathbb{R}^+$ is a symmetric pdf (in the sense of previous lemma) and $\pi : \mathbb{R}^p \to [0,1]$ is a skewing function such that $\pi(-x) = \pi(x)$. Conversely, any function $g$ of the kind (52) is a valid pdf. Multivariate distribution such as skew-Cauchy (Arnold and Beaver, 2000), skew-t (Branco and Dey, 2001; Azzalini and Capitanio, 2003) and other skew-elliptical distributions can be represented using previous formulations (51)-(52).

Finally, we mention some applications of the distribution described in this section: compositional data, financial market and insurance (Vernic, 2005), selective sampling, stochastic frontier models and modelling of environmental data.

## 6 The Variables in Common Method

This method, also known as “trivariate reduction”, is a popular and old technique used for building dependent variables, both in continuous and discrete cases. Our attention focuses on the bivariate case.

The method consists of building a pair of dependent random variables starting from three (or more) random variables. These initial random variables are usually independent. The functions that connect initial variables are generally elementary functions, or are given by the structure of the variables that we want to generate. A broad definition can be

$$\begin{cases} X = \upsilon_1(e_X, c_{XY}), \\ Y = \upsilon_2(e_Y, \tilde{c}_{XY}), \end{cases}$$

where $e_X$, $e_Y$ represent two sets containing the specific variables of $X$ and $Y$ respectively, and $c_{XY}, \tilde{c}_{XY}$ sets containing the common or latent variables.

According to Marshall and Olkin (2007), many of the couples $(X, Y)$ here presented are associated (formula (11)), and then only positive correlations are possible.

Over the last few years, several new dependent distributions using this method have been proposed. All the models presented in this section can be extended to higher dimensions. We present some relevant models.

### 6.1 Bivariate generalized Poisson distribution

Let $X_i$, $i = 1, 2, 3$ be mutually independent random variables. An usual trivariate reduction scheme is defined as

$$\begin{aligned} X &= X_1 + X_3, \\ Y &= X_2 + X_3. \end{aligned}$$

A disadvantage of this model is that only positive correlations are possible. If the $X_i$'s are discrete, the joint pgf is

$$g_{X,Y}(u, v) = g_{X_1}(u) g_{X_2}(v) g_{X_3}(uv).$$

If the $X_i$ are Poisson random variables, we obtain the classical bivariate Poisson distribution, which is often used for obtaining compound bivariate Poisson distributions. If we consider for the $X_i$'s random variables a generalized Poisson distribution, we obtain the model considered by Vernic (1997, 2000).

### 6.2 Bivariate beta distribution

In a Bayesian context, when we work with independent or correlated binomial distributions, a density defined over $\{0 \le x_i \le 1; i = 1, \ldots, p\}$ on the unit cube is needed. Olkin and Liu (2003) proposed the following method for constructing this distribution. Let $X_i \sim \mathcal{G}(a_i, 1)$, $i = 1, 2, 3$ be independent gamma variables with unit scale parameters, and define

$$\begin{aligned} X &= \frac{X_1}{X_1 + X_3}, \\ Y &= \frac{X_2}{X_2 + X_3}. \end{aligned}$$

Now, we have correlated beta distributions $\mathcal{B}e(a_1, a_3)$ and $\mathcal{B}e(a_2, a_3)$ over $0 \leq x, y \leq 1$ with joint pdf,

$$f(x, y; a_1, a_2, a_3) = \frac{x^{a_1-1} y^{a_2-1} (1-x)^{a_2+a_3-1} (1-y)^{a_1+a_3-1}}{B(a_1, a_2, a_3)(1-xy)^{a_1+a_2+a_3}}, \tag{53}$$

where $B(a_1, a_2, a_3) = \prod_{i=1}^{3} \Gamma(a_i)/\Gamma(\sum_{i=1}^{3} a_i)$. The bivariate density (53) is positively likelihood ratio dependent and hence positive quadrant dependent. Sarabia and Castillo (2006) have considered a generalization of (53) under a conditional specification. With this specification, they obtain a broad class of distributions, where an important submodel is

$$f(x, y; a_1, a_2, b_1, m) = \frac{x^{a_1-1}(1-x)^{b_1-1} y^{a_2-1} (1-y)^{a_1+b_1-a_2-1}}{n(a_1, a_2, b_1, m)(1-mxy)^{a_1+b_1}}, \tag{54}$$

where $a_1, b_1, a_1 + b_1 - a_2 > 0$, $m \leq 1$ and where $1/n$ is the normalizing constant. This model contains the Olkin and Liu (2003) proposal for $m = 1$ and $X$ is stochastically increasing or decreasing with $Y$, so, consequently

$$\text{sign}\rho(X, Y) = \text{sign}(m).$$

Then, if $0 < m \leq 1$ we have positive correlation and if $m < 0$, negative correlations. The marginal distributions are of the Gauss hypergeometric type.

### 6.3 Bivariate *t* distribution

The usual bivariate spherically symmetric distribution on $n_1$ degrees of freedom is defined as (Fang *et al.*, 1990)

$$\begin{aligned} X &= X_1/\sqrt{X_3/n_1}, \\ Y &= X_2/\sqrt{X_3/n_1}, \end{aligned}$$

where $X_1, X_2, X_3$ are mutually independent random variables with distributions $X_1, X_2 \sim \mathcal{N}(0, 1)$ standard normal and $X_3 \sim \chi^2_{n_1}$ chi-squared distribution on $n_1$ degrees of freedom. Note that the marginal distributions are both (dependent) Student $t$ distributions on $n_1$ degrees of freedom. If we need a bivariate distribution with Student $t$ marginals with different degrees of freedom $\nu_1$ and $\nu_2$, one possibility is defined,

$$\begin{aligned} X &= X_1/\sqrt{X_3/\nu_1}, \\ Y &= X_2/\sqrt{(X_3 + X_4)/\nu_2}, \end{aligned}$$

where $\nu_1 = n_1$, $\nu_2 = n_1 + n_2$ and $X_4 \sim \chi^2_{n_2}$ is a new independent chi-squared distribution on $n_2$ degrees of freedom. This model has been proposed by Jones (2002). Note that last model includes the previous one by taking $n_4 = 0$ degrees of freedom.

An alternative bivariate $t$ distribution (including the independence case) has been proposed by Shaw and Lee (2008).

### 6.4 Bivariate Marshall-Olkin type distributions

Let $X_1, X_2$ and $X_3$ be mutually independent random variables with cdf $G_i(\cdot)$, $i = 1, 2, 3$. Define the random variable $(X, Y)$ as

$$X = \min\{X_1, X_3\},$$
$$Y = \min\{X_2, X_3\}.$$

With this scheme, $X$ and $Y$ are dependent, through the common random latent variable $X_3$. The joint survival function is

$$\Pr(X > x, Y > y) = \overline{G}_1(x)\overline{G}_2(y)\overline{G}_3(z), \tag{55}$$

where $z = \max\{x, y\}$ and $\overline{G} = 1 - G$. Note that (55) has a singular component. If the components correspond to exponential distribution, we obtain the Marshall-Olkin distribution (Marshall and Olkin, 1967). Other survival models have been considered by Sarhan and Balakrishnan (2007) with the exponentiated exponential distribution, as well as a mixture of the proposed bivariate distribution. Arnold and Brockett (1983) have obtained a bivariate Gompertz-Makeham distribution using a similar construction.

### 6.5 Bivariate $F$ distribution

Now, let $X_1, X_2$ and $X_3$ be mutually independent chi-squared random variables with degrees of freedom $n_i > 0$, $i = 1, 2, 3$. The classical bivariate $F$ distribution is defined as (Kotz, Balakrishnan and Johnson, 2000)

$$X = \frac{X_1/n_1}{X_3/n_3}, \quad Y = \frac{X_2/n_2}{X_3/n_3}.$$

We have $X \sim \mathcal{F}_{n_1,n_3}$ and $Y \sim \mathcal{F}_{n_2,n_3}$, which share the degrees of freedom on the denominator. In order to obtain a bivariate $F$ distribution with arbitrary degrees of freedom, El-Bassiouny and Jones (2007) have proposed the model

$$X = \frac{X_1/n_1}{X_3/n_3}, \quad Y = \frac{X_2/n_2}{(X_3 + X_4)/(n_3 + n_4)},$$

where $X_4 \sim \chi^2_{n_4}$ is a new independent chi-squared distribution and now $X \sim \mathcal{F}_{n_1,n_3}$ and $Y \sim \mathcal{F}_{n_2,n_3+n_4}$, which includes the previous model. The joint pdf can be expressed as a function of the Gauss hypergeometric function and positive correlation still arises.

## 7 Other Methods

In this last section multivariate weighted distributions, graphical models based on vines and multivariate Zipf distributions are briefly commented upon.

### *7.1 Multivariate weighted distributions*

The usual weighted distributions can be introduced in the following way. Let $F$ be a distribution function of a random variable $X$ and $w$ a positive function. The univariate weighted distribution associated with $F$ and $w$ is defined as (Rao, 1965)

$$dF^w(x) = \frac{w(x)}{E[w(X)]} dF(x),$$

if $E[w(X)] < \infty$. If $F$ is absolutely continuous, the density $f^w$ associated to $F$ is called the weighted density, and the corresponding random variable is denoted by $X^w$. Weighted random variables are used to model sampling procedures with unequal sampling probabilities proportional to a weighted function $w$, that is, when we want to study $X$ with a sample from $X^w$. In a multivariate setting, let $F$ be an absolutely continuous distribution of a $p$-dimensional random vector with density $f$ and $w : \mathbb{R}^p \to \mathbb{R}$ a positive function. The multivariate weighted or biased distribution associated with $F$ and $w$ is defined by the $p$-dimensional probability density function

$$f^w(x_1, \ldots, x_p) = \frac{w(x_1, \ldots, x_p)}{E[w(X_1, \ldots, X_p)]} f(x_1, \ldots, x_p), \tag{56}$$

if $E[w(X_1, \ldots, X_p)] < \infty$. In the particular case $w(x_1, \ldots, x_p) = x_1 \cdots x_p$, it is called the multivariate size biased distribution, with density

$$f^{sb}(x_1, \ldots, x_p) = \frac{x_1 \cdots x_p}{E(X_1 \cdots X_p)]} f(x_1, \ldots, x_p). \tag{57}$$

Classes of distributions (56) and (57) have been compiled and studied by Navarro *et al.* (2006), paying special attention to reliability aspects, ordering and equilibrium distributions in renewal processes. The multivariate distribution (57) represents sampling methods where a vector $\underline{X} = (X_1, \ldots, X_p)$ has a sampling probability proportional to $x_i$, $i = 1, \ldots, p$. In the event that $\underline{X}$ represents the life lengths of items in a system, then the sampling probability for a system is proportional to the life lengths of its units. Additional applications included aerial sampling methods and tourism studies (see Navarro *et al.*, 2006).

Many of the multivariate distributions introduced in the previous sections respond to general scheme (56), for example, distributions (39), (40) and (44).

### 7.2 Graphical models: vines

In the context of graphical dependency models, a new methodology called vines has been introduced recently by Berdford and Cooke (2002) and Kurowicka and Cooke (2006) to build complex multivariate highly dependent models satisfying conditional dependence specifications. This methodology can be considered as an alternative of the simple Markov trees to belief networks and influence diagrams. The definition of conditional independence is weakened to allow for several kinds of conditional dependence.

Copulae construction is the usual way to build a model with dependence structure. However, in high dimensional distributions, this methodology is complicated, since it requires a large number of possible pair-copulae constructions. Vines let us organize this large amount of information through the regular vine or other particular cases of regular vines, the canonical vine and the D-vine. This new methodology has proved useful, for example, in the analysis of financial data sets (Aas *et al.*, 2008). A connection between vines and other types of related works has been obtained in the specification of a multivariate normal distribution using partial correlations, from a generalization of a problem dealt with by Joe (1996).

### 7.3 Multivariate Zipf distributions

The multivariate Zipf distributions correspond to the discrete version of the multivariate Pareto distributions, introduced by Arnold (1983). In the univariate case, the Zipf distribution is the discrete version of the usual Pareto distribution. A discrete random variable $X$ is said to have a Zipf(IV) distribution with positive parameters $k_0, \sigma, \gamma$ and $\alpha$ if its survival function is

$$\Pr(X \geq k) = \left[1 + \left(\frac{k - k_0}{\sigma}\right)^{1/\gamma}\right]^{-\alpha}, \; k = k_0, k_0 + 1, \ldots \tag{58}$$

***Table 1:*** *Multivariate Zipf distributions*

| Multivariate Zipf | Survival function, $\bar{F}_{\underline{X}}(\underline{k}) = \Pr\left(\underline{X} \geq \underline{k}\right)$ |
|---|---|
| Type (I) <br> $\underline{X} \sim M^{(p)}\text{Zipf(I)}(\underline{\sigma}, \alpha)$ | $\left[1 + \sum_{i=1}^{p} (k_i/\sigma_i)\right]^{-\alpha}, \quad k_i \in \{0, 1, \ldots\},$ <br> $1 \leq i \leq p,$ <br> $\alpha > 0, \ \sigma > 0$ |
| Type (II) <br> $\underline{X} \sim M^{(p)}\text{Zipf(II)}(\underline{\mu}, \underline{\sigma}, \alpha)$ | $\left[1 + \sum_{i=1}^{p} ((k_i - \mu_i)/\sigma_i)\right]^{-\alpha}, \quad k_i \geq \mu_i,$ <br> $1 \leq i \leq p,$ <br> $\alpha > 0, \ \sigma > 0; k_i, \ \mu_i$ integers |
| Type (III) <br> $\underline{X} \sim M^{(p)}\text{Zipf(III)}(\underline{\mu}, \underline{\sigma}, \underline{\gamma})$ | $\left[1 + \sum_{i=1}^{p} ((k_i - \mu_i)/\sigma_i)^{1/\gamma_i}\right]^{-1}, \quad k_i \geq \mu_i,$ <br> $1 \leq i \leq p,$ <br> $\sigma > 0, \ \gamma_i > 0; k_i, \ \mu_i$ integers |
| Type (IV) <br> $\underline{X} \sim M^{(p)}\text{Zipf(IV)}(\underline{\mu}, \underline{\sigma}, \underline{\gamma}, \alpha)$ | $\left[1 + \sum_{i=1}^{p} ((k_i - \mu_i)/\sigma_i)^{1/\gamma_i}\right]^{-\alpha}, \quad k_i \geq \mu_i,$ <br> $1 \leq i \leq p,$ <br> $\alpha > 0, \ \sigma > 0; k_i, \ \mu_i$ integers |

Yeh (2002) introduces six different multivariate Zipf distributions in terms of the joint survival functions of $\underline{X} = (X_1, \ldots, X_p)$ having Zipf marginals as (58). Four of them are presented in Table 1.

The four distributions in Table 1 are analogous to the continuous multivariate Pareto distributions in Arnold (1983), also studied and extended in Yeh (2004, 2007). The standard multivariate Zipf distribution, Zipf(IV) $(\underline{0}, \underline{1}, \underline{1}, 1)$ arises as a unifom or beta mixture of conditional independent geometric distributions according to mixture (10).

The order statistics of discrete random variables are difficult to work with except for the extremes. However, in the case of families in Table 1 several results can be obtained (Yeh, 2002). Note that Zipf(I), II and III are special cases of the multivariate Zipf(IV) family as follows:

$$
\begin{aligned}
M^{(p)} \text{ Zipf(I) } (\underline{\sigma}, \alpha) &= M^{(p)} \text{ Zipf(IV) } (\underline{0}, \underline{\sigma}, \underline{1}, \alpha), \\
M^{(p)} \text{ Zipf(II) } (\underline{\mu}, \underline{\sigma}, \alpha) &= M^{(p)} \text{ Zipf(IV) } (\underline{\mu}, \underline{\sigma}, \underline{1}, \alpha), \\
M^{(p)} \text{ Zipf(III) } (\underline{\mu}, \underline{\sigma}, \underline{\gamma}) &= M^{(p)} \text{ Zipf(IV) } (\underline{\mu}, \underline{\sigma}, \underline{\gamma}, 1).
\end{aligned}
$$

## Acknowledgements

The authors wish to thank the Editor Monserrat Guillén for her kind invitation to write this paper and are grateful for her valuable comments which have improved the presentation of the paper. The authors also thank the Ministerio de Educación y Ciencia (projects SEJ2007-65818 (JMS) and SEJ2006-12685 (EGD)) for partial support of this work.

## References

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2008). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*. In press.

Ahsanullah, M. and Nevzorov, V. B. (2005). *Order Statistics: Examples and Exercises*. Nova Science Publishers.

Aitchison, J. and Ho, C. H. (1989). The multivariate Poisson-lognormal distribution. *Biometrika*, 76, 4, 643-653.

Arnold, B. C. (1983). *Pareto Distributions*. International Co-operative Publishing House, Fairland, MD.

Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1992). *A First Course in Order Statistics*. John Wiley and Sons, New York.

Arnold, B. C. and Beaver, R. J. (2000). The skew-Cauchy distribution. *Statistics and Probability Letters*, 49, 285-290.

Arnold, B. C. and Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test*, 11, 7-54.

Arnold, B. C. and Brockett, P. L. (1983). Identifiability for dependent multiple decrement/competing risk models. *Scandinavian Actuarial Journal*, 10, 117-127.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (1992). *Conditionally specified distributions*, Lecture Notes in Statistics, vol. 73, Springer Verlag, New York.

Arnold, B. C, Castillo, E. and Sarabia, J. M. (1999). *Conditional Specification of Statistical Models*, Springer Series in Statistics, Springer Verlag, New York.

Arnold, B. C., Castillo, E. Sarabia, J. M. and González-Vega, L. (2000). Multiple modes in densities with normal conditionals. *Statistics and Probability Letters*, 49, 355-363.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2001). Conditionally specified distributions: an introduction (with discussion), *Statistical Science*, 16, 249-274.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2002). Conditionally specified multivariate skewed distributions. *Sankhya, Ser. A*, 64, 1-21.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2006). Families of multivariate distributions involving the Rosenblatt construction. *Journal of the American Statistical Association*, 101, 1652-1662.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2007a). Distributions with generalized skewed conditionals and mixtures of such distributions. *Communications in Statistics, Theory and Methods*, 36, 1493-1503.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2007b). Variations on the classical multivariate normal theme. *Journal of Statistical Planning and Inference*, 137, 3249-3260.

Arnold, B. C. and Press, S. J. (1989). Compatible conditional distributions. *Journal of the American Statistical Association*, 84, 152-156.

Arnold, B. C. and Strauss, D. (1991). Bivariate distributions with conditionals in prescribed exponential families, *Journal of the Royal Statistical Society, Ser. B*, 53, 365-375.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32, 159-188.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83, 715-726.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distributions. *Journal of the Royal Statistical Society, Ser. B*, 61, 579-602.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew $t$ distribution. *Journal of the Royal Statistical Society, Ser. B*, 65, 367-389.

Balakrishnan, N. (2002). Discussion of "Skewed multivariate models related to hidden truncation and/or selective reporting", by B. C. Arnold and R. J. Beaver. *Test*, 11, 37-39.

Balakrishnan, N. (2004). Discrete multivariate distributions. In: *Encyclopedia of Actuarial Sciences* (Eds. J. L. Teugels and B. Sundt), pp. 549-571. John Wiley and Sons, New York.

Balakrishnan, N. (2005). Discrete multivariate distributions. In: *Encyclopedia of Statistical Sciences*, 2nd ed. (Eds. N. Balakrishnan, C. Read and B. Vidakovic). John Wiley and Sons, New York.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24, 1-13.

Bedford, T. and Cooke, R. (2002). Vines-a new graphical model for dependent random variables. *The Annals of Statistics*, 30, 4, 1031-1068.

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79, 99-113.

Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*, John Wiley and Sons, New York.

David, H. A. and Nagaraja, H. N. (2003). *Order Statistics*. Third Edition. John Wiley, New York.

El-Bassiouny, A. H. and Jones, M. C. (2007). A bivariate $F$ distribution with marginals on arbitrary numerator and denominator degrees of freedom, and related bivariate beta and $t$ distributions. *Technical Report*, The Open University, UK.

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London.

Genton, M. G. (Ed.) (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Chapman and Hall/CRC, Boca Raton, FL.

Gómez-Déniz, E., Sarabia, J. M. and Calderín-Ojeda, E. (2008). Univariate and multivariate versions of the negative binomial inverse Gaussian distribution with applications. *Insurance: Mathematics and Economics*, 42, 39-49.

González-Farías, G., Domínguez-Molina, J. A. and Gupta, A. K. (2003). Additive properties of skew-normal random variables. *Journal of Statistical Planning and Inference*, 126, 521-534.

González-Farías, G., Domínguez-Molina, J. A. and Gupta, A. K. (2004). The closed skew-normal distribution. In: *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Genton, M.G., Ed., Chapman and Hall/CRC, Boca Raton, FL, 25-42.

Gupta, A. K., González-Farías, G. and Domínguez-Molina, J. A. (2004). A multivariate skew normal distribution. *Journal of Multivariate Analysis*, 89, 181-190.

Gupta, R. C. and Gupta, R. D. (2004). Generalized skew normal model. *Test*, 13, 501-524.

Gurland, J. (1957). Some interrelations among compound and generalized distributions. *Biometrika*, 44, 265-268.

Hutchinson, T. P. and Lai, C. D. (1990). *Continuous Bivariate Distributions, Emphasizing Applications*, Rumsby Scientific Press, Adelaide, Australia.

Joe, H. (1996). Families of $m$-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In: *Distributions with fixed margins and related topics*, IMS Lecture Notes Monograph Series, 120-141.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, New York.

Joe, H. and Liu, Y. (1996). A model for a multivariate binary response with covariates based on conditionally specified logistic regressions. *Statistics and Probability Letters*, 31, 113-120.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. John Wiley and Sons, New York.

Jones, M. C. (2002). A dependent bivariate *t* distribution with marginals on different degrees of freedom. *Statistics and Probability Letters*, 56, 163-170.

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics (with discussion), *Test*, 13, 1-43.

Jones, M. C. and Larsen, P. V. (2004). Multivariate distributions with support above the diagonal. *Biometrika*, 91, 975-986.

Kamps, U. (1995). A concept of generalized order statistics. *Journal of Statistical Planning and Inference*, 48, 1-23.

Katti, S. K. (1966). Interrelations among generalized distributions and their components. *Biometrics*, 22, 44-52.

Kemp, C. D. and Papageorgiou, H. (1982). Bivariate Hermite distributions. *Sankhya, Ser. A*, 44, 269-280.

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. Dekker, New York.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions*, vol. 1: Models and Applications. John Wiley and Sons, New York.

Kotz, S. and Nadarajah, S. (2004). *Multivariate T-Distributions and Their Applications*. Cambridge University Press, Cambridge.

Kurowicka, D. and Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. John Wiley and Sons, New York.

Lai, C. D. (2004). Constructions of continuous bivariate distributions. *Journal of the Indian Society for Probability and Statistics*, 8, 21-43.

Lai, C. D. (2006). Constructions of discrete bivariate distributions. In: *Advances on Distribution Theory, Order Statistics and Inference*, Birkhäuser, Boston., N. Balakrishnan, E. Castillo, J.M. Sarabia Editors, 29-58.

Lee, M.-L. T. (1996). Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics, Theory and Methods*, 25, 1207-1222.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62, 30-44.

Marshall, A. W. and Olkin, I. (2007). *Life Distrubutions. Structure of Nonparametric, Semiparametric, and Parametric Families*. Springer, New York.

Mikosch, T. (2006). Copulas: tales and facts. *Extremes*, 9, 3-20.

Navarro, J., Ruiz, J. M. and Del Aguila, Y. (2006). Multivariate weighted distributions: a review and some extensions. *Statistics*, 40, 51-64.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Second Edition. Springer, New York.

Øigard, T. A. and Hanssen, A. (2002). The multivariate normal inverse Gaussian heavy-tailed distribution: simulation and estimation. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2, 1489-1492.

Olkin, I. and Liu, R. (2003). A bivariate beta distribution. *Statistics and Probability Letters*, 62, 407-412.

Papadatos, N. (1995). Intermediate order statistics with applications to nonparametric estimation. *Statistics and Probability Letters*, 22, 231-238.

Protassov, R. S. (2004). EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed $\lambda$. *Statistics and Computing*, 14, 67-77.

Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment. *Sankhya Series A*, 27, 311-324.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23, 470-472.

Sarabia, J. M. (2002). Discussion of "Skewed Multivariate Models Related to Hidden Truncation and/or Selective Reporting", by B. C. Arnold and R. J. Beaver. *Test*, 11, 49-52.

Sarabia, J. M. and Castillo, E. (2006). Bivariate distributions based on the generalized three-parameter beta distribution. In: *Advances on Distribution Theory, Order Statistics and Inference*, Birkhäuser, Boston., N. Balakrishnan, E. Castillo, J. M. Sarabia Editors, 85-110.

Sarabia, J. M., Castillo, E., Gómez, E. and Vázquez, F. (2005). A class of conjugate priors for log-normal claims based on conditional specification. *Journal of Risk and Insurance*, 72, 479-495.

Sarabia, J. M., Castillo, E., Pascual, M. and Sarabia, M. (2007). Bivariate income distributions with lognormal conditionals. *Journal of Economic Inequality*, 5, 371-383.

Sarabia, J. M. and Gómez-Déniz, E. (2008). Univariate and multivariate Poisson-beta distributions with applications. *Submitted*.

Sarabia, J. M., Gómez-Déniz, E. and Vázquez-Polo, F. (2004). On the use of conditional specification models in claim count distributions: an application to bonus-malus systems. *ASTIN Bulletin*, 34, 85-98.

Sarabia, J. M. and Guillén, M. (2008). Joint modelling of the total amount and the number of claims by conditionals. *Submitted*.

Sarhan, A. M. and Balakrishnan, N. (2007). A new class of bivariate distributions and its mixture. *Journal of Multivariate Analysis*, 98, 1508-1527.

Sarmanov, O. V. (1966). Generalized normal correlation and two-dimensional Frechet classes. *Doklady, Soviet Mathematics*, 168, 596-599.

Shaw, W. T. and Lee, K. T. A. (2008). Bivariate Student *t* distributions with variable marginal degrees of freedom and independence. *Journal of Multivariate Analysis*, In press.

Sichel, H. S. (1971). On a family of discrete distributions particularly suited to represent long-tailed frequency data. In: *Proceedings of the third Symposium on Mathematical Statistics*, Eds. N.F. Laubscher, Pretoria, South Africa: Council for Scientific and Industrial Research, 51-97.

Stein, G., Zucchini, W. and Juritz, J. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, 82, 938-944.

Stigler, S. M. (1977). Fractional order statistics, with applications. *Journal of the American Statistical Association*, 72, 544-550.

Tonda, T. (2005). A class of multivariate discrete distributions based on an approximate density in GLMM. *Hiroshima Math. J.*, 35, 327-349.

Vernic, R. (1997). On the bivariate generalized Poisson distribution. *ASTIN Bulletin*, 27, 23-31.

Vernic, R. (2000). A multivariate generalization of the generalized Poisson distribution. *ASTIN Bulletin*, 30, 57-67.

Vernic, R. (2005). Multivariate skew-normal distributions with applications in insurance. *Insurance: Mathematics and Economics*, 38, 413-426.

Walker, S. G. and Stephens, D. A. (1999). A multivariate family on $(0, \infty)^p$. *Biometrika*, 86, 703-709.

Wang, J., Boyer, J. and Genton, M. G. (2004). A skew-symmetric representation of multivariate distributions. *Statistica Sinica*, 14, 1259-1270.

Wesolowski, J. (1996). A New conditional specification of the bivariate Poisson conditionals distribution. *Statistica Neerlandica*, 50, 390-393.

Yadegari, I., Gerami, A. and Khaledi, M. J. (2008). A generalization of the Balakrishnan skew-normal distribution. *Statistics and Probability Letters*, In Press.

Yeh, H. C. (2002). Six multivariate Zipf distributions and their related properties. *Statistics and Probability Letters*, 56, 131-141.

Yeh, H. C. (2004). Some properties and characterizations for generalized multivariate Pareto distributions. *Journal of Multivariate Analysis*, 88, 47-60.

Yeh, H. C. (2007). Three general multivariate semi-Pareto distributions and their characterizations. *Journal of Multivariate Analysis*, 98, 1305-1319.

# Discussion of “Construction of multivariate distributions: a review of some recent results” by José María Sarabia and Emilio Gómez-Déniz

**M. C. Pardo**
Deparment of Statistics and O.R (I), Complutense University of Madrid,
28040-Madrid, SPAIN, Phone: 34 91 3944473, fax: 34 91 3944606.
E-mail: mcapardo@mat.ucm.es,
http://www.mat.ucm.es/˜mcapardo

I congratulate the authors on this excellent review. In this review paper they present a nice overview on construction of multivariate distributions. Due to it being such an active field of research, new models are constantly being discovered. However, they have been able to present some of the most recent methods in a very clear manner, and many of those omitted can be found in the references mentioned. It is my pleasure to comment on this article.

I agree with professors Sarabia and Gómez-Déniz that is not possible to mention all the methods for constructing distributions that exist. However, owed perhaps to my own field of research, I miss the *Maximum Entropy Principle* used to construct probability distributions. Therefore, my discussion focuses on presenting the practical usefulness of this method.

Let $(\mathcal{X}, \beta_{\mathcal{X}},\ P)$ be the statistical space associated with the random variable $\boldsymbol{X}$, where $\beta_{\mathcal{X}}$ is the $\sigma$-field of Borel subsets $A \subset \mathcal{X}$ and $\{P\}$ is a family of probability distributions defined on the measurable space $(\mathcal{X}, \beta_{\mathcal{X}})$. We assume that the probability distributions $P$ are absolutely continuous with respect to $\sigma$-finite measure $\mu$ on $(\mathcal{X}, \beta_{\mathcal{X}})$. The Shannon entropy is defined by

$$H = -\int_{\mathcal{X}} f(\mathbf{x}) \log f(\mathbf{x}) d\mu(\mathbf{x}) \tag{1}$$

where $f(\mathbf{x}) = \frac{dP}{d\mu}(\mathbf{x})$.

The Maximum Entropy Principle states that, maximizing entropy subject to a set of constraints can be regarded as deriving a distribution that is consistent with the information specified in the constraints while making minimal assumptions about the form of the distribution other than those embodied in the constraints. Numerous distributions have been obtained in this manner (Kapur, 1994; Ebrahimi, 2000, Asadi *et al.*, 2004). For example, the normal distribution may be obtained as the distribution on the real line that has maximum entropy subject to having specified mean and variance; see Rao (1965, p. 132). An earlier result by Goldman (1955) characterized $\mathcal{N}(0, \sigma^2)$ as the MED with specified value of $E[Z^2]$ being $Z$ a continuous random variable with support $(-\infty, \infty)$; the MED with specified value of $E\left[(Z-a)^2\right]$ was shown to be $\mathcal{N}(a, \sigma^2)$ by Lisman and van Zuylen (1972). More generally, if a set of moments

$E[X^r], r = 1, \ldots, R$, is specified, the distribution on the real line that has the maximum entropy subject to these constraints has probability density

$$f(x) \propto \exp\left(\sum_{i=1}^{R} \alpha_r x^r\right)$$

for suitable constants $\alpha_r, r = 1, \ldots$, R. Hosking (2007) derived the distribution that has maximum entropy conditional on having specified values of its first r L-moments. Note that L-moments are now widely used in the environmental sciences to summarize data and fit frequency distributions. This maximum entropy distribution has a polynomial density-quantile distribution (PDQ distribution). Some special cases of the PDQ distribution are: On a finite interval, the MED is the uniform distribution; on a semi-infinite interval, the MED with the first L-moments specified is the exponential distribution and on an infinite interval, the MED with the first two L-moments specified is the logistic distribution. Maximum entropy distributions conditional on specified L-moments of orders $\{1, 2, 3\}$ and $\{1, 2, 4\}$ generate families of distributions that generalize the logistic distribution and may be useful for modelling data.

There is a lot of work devoted to the maximum entropy characterization of the most well-known univariate probability distributions. Although available literature is significantly less for the multivariate distributions, the book of Kapur (1989) considers several usual multivariate distributions and Zografos (1999) considered the cases of Pearson's types II and VII multivariate distributions (t-distribution and generalized Cauchy distribution are obtained from an application of Pearson's types VII distribution). Aulogiaris and Zografos (2004) considered symmetric Kotz type and Burr multivariate distributions. Later Bhattacharya (2006) derived appropriate constraints which establish the maximum entropy characterization of the Liouville distributions among all multivariate distributions.

Amongst discrete distributions, the geometric distribution with support 1, 2, ... is the MED given a specified arithmetic mean. The Riemann zeta distribution, also called the discrete Pareto distribution, is the MED for a specific geometric mean. In linguistics, it is called the Zipf distribution. It has also been used to model numbers of insurance policies, the distribution of surnames and scientific productivity. The Good type-I distribution is the MED when the arithmetic and geometric means are both specified. If $x = 1, 2, \ldots, n$ and there is no restriction on the probabilities, then the MED is a discrete rectangular distribution. Given finite support and specified arithmetic or geometric means, or both arithmetic and geometric mean, the MEDs are the right-truncated geometric, right-truncated Rienmann zeta, and right-truncated Good type-I distribution, respectively. Kemp (1997) obtained a discrete analogue of the normal distribution as the distribution that is characterized by maximum entropy, specified mean and variance, and integer support on $(-\infty, \infty)$. Binomial and Poisson distributions are also MEDs of suitable defined sets (Harremoës, 2001).

The Maximum Entropy Principle has applications in many domains, but was originally motivated by statistical physics (Jaynes, 1957), which attempts to relate macroscopic measurable properties of physical systems to a description at an atomic or molecular level. Applications in econometrics can be seen in several works of Theil (see for example, Theil and Fiebig, 1984). A popular method for estimation of spectral densities was given by Burg (1975) based on Maximum Entropy method. Many works and books following this idea have appeared, see for instance Golan *et al.* (1996). In Finance, this principle is applied to infer a probability density from option prices. Buchen and Kelly (1996) showed that, with a set of well-spread simulated exact-option prices, the MED approximates a risk-neutral distribution to a high degree of accuracy. Guo (2001), motivated by the characteristic that a call price is a convex function of the option's strike price, suggests a simple convex-spline procedure to reduce the impact of noise on observed option prices before inferring the MED.

Apart from density estimation, many statistical problems have been studied on the basis of the Maximum Entropy Principle. Using sample quantiles, Menéndez *et al.* (1997) proposed a point estimation procedure as well as a goodness-of-fit test statistic based on the he Maximum Entropy Principle. But there are many important different entropy measures (see Chapter 2 of Pardo, 2006), and in a similar manner the Maximum Entropy Principle associated with these others entropies can be defined. Menéndez *et al.* (1997) generalized the previous work using a general family of entropies that contains (1).

## References


Asadi, M., Ebrahimi, N., Hamedani, G. G. and Soofi, E. S. (2004). Maximum dynamic Entropy Models. *IEEE Transactions on Information Theory*, 50, 177-183.

Aulogiaris, G. and Zografos, K. (2004). A Maximum Entropy characterization of symmetric Kotz type and multivariate Burr distributions. *TEST*, 13, 65-83.

Bhattacharya, B. (2006). Maximum entropy characterizations of the multivariate Liouville distributions. *Journal of Multivariate Analysis*, 97, 1272-1283.

Ebrahimi, N. (2000). The maximum entropy method for life time distributions. *Sankhyā*, A, 62, 236-243.

Burg, J. P. (1975). Maximum Entropy Spectral Analysis. *PhD. dissertation, Department of Geophysics, Standford University.*

Buchen, P. W. and Kelly, M. (1996). The Maximum Entropy Distribution of an Asset Inferred from OPtion Prices. *Journal of Financial and Quantitative Analysis*, 31, 143-159.

Golan, A., Judge, G. G. and Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Wiley, New York.

Goldman, S. (1955). *Information Theory*, Prentice-Hall, New York.

Guo, W. (2001). Maximum Entropy in Option Pricing: A convex-Spline Smoothing Method. *The Journal of Future Markets*, 21, 819-832.

Harremoës, P. (2001). Binomial and Poisson Distributions as Maximum Entropy Distributions. *IEEE Transactions on Information Theory*, 47, 2039-2041.

Hosking, J. R. M. (2007). Distributions with maximum entropy subject to constraints on their L-moments or expected order statistics. *Journal of Statistical Planning and Inference*, 137, 2840-2891.

Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, 106, 620-630.

Kapur, J. N. (1989). *Maximum Entropy Models in Engineering*. Wiley, New York.

Kapur, J. N. (1994). *Measures of Information and their applications*. Wiley, New York.

Kemp, A. W. (1997). Characterizations of a discrete normal distributions. *Journal of Statistical Planning and Inference*, 63, 223-229.

Lisman, J. H. C. and van Zuylen, M. C. A. (1972). Note on the generation of most probable frequency distriburions. *Statistica Neerlandica*, 26, 19-23.

Ménendez, M. L., Morales, D. and Pardo, L. (1997a). Maximum entropy principle and statistical inference on condensed data. *Statistics and Probability Letters*, 34, 85-93.

Ménendez, M. L., Pardo, J. A. and Pardo, M. C. (1997b). Estimators based on sample quantiles using $(h, \phi)-$ entropy measures. *Applied Mathematics Letters*, 11, 99-104.

Pardo, L. (2006). *Statistical Inference Based on Divergence Measures.* Chapman & Hall.

Rao, C. R. (1965). *Linear Statistical Inference and Its Application*. Wiley, New York.

Theil, H. and Fiebig, D. G. (1984). *Maximun Entropy Estimation of Continuous Distributions.* Ballinger, Cambridge.

Zografos, K. (1999). On maximum entropy characterization of Pearson's type II and VII multivariate distributions. *Journal of Multivariate Analysis*, 71, 67-75.

**Jorge Navarro**
Facultad de Matemáticas, Universidad de Murcia, 30100 Murcia, Spain

The construction of Multivariate Distributions which can be fitted to multivariate data sets is a very relevant topic of research in probability and statistics. First of all I would like to warmly congratulate Professors Sarabia and Gómez-Déniz for an excellent and stimulating review of some recent results on this topic. The different methods presented can be classified in two groups: (*i*) multivariate distributions arising out from univariate distributions, and (*ii*) multivariate distributions obtained from other multivariate distributions. In the first group we can include the techniques based on (*a*) order statistics, (*b*) mixtures, (*c*) conditional specification and (*e*) the method of the variables in common, while, in the second group, we can include the methods of (*d*) skew distributions and (*f*) weighted distributions.

The distribution of order statistics (OS) or other generalizations such as the Generalized Order Statistics (GOS) only depends on the univariate parent distribution from which the sample of IID (independent and identically distributed) random variables is obtained. Two possible extensions can be considered here. If we consider (or we have) a sample $X_1, X_2, \ldots, X_n$ of INID (independent non-necessarily identically distributed) random variables, then the joint distribution only depend on the univariate distributions $F_i(x) = \Pr(X_i \leq x)$ $i = 1, 2, \ldots, n$. In this case, the joint distribution of the OS and the joint distribution of a subset of OS can be represented in terms of permanents (see Balakrishnan (2007)). The second option is to consider the OS obtained from a random vector $(X_1, X_2, \ldots, X_n)$, where the possible dependence between the random variables is modelled through their joint distribution. This case has special interest when $(X_1, X_2, \ldots, X_n)$ represent the lifetimes of some components in a system. This case will be included in the second group (*ii*) since we obtain multivariate distributions (that of subsets of OS) from a parent multivariate distribution. In the three cases, it is of special interest to study the distribution of the $k$ first OS $(X_{1:n}, X_{2:n}, \ldots, X_{k:n})$ (for $k < n$) since in many practical situations, when we put-on-test some devices (with lifetimes $X_i$, $i = 1, 2, \ldots, n$), at the end of the test period we only have information about the 'early failures' (see, *e.g.*, Balakrishnan, Ng and Panchapakesan (2006)). In other cases we only have information about the series system $(X_{1:n})$ or the parallel system $(X_{n:n})$. It these cases, it is interesting to note how multivariate models can also be used to obtain new relevant univariate models (see Navarro, Ruiz and Sandoval (2006)).

With respect to the methods based on mixtures, first we must note that they are not the usual mixtures used to represent heterogeneous populations obtained by mixing some groups with different characteristics (*e.g.* a mixture of two multivariate normal distributions). This case will be included in group (*ii*). Actually, the multivariate distributions obtained by (9) or (12) Sarabia and Gómez-Déniz' paper are the joint distributions of IID random variables that share one (or more) parameter with a known distribution. The dependence is due to this common parameter. These models have special interest in reliability and survival studies when $(X_1, X_2, \dots, X_n)$ represent the lifetimes of some components in a system. Usually, the components are independent but they share the same environment and hence their distributions depends on some common parameters which induce a dependence between them.

It is not easy to add more on conditional specification since Sarabia (jointly with Arnold and Castillo) is one of the fathers of this technique. I would only like to say that, in my opinion, this is a very reasonable technique to obtain multivariate models from univariate models when the conditional distributions are known. In practice, this is quite common in reliability or survival studies where, for example, we can suppose that if $(X, Y)$ are the lifetimes of two units in a parallel system, when a unit have failed at age $t$, the distribution of the other component has a known distribution (*e.g.* exponential) with some parameters depending on $t$. It is also important to note other possible situations (also studied in the book by Arnold, Castillo and Sarabia (1999)) as, for example, when we know the conditional distributions of $(X|Y \leq y)$ or $(X|Y \geq y)$. The distributions obtained by this method can be included in the distributions obtained from characterization methods, that is, we look for all the multivariate models which satisfy a certain property (in this case to have some specified conditional distributions). This is a classical method to obtain multivariate distributions. I would like to note here that another (related) option is to obtain probability models by characterizations based on 'ageing measures' such as the hazard rate or the mean residual life functions and their corresponding multivariate generalizations. For example, Ruiz, Marín and Zoroa (1993) gave a general way to obtain multivariate models from $m(\mathbf{x}) = E(\mathbf{X}|\mathbf{X} \geq \mathbf{x})$, where $\mathbf{X} = (X_1, X_2, \dots, X_n)$ and $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Some recent results are given in Navarro and Ruiz (2004), Kotz, Navarro and Ruiz (2007), Navarro, Ruiz and Sandoval (2007) and Navarro (2008). For example, in Navarro and Ruiz (2004), the multivariate normal distribution with mean vector $\mu$ and variances-covariances matrix $V$ is characterized by $m(\mathbf{x}) = \mu + Vh(\mathbf{x})$, where $h(\mathbf{x})$ is the hazard gradient (the multivariate version of the hazard rate function which contains the hazard rate functions of the conditional distributions $(X_i|X_j \geq x_j, j \neq i)$). A general method is given in Kotz, Navarro and Ruiz (2007) where, for example, Arnold and Strauss bivariate exponential distribution is obtained from $m(x_1, x_2) = (k_1, k_2)' + Vh(x_1, x_2)$ for $x_1, x_2 \geq 0$ where

$$V = \begin{pmatrix} 0 & c \\ c & 0 \end{pmatrix}.$$

I also would like to mention here the interesting univariate and multivariate models obtained from maximum entropy characterization techniques.

The last method in this first group ($i$) is the method based on variables in common which has some similarities with the method of mixtures (or parameters in common). I would like to remark here the relationships of this method with censoring in sampling procedures where, for example, some independent random variables $X_1, X_2, \ldots, X_n$ are observable if, and only if, they occur before an independent random variable $Y$ representing the testing period. Hence the observations $Z_i = \min(X_i, Y)$, $i = 1, 2, \ldots, n$, are dependent due to the common variable $Y$. They are also related with shock or step-stress models where the independent component lifetimes $X_i$ in a system are observable if, and only if, they pass a common stress level due to the shared environment. It is important to note here that some of these models have a singular part due to the fact that several components can fail at the same time. These 'natural' nonabsolutely continuous models as, for example, the Marshall and Olkin bivariate exponential stress (or shock) model, are very important in practice.

There are few techniques included in the second group ($ii$) since we must start from a multivariate model and there are few multivariate models commonly accepted as unique extensions of univariate models. Actually, we can say that the only one might be the multivariate normal distribution. In a practical situation, the main disadvantage of the normal distribution is the symmetry (with respect to the mean). Thus, it is is natural to consider in the multivariate set-up the skew techniques ($d$) used in the univariate case to obtain asymmetric models. I would like to note here that the normal skew distribution can also be obtained as the distribution of the OS from a random vector $(X, Y)$ having a bivariate Normal distribution (see Loperfido, 2002). Also, we must note here how the univariate skew distribution can be generalized by considering the minimum (the maximum or, in general, the OS) from a random vector having a multivariate normal distribution (see Loperfido *et al.* (2007)).

Another option is to consider weighted models due to a biased random sampling procedure where a sample value $\mathbf{x}$ from a random vector $\mathbf{X}$ is observed with a probability proportional to $w(\mathbf{x})$, where $w$ is a positive (weight) function. For example, if we survey tourists randomly at the hotels, the more the length of stay, the higher sampling probability. Hence, the variable 'length-of-stay' is biased and so are other related variables. Therefore some new techniques should be developed to correct this sampling bias (see, *e.g.*, Cristobal, Ojeda and Alcalá, 2004). This method can also be included in the first group when we consider independent random variables and a dependence factor due only to the common sampling procedure, that is, we have a sample from the multivariate density function

$$f^w(x_1, x_2, \ldots, x_n) = cw(x_1, x_2, \ldots, x_n) f_1(x_1) f_2(x_2) \ldots f_n(x_n),$$

where $f_i$ are the (given) density functions of the random variables $X_i$, $i = 1, 2, \ldots, n$, and $c = 1/E(w(X_1, X_2, \ldots, X_n))$.

To finish I would like to mention that as well as obtaining new multivariate models that can represent data sets, it is also very important to develop fit techniques to measure the accuracy of these models to data in several practical situations. In my opinion, both aspects will be relevant research fields in the future in probability and statistics studies.

## Acknowledgements

Jorge Navarro is partially supported by the Ministerio de Ciencia y Tecnología under grant MTM2006-12834. I would like to thank the Editor in Chief Professor Montserrat Guillén for inviting me to provide this comment on the paper by Jose M. Sarabia and Emilio Gómez-Déniz.

## References

Arnold, B. C., Castillo, E. and Sarabia, J. M. (1999). *Conditional Specification of Statistical Models.* Springer Series in Statistics. Springer, Heidelberg, New York.

Balakrishnan, N. (2007). Permanents, order statistics, outliers, and robustness. *Revista Matemática Complutense*, 20, 7-107.

Balakrishnan, N., Ng. H. K. T. and Panchapakesan, S. (2006). A nonparametric procedure based on early failures for selecting the best population using a test for equality. *Journal of Statistical Planning and Inference*, 136, 2087-2111.

Cristóbal, J. A., Ojeda, J. L. and Alcalá, J. T. (2004). Confidence bands in nonparametric regression with length biased data. *Annals of the Institute of Statistical Mathematics*, 56, 475-496.

Kotz, S., Navarro, J. and Ruiz, J. M. (2007). Characterizations of Arnold and Strauss' and related bivariate exponential models. *Journal of Multivariate Analysis*, 98, 1494-1507.

Loperfido, N. (2002). Statistical implications of selectively reported inferential results. *Statistics and Probability Letters*, 56, 13-22.

Loperfido, N., Navarro, J., Ruiz, J. M. and Sandoval, C. J. (2007). Some Relationships between Skew-normal Distributions and Order Statistics from Exchangeable Normal Random Vectors. *Communications in Statistics Theory and Methods*, 36, 1719-1733.

Navarro, J. and Ruiz, J. M. (2004). A characterization of the multivariate normal distribution by using the hazard gradient. *Annals of the Institute of Statistical Mathematics*, 56, 361-367.

Navarro, J., Ruiz, J. M. and Sandoval, C. J. (2006). Reliability properties of systems with exchangeable components and exponential conditional distributions. *TEST*, 15, 471-484.

Navarro, J., Ruiz, J. M. and Sandoval, C. J. (2007). Some characterizations of multivariate distributions using product of the hazard gradient and mean residual life components. *Statistics*, 41, 85-91.

Navarro, J. (2008). Characterizations using the bivariate failure rate function. To appear in *Statistics and Probability Letters*.

Ruiz, J. M., Marín, J. and Zoroa, P. (1993). A characterization of continuous multivariate distributions by conditional expectations. *Journal of Statistical Planning and Inference*, 37, 13-21.

# Rejoinder

We are extremely grateful to the two discussants of the paper for their positive and thoughtful comments and remarks.

Professor Navarro begins with an interesting classification in two groups of the methods for construction of multivariate distributions. Prof. Navarro points out the construction of order statistics obtained from a random vector, where the possible dependence between the components is modelled by their joint distribution. In this sense, Balakrishnan (2007) has obtained several results in the independent and non-identically distributed case and Arellano-Valle and Genton (2008) have obtained the exact distribution of the maximum of absolutely continuous dependent random variables.

Other strategies of construction of multivariate distribution based on conditional specification are commented by Prof. Navarro. In reliability contexts, other modelling approaches are used. For example, a dynamic construction prescribes the joint distribution of $(X, Y)$ by specifying the conditional distribution of $Y$ given $\min\{X, Y\} = X = t$ and the conditional distribution of $X$ given $\min\{X, Y\} = Y = t$.

Professor Pardo focuses her comments on the maximum entropy principle and characterization problems. This principle and its different variants is an alternative method for generating multivariate distributions.

In the context of conditional specification, Gokhale (1999) has shown that if the conditional densities of a bivariate random variable have maximum entropies, subject to certain constraints, then the bivariate density also maximizes entropy, subject to appropriate constraints. An important example of this situation is given by distribution (29). This result provides an interesting insight in the structure of joint maximum entropy distributions when conditional maximum entropy distributions are specified. An application of this kind of distribution in hydrology has been provided by Agrawal, Singh and Kumar (2005).

Minimum cross-entropy methods are also used to recover a joint density function from information about the joint and marginal moments and the marginal density function (see Miller and Liu, 2002).

Finally, both Professors Navarro and Pardo have commented the importance of characterization problems in the construction of multivariate distributions. This topic is presently receiving a lot of attention in the statistics and probabilistic research. In this sense, there is a considerable work dealing with the problem of characterizing distributions by means of conditional moments (see Wesolowski, 1995, and Arnold, Castillo and Sarabia, 1999, Chapter 7). For example, we might be interested in identifying all distributions with linear regression and constant conditional variances.

## References

Agrawal, D., Singh, J. K. and Kumar, A. (2005). Maximum entropy-based conditional probability distribution runoff model. *Biosystems Engineering*, 90, 103-113.

Arellano-Valle, R. B. and Genton, M. G. (2008). On the exact distribution of the maximum of absolutely continuous dependent random variables. *Statistics and Probability Letters*, 78, 27-35.

Gokhale, D. V. (1999). On joint and conditional entropies. *Entropy*, 1, 21.24.

Miller, D. J. and Liu, W-H. (2002). On the recovery of joint distributions from limited information. *Journal of Econometrics*, 107, 259-274.

Wesolowski, J. (1995). Bivariate discrete measures via a power series conditional distribution and a regression function. *Journal of Multivariate Analysis*, 55, 219-229.