Distributions with Fixed Marginals and Related Topics
IMS Lecture Notes – Monograph Series Vol. 28, 1996

DISTRIBUTIONS WITH LOGISTIC MARGINALS AND/OR CONDITIONALS

BY BARRY C. ARNOLD
University of California, Riverside

Bell shaped marginals and conditionals are not uniquely associated with multivariate normality. A trained eye is required to distinguish logistic and normal densities. Consequently it is of interest to study the variety of multivariate distributions with logistic rather than normal marginals and/or conditionals. A brief survey is provided. Selection of the particular model, as always, should be driven by some knowledge of the stochastic mechanism generating the data at hand.

1. Introduction. A random variable X will be said to have a logistic distribution with location parameter $\mu(\epsilon\mathbf{R})$ and scale parameter $\sigma(\epsilon\mathbf{R}^+)$ if its survival function is of the form

$$\bar{F}_X(x) = P(X > x) = [1 + \exp(\frac{x-\mu}{\sigma})]^{-1}, \quad x\epsilon\mathbf{R}. \tag{1.1}$$

If (1.1) holds, we write $X \sim \mathcal{L}(\mu, \sigma)$. The standard logistic corresponds to the choice $\mu = 0, \sigma = 1$ and we typically use Z to denote a standard logistic variable. Evidently $E(Z) = 0$ and, not so evidently, $var(Z) = \pi^2/3$. Our development of multivariate logistic distributions (distributions with marginals and/or conditionals of the form (1.1)) will exploit a variety of special features and representations of the univariate logistic distribution. We begin by reviewing these univariate facts. For more details, see Johnson and Kotz (1970, Chapter 23).

The quantile or inverse distribution function of the standard logistic distribution is of the form

$$F_Z^{-1}(u) = \log(\frac{u}{1-u}), \quad 0 < u < 1 \tag{1.2}$$

AMS 1991 Subject Classification: Primary 62H05, 60E05.

Key words and phrases: Mixtures, elliptically contoured distributions, conditional specification, geometric minima, Pareto conditionals.

(here and henceforth all logarithms are natural logarithms). If U has a uniform $(0,1)$ distribution then using (1.2) it follows that

$$\mu + \sigma \log(\frac{U}{1-U}) \sim \mathcal{L}(\mu, \sigma) . \tag{1.3}$$

The defining differential equation for the standard logistic distribution is

$$f_Z(z) = \frac{d}{dz} F_Z(z) = F_Z(z)\bar{F}_Z(z) . \tag{1.4}$$

Insight into alternative stochastic models of random variables with logistic distributions is provided by study of the moment generating function.

$$M_Z(t) = E(e^{tZ}) = \Gamma(1+t)\Gamma(1-t), \quad |t| < 1 . \tag{1.5}$$

One can recognize the term $\Gamma(1-t)$ as the moment generating function of an extreme value random variable with density

$$f_Y(y) = e^{-y} \exp(e^{-y}), \quad y \in \mathbf{R} \tag{1.6}$$

and so if $Z \sim \mathcal{L}(0,1)$ then

$$Z \stackrel{d}{=} Y_1 - Y_2 \tag{1.7}$$

where Y_1, Y_2 are i.i.d. extreme value variables with common density (1.6). A famous convergent infinite product

$$\Gamma(1-it)\Gamma(1+it) = \prod_{j=1}^{\infty} (1 + \frac{t^2}{j^2})^{-1} \tag{1.8}$$

gives us immediately an alternative representation of a logistic random variable in terms of independent Laplace (or double exponential) variables. Thus if $Z \sim \mathcal{L}(0,1)$

$$Z \stackrel{d}{=} \sum_{j=1}^{\infty} W_j/j \tag{1.9}$$

where the W_j's are i.i.d. standard Laplace variables with common density

$$f_W(w) = \frac{1}{2} e^{-|w|}, \quad w \in \mathbf{R} . \tag{1.10}$$

Since a Laplace variable is representable as a difference of two i.i.d. standard exponential variables we can give a conditionally convergent expression in terms of exponential variables

$$Z \stackrel{d}{=} \sum_{j=1}^{\infty} \frac{E_j}{j} - \sum_{j=1}^{\infty} \frac{E_j'}{j} \tag{1.11}$$

where the E_j's and E_j''s are independent with common density

$$f_E(x) = e^{-x}, \quad x > 0 \ . \tag{1.12}$$

Representations (1.9) and (1.11) are discussed in some detail and developed further in George and Devidas (1992).

A characteristic feature of the logistic distribution is its closure under geometric minimization. Suppose $X_1, X_2, \ldots$ are i.i.d. $\mathcal{L}(\mu, \sigma)$ variables and N, independent of the X_i's, is a geometric(p) random variable (i.e. $P(N = n) = p(1-p)^{n-1}, \quad n = 1, 2, \ldots$). Define

$$Y = \min_{i \leq N} X_i \ . \tag{1.13}$$

By conditioning on N, it is readily verified that Y, the geometric minimum of the X_i's, is again a logistic variable (Arnold and Laguna (1976)). Specifically we have

$$Y \sim \mathcal{L}(\mu + \sigma \log p, \sigma) \ . \tag{1.14}$$

The symmetry of the logistic distribution yields an analogous result for geometric maxima. If we define

$$Z = \max_{i \leq N} X_i \tag{1.15}$$

then

$$Z \sim \mathcal{L}(\mu - \sigma \log p, \sigma) \ . \tag{1.16}$$

From (1.14), following Arnold and Laguna (1976), we can observe that

$$Y - \sigma \log p \stackrel{d}{=} X_1 \ . \tag{1.17}$$

If we begin with i.i.d. X_i's and define Y as in (1.13) then if (1.17) holds for every $p \in (0,1)$ then necessarily X_1 is logistic. If (1.17) holds for only one value of p, then a so-called semi-logistic distribution is possible for X. A random variable X will be said to have a semi-logistic (p, ψ) distribution if

$$\bar{F}_X(x) = [1 + \psi(x)]^{-1}, \quad x \in \mathbf{R} \tag{1.18}$$

where ψ is nondecreasing and right continuous and satisfies

$$\psi(x) = \frac{1}{p} \psi(x + \sigma \ \log \ p) \tag{1.19}$$

for some $\sigma > 0$. The nicest solution to (1.19) is of the form $\psi(x) = e^{a+bx}$ which takes us back to the logistic distribution.

A final observation which will prove useful in Section 3 is that the logistic distribution is representable as a scale mixture of normal random variables (Andrews and Mallows, 1974). In this setting, let Z be standard logistic and Z^* standard normal. Now assume V, independent of Z^*, has a Kolmogorov-Smirnov distribution, i.e.

$$P(V < v) = 1 - 2\sum_{j=1}^{\infty}(-1)^{j+1}\exp(-2j^2v^2) \ . \tag{1.20}$$

It follows that

$$Z \stackrel{d}{=} 2VZ^* \tag{1.21}$$

2. Multivariate Logistic Distributions. When we speak of a multivariate logistic distribution we usually refer to a distribution with logistic marginals. In Section 7 we will broaden the definition to include distributions with logistic conditionals, but in Sections 3–6 the requirement of logistic marginals will be implicitly or explicitly invoked. Curiously it is possible to encounter in the literature, multivariate logistic distributions which have neither marginals nor conditionals of logistic form (for example the elliptically symmetric k-dimensional logistic distribution introduced by Jensen (1985) and discussed in some detail in Fang, Kotz and Ng (1990) is of this kind). Arnold (1992) provides a broad coverage of methods which have been used to generate multivariate distributions with logistic marginals. In this paper we focus on a detailed discussion of just three techniques: Mixtures, differences of extremes and geometric minima. In the final section we summarize results on distributions with logistic conditionals.

3. Mixture Representations. Suppose that we begin with a joint density $f_{Z,W}(z,w)$ with a standard logistic marginal density for Z. We can write

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z|W}(z|w)dF_W(w) \tag{3.1}$$

and view this as a mixture representation of the univariate logistic density. An enormous variety of conditional densities $f_{Z|W}(z|w)$ and of mixture distributions $F_W(w)$ can be used in the form (3.1). Equivalent but sometimes more convenient representations might involve distribution functions, survival functions, or moment generating functions.

$$F_Z(z) = \int_{-\infty}^{\infty} F_{Z|W}(z|w)dF_W(w) \tag{3.2}$$

$$\bar{F}_Z(z) = \int_{-\infty}^{\infty} \bar{F}_{Z|W}(z|w)dF_W(w) \tag{3.3}$$

$$M_Z(t) = \int_{-\infty}^{\infty} M_{Z|W}(t|w)dF_W(w) \; . \tag{3.4}$$

Any of the above mixture representations can be readily employed to generate a multivariate logistic distribution, which is conditionally independent (and hence exchangeable in the strong deFinetti sense). Thus the k-dimensional versions are

$$f_{\underline{Z}}(\underline{z}) = \int_{-\infty}^{\infty} [\prod_{i=1}^{k} f_{Z|W}(z_i|w)]dF_W(w) \tag{3.5}$$

$$F_{\underline{Z}}(\underline{z}) = \int_{-\infty}^{\infty} [\prod_{i=1}^{k} F_{Z|W}(z_i|w)]dF_W(w) \tag{3.6}$$

$$\bar{F}_{\underline{Z}}(\underline{z}) = \int_{-\infty}^{\infty} [\prod_{i=1}^{k} \bar{F}_{Z|W}(z_i|w)]dF_W(w) \tag{3.7}$$

and

$$M_{\underline{Z}}(\underline{t}) = \int_{-\infty}^{\infty} [\prod_{i=1}^{k} M_{Z|W}(t_i|w)]dF_W(w) \tag{3.8}$$

In all these formulations, the random variables $(Z_1, \ldots, Z_k)$ are conditionally independent given the mixing random variable W.

Scale mixtures provide an example of the use of representations (3.5). For them, it is assumed that

$$f_{Z|W}(z|w) = f_0(z/w) \tag{3.9}$$

and that $F_W(w)$ has support on the positive half line. If we accept (3.9) as our model, we can alternatively write

$$Z_i = WV_i \tag{3.10}$$

where the V_i's are i.i.d. and W is independent of the V_i's, chosen of course such that $WV_1 \sim \mathcal{L}(0,1)$. Of course there are enormous numbers of choices for the distributions of the independent variables, the V's and W. There are some restrictions. Arnold (1992) observed that one cannot have such a representation with $W \sim \Gamma(\alpha, 1)$ where $\alpha < 1.5$. It is possible to have a representation with $W \sim$ Uniform $(0,1)$. As is shown in Arnold and Robertson (1989), the appropriate density for V in that case (chosen so that WV is standard logistic) is

$$f_V(v) = \frac{|v|}{8} \tanh\frac{|v|}{2} \operatorname{sech}^2\frac{|v|}{2}, \quad v \in \mathbf{R} \; . \tag{3.11}$$

A more attractive alternative is available using the representation (1.21). In that framework (with notational changes to fit the present discussion) we can choose

$$V_i \sim N(0,1), \quad i = 1, 2, \cdots, k \tag{3.12}$$

and

$$W = 2W' \tag{3.13}$$

where W' has the Kolmogorov-Smirnov distribution. Since in this case $\underline{V} = (V_1, \ldots, V_k)$ is spherically symmetric, so is any scale mixture. It follows that we have at hand the means to construct an elliptically contoured multivariate logistic distribution of the form

$$\underline{X} = \underline{\mu} + 2W' A \underline{V} \tag{3.14}$$

where $\underline{\mu} \in \mathbf{R}^k$, A is a non-singular $k \times k$ matrix, $\underline{V} \sim N(0, I)$ and W', independent of $\underline{V}$, has the Kolmogorov-Smirnov distribution. In general we cannot write down the density for the model (3.14). An exception occurs in the 3 dimensional case. When $k = 3$ and $A = I$ we can use properties of the inverse Gaussian density to eventually obtain the following representation for $f_{\underline{X}}(\underline{x})$.

$$f_{(X_1, X_2, X_3)}(x_1, x_2, x_3) = \frac{\tanh(\frac{\sqrt{x_1^2+x_2^2+x_3^2}}{2}) \operatorname{sech}^2(\frac{\sqrt{x_1^2+x_2^2+x_3^2}}{2})}{8\pi\sqrt{x_1^2 + x_2^2 + x_3^2}} \ . \tag{3.15}$$

The joint density corresponding to the model (3.14) was developed earlier in the form $\underline{\mu} + RA\underline{U}$ where $\underline{U}$ is a k-dimensional random vector uniformly distributed over the unit k-sphere and R, independent of $\underline{U}$ was a positive random variable whose density involved modified Bessel functions of the third kind. (Arnold and Robertson (1989)). The present development is clearly more aesthetically appealing.

Instead of a scale mixture, one could consider a location mixture (an additive rather than a multiplicative model). Here we seek independent random variables V and W such that $V + W \sim \mathcal{L}(0,1)$. Our k- dimensional logistic distribution would then be of the form

$$Z_i = V_i + W \tag{3.16}$$

where the V_i's are i.i.d., independent of W. A plethora of choices for the distributions of V and W exist. Referring to equation (1.7), one obvious possibility is to choose V_i's to be i.i.d. extreme value variables (with density (1.6)) and assume W is such that $-W$ is an extreme value variable (independent of the V_i's). This leads to the classical Gumbel-Malik-Abraham k-variate logistic distribution (Gumbel (1961), Malik and Abraham (1973)). This model will be discussed in more detail and generalized in Section 4. Alternative additive models for the logistic can be derived from the "sums of Laplace " representation (1.9) or the conditionally convergent "linear combinations of exponentials" representation (1.11). So, for example, the V_i's could be exponential with an appropriate choice of distribution for W. Alternatively they could be Laplace variables. Further discussion of such models is deferred to the next section.

We turn now to mixture models generated using (3.2) or (3.3). Perhaps the simplest examples involve powers of distribution functions (or of survival functions). Thus to use (3.2) we might postulate that for every $w > 0$,

$$F_{Z|W}(z|w) = [F_0(z)]^w \tag{3.17}$$

where $F_0(z)$ is a fixed distribution function. Then, provided that the support of W is $(0, \infty)$, we may seek choices for the base distribution function F_0 and for the distribution of the mixing variable W so that

$$(1 + e^{-z})^{-1} = \int_0^\infty [F_0(z)]^w dF_W(w) \ . \tag{3.18}$$

In this expression, the distribution F_W can be chosen quite arbitarily and the corresponding form of F_0 can be expressed in terms of the inverse of the Laplace transform of F_W. Specifically, $F_0(z) = exp[-M_F^{-1}((1 + e^z)^{-1})]$ where M_F is the Laplace transform of F_W as defined following equation (3.21).

Our resulting k-variate distribution with standard logistic marginals would then have joint distribution function

$$F_{\underline{Z}}(\underline{z}) = \int_0^\infty [\prod_{i=1}^k F_0(z_i)]^w dF_W(w) \ . \tag{3.19}$$

An analogous development involving survival functions using (3.3) provides closely related distributions. For these we choose a survival function $\bar{G}_0(z)$ and a mixing distribution $F_W(w)$ such that

$$(1 + e^z)^{-1} = \int_0^\infty [\bar{G}_0(z)]^w dF_W(w) \tag{3.20}$$

and obtain a joint survival function of the form

$$\bar{F}_{\underline{Z}}(\underline{z}) = \int_0^\infty [\prod_{i=1}^k \bar{G}_0(z_i)]^w dF_W(w) \ . \tag{3.21}$$

Such models are typically described as frailty models (see Oakes (1989)). If we denote the Laplace transform of F by

$$M_F(t) = \int_0^\infty e^{-tw} dF(w)$$

we can write (3.21) in the form

$$\bar{F}_{\underline{Z}}(\underline{z}) = M_F(\sum_{i=1}^{k} M_F^{-1}((1+e^{z_i})^{-1})) \tag{3.22}$$

where $F = F_W$. This expression yields a valid joint survival function even if M_F is not a Laplace transform. It must however satisfy $M_F(0) = 1, M_F'(t) \leq 0$ and $M_F''(t) \geq 0$, for every $t > 0$. In this generality it is an Archimedean distribution (Genest and MacKay (1986)). The distribution (3.19) admits the parallel representation

$$F_{\underline{Z}}(\underline{z}) = M_F(\sum_{i=1}^{k} M_F^{-1}((1+e^{-z_i})^{-1})) \ . \tag{3.23}$$

The choice $M_F(t) = (1+t)^{-\alpha}$ (corresponding to a gamma distribution for W), in (3.23) yields a multivariate logistic distribution of the form

$$F_{\underline{Z}}(\underline{z}) = (\sum_{i=1}^{k}(1+e^{-z_i})^{1/\alpha} - k + 1)^{-\alpha} \tag{3.24}$$

(the choice $\alpha = 1$ in (3.24) corresponds to the Gumbel-Malik-Abraham distribution). Other examples may be found in Arnold (1992) and some extensions are discussed in the section on frailty models in this paper.

Note that, in the above development, it is not required that the distribution of W be continuous. If W has a geometric distribution, then (3.2) and (3.3) can be viewed as dealing with geometric maxima and minima respectively. Referring to (1.17), we see that an admissible choice for the distribution function F_0 is translated logistic. Further discussion of such geometric extreme models is deferred to Section 5.

The final class of exchangeable models obtained via mixtures involves the use of moment generating functions. In (3.8) we assume that $F_W(w)$ has support $(0, \infty)$ and our joint moment generating function is of the form

$$M_{\underline{Z}}(\underline{t}) = \int_0^\infty [\prod_{i=1}^{k} M_0(t_i)]^w dF_W(w) \tag{3.25}$$

where $M_0(t)$ is an arbitrary infinitely divisible moment generating function (so that $[M_0(t)]^w$ is an m.g.f. for every w) and $F_W(w)$ is chosen so that

$$\Gamma(1+t)\Gamma(1-t) = \int_0^\infty [M_0(t)]^w dF_W(w) \; . \tag{3.26}$$

Arguing in a manner analogous to that used in the frailty discussion, we can write the resulting joint moment generating function in terms of the moment generating function $M_F(t)$ corresponding to F_W in (3.26). Thus we have

$$M_{\underline{Z}}(\underline{t}) = M_F(\sum_{i=1}^k M_F^{-1}(\Gamma(1+t_i)\Gamma(1-t_i))) \; . \tag{3.27}$$

For example if we choose $M_F(t) = (1+t)^{-\alpha}$ for $\alpha > 0$, we obtain the following joint moment generating function (with standard logistic marginals).

$$M_{\underline{Z}}(\underline{t}) = [\sum_{i=1}^k [\Gamma(1+t)\Gamma(1-t)]^{1/\alpha} - k + 1]^\alpha \; . \tag{3.28}$$

The use of such models is complicated by the difficulty of "inverting" joint moment generating functions. Having developed a spectrum of exchangeable multivariate logistic models using the mixture paradigms (3.5)–(3.8), it is immediately evident that more general families can be developed by mixing dependent rather than independent variables. Thus if we have determined a mixing distribution $F_W(w)$ and a family of conditional densities $f_{Z|W}(z|w)$ such that (3.5) yields a multivariate logistic distribution, then, instead of using $\prod_{j=1}^k f_{Z|W}(z_i|w)$, we could substitute any k dimensional joint density with marginals given by $f_{Z|W}(z_i|w)$, $i = 1, 2, \ldots, k$. Analogous modifications of (3.6) - (3.8) are of course possible.

Some of the distributions to be described in subsequent sections can be viewed as mixtures of dependent distributions and so provide examples of this technique. Note that in the extension of (3.19) to dependent mixtures by replacing $\prod_{i=1}^k F_0(z_i)$ by $F_0(\underline{z})$ we must be sure that $F_0(\underline{z})$ is max-infinitely divisible to guarantee that $[F_0(\underline{z})]^w$ is a valid k-dimensional distribution function for every w. An analogous requirement of min-infinite divisibility is needed to extend (3.21).

4. Differences of Extremes. As observed in Section 1, if Y_1 and Y_2 are i.i.d. extreme value random variables (with density (1.6)) then $Y_1 - Y_2$ has a standard logistic distributrion. This immediatedly permits us to construct k-dimensional logistic distributions of the form

$$\underline{Z} = \underline{U} - \underline{V} \tag{4.1}$$

where $\underline{U}$ and $\underline{V}$ are independent random vectors each with extreme value marginals. The case where $V_1 = \cdots = V_k$ and where $U_1, \ldots, U_k$ are independent corresponds to the classical Gumbel-Malik-Abraham distribution with joint moment generating function of the form

$$M_{\underline{X}}(\underline{t}) = \Gamma(1 + \sum_{i=1}^{k} t_i) \prod_{i=1}^{k} \Gamma(1 - t_i) . \tag{4.2}$$

If instead we allow $\underline{U}$ to have Gumbel's multivariate extreme distribution, i.e.

$$F_{\underline{U}}(\underline{u}) = \exp[-(\sum_{i=1}^{k} e^{-\alpha u_i})^{1/\alpha}] , \tag{4.3}$$

(where $\alpha > 1$) and assume $V_1 = \cdots = V_k$ (with density (1.6)) then, with $\underline{Z}$ defined by (4.1) we have

$$F_{\underline{Z}}(\underline{z}) = [1 + (\sum_{i=1}^{k} e^{-\alpha z_i})^{1/\alpha}]^{-1} , \tag{4.4}$$

a distribution introduced by Strauss (1979). A serious problem with such a distribution is that $\rho(Z_i, Z_i') \geq \frac{1}{2}$ for every $i \neq i'$ and such strong correlation may not be desirable. An easy alternative with a full range of non-negative correlations is provided by choosing $\underline{U}$ and $\underline{V}$ to be independent, each with Gumbel multivariate extreme distributions (given by (4.3)), perhaps with different values of α for $\underline{U}$ and $\underline{V}$ respectively. A technique to construct a k-variate logistic distribution with completely arbitrary covariance structure was outlined in Arnold (1992). The trick is to use the conditionally convergent representation (1.11) for logistic variables or, alternatively, the absolutely convergent representation

$$Z \stackrel{d}{=} \sum_{j=1}^{J} \frac{E_j}{j} - \sum_{j=1}^{J} \frac{E_j'}{j} + \sum_{j=J+1}^{\infty} \frac{W_j}{j} \tag{4.5}$$

for any integer J, where the E_j's, E_j''s and W_j's are independent, the E_j's and E_j''s having a common exponential density (1.12) and the W_j's having a Laplace density (1.10). We use representations (4.5) for $Z_1, Z_2, \ldots, Z_k$ and achieve the desired correlation by allowing some of the exponential variables to appear in the expressions for more than one Z_i. For example if

$$Z_{i_k} = \sum_{j=1}^{J} \frac{E_j^{(i_k)}}{j} - \sum_{j=1}^{J} \frac{E_j'^{(i_k)}}{j} + \sum_{j=J+1}^{\infty} \frac{W_j^{(i_k)}}{j}$$

$k = 1, 2$ and if for $j = 1, 2, \ldots, J$

$$E_j^{(i_1)} = E_j'^{(i_2)}$$

and

$$E_j^{(i_2)} = E_j'^{(i_1)}$$

then

$$\rho(Z_{i_1}, Z_{i_2}) = -(\sum_{j=1}^{J} \frac{1}{j^2})/(\sum_{j=1}^{\infty} \frac{1}{j^2})$$

which will be arbitrarily close to -1 for J sufficiently large.

5. Geometric Extremes. It was remarked in Section 1 that if $X_1, X_2, \ldots$ were i.i.d. logistic (μ, σ) variables and if $N \sim$ geometric(p) is independent of X_i's then

$$Y = \min_{i \leq N} X_i \sim \mathcal{L}(\mu + \sigma \log p, \sigma) . \tag{5.1}$$

Thus in order to get Y to have a standard logistic distribution we may begin with X_i's with a $\mathcal{L}(-\log p, 1)$ distribution. To generate multivariate logistic distributions with standard logistic marginals we may define for $i = 1, 2, \ldots, k$,

$$Z_i = \min_{j \leq N_i} X_{ij} \tag{5.2}$$

where $\underline{N}$ is a random vector with geometric(p_i) marginals $(i = 1, 2, \ldots, k)$ and $(X_{1j}, X_{2j}, \ldots, X_{kj})$ are i.i.d. k dimensional random variables with logistic $(-\log p_i, 1)$ marginals, $i = 1, 2, \ldots, k$. Any multivariate geometric distribution and any multivariate logistic distribution can be used. A hierarchy of multivariate logistic distributions can be constructed paralleling the hierarchy of multivariate exponential distributions described in Arnold (1975). A complete listing is impossible. Certain special cases merit attention.

If $N_1 = N_2 = \cdots = N_k (= N \text{say})$ where $N \sim$ geometric (p), and if $\bar{F}_{\underline{X}}(\underline{x})$ is a k-dimensional survival function with logistic $(-\log p, 1)$ marginals then the joint survival function of the random vector $\underline{Z}$ defined by (5.2) is

$$\bar{F}_{\underline{Z}}(\underline{z}) = p\bar{F}_{\underline{X}}(\underline{z})/[1 - (1-p)\bar{F}_{\underline{X}}(\underline{z})] . \tag{5.3}$$

From experience in the univariate case we are aware that it is possible that the distributions $F_{\underline{X}}$ and $F_{\underline{Z}}$ in (5.3) could be of the same type for every p. When $\underline{Z} \stackrel{d}{=} \underline{X} - \underline{c}(p)$ in (5.3) we say that the distribution is min-geometric stable (following, for example, Rachev and Resnick (1991)). As outlined in Arnold (1992), using results of Rachev and Resnick, the general form of min-geometric stable distributions with standard logistic marginals is

$$\bar{F}_{\underline{Z}}(\underline{z}) = \left[1 + \int_0^1 \max_{1 \le i \le k} [f_i(s)e^{z_i}]ds\right]^{-1}, \quad \underline{z} \in \mathbf{R}^k \tag{5.4}$$

where $f_1(s), \ldots, f_k(s)$ are non-negative functions on $[0,1]$ satisfying $\int_0^1 f_i(s)ds = 1$, $i = 1, 2, \ldots, k$. Clearly a broad spectrum of such distributions exists, since the $f_i(s)'s$ are quite arbitrary in (5.4). The semi-logistic version of (5.4) results if we replace each e^{z_i} by $g_i(z_i)$ where the $g_i(z_i)$'s are non-decreasing right continuous functions satisfying $g_i(z_i) = \frac{1}{p^*} g_i(z_i + logp^*)$. The type of such distributions is preserved under geometric(p^*) minimization for one fixed choice p^*.

Examples of the min-geometric stable paradigm (5.4) are the "Strauss" model

$$\bar{F}_{\underline{Z}}(\underline{z}) = [1 + (\sum_{i=1}^k e^{\alpha z_i})^{1/\alpha}]^{-1} \tag{5.5}$$

(cf. (4.4)) and

$$\bar{F}_{\underline{Z}}(\underline{z}) = \left[1 + \sum_{i=1}^k e^{z_i} - (\sum_{i=1}^k e^{-z_i})^{-1}\right]^{-1}, \quad \underline{z} \in \mathbf{R}^k . \tag{5.6}$$

A random variable $\underline{Z}$ will be said to be max-geometric stable if $-\underline{Z}$ is min-geometric stable. It has the property that geometric maxima will be of the same type. Trivially $\underline{Z}$ is max-geometric stable with standard logistic marginals if and only if $-\underline{Z}$ has its survival function of the form (5.4). An interesting open question is that of estimating the structure functions $f_1(s), \ldots, f_k(s)$ in (5.4) based on a sample from the distribution $\underline{Z}$.

Of course the model (5.3) can be used to generate new multivariate logistic distributions beginning with a non-min-stable distribution for $\underline{X}$. For example if $\underline{X}$ has independent standard logistic marginals, then the version of (5.3) with standard logistic marginals is

$$\begin{aligned} \bar{F}_{\underline{Z}}(\underline{z}) = \Big[1 &+ \sum_{i=1}^k e^{z_i} \\ &+ p \sum_{i \ne j} e^{z_i + z_j} + p^2 \sum_{i \ne j \ne \ell} e^{z_i + z_j + z_\ell} \\ &+ \ldots + p^{k-1} e^{\sum_{i=1}^k z_i}\Big]^{-1}, \underline{z} \in \mathbf{R}^k . \end{aligned} \tag{5.7}$$

Further examples can be generated if we allow $\underline{N}$ to be a more general multivariate geometric distribution. For example, in two dimensions for notational economy, if we assume that (N_1, N_2) has a bivariate distribution of the form

$$\begin{aligned} P(N_1 \geq n_1, N_2 \geq n_2) &= p_{00}^{n_1}(p_{10} + p_{00})^{n_2 - n_1}, \ n_1 \leq n_2 \\ &= p_{00}^{n_2}(p_{01} + p_{00})^{n_1 - n_2}, \ n_1 > n_2 \end{aligned} \tag{5.8}$$

and if we assume $\underline{X}$ has independent logistic marginals, then (5.2) yields a non-min-stable bivariate distribution with standard logistic marginals of the form

$$\bar{F}_{Z_1,Z_2}(z_1, z_2) = \frac{p_{11} + \frac{p_{01}}{1+e^{z_1}} + \frac{p_{10}}{1+e^{z_2}}}{[1 + (p_{10} + p_{11})e^{z_1}][1 + (p_{01} + p_{11})e^{z_2}] - p_{00}} \ . \tag{5.9}$$

Following Arnold (1990) if we consider a sequence of independent trials with $k + 1$ possible outcomes $0, 1, 2, \ldots, k$ and associated probabilities $p_0, p_1, \ldots, p_k (\sum_{i=0}^{k} p_i = 1)$, we can define a multivariate geometric random vector $\underline{N} = (N_1, \ldots, N_k)$ such that N_i represents the number of outcomes of type i which precede the first outcome of type 0. If we use this distribution in (5.2) and assume $\underline{X}$ has independent logistic marginals, we are eventually led to a flexible k-dimensional distribution with standard logistic marginals of the form

$$\begin{aligned} P(\underline{Z} \geq \underline{z}) = \Bigg[1 &+ \sum_{j=1}^{k} e^{z_j} + \sum\sum_{j_1 \neq j_2} c_{j_1 j_2} e^{z_{j_1} + z_{j_2}} \\ &+ \sum\sum\sum_{j_1 \neq j_2 \neq j_3} c_{j_1 j_2 j_3} e^{z_{j_1} + z_{j_2} + z_{j_3}} \\ &+ \cdots + c_{1 \cdots k} e^{z_1 + z_2 \cdots + z_k} \Bigg]^{-1} . \end{aligned} \tag{5.10}$$

As Arnold (1990) points out, although in the above development the c's in (5.10) are functions of $p_0, p_1, \ldots, p_k$, the expression (5.10) continues to represent a genuine survival function with standard logistic marginals for a much broader spectrum of choices of c's.

6. Frailty. The general class of frailty type multivariate logistic distributions may be defined as follows. Choose a univariate survival function $\bar{G}_0(z)$ and a mixing distribution $F_W(w)$ so that (3.20) holds, i.e.

$$(1 + e^z)^{-1} = \int_0^1 [\bar{G}_0(z)]^w dF_W(w) \ . \tag{6.1}$$

Now take any k-dimensional survival function $\bar{G}(\underline{z})$ with marginal survival functions of the form $\bar{G}_0(z_i)$ (independent marginals is a possibility as discussed in Section 3). Then define a k-variate logistic distribution by

$$\bar{F}_{\underline{Z}}(\underline{z}) = \int_0^\infty [\bar{G}(\underline{z})]^w dF_W(w) \ . \tag{6.2}$$

If we denote the Laplace transform of F_W by M_F, then (6.2) can be written as

$$\bar{F}_{\underline{Z}}(\underline{z}) = M_F(-log\bar{G}(\underline{z})) \tag{6.3}$$

where the marginal survival functions of $\bar{G}(\underline{z})$ must be of the form

$$\bar{G}_0(z_i) = \exp[-M_F^{-1}((1+e^{z_i})^{-1})] \tag{6.4}$$

to guarantee standard logistic marginals in (6.3). If for example $M_F(t) = (1+t)^{-1}$, then

$$\bar{G}_0(z_i) = \exp(-e^{z_i}) \ . \tag{6.5}$$

Any joint survival function $\bar{G}(\underline{z})$ with marginals of the extreme type (6.5) can be used to generate a k-variate logistic distribution of the frailty type (6.3). For example if we take a Gumbel multivariate extreme survival function of the form

$$\bar{G}_0(\underline{z}) = \exp\left(-\left(\sum_{i=1}^k e^{\alpha z_i}\right)^{1/\alpha}\right) \tag{6.6}$$

for some $\alpha \geq 1$, then (6.3) yields

$$\bar{F}_{\underline{Z}}(\underline{z}) = \left[1 + \left[\sum_{i=1}^k e^{\alpha z_i}\right]^{1/\alpha}\right]^{-1} \ , \quad \underline{z} \in \mathbf{R}^k \ , \tag{6.7}$$

which is the survival function of minus 1 times a Strauss multivariate logistic variable (cf. (4.4)). By choosing $\bar{G}_0(z)$ to be an arbitrary min-stable distribution with marginals given by (6.5) we can recognize that all the min-geometric stable distributions with standard logistic marginals (displayed in (5.4)) may be viewed as frailty models. Marshall and Olkin (1990) describe several multivariate logistic distributions using what they call product survival functions. This technique is closely related to the frailty approach, using (6.4).

7. Distributions with Logistic Conditionals. The preoccupation with development of models with specified marginals overlooks the inherent difficulty of envisioning marginal densities of dependent distributions. With this in mind, Arnold, Castillo and Sarabia (1992) have advocated the use of distributions with specified conditional distributions. They provided some material on distributions with logistic conditionals. That material will be reviewed and extended somewhat. The general problem of determining all k-variate distributions for which all conditionals are logistic remains tantalizingly open.

We consider a k-dimensional random vector $\underline{X} = (X_1, \ldots, X_k)$. For each i, let $\underline{X}^{(i)}$ denote the vector $\underline{X}$ with the i'th coordinate deleted. Analogously for a vector $\underline{x}$ in $\mathbf{R}^k, \underline{x}^{(i)}$ is $\underline{x}$ with x_i deleted. We seek to identify all joint densities for $\underline{X}$ such that, for each i and each $\underline{x}^{(i)} \epsilon \mathbf{R}^{k-1}$, the conditional density of X_i given $\underline{X}^{(i)} = \underline{x}^{(i)}$ is logistic with location parameter $\mu_i(\underline{x}^{(i)})$ and scale parameter $\sigma_i(\underline{x}^{(i)})$. If a distribution has such properties, then there will exist corresponding marginal densities

$$g_i(\underline{x}^{(i)}) = f_{\underline{X}^{(i)}}(\underline{x}^{(i)}) \tag{7.1}$$

and, writing the joint density as a product of conditional and marginal densities in k available ways, we have

$$\begin{aligned} &\frac{g_1(\underline{x}^{(1)})}{\sigma_1(\underline{x}^{(1)})} \frac{\exp[(x_1 - \mu_1(\underline{x}^{(1)}))/\sigma_1(\underline{x}^{(1)})]}{[1 + \exp[(x_1 - \mu_1(\underline{x}^{(1)}))/\sigma_1(\underline{x}^{(1)})]]^2} \\ &= \frac{g_2(\underline{x}^{(2)})}{\sigma_2(\underline{x}^{(2)})} \frac{\exp[(x_2 - \mu_2(\underline{x}^{(2)}))/\sigma_2(\underline{x}^{(2)})]}{[1 + \exp[(x_2 - \mu_2(\underline{x}^{(2)}))/\sigma_2(\underline{x}^{(2)})]^2} \\ &= \cdots \\ &= \frac{g_k(\underline{x}^{(k)})}{\sigma_k(\underline{x}^{(k)})} \frac{\exp[(x_k - \mu_k(\underline{x}^{(k)}))/\sigma_k(\underline{x}^{(k)})]}{[1 + \exp[(x_k - \mu_k(\underline{x}^{(k)}))/\sigma_k(\underline{x}^{(k)})]^2} \ . \end{aligned} \tag{7.2}$$

If we define

$$\phi_i(\underline{x}^{(i)}) = 2\sqrt{\sigma_i(\underline{x}^{(i)})/g_i(\underline{x}^{(i)})} \quad i = 1, 2, \ldots, k \tag{7.3}$$

we can rewrite (7.2) in the form

$$\begin{aligned} &\phi_1(\underline{x}^{(1)}) \cosh \left(\frac{x_1 - \mu_1(\underline{x}^{(1)})}{\sigma_1(\underline{x}^{(1)})} \right) \\ &= \phi_2(\underline{x}^{(2)}) \cosh \left(\frac{x_2 - \mu_2(\underline{x}^{(2)})}{\sigma_2(\underline{x}^{(2)})} \right) \\ &= \cdots \\ &= \phi_k(\underline{x}^{(k)}) \cosh \left(\frac{x_k - \mu_k(\underline{x}^{(k)})}{\sigma_k(\underline{x}^{(k)})} \right) \ . \end{aligned} \tag{7.4}$$

This harmless looking system of functional equations awaits solution. Arnold, Castillo and Sarabia (1992) pointed out that if we are willing to assume that $\sigma_i(\underline{x}^{(i)}) = \sigma_i$, $i = 1, 2, \ldots, k$ (constant conditional scedastic functions), then a solution can be obtained. If $\sigma_i(\underline{x}^{(i)}) = \sigma_i$, then defining $Y_i = e^{X_i/\sigma_i}$, $i = 1, 2, \ldots, k$ we find that $\underline{Y}$ has Pareto conditionals. It then follows that the joint density of $\underline{Y}$ must be of the form

$$f_{\underline{Y}}(\underline{y}) = \left[\sum_{\underline{s} \epsilon \xi_k} \delta_{\underline{s}} \prod_{i=1}^{k} y_i^{s_i} \right]^{-2}, \quad \underline{y} > \underline{0} \tag{7.5}$$

where ξ_k is the set of all vectors of 0's and 1's of dimension k. All the $\delta_{\underline{s}}$'s are non-negative in (7.5). Some but not all can be zero. Transforming back we get the following family of logistic conditionals densities

$$f_{\underline{X}}(\underline{x}) = (\prod_{i=1}^{k} \sigma_i)^{-1} e^{\sum_{i=1}^{k} x_i/\sigma_i} \left[\sum_{\underline{s} \epsilon \xi_k} \delta_{\underline{s}} e^{\sum_{i=1}^{k} s_i x_i/\sigma_i} \right]^{-2}, \quad \underline{x} \in \mathbf{R}^k . \tag{7.6}$$

We turn next to discuss the possible existence of centered logistic conditionals distributions, i.e. distributions for which $X_i | \underline{X}^{(i)} = \underline{x}^{(i)} \sim \mathcal{L}(0, \sigma_i(\underline{x}^{(i)}))$, i.e. those in which $\mu_i(\underline{x}^{(i)}) \equiv 0$, for all i's. A trivial example is provided by the case in which the X_i's are independent so that $\sigma_i(\underline{x}^{(i)}) = \sigma_i$, $i = 1, 2, \ldots, k$. It is conjectured that no other examples exist. To buttress this claim consider the bivariate case. Equation (7.4) reduces to

$$\phi_1(x_2) \cosh(\frac{x_1}{2\sigma_1(x_2)}) = \phi_2(x_1) \cosh(\frac{x_2}{2\sigma_2(x_1)}) . \tag{7.7}$$

Set $x_1 = 0$ in (7.7) to obtain

$$\phi_1(x_2) = \phi_2(0) \cosh(\frac{x_2}{2\sigma_2(0)}) .$$

Analogously, setting $x_2 = 0$,

$$\phi_2(x_1) = \phi_1(0) \cosh(\frac{x_1}{2\sigma_1(0)}) .$$

Evidently also $\phi_1(0) = \phi_2(0)$. Substituting back in (7.7) yields

$$\cosh(\frac{x_2}{2\sigma_2(0)}) \cosh(\frac{x_1}{2\sigma_1(x_2)}) = \cosh(\frac{x_1}{2\sigma_1(0)}) \cosh(\frac{x_2}{2\sigma_2(x_1)}) . \tag{7.8}$$

If we change variables letting

$$y_i = x_i/2\sigma_i(0), \quad i = 1, 2$$

and define

$$\alpha_1(y_1) = \sigma_2(0)/\sigma_2(y_1 2\sigma_1(0))$$
$$\alpha_2(y_2) = \sigma_1(0)/\sigma_1(y_2 2\sigma_2(0))$$

then (7.8) can be written in the form

$$\cosh(y_1)\cosh(\alpha_1(y_1)y_2) = \cosh(y_2)\cosh(\alpha_2(y_2)y_1)\ . \qquad (7.9)$$

Taking logarithms in (7.9) and differentiating we conclude eventually that only solutions of the form $\alpha_1(y_1) \equiv \alpha_1$ and $\alpha_2(y_2) \equiv \alpha_2$ are possible. Then $\sigma_1(x_1) \equiv \sigma_1$ and $\sigma_2(x_2) \equiv \sigma_2$. Consequently the only centered bivariate logistic conditionals distributions are those with independent marginals. If this result can be extended to k dimensions, it suggests the possibility that (7.6) may actually exhaust the possibilities for logistic conditionals distributions. Finally, it may be observed that in the class of distributions (7.6), logistic marginals are only encountered in the case of independence. It is consequently conjectured that an assumption of logistic marginals and logistic conditionals can only be satisfied in the trivial case in which the marginals are independent.

Acknowledgement. I am grateful to a referee for pointing out an error in the original version of this paper.

References

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions, *J. Roy. Statist. Soc. B* **36**, 99–102.

Arnold, B. C. (1975). Multivariate exponential distributions based on hierarchical successive damage, *J. Appl. Probab.* **12**, 142–147.

Arnold, B. C. (1990). A flexible family of multivariate Pareto distributions, *J. Statist. Plann. Inference* **24**, 249–258.

Arnold, B. C. (1992). Multivariate logistic distributions, In *Handbook of the Logistic Distributions*, N. Balakrishnan, ed., Decker, New York, 237–261.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (1992). Conditionally Specified Distributions, *Lecture Notes in Statistics*, Vol. 73, Springer-Verlag, Berlin.

Arnold, B. C. and Laguna, L. (1976). *A stochastic mechanism leading to asymptotically Paretian distributions*, Business and Economic Statistics Section, Proceedings of the American Statistical Association, 208–210.

Arnold, B. C. and Robertson, C. A. (1989). *Elliptically contoured distributions with logistic marginals*, Technical Report #180, Department of Statistics, University of California, Riverside, CA.

FANG, K. T., KOTZ, S. and NG, K. W. (1990). Symmetric multivariate and related distributions, *Monographs in Statistics and Applied Probability*, Vol. 36, Chapman and Hall, London.

GENEST, C. and MACKAY, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals, *Amer. Statist.* **40**, 280–283.

GEORGE, E. O. and DEVIDAS, M. (1992). *Some related distributions*, In Handbook of the Logistic Distribution, N. Balakrishnan, ed., Decker, New York, 223–235.

GUMBEL, E. J. (1961). Bivariate logistic distributions, *J. Amer. Statist. Assoc.* **56**, 335–349.

JENSEN, D. R. (1985). Multivariate distributions, *Encyclopedia of Statistical Sciences* **6**, (eds. S. Kotz, N. L. Johnson and C. B. Read), Wiley, New York, 43—55.

JOHNSON, N. L. and KOTZ, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2, John Wiley and Sons, New York.

MALIK, H. J. and ABRAHAM, B. (1973). Multivariate logistic distributions, *Ann. Statist.* **1**, 588–590.

MARSHALL, A. W. and OLKIN, I. (1990). Multivariate distributions generated from mixtures of convolution and product families . *Topics in Statistical Dependence*, H. W. Block et al. eds. Institute of Mathematical Statistics Lecture Notes – monograph series, Volume 16, 371–393.

OAKES, D. (1989). Bivariate survival models induced by frailties, *J. Amer. Statist. Assoc.* **84**, 487–493.

RACHEV, S. T. and RESNICK, S. (1991). Max-geometric infinite divisibility and stability, *Comm. Statist. - Stochastic Models* **7**, 191–218.

STRAUSS, D. J. (1979). Some results on random utility, *J. Math. Psych.* **20**, 35–52.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
RIVERSIDE, CA 92521-0138
barnold@ucrstat2.urc.edu