Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

MULTIVARIATE NON-NORMAL DISTRIBUTIONS AND MODELS OF DEPENDENCY

BY INGRAM OLKIN

Stanford University

The univariate and multivariate normal distributions play a central role in statistical modeling. However, there are many natural phenomena that do not behave according to the normal law. In particular, there is a need to model multivariate binomial, Poisson, exponential, gamma, and beta distributions, for example. There are many ways to create bivariate (or multivariate) distributions with given marginals, so that it is important to understand the underlying rationale for these extensions. The present review outlines a number of methods that have been used to create "natural" multivariate non-normal distributions.

1. Introduction. Although independence plays a central role in random sampling, there are may phenomena in which multiple observations are dependent. Whereas independence is a singular concept, dependence provides a rich outlet for alternative ways in which dependency can be generated. Among the many forms that lead to dependence are sampling plans from urn models, of which sampling without replacement is the simplest and most elementary. Exchangeability, Markov chains, autocorrelated time series, contingency tables are other examples of dependency.

In this review we discuss several avenues that have been used to create multivariate dependent models. This review is not designed to be exhaustive, and depends to a great degree on subjective choices. We also do not provide an exhaustive set of references; rather we give a few references that will permit the reader to trace other results.

For simplicity of exposition we limit ourselves to the bivariate case; the extension to higher dimensions is often clear from the bivariate case.

2. Characterizations.

There is a large literature on characterizations of distributions. The books

---

AMS 1980 Subject Classifications: Primary 62E05, 62H10, 62N05.

Key words and phrases: Characterization of distribution, Dirichlet-Liouville distribution, ecological correlations, families with given marginals, Fréchet bounds, multivariate beta distribution, multivariate exponential distributions, mixtures of distributions, sampling plans, simulations, urn models.

by Kagan, Linnik and Rao (1973), Calambos and Kotz (1978), and Kakosyan, Klebanov and Melamed (1974) provide good initial surveys. The underlying theme is to extend a univariate characterization to a bivariate version. In some instances there is no single such extension, so that several bivariate versions may result. The extension may yield a known distribution, but may also yield a new distribution. We single out several examples to illustrate the central ideas. These relate to characterizations of the normal and exponential distributions.

2.1 The normal distribution

2.1a Characterization by independence

One of the oldest characterization univariate results is that the independence of the mean $\overline{x} = \sum x_i/n$ and the variance $s^2 = \Sigma(x_i - \overline{x})^2/n$ of a random sample $x_1, \cdots, x_n$ characterizes the normal distribution.

To generate a multivariate version let $\underset{\sim}{x}_1 \cdot \cdots, \underset{\sim}{x}_n$, where $\underset{\sim}{x}_j = (x_{j1}, \cdots, x_{jp})$, be a random sample from a p-variate distribution. The independence of the mean vector $\underset{\sim}{\overline{x}} = (\overline{x}_1, \cdots, \overline{x}_p)$, $\overline{x}_j = \Sigma_\alpha x_{j\alpha}/n$, $j = 1, \cdots, p$, and the covariance matrix $S = (s_{ij})$, $s_{ij} = \Sigma_\alpha(x_{i\alpha} - \overline{x}_i)(x_{j\alpha} - \overline{x}_j)/n$, characterizes the multivariate normal distribution.

Other characterizations involving independence relate to the independence of linear forms. Typical of these is the following. Let $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_n$ be mutually independent p-dimensional random row vectors, and let $A_1, \cdots, A_n$, $B_1, \cdots, B_n$ be non-singular $p \times p$ matrices. If $\Sigma A_i X_i'$ is independent of $\Sigma B_i X_i'$, then the X_i are normally distributed (see Ghurye and Olkin, 1962).

For a history and details of related characterizations, see Kagan, Linnik and Rao (1973).

2.1b Characterization by maximum likelihood

If $\{F(x - \mu)\}$ is a translation family of distributions such that for all samples of size 2 and 3, a maximum likelihood estimator (MLE) of μ is the mean vector, then F is a multivariate normal distribution with mean zero. (A lower semi-continuity condition on the density f is also required.)

This result in the univariate case is due to Teicher (1961); the multivariate case is contained in Marshall and Olkin (1993). The proof is based on showing that the inequality

$$\prod_1^n f(x_i - \overline{x}) \geq \prod_1^n f(x_i - \mu), \qquad \forall \mu,\ n = 2, 3,$$

implies that f is the multivariate normal density.

2.1c A "characterization" based on the identical distribution of the maximum

Anderson and Ghurye (1977) answer the following questions. Suppose $\max\{X_1, \cdots, X_n\}$ and $\max\{Y_1, \cdots, Y_n\}$ have identical density functions, under what conditions will the set of densities of the X's and Y's be the same.

They also treat a bivariate case that is illustrative of their paper. Suppose $F_1, \cdots, F_n$ and $G_1, \cdots, G_m$ are nonsingular bivariate normal c.d.f.'s, each with zero means and correlations $\rho_1, \cdots, \rho_n$ and $\tau_1, \cdots, \tau_m$, respectively. If $\tau_i > 0$, $i = 1, \cdots, m$, and $\prod_1^n F_i(x, y) = \prod_1^m G_i(x, y)$, then $m = n$ and $\{\rho_1, \cdots, \rho_n\} = \{\tau_1, \cdots, \tau_n\}$. (Anderson and Ghurye (1993) treat a slightly more general version than the above.)

2.2 The exponential distribution.

The exponential distribution has a rich range of characterizations. We discuss only two such that exhibit some directions that have been studied.

2.2a Loss of memory

Perhaps the oldest result concerning the exponential distribution states that, given that a system has survived to time s, the conditional probability of surviving an additional time t is equal to the unconditional probability of surviving to time t. That is,

$$P\{X > t + s \mid X > s\} = P\{X > t\}, \qquad \forall s,\, t.$$

This yields the functional equation

$$\overline{F}(s + t) = \overline{F}(s)\overline{F}(t),$$

where $\overline{F}(s) = P\{X > s\}$. This functional equation is the well-known Cauchy equation, for which, under a variety of regularity conditions, yields the solution $\overline{F}(s) = \exp\{-s\}$.

A number of multivariate extensions of the univariate case have been proposed. If we extend the univariate version in a straightforward manner, we obtain

$$P\big\{X > t_1 + s_1,\ Y > t_2 + s_2 \mid X > s_1,\ Y > s_2\big\} = P\big\{X > t_1,\ Y > t_2\big\},$$

or equivalently,

$$\overline{F}(s_1 + t_1,\ s_2 + t_2) = \overline{F}(t_1,\ t_2)\overline{F}(s_1,\ s_2) \qquad \forall\, s_i,\ t_i.$$

This is a stringent set of conditions, and the only solution is the case of independence of X and Y. However, a weakening on the conditioning to

$$P\big\{X > tt_1 + s,\ Y > t_2 + s \mid X > s,\ Y > s\big\} = P\big\{X > t_1,\ Y > t_2\big\},$$

or equivalently,

$$\overline{F}(s + t_1,\ s + t_2) = \overline{F}(s,\ s)\overline{F}(t_1,\ t_2)$$

does yield a solution. With marginal exponential distributions we obtain the bivariate exponential distribution

$$\overline{F}(x,y) = \{\exp -[\lambda_1 x + \lambda_2 y + \lambda_{12}\max(x,y)]\}.$$

This distribution was obtained by Marshall and Olkin (1967). For a further discussion of this distribution and other characterizations see Azlarov and Volodin (1986); also see Ghurye and Marshall (1984) for further developments concerning the functional equation.

2.2b Minima

The minimum for the exponential distribution plays a role similar to that of the mean for the normal distribution. The minimum of independent exponential random variables is exponential, and this property characterizes the exponential distribution.

This fact can be used to generate a bivariate version. Let U, V, W be independent exponential random variables, then

$$X = \min(U, W), \qquad Y = \min(V, W)$$

has a bivariate exponential distribution. Indeed, it is surprising that the result of this construction leads to the same distribution as obtained by the loss of memory property.

2.2c Order statistics

There are a host of characterizations of the exponential distribution based on order statistics. For example, Ferguson (1964) shows that if X and Y are independent (with absolutely continuous distributions), then the independence of min (X, Y) and $X - Y$ characterizes the exponential distribution.

The independence of spacings of order statistics implies that the underlying distribution is exponential. These characterizations have not yielded multivariate versions; the extension of univariate order statistics to vector order statistics appears illusive in this context.

3. Multivariate Beta Distributions. Whereas the normal and exponential distributions seem to have "natural" extensions, there is no obvious generalization of the univariate beta distribution.

3.1 A matrix extension

The analysis of variance makes use of the fact that if U and V are independent, each having a gamma distribution, then U/V has an F-distribution (ignoring constants), and $U/(U+V)$ has a beta distribution, independent of $U+V$.

Because the Wishart distribution is the distribution of the sample covariance matrix, it is a multivariate version of the gamma, or chi-quare, distribution. This permits the following extension: If U and V are independent $p \times p$ random matrices, each having a Wishart distribution with the same scale, then $W = V^{-1/2}UV^{-1/2}$ has a multivariate F-distribution, and $Z = (U+V)^{1/2}U(U+V)^{-1/2}$, independent of $U+V$, has a multivariate beta distribution with density function

$$f(Z) = \text{constant } |Z|^a |I-Z|^b, \qquad 0 < Z < I, \qquad a, b > 0.$$

(Here $U^{1/2}$ is the unique positive definite square root of U, and $A < B$ means that the matrix $B - A$ is positive definite.)

3.2 Bivariate gamma variates.

An alternative direct extension is to start with (U_1, V_1) and (U_2, V_2) independent, each having a bivariate gamma distribution. Now define

$$(X, Y) = \left(\frac{U_1}{U_1+V_1}, \frac{U_2}{U_2+V_2}\right).$$

Then (X, Y) has a bivariate beta distribution. There is considerable flexibility in this formulation in that there are many bivariate gamma distributions. However, the form of this joint distribution has not been studied even for the simplest bivariate gamma distribution.

3.3 The Dirichlet-Liouville distribution

The Dirichlet density

$$f(x_1, \cdots, x_k) = \text{constant} \prod_l^k x_i^a \left(1 - \sum x_i\right)^b, \qquad 0 \le x_i, \sum x_i < 1$$

has surfaced from many contexts as being an important distribution. One of its distributional origins is the following: If $U_0, U_1, \cdots, U_k$ are independent gamma variates with the same scale parameter, then $X_j = U_j / \sum_0^k U_i$, $j = 1, \cdots, k$, has a multivariate Dirchlet distribution.

A matrix version arises in a parallel manner: If $U_0, U_1, \cdots, U_k$ are independent random matrices, each having a Wishart distribution with the same scaling matrix, then $X_j = (\sum_0^k U_i)^{-1/2} U_j (\sum_0^k U_i)^{-1/2}$ has a Dirichlet distribution:

$$f(X_1, \cdots, X_k) = (\text{constant}) \prod_1^k |X_j|^{a_j} \left|I - \sum X_j\right|^b,$$

where $0 < X_i$, $\sum X_i < I$.

The construction $Y_i = U_0^{-1/2} U_i U_0^{-1/2}$ yields a matrix version of the F-distribution:

$$f(Y_1, \cdots, Y_k) = (\text{constant}) \prod_1^k Y_i^{a_i} / |I + \Sigma Y_i|^b, \qquad 0 < Y_i.$$

Other general versions arising from Wishart matrices are given in Olkin (1959); see also Gupta and Richards (1992).

4. Sampling Methods. The Bernoulli distribution plays a central role in sampling procedures. Different sampling procedures lead to the binomial, hypergeometric, geometric, negative binomial distributions. Limits can then be invoked to yield the Poisson, geometric, gamma, and normal distributions. In this sense, this family has a cohesiveness.

A bivariate family can be generated by starting with a bivariate Bernoulli distribution and employing similar sampling plans, and then invoking limits. In this way we generate bivariate binomial, bivariate geometric,bivariate Poisson distributions, and so on.

We now display but two of these. Our starting point is a bivariate Bernoulli distribution:

$$P\{(U,V)=(i,j)\}=p_{ij}, \qquad i,j=0,1; \quad 0\le p_{ij}, \quad \sum p_{ij}=1.$$

For a sample of size n, let $X=\sum U_i$, $Y=\sum V_i$. Clearly, X and Y each has a binomial distribution. The joint distribution is

$$P(X=x,\ Y=y)=\sum_{\alpha} c(\alpha,x,y)p_{11}^{\alpha}p_{10}^{x-\alpha}p_{01}^{y-\alpha}P_{00}^{n-x-y+\alpha},$$

where $x,y=0,1,\cdots,n,$

$$c(\alpha,x,y)=n!/\left[\alpha!(x-a)!(y-\alpha)!(n-x-y+\alpha)!\right].$$

and the sum is over all integers α for which $\alpha\ge 0$, $x-\alpha\ge 0$, $y-\alpha\ge 0$, $n-x-y+\alpha\ge 0$. By letting $p_{ij}=\lambda_{ij}/n$ and taking limits we obtain a bivariate Poisson distribution:

$$P\{X=x,\ Y=y\}=\sum_{\alpha}\frac{\lambda_{11}^{\alpha}\lambda_{10}^{x-\alpha}\lambda_{01}^{y-\alpha}}{\alpha!(x-\alpha)!(y-\alpha)!}e^{-\lambda_{11}-\lambda_{10}-\lambda_{01}}.$$

It is interesting that this distribution was developed early, M'Kendrick (1926), in the context of two types of infrequent injuries inflicted on soldiers. (See Marshall and Olkin (1985) for more details using this construction.)

This bivariate distribution also arises from the construction

$$X=U+W, \qquad Y=U+W,$$

where U,V,W are independent, each having a Poisson distribution with parameters $\lambda_{10},\lambda_{01},\lambda_{11}$, respectively.

This latter construction points out a fundamental problem with many multivariate non-normal distributions, namely, that the number of parameters becomes large. To see this, note that we require seven (generally 2^1)

independent Poisson random variables to create a general trivariate Poisson distribution:

$$\begin{aligned} X &= U_1 + V_{12} \qquad\quad + V_{13} + W, \\ Y &= U_2 + V_{12} + V_{23} \qquad\quad + W, \\ Z &= U_3 \qquad\quad + V_{23} + V_{13} + W. \end{aligned}$$

Here $U_1, U_2, U_3, V_{12}, V_{13}, V_{23}$, and W are independent Poisson random variables.

This problem also occurs in generating a trivariate exponential distribution using minima. Now we require

$$\begin{aligned} X &= \min(U_1, V_{12}, V_{13}, W), \\ Y &= \min(U_2, V_{12}, V_{23}, W), \\ Z &= \min(U_3, V_{23}, V_{13}, W), \end{aligned}$$

where $U_1, U_2, U_3, V_{12}, V_{13}, V_{23}$ and W are independent exponential random variables.

These two examples illustrate why normality, which has $p(p+3)/2$ parameters, is easier to deal with. Problems of estimation become considerably more difficult for many non-normal distributions.

4.1 Urn models

The basic Pólya-Eggenberger sampling procedure starts with an urn containing a red and b black balls. A ball is drawn, replaced together with s additional balls of the same color. For simplicity let $s = 1$.

If X is the number of red balls drawn in the first n trials, then

$$P\{X = x\} = \binom{n}{x} \frac{B(x+a,\ n-x+b)}{B(a,\ b)},$$

where $B(\,\cdot\,,\,\cdot\,)$ is the beta function.

It is interesting that by changing the sampling scheme and by taking limits, this model leads to the binomial, negative binomial, beta, Poisson, gamma, and normal distribution.

There are many ways to generate bivariate versions. Perhaps the most direct is to let the urn consist of a red, b black and c white balls. With X the number of red balls, and Y the number of black balls in the first n trials, we obtain the distribution

$$P\{X = x,\ Y = y\} = \binom{n}{x, y} \frac{B(a+x,\, b+y,\, c+n-x-y)}{B(a,b,c)},$$

where $B(a_1, \cdots, a_k) = \prod_1^k \Gamma(a_i)/\Gamma(\sum_1^k a_i)$.

In an alternative construction, let the urn contain a_{ij} balls marked (i,j), $i,j = 0,1$, and let Z_{ij} denote the number of balls of type (i,j) in the first n trials. Now define X to be the number of balls with first digit 1, and Y the number of balls with second digit 1, that is,

$$X = Z_{11} + Z_{10}, \qquad Y = Z_{11} + Z_{01}.$$

Then the joint distribution of (X,Y) is

$$P(X = x,\ Y = y)$$
$$= \sum_{\alpha} c(x,y,\alpha,n) \frac{B\big(a_{11} + \alpha,\ a_{10} + x - \alpha,\ a_{01} + y - \alpha,\ a_{00} + n - x - y + \alpha\big)}{B(a_{11}, a_{10}, a_{01}, a_{00})},$$

where

$$c(x,y;\alpha,n) = a!\,/\,[\alpha!\,(x-\alpha)!\,(y-\alpha)!\,(n-x-y+\alpha)!]$$

.

$\alpha \geq 0, x - \alpha \geq 0, y - \alpha \geq 0, n - x - y + \alpha \geq 0.$

These, and other distributions arising from the Pólya-Eggenberger urn model are discussed in Marshall and Olkin (1990a, 1993b). Other aspects of urn models are provided by Johnson and Kotz(1977).

5. Families with Given Marginals. Considerable effort has been expended to create usable general bivariate families with given marginals. That these families have limitations is to be expected, since they apply to many different situations.

Two such general families are those of Farlie-Gumbel-Morgenstern:

$$H(x,y) = F(x)G(y)(1 + \alpha\overline{F}(x)\overline{G}(y))$$

and Gumbel:

$$[-\log \overline{H}(x,y)]^m = [-\log \overline{F}(x)]^m + [-\log \overline{G}(y)]^m,$$

where $\overline{F} = 1 - F$.

One mechanism by which many families can be uncovered is provided in papers by Genest and Mackay (1986) and Marshall and Olkin (1988). We state one such result: Let $\phi(s,t)$ denote a bivariate Laplace transform and $\phi_1(s) = \phi(s,0)$ $\phi_2(t) = \phi(0,t)$ the marginal Laplace transforms.Then

$$H(x,y) = \phi(\phi_1^{-1} H_1(x), \phi_2^{-1} H_2(y))$$

is a bivariate distribution with marginals H_1, H_2. A simpler version is the case that ϕ is a univariate Laplace transform, which yields that

$$H(x,y) = \phi(\phi^{-1} H_1(x) + \phi^{-1} H_2(y))$$

is a bivariate distribution with marginals H_1, H_2. Genest and Mackay (1986) given examples that yield a number of known general families. In particular, the choice of $\phi(s) = \exp\{-s^{1/m}\}$, $m \geq 1$, yields Gumbel's family above.

For a compendium and discussion of families of distributions, see Hutchinson and Lai (1990).

6. A Hierarchy Using Mixtures. Mixture models provide a simple device for generating multivariate distributions. For any density $f(x, \theta)$ depending on a parameter θ, we can generate a new density by mixing on θ:

$$h(x) = \int f(x \mid \theta) g(\theta)\, d\theta.$$

For example, with

$$f(x \mid \theta) = \binom{n}{x} \theta^x \overline{\theta}^{n-x}, \qquad \overline{\theta} = 1 - \theta,$$

$$g(\theta) = \theta^{a-1} \overline{\theta}^{b-a} / B(a, b),$$

we obtain

$$h(x) = \binom{n}{x} \frac{B(a + x,\ n - x + b)}{B(a, b)}.$$

A direct generalization of this construction is obtained from

$$h(x, y) = \int f_1(x \mid \theta) f_2(y \mid \theta) g(\theta)\, d\theta.$$

For example, with $f_1(x \mid \theta)$ and $g(\theta)$ as above and

$$f_2(y \mid \theta) = \binom{m}{y} \theta^y \overline{\theta}^{m-y}$$

we obtain

$$h(x,\, y) \binom{n}{x} \binom{m}{y} \frac{B(a + x + y,\, n + m - x - y + b)}{B(a, b)}.$$

Of course, in general θ need not be univariate, so that we could have

$$h(x, y) = \int f_1(x \mid \underset{\sim}{\theta}) f_2(y \mid \underset{\sim}{\theta}) g(\underset{\sim}{\theta})\, d\underset{\sim}{\theta},$$

where $\underset{\sim}{\theta}$ is a vector and g is a multivariate distribution. For example, let x and y have Poisson distributions

$$f_1(x \mid \underset{\sim}{\theta}) = \theta_1^x e^{-\theta_1} / x\,!,$$

$$f_2(x \mid \underset{\sim}{\theta}) = \theta_2^y e^{-\theta_2} / y\,!,$$

and $g(\underset{\sim}{\theta})$ has a bivariate gamma distribution:

$$g(\underset{\sim}{\theta}) = (\text{constant})e^{-\theta_1-\theta_2} \int_0^{\min(\theta_1,\theta_2)} (\theta_1 - t)^{\alpha-1}(\theta_2 - t)^{b-1}t^{c-1}e^t dt.$$

Although these constructions readily arise from this mixture construction, the resulting distributions may not be useful in practice.

Finally, we note that x and y are independent in the above constructions. Instead, we could start with x, y dependent:

$$h(x, y) = \int f(x, y \mid \underset{\sim}{\theta})g(\underset{\sim}{\theta})\, d\underset{\sim}{\theta}.$$

An example is the case that $f(x, y \mid \underset{\sim}{\theta}) \sim \mathcal{N}(\underset{\sim}{\theta}, \Sigma)$ and $g(\underset{\sim}{\theta}) \sim \mathcal{N}(\underset{\sim}{\mu}, \Psi)$.

The multivariate F-distribution of Section 3.1 can be generated as a mixture. Let S be a $p \times p$ covariance matrix having a Wishart distribution

$$f(s \mid \Theta) = \text{constant}|S|^{\frac{n-p-1}{2}}|\Theta|^{\frac{n}{2}}e^{\frac{1}{2}\text{tr}S\Theta},$$

and let

$$g(\Theta) = \text{constant}|\theta|^{\frac{m-p-1}{2}}|\Psi|^{\frac{m}{2}}e^{\frac{1}{2}\text{tr}\Theta\Psi},$$

then

$$\begin{aligned} h(S) &= \int f(S \mid \Theta)g(\Theta)\, d\Theta \\ &= \text{constant}\frac{|S|^{\frac{n-p-1}{2}}|\Psi|^{\frac{m}{2}}}{|S + \Psi|^{\frac{n+m}{2}}}. \end{aligned}$$

The multivariate t-distribution can also be generated using mixtures.

There are some byproducts of mixtures that are important; one is the notion of association. Random variables $T_1, \cdots, T_n$ are said to be associated, or their joint distribution is associated, if

$$\text{cov}\left(f(T_1, \cdots, T_n),\ g(T_1, \cdots, T_n)\right) \geq 0$$

for all f and g increasing in each argument. Consider the mixture

$$\overline{H}(x) = \int \overline{F}(x \mid \theta)dG(\theta).$$

It was shown by Ahmed, León and Proschan (1981) that under certain regularity conditions on F, if $F(x \mid \theta)$ is associated for each fixed θ, and if $g(\theta)$ is associated, then H is associated. This result provides a powerful general mechanism to show association.

A general discussion of mixture models is contained in Marshall and Olkin (1990b).

7. Ecological Correlations. There is a considerable literature in the social sciences concerned with aggregation and models for disaggregation. The term ecological correlation appeared in a paper of Robinson (1950) in which he demonstrates that correlations can vary for different levels of aggregation, which can lead to errors in inferring individual behavior from group behavior.

We describe one such model due to Goodman (1953) which provides a method for estimating individual behavior from grouped data. Consider k groups in which each group yields a 2×2 table

	Success	Failure	
C			x_i
$\overline{C}$			$1 - x_i$
	y_i	$1 - y_i$	

We observe the proportions x_i and y_i, $i = 1, \cdots, k$. (In some instances Success may refer to the proportion voting for a candidate, C is a characteristic, such as race or gender. Here x mat not be known, but obtained from census data.) If we assume that the group cell proportions are constant

	Success	Failure	
C	θ	$1 - \theta$	x_i
$\overline{C}$	t	$1 - t$	$1 - x_i$
	y_i	$1 - y_i$	

then we can generate a regression model

$$y_i = \theta x_i + t(1 - x_i) \equiv \alpha + \beta x_i$$

A general discussion of issues and methods of aggregation and disaggregation is provided by Hannan (1971).

8. Scaling of Matrices and Joint Distributions of Discrete Random Variables. The purpose of scaling arises in many different contexts. One of the early origins is to determine where it is preferable to solve the linear system $Ax = b$ or one of the alternative systems $A'Ax = b$, or $D_1AD_2x = b$, where D_1 and D_2 are diagonal scaling matrices.

The first comparison for solving $Ax = b$ versus $A'Ax = b$ can be answered by comparing the condition number of the two systems, namely, $c(A) = \Phi(A)\Phi(A^{-1})$ with $c(A'A) = \Phi(A'A)\Phi((A'A)^{-1})$, where Φ is a norm. For a broad class of norms Φ, it can be shown that $c(A) \leq c(A'A)$, which suggests that it is preferable to solve $Ax = b$. (See Marshall and Olkin (1979) for further details.)

In a probabilistic context suppose that we have a discrete bivariate distribution as exemplified by an $m \times n$ contingency table. The probabilities are p_{ij} with marginal distributions p_{i+} and p_{+j}, $i = 1, \cdots, m$; $j = 1, \cdots, n$. The question is whether there exists scaling values a_i and b_j so that the bivariate distribution $a_i b_j p_{ij}$ has target marginals $s_1, \cdots, s_m$ and $t_1, \cdots, t_n$. When $s_i = t_j = 1$, the new matrix $(a_i b_j p_{ij})$ is doubly stochastic.

This problem can be formulated as follows: Given a matrix $P = (p_{ij})$, $p_{ij} \geq 0$, do there exist diagonal matrices $D_a = \text{diag}(a_1, \cdots, a_m)$ and $D_b = \text{diag}(b_1, \cdots, b_n)$ such that $D_a P D_b$ is doubly stochastic. Under mild conditions the answer is in the affirmative, as shown by Sinkhorn (1964). There is now a considerable history on this problem (see e.g., Marshall and Olkin (1979), Rothblum and Schneider (1989)).

9. Fréchet Bounds. The determination of distributional bounds plays a central role in generating multivariate distributions. The classical result of Fréchet and Hoeffding (see Marshall and Olkin (1979)) states that every bivariate distribution $H(x, y)$ with given marginals $F_1(x)$ and $F_2(y)$ is bounded by two bivariate distributions:

$$H_L = \max\{0, F_1 + F_2 - 1\}, \qquad H_U = \min\{F_1, F_2\},$$

namely,

$$H_L \leq H \leq H_U.$$

An implication of this is that the respective correlations ρ_L, ρ_U, ρ satisfy the same inequality:

$$\rho_L \leq \rho \leq \rho_U.$$

The correlation inequality holds for other measures of dependence such as Kendall's tau. Consequently, we can determine the range of correlations of a bivariate gamma family, say, by determining the correlations of H_L and H_U, which depend only on the marginal distributions.

A direct extension to the multivariate case is that $H_L = \max\{0, F_1 + \cdots + F_k - (k-1)\}$, $H_U = \min\{F_1, \cdots, F_k\}$. However, now H_L need not be a multivariate distribution. This has led to a range of formulations and results. In particular, a reformulation in terms of a transportation problem has been fruitful.

The multivariate version also leads to new questions. For example, given compatible and consistent bivariate distributions F_{12}, F_{13}, F_{23} what are bounds for H. Another class of problems is to determine bounds for the joint distribution H with given marginals F_1, F_2 that possesses an additional property, such as, unimodality or with a given correlation.

A problem that has recently received attention is given marginal distributions F_1, F_2 and any bivariate distribution $G(x,y)$ find the bivariate distribution $H(x,y)$ with marginals F_1, F_2 such that

$$H(x,y) \leq G(x,y).$$

The solution is given by Hoffman and Veinott (1990):

$$H(x,y) = \inf_{\substack{u \leq x \\ v \leq y}} [G(u,v) + F_1(x) + F_2(y) - F_1(u) - F_2(v)].$$

(See also Olkin and Rachev (1990).)

10. Generating Correlation Matrices. There are now algorithms for generating a wide variety of non-normal distributions. Some of the constructions discussed in this paper suggest methods for generating samples from non-normal populations. For example, the method of Section 2.2b can be used to generate bivariate exponential distributions. The sampling methods of Section 4 and the mixture models of Section 6 can also be used to generate bivariate families.

Some numerical analysis simulations require generating a random correlation matrix R with the constraint that its expected value is a given matrix C, and that it be positive definite with minimum eigenvalue λ. A construction is provided by Marsaglia and Olkin (1983) as follows. Define $R = C + X$, where $X = (x_{ij})$ is symmetric, $x_{ii} = 0, x_{ij}, i \neq j$, are generated in the interval $|x_{ij}| < a_{ij}$, where $A = (a_{ij})$ is any given symmetric matrix, $a_{ii} = 0, max_i \sum_j |a_{ij}| < \lambda$.

In another context Anderson, Olkin and Underhill (1987) discuss methods for generating random orthogonal matrices. This problem has a history of alternative procedures.

An excellent compendium of methods for constructing non-uniform random variables is contained in Devroye (1986).

11. Physical Models. Determining distributions from physical models is a satisfying method in that the distribution is designed for a particular situation. Because of the complexity of physical phenomena, this course is more unusual than one might hope for. However, there are a number of examples in which distributions have been tailored specifically to applications.

An early example is the work of M'Kendrick (1926), discussed in Section 4. The bivariate exponential distribution of Section 2.2 was used in the study of cancer at two sites (see Klein, Keiding and Kamby, 1989). Freund (1961) considered a model in which there are two components A and B, each having an exponential life distribution with parameters α and β, respectively. The failure of component A alters the β parameter ot β', and similarly, the failure of B alters α to α'.

A general method for generating physical models is via shock models. In such models a shock may destroy a component, or it may destroy it with some probability. It is also possible to impose intensities on the shock. This permits considerable latitude in the dependency relation between the two variables. The bivariate exponential distribution of Section 2.2 can be generated in this way. For a general discussion of shock models, see Friday and Patil (1977).

12. Summary. We have provided a brief panorama of problems inherent in the generation of bivariate (or multivariate) distributions, and a number of general procedures used to produce bivariate distributions. An excellent compendium of bivariate distributions and methods used to generate them is contained in Hutchinson and Lai (1990). Of course, the now classical reference on distributions by Johnson and Kotz (1970, 1972, 1992) is indispensable in any study of special distributions.

REFERENCES

AHMED, A. N., LEÓN, R. and PROSCHAN, F. (1981). Generalized association with applications in multivariate statistics, *Ann. Statist.* **9**, 168–176.

ANDERSON, T. W. and GHURYE, S.G. (1977). Identification of parameters by the distribution of a maximum random variable, *J. Roy. Statist. Soc.* Ser. B **39**, 337–342.

ANDERSON, T. W., OLKIN, I. and UNDERHILL, L. (1987). Generation of random orthogonal matrices, *SIAM J. Scientific and Statist. Computing* **8**, 625–629.

AZLAROV, T. A. and VOLODIN, N. A. (1986), Characterization problems associated with the exponential distribution (translated by M. Stein and edited by I. Olkin). Springer-Verlag, New York.

DEVROYE, L. (1986). *Non-uniform Random Variate Generation.* Springer-Verlag, New York.

FREUND, J. E. (1961). A bivariate extension of the exponential distribution, *J. Amer. Statist. Assoc.* **56**, 971–977.

FRIDAY, D. S. and PATIL, G. P. (1977). A bivariate exponential model with applications to reliability and computer generation of random variables, p. 527–549 in The Theory and Applications of Reliability, Vol. I, with Emphasis on Bayesian and Nonparametric Methods (ed. by C. Tsokos and I.N. Shimi). Academic Press, New York.

GALAMBOS, J. and KOTZ, S. (1978). Characterizations of Probability Distributions. Lecture Notes Math 675, Springer, New York.

GENEST, C. and MACKAY, R. J. (1986). The joy of copulas: bivariate distributions with uniform marginals, *The American Statistician* **40**, 280–283.

GHURYE, S. G. and MARSHALL, A. W. (1984). Shock processes with after effects and multivariate lack of memory, *J. Appl. Prob.* **21**, 786–801.

GHURYE, S. G. and OLKIN, I. (1962). A characterization of the multivariate normal distribution, *Ann. Math. Statist.* **33**, 533–541.

GOODMAN, L. (1953). Ecological regression and the behavior of individuals, *Amer. J. Sociology* **64**, 610–625.

GUPTA, R.D. and RICHARDS, D. St. P. (1992). Multivariate Liouville distributions, III *J. Multivariate Analysis* **43**, 29–57.

HANNAN, M. T. (1971). Aggregation and Disaggregation in Sociology. Lexington Books, D.C. Heath & Co., Lexington, MA.

HOFFMAN, A. J. and VEINOTT, A. F. (1990). Staircase transportation problems with superadditive rewards and cumulative capacities, to appear in *J. Math. Programming.*

HUTCHINSON, T. P. and LAI, C. D. (1990). Continuous Bivariate Distributions, Emphasizing Applications. Rumsby Scientific Publishing Adelaide, South Australia.

JOHNSON, N. L. and KOTZ, S. (1970). Continuous Univariate Distributions, Vol. 1 and 2. Houghton Mifflin Co., Boston.

JOHNSON, N. L. and KOTZ, S. (1972). Distributions in Statistics; Continuous Multivariate Distributions. Wiley, New York.

JOHNSON, N. L. and KOTZ, S. (1977). Urn Models and Their Application. Wiley, New York.

JOHNSON, N. L., KOTZ, S. and KEMP, A. W. (1992), Univariate Discrete distributions. Wiley, New York.

KAGAN, A. M., LINNIK, YU. V. and RAO, C. R. (1973). Characterization Problems in Mathematical Statistics. Wiley, New York.

KAKOSYAN, A. V., KLEBANOV, L. B. and MELAMED, J. A. (1984) . Characterization of Distributions by the Method of Intensively Monotone Operators. Lecture Notes Math 1088, Springer-Verlag, Berlin.

KLEIN, J. P., KEIDING N. and KAMBY, C. (1989). Semiparametric Marshall-Olkin models applied to the occurrence of metasteses of multiple sites after breast cancer, *Biometrics* **45**, 1073–1086.

MARSAGLIA, G. and OLKIN, I. (1983). Generating correlation matrices, *SIAM J. Sci. Statist. Comput.* **5**, 470–475.

MARSHALL, A. W. and OLKIN, I. (1967). A multivariate exponential distribution, *J. Amer. Statist. Assoc.* **62**, 30–44.

MARSHALL, A. W. and OLKIN, I. (1979). Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.

MARSHALL, A. W. and OLKIN, I. (1985). A family of bivariate distributions generated by the bivariate Bernoulli distribution, *J. Amer. Statist. Assoc.* **80**, 332–338.

MARSHALL, A. W. and OLKIN, I. (1988). Families of multivariate distributions, *J. Amer. Statist. Assoc.* **83**, 834–841.

MARSHALL, A. W. and OLKIN, I. (1990a). Bivariate distributions generated from Pólya-Eggenberger urn models, *J. Multivariate Anal.* **35**, 48–65.

MARSHALL, A. W. and OLKIN, I. (1990b). Multivariate distributions generated from mixtures of convolution and product families, p.371–393 in Topics in Statistical Dependence (ed. by H.W. Black, A.R. Sampson, and T.Savits), Volume 16, Lecture Notes/Monograph Series, Institute of Mathematical Statistics, Hayward, CA.

MARSHALL, A. W. and OLKIN, I. (1991). Functional equations for multivariate distributions, *J. Multivariate Anal.* **39**, 209–215.

MARSHALL, A. W. and OLKIN, I. (1993a). Maximum likelihood characterization of distributions, *Statistica Sinica* **3**, 157–171.

MARSHALL, A. W. and OLKIN, I. (1993b). Bivariate life distributions from Pólya's urn model for contagion, *J. Appl. Prob.*

M'KENDRICK, A. G. (1926). Applications of mathematics to medical problems, *Proc. Edinburgh Math. Soc.* **44**, 98–130.

OLKIN, I. (1959). A class of integral identities with matrix argument, *Duke Math. J.* **26**, 207–213.

OLKIN, I and RACHEV, S. T. (1990). Distributions with given marginals subject to certain constraints. Technical Report No. 270, Stanford University.

ROTHBLUM U. G. and SCHNEIDER, H. (1989). Scalings of matrices which have prespecified row sums and column sums via optimization, *Linear Alg. Appl.* 114/115, 737–764.

SINKHORN, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices, *Ann. Math. Statist.* **35**, 876–879.

TEICHER, H. (1961). Maximum likelihood characterization of distributions, *Ann. Math. Statist.* **32**, 1214–1222.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
SEQUOIA HALL
STANFORD, CALIFORNIA 94305
USA