3

Constructions of Discrete Bivariate Distributions

C. D. Lai

Massey University, Palmerston North, New Zealand

Abstract: Various techniques for constructing discrete bivariate distributions are scattered in the literature. We review these methods of construction and group them into some loosely defined clusters.

Keywords and phrases: Bernoulli, bivariate distributions, conditioning; canonical correlation, clustering, constructions, compound, discrete, extreme points, Fréchet bounds, marginal transformation, mixing, sampling, trivariate, truncations, urn models, weighting functions

3.1 Introduction

Over the last two or three decades, a vast amount of literature on discrete bivariate and multivariate distributions has been accumulated. For an extensive account of these distributions, we refer our readers to the books by Kocherlakota and Kocherlakota (1992) and Johnson *et al.* (1997), and the review articles by Papageorgiou (1997), Kocherlakota and Kocherlakota (1998), and Balakrishnan (2004, 2005).

In this chapter, we restrict ourselves to reviewing methods of constructing discrete bivariate distributions. A review on constructions of continuous bivariate distributions is given by Lai (2004). Unlike their continuous analogues, discrete bivariate distributions appear to be harder to construct. One of the problems is highlighted in Kemp and Papageorgiou (1982) in which they said, "Various authors have discussed the problem of constructing meaningful and useful bivariate versions of a given univariate distribution, the main difficulty being 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 bivariate version." Many bivariate distributions arise without having prespecified the marginals. There is no satisfactory unified mathematical scheme of classifying these methods. What we hope to achieve is to group them into semicoherent clusters. The clusters may be listed as

- Mixing and compounding
- Trivariate reduction
- One conditional and one marginal given
- Conditionally specified method
- Construction of discrete bivariate distributions with given marginals and correlation
- Sums and limits of Bernoulli trials models
- Sampling from urn models
- Clustering (bivariate distributions of order k)
- Construction of finite bivariate distributions via extreme points of convex sets
- Generalized distributions method
- Canonical correlation coefficients and semi-groups
- Distributions arising from accident theory
- Bivariate distributions generated from weight functions
- Marginal transformations method
- Truncation method
- Constructions of positively dependent discrete bivariate distributions.

Several of these are also common methods for constructing continuous bivariate distributions. We refer the reader to Lai (2004) for a review of these and other methods of constructing continuous bivariate distributions. We note that for discrete bivariate distributions, the probability generating function is often used as a tool for construction as well as for studying their properties.

We have not discussed computer generation of discrete bivariate random variables. We refer interested readers to the works by Professors A. W. Kemp and C. D. Kemp on this subject. Kocherlakota and Kocherlakota (1992) present several such references by the Kemps.

3.2 Mixing and Compounding

3.2.1 Mixing

As for continuous bivariate distributions, an easy way to construct a discrete bivariate distribution is to use the method of mixing two or more distributions. Suppose H_1 and H_2 are two discrete bivariate distributions; then

$$H(x,y) = \alpha H_1(x,y) + (1-\alpha)H_2(x,y) \tag{3.1}$$

($0 \leq \alpha \leq 1$) is a new bivariate distribution.

Example: Consider the problem of describing the sex distribution of twins. Twin pairs fall into three classes: MM, MF, and FF where M denotes male and F female. This leads to the trinomial distribution. As twins may be dizygotic or monozygotic, a mixture of trinomials results. For more details, see Blischke (1978), Goodman and Kruskal (1959), and Strandskov and Edelen (1946).

Papageorgiou and David (1994) studied several countable mixtures of binomial distributions.

3.2.2 Compounding

Compounding is perhaps the most common method of constructing discrete bivariate distributions. Let X and Y be two random variables with parameters θ_1 and θ_2, respectively. For a given value of (θ_1, θ_2), X and Y may be either independent or correlated.

(i) X and Y are conditionally independent.

If θ_1 and θ_2 are independent, then the resulting pair $X$and Y are also independent. For example, for given (θ_1, θ_2), X and Y are independent Poissons. If θ_1 and θ_2 are independent gammas, then the resulting X and Y are independent negative binomials.

- θ_1 and θ_2 may have a bivariate distribution such as the case of Consael's bivariate Poisson distribution [Consael (1952)].
- David and Papageorgiou (1994) presented several compounded bivariate Poisson distributions that can be derived in this manner.

(ii) X and Y are dependent for given values of the compounding parameters.

- The compounded bivariate Poisson distributions given by Kocherlakota (1988) are obvious examples.
- Another example is the generalized Consael distribution obtained by

  $$(X, Y) \sim \mathrm{Biv\,P}(\lambda_1, \lambda_2, \lambda_3) \underset{(\lambda_1, \lambda_2, \lambda_3)}{\wedge} F(\lambda_1, \lambda_2, \lambda_3)$$

  where the symbol $\wedge$ denotes compounding. Here $\mathrm{Biv\,P}(\lambda_1, \lambda_2, \lambda_3)$ has a bivariate Poisson distribution with a probability-generating function given by

  $$g(s, t) = \exp\{\lambda_1(s-1) + \lambda_2(t-1) + \lambda_3(st-1)\}, \tag{3.2}$$

and $(\lambda_1, \lambda_2, \lambda_3)$ has a trivariate distribution function F.

For example, H_8 distribution [Kemp and Papageorgiou (1982)] is obtained when $(\lambda_1,\lambda_2, \lambda_3)$ has a trivariate normal distribution.

There are other variants of compounding; see, for example, Chapter 8 of Kocherlakota and Kocherlakota (1992).

3.3 Trivariate Reduction

This is also known as "the variables in common method." The idea here is to create a pair of dependent random variables from three or more random variables. In many cases, these initial random variables are independent, but occasionally they may be dependent. An important aspect of this method is that the functions connecting these random variables to the two dependent random variables are generally elementary ones; random realizations of the latter can therefore be generated easily from random realizations of the former. A broad definition of the variables-in-common technique is as follows. Set

$$\left.\begin{aligned} X &= \tau_1(X_1, X_2, X_3), \\ Y &= \tau_2(X_1, X_2, X_3), \end{aligned}\right\} \tag{3.3}$$

where X_1, X_2, X_3 are not necessarily independent or identically distributed. A narrow definition is

$$\left.\begin{aligned} X &= X_1 + X_3, \\ Y &= X_2 + X_3, \end{aligned}\right\} \tag{3.4}$$

with X_1, X_2, X_3 being i.i.d. Another possible definition is

$$\left.\begin{aligned} X &= \tau(X_1, X_3), \\ Y &= \tau(X_2, X_3), \end{aligned}\right\} \tag{3.5}$$

with (i) the X_i being independently distributed and having c.d.f. $F_0(x_i; \lambda_i)$, and (ii) X and Y having distributions $F_0(x; \lambda_1+\lambda_2)$ and $F_0(y; \lambda_1+\lambda_3)$, respectively.

Example: Suppose $X_i \sim \text{Poisson}(\lambda_i)$, $i = 1, 2, 3$. Define $X = X_1 + X_3$, $Y = X_2 + X_3$ so that the joint pgf of (X, Y) is given by

$$g(s, t) = \exp\{\lambda_1(s-1) + \lambda_2(t-1) + \lambda_3(st-1)\} \tag{3.6}$$

which is called the bivariate Poisson distribution. This distribution is often used as a basis for obtaining a compound bivariate Poisson distribution. More specifically, if each independent $\lambda_i \sim \text{Gamma}(\alpha_i, \beta)$, then the resulting distribution is a bivariate negative binomial [see, e.g., Stein and Juritz (1987)]. If each independent $\lambda_i \sim \text{GIG}(\alpha_i, \zeta_i, \frac{1}{2})$ (GIG = generalized inverse Gaussian), then (X, Y) has a bivariate inverse Gaussian–Poisson distribution.

(**Note:** $\lambda_1 + \lambda_2 \sim \text{GIG}(\alpha_1 + \alpha_2, \zeta_1 + \zeta_2, \frac{1}{2})$. The inverse Gaussian–Poisson distribution is a special case of Sichel distribution.)

An obvious disadvantage of this method is that the correlation is restricted to be strictly positive.

Zheng and Matis (1993) generalized the trivariate reduction method by considering a random rewarding system so that

$$X = \begin{cases} X_1 + X_2 & \text{with prob } \pi_1 \\ X_1 & \text{with prob } 1 - \pi_1 \end{cases}$$

and

$$Y = \begin{cases} X_1 + X_3 & \text{with prob } \pi_2 \\ X_3 & \text{with prob } 1 - \pi_2. \end{cases}$$

Several discrete bivariate distributions were constructed, whose marginal distributions are mixtures of negative binomial distributions.

Lai (1995) proposed an extension to the model of Zheng and Matis (1993) by setting

$$\left. \begin{aligned} X &= X_1 + I_1 X_2, \\ Y &= X_3 + I_2 X_2, \end{aligned} \right\} \tag{3.7}$$

where $I_i\,(i = 1, 2)$ are indicator random variables which are independent of X_i, but (I_1, I_2) has a joint probability function.

3.4 One Conditional and One Marginal Given

A discrete bivariate distribution can be expressed as the product of a marginal distribution and a conditional distribution as

$$\Pr\{X = x, Y = y\} = \Pr\{Y = y | X = x\} \Pr\{X = x\}. \tag{3.8}$$

This is an intuitively appealing approach, especially when Y can be thought of caused by, or predictable from, X.

Moreover, given positive $\Pr\{X = x|Y = y\}$ for all x, y, and $\Pr\{Y = y|X = x_0\}$, for all y and a fixed x_0, the joint distribution can be determined uniquely [Patil (1965)]:

$$\Pr\{X = x, Y = y\} \propto \frac{\Pr\{X = x|Y = y\} \Pr\{Y = y|X = x_0\}}{\Pr\{X = x_0|Y = y\}}. \tag{3.9}$$

Normalization determines the proportional constant; see, for example, Gelman and Speed (1993).

Furthermore, discrete bivariate distributions can be generated from given conditional distributions and regression functions. We will discuss this in the next section dealing with conditionally specified distributions below.

Examples: Korwar (1975), Dahiya and Korwar (1977), Cacoullos and Papageorgiou (1983), Papageorgiou (1983, 1984, 1985a), Kyriakoussis (1988), and Kyriakoussis and Papageorgiou (1989).

3.5 Conditionally Specified Method

Suppose in the preceding section, both $\Pr(Y = y|X = x)$ and $\Pr(X = x|Y = y)$ are given for all x and y. We may have then overspecified the conditions as the two conditional distributions may not be compatible. In cases in which compatibility is confirmed, the question of possible uniqueness of the compatible distribution need to be addressed. The book of Arnold *et al.* (1999) has revolutionized this subject area as it provides a rich mechanism for generating bivariate distributions. This book focuses on those conditional distributions that are members of some well-defined parametric families such as the exponential families. Three discrete distributions are from exponential families, that is, binomial, geometric, and Poisson. Section 4.12 of the above mentioned monograph devotes a discussion to constructions of bivariate binomial, geometric, and Poisson distributions.

Section 7.7 of Arnold *et al.* (1999) discusses generation of bivariate discrete distributions (as well as continuous bivariate distributions) for a given conditional distribution of X given Y and the regression function of Y on X. In particular, Wesolowski (1995) has shown that if $X|Y = y$ has a power series distribution, that is,

$$\Pr(X = x|Y = y) = c(x)y^x/c^*(y),$$

then the joint distribution of (X, Y) will be uniquely determined by the regression function of Y on X provided $c(\cdot)$ is reasonably well behaved.

3.6 Construction of Discrete Bivariate Distributions with Given Marginals and Correlation

3.6.1 Discrete Fréchet bounds

For given marginals F and G, Hoeffding (1940) and Fréchet (1951) have proved that there exist bivariate distribution functions, H_L and H_U, called the lower and upper Fréchet bounds, respectively, having minimum and maximum correlation. Specifically, we have

$$H_L(x,y) = \max[F(x) + G(y) - 1, 0] \tag{3.10}$$

$$H_U(x,y) = \min[F(x), G(y)] \tag{3.11}$$

satisfying

$$H_L(x,y) \leq H(x,y) \leq H_U(x,y) \tag{3.12}$$

and that

$$\rho_L \leq \rho \leq \rho_U \tag{3.13}$$

where ρ_L, ρ and ρ_U denote the Pearson product-moment correlation coefficients for H_L, H and H_U, respectively.

3.6.2 Probability functions of Fréchet bounds

We now assume that X and Y are discrete with ranges that are subsets of $N = \{0, 1, 2, \ldots\}$. Let h, f, and g be the probability functions that correspond to H, F, and G, respectively. Our aim now is to construct the probability functions h_L and h_U that correspond to H_L and H_U, respectively. In the following, we adopt the notations given in Nelsen (1987).

Let D denote the portion of N^2 where $H_L(x,y) > 0$, D' denote the complement of D in N^2, and ∂D denote the border of D; that is,

$$D = \{(x,y) \in N^2 \,|\, F(x) + G(y) - 1 > 0\}$$
$$D' = \{(x,y) \in N^2 \,|\, F(x) + G(y) - 1 = 0\},$$

and

$$\partial D = \{(x,y) \in D | (x-1,y), (x,y-1) \text{ or } (x-1,y-1) \notin D\}.$$

Nelsen (1987) has shown that

$$h_L(x,y) = \begin{cases} f(x) & (x,y) \in \partial D, (x,y-1) \notin D, (x-1,y) \in \partial D \\ g(y) & (x,y) \in \partial D, (x-1,y) \notin D, (x,y-1) \in \partial D \\ F(x) + G(y) - 1 & (x,y) \in \partial D, (x,y-1) \notin D, (x-1,y) \notin D \\ 1 - F(x-1) - G(y-1) & (x,y) \in \partial D, (x,y-1) \in \partial D, (x-1,y) \in \partial D \\ 0 & \text{otherwise.} \end{cases}$$

In order to obtain h_U, we set

$$S = \{(x, y) \in N^2 | F(x) = G(y)\},$$
$$T = \{(x, y) \in N^2 | F(x) > G(y)\},$$

and

$$\partial S = \{(x, y) \in S | (x, y - 1) \notin S\},$$
$$\partial T = \{(x, y) \in T | (x - 1, y) \notin T\}.$$

Nelsen (1987) has shown that

$$h_U(x, y) = \begin{cases} f(x) & (x, y) \in \partial S, (x-1, y-1) \in T, \text{ or } y = 0 \\ g(y) & (x, y) \in \partial S, (x-1, y-1) \in S, \text{ or } x = 0,\ y \neq 0 \\ F(x) + G(y) - 1 & (x, y) \in \partial T, (x-1, y-1) \in S, \text{ or } x = 0 \\ 1 - F(x-1) - G(y-1) & (x, y) \in \partial T, (x-1, y-1) \in T, \text{ or } y = 0,\ x \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The author has also presented two examples of finding h_L and another two of finding h_U.

3.6.3 Construction of bivariate distributions

Having obtained h_L and h_U, we are now in a position to generate one-parameter or two-parameter families of bivariate distributions with given marginals:

$$h_{\theta,\phi} = \theta h_L(x, y) + (1 - \theta - \phi) f(x) g(y) + \phi h_U(x, y), \theta, \phi \geq 0, \theta + \phi \leq 1. \quad (3.14)$$

Upon setting $\theta = 0, \phi > 0$, we obtain a one-parameter family with positive correlation; and upon setting $\phi = 0, \theta > 0$, a one-parameter family with negative correlation; and correlation coefficients for members of these families are functions of $\theta, \phi, \rho_{\mathrm{L}}$ and ρ_{U}.

Mardia (1970, p. 33) has noted that if we let $\theta^2 = \frac{\gamma^2}{2}(1-\gamma)$ and $\phi = \frac{\gamma^2}{2}(1+\gamma)$, then (3.14) becomes

$$h_\gamma = \frac{\gamma^2}{2}(1 - \gamma) h_L(x, y) + (1 - \gamma^2) f(x) g(y) + \frac{1}{2}\gamma^2(1 + \gamma) h_U(x, y) \quad (3.15)$$

It is worth noting that for $\phi = 0$,

$$h_\theta = \theta h_L(x, y) + (1 - \theta) f(x) g(y) \quad (3.16)$$

and that the correlation coefficient ρ is given by

$$\rho = \theta \rho_L, \ 0 \leq \theta \leq 1 \quad (3.17)$$

which has values between ρ_L and 0. Thus for any desired correlation ρ between ρ_L and 0, we can find the required value of θ in [0, 1] to satisfy (3.17).

Similarly, for $\theta = 0, \phi > 0$, we have

$$h_\phi = (1 - \phi)f(x)g(y) + \phi h_U(x, y) \tag{3.18}$$

and that the correlation coefficient ρ is given by

$$\rho = \phi\rho_U. \tag{3.19}$$

For any desired correlation between 0 and ρ_U, we can find the required value of ϕ in [0, 1].

Nelsen (1987) presented two examples:

1. both marginals are Poisson but with different parameters, $\rho = -0.5$ and
2. one marginal is binomial ($n = 4, p = 0.8$) and the other discrete uniform on $\{1, 2, 3, 4, 5\}$; and ρ positive.

If we wish to use Mardia's one-parameter family (3.15), then the correlation coefficient ρ for h_γ is given by

$$\rho = \frac{\gamma^2}{2}(1 - \gamma)\rho_L + \frac{\gamma^2}{2}(1 + \gamma)\rho_U.$$

To find the required value ρ between ρ_L and ρ_U, we need to solve for γ in the following cubic equation

$$(\rho_U - \rho_L)\gamma^3 + (\rho_U + \rho_L)\gamma^2 - 2\rho = 0.$$

Then, we can construct the probability function by substituting the value γ into (3.14).

3.6.4 Construction of bivariate Poisson distributions

Griffiths *et al.* (1979) gave procedures for constructing bivariate Poisson distributions having negative correlations when the two marginals are specified. For given Poisson marginals F and G having parameters λ_1 and λ_2, respectively, they calculated and tabulated the minimum and maximum correlation coefficients (i.e., the correlation coefficients of H_L and H_U defined, respectively, by (3.10) and (3.11)).

3.7 Sums and Limits of Bernoulli Trials

3.7.1 The bivariate Bernoulli distribution

Suppose (X, Y) has Bernoulli marginals; then it has only four possible values: $(1,1), (1,0), (0,1), (0,0)$ with probabilities $p_{11}, p_{10}, p_{01}, p_{00}$, respectively. The marginal probabilities are given by

$$\left.\begin{array}{l} p_{1+} = p_{11} + p_{10} = 1 - p_{0+}, \\ p_{+1} = p_{11} + p_{01} = 1 - p_{+0} \end{array}\right\}. \tag{3.20}$$

It is easy to show that the correlation coefficient is given by

$$\rho = \frac{p_{11} p_{1+} p_{+1}}{\sqrt{p_{1+} p_{0+} p_{+1} p_{+0}}}. \tag{3.21}$$

It takes on values -1 and +1 when $\rho_{11} = \rho_{00} = 0$ and $\rho_{10} = \rho_{01} = 0$, respectively. Here, $\rho = 0$ implies X and Y are independent.

3.7.2 Construction of bivariate Bernoulli distributions

It is well known that in the univariate case, the binomial, negative binomial (including geometric), hypergeometric and Poisson distributions are obtainable from the univariate Bernoulli distribution. Marshall and Olkin (1985) showed that these methods of derivation (using sums and limits) can be extended to twodimensions to obtain many bivariate distributions with binomial, negative binomial, geometric, hypergeometric, or Poisson marginals.

3.8 Sampling from Urn Models

Many discrete bivariate distributions are constructed by sampling from urn models. There are two types of sampling: (i) direct sampling and (ii) inverse sampling. By inverse sampling, we mean the sampling is continued until k individuals of a certain type are observed. For both types, sampling may be with or without replacement.

Suppose a population has three distinct characters and let the population size be N. Let N_i, $i = 0, 1, 2$, be the number of individuals having character i, for $i = 0, 1, 2$ such that $N_0 + N_1 + N_2 = N$ (alternatively, an urn contains N balls of three different colours, N_i being of i^{th} colour $(i = 0, 1, 2)$ such that $N_0 + N_1 + N_2 = N$). Suppose that n individuals (balls) are drawn from the population (urn) with various forms of sampling schemes, and let X and Y

Table 3.1: Bivariate distributions from direct and inverse samplings

No	Name	Type of Sampling	Replace (Yes/No)	Special Features
(i)	Bivariate Binomial	Direct	Yes	N_i finite
(ii)	Bivariate Negative Binomial	Inverse	Yes	N_i infinite
(iii)	Bivariate Hypergeometric	Direct	No	—
(iv)	Bivariate Inverse Hypergeometric	Inverse	No	—
(v)	Bivariate Negative Hypergeometric	Direct	—	Trinomial compounded by bivariate beta
(vi)	Bivariate Inverse Negative Hypergeometric	Inverse	—	Negative trinomial compounded by bivariate beta
(vii)	Bivariate Polya	Direct		Add c additional individuals
(viii)	Bivariate Inverse Polya	Inverse		Add c additional individuals

denote the number of type 1 character and type 2 character, respectively, in the sample. We can then construct various kinds of bivariate distributions which are summarized below:

- Distribution (i) is also known as type 1 bivariate binomial distribution; see, for example, Section 3.3 of Kocherlakota and Kocherlakota (1992).
- For distribution (ii), see, for example, Section 5.2 of Kocherlakota and Kocherlakota (1992).
- For distributions (iii)–(vi), see Janardan (1972, 1973, 1975, 1976), Janardan and Patil (1970, 1971, 1972). See also Chapter 6 of Kocherlakota and Kocherlakota (1992).
- For distributions (vii) and (viii), see Janardan and Patil (1970, 1971) and Patil *et al.* (1986).

For other references and other distributions generated from urn models, see Johnson and Kotz (1977), Korwar (1988), and Marshall and Olkin (1990).

3.9 Clustering (Bivariate Distributions of Order k)

In recent years, several bivariate generalizations of the binomial, negative binomial, hypergeometric, Poisson, logarithmic, and other distributions were obtained. These are often called bivariate distributions of order k or bivariate cluster distributions; see Balakrishnan and Koutras (2002). As they bear the names binomial, negative binomial, hypergeometric, and negative hypergeometric, it is not surprising that they also have the origin of sampling from an urn with and without replacements.

3.9.1 Preliminary

Consider an urn that contains balls of $k+1$ types such that α balls bear the number 0 and β_{i} balls bear the number $i, i = 1, 2, \ldots, k$.

(i) Suppose a sample of n balls is drawn with replacement. Let X denote the sum of the numbers shown on the balls drawn and p_i, $i = 1, 2, \ldots, k$ be the probability that a ball bearing the number i will be drawn: $\sum_{i=1}^{k} p_i = p$ and $q = 1 - p$ is the probability that a ball bearing a zero will be drawn. Then, X has a cluster binomial distribution.

(ii) If the sampling scheme above is without replacement, then a cluster hypergeometric distribution results.

(iii) If as in (i) above, but with n not fixed and letting X be the sum of numbers sampled before the r^{th} zero, then X has a cluster negative binomial distribution.

(iv) If as in (ii) above but the compositions of balls is to be altered at each stage by adding a ball of the same type as the sampled one before the next draw is made, then X has a cluster Polya distribution.

3.9.2 Bivariate Distributions of order k

Now we may generalize this idea to the bivariate case.

Suppose an urn contains balls of two different colours (say colour 1 and colour 2). The balls of colour i are numbered from 0 to k_i, $i = 1, 2$. n balls are drawn with replacement. Let p_{ij} denote the probability that a ball of colour i will bear number j, $j = 0, 1, 2, \ldots, k_i$. Let X and Y denote the sum of the numbers of the first and second colour, respectively; then (X, Y) has a cluster bivariate binomial distribution [Panaretos and Xekalaki (1986)].

Suppose now in the above example, $k_1 = k_2$ and another ball is added and labelled by (0, 0) with proportion p such that $p + \sum_{i=1}^{2} \sum_{j=1}^{k} p_{ij} = 1$. Balls are drawn with replacement until the r balls ($r \geq 1$) bearing the number (0, 0) appear. Let X and Y denote the sum of the numbers on colour 1 and colour 2, respectively. Then (X, Y) has the bivariate negative binomial distribution of order k [Philippou *et al.* (1989) and Antzoulakos and Philippou (1991)].

Philippou *et al.* (1989) obtained a bivariate Poisson distribution of order k by taking limits from the above model such that

$$p_{ij} \to 0 \quad \text{and} \quad rp_{ij} \to \lambda_{ij} \quad (0 < \lambda_{ij} < \infty, \ \text{for } 1 \leq i \leq 2, 1 \leq j \leq k).$$

For construction of bivariate logarithmic series distribution of order k, also a limiting case of bivariate negative binomial of order k, see Philippou *et al.* (1989, 1990). For constructions of bivariate Polya and inverse Polya distributions of order k, see Philippou and Tripsiannis (1991).

Aki and Hirano (1994, 1995) have constructed multivariate geometric distributions of order k. For a review on this subject, see Chapter 42 of Johnson *et al.* (1997) and Balakrishnan and Koutras (2002).

Philippou and Antzoulakos (1990) have obtained several bivariate distributions of order k through a "generalised sequence of order k" which was first introduced by Aki (1985). For other types of bivariate binomial distributions of order k, see Ling and Tai (1990).

3.10 Construction of Finite Bivariate Distributions via Extreme Points of Convex Sets

In this section, we consider the construction of bivariate distributions with finite support. The key reference for the following discussion is that of Rao and Subramanyam (1990).

Let $M(F, G)$ be the collection of all bivariate distributions with finite support and marginals F and G. Then M is a compact convex set. In order to give an insight of the problem, we begin by considering joint probabilities of X and Y: $p_{ij} = \Pr(X = i, Y = j)$, $p_i = \Pr(X = i)$, $q_j = \Pr(Y = j)$, $i, = 1, 2$; $j = 1, 2, 3$.

It is easy to see that the following set of equations hold (assuming for the time being that p_{11} and p_{12} are known):

$$\left.\begin{array}{lcl} p_{13} & = & p_1 - p_{11} - p_{12} \\ p_{21} & = & q_1 - p_{11} \\ p_{22} & = & q_2 - p_{12} \\ p_{13} + p_{23} & = & q_3 \end{array}\right\}. \tag{3.22}$$

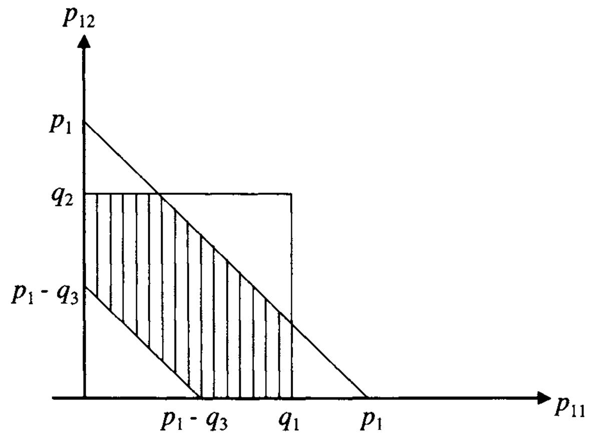

Figure 3.1: Feasible region

There are five equations and four unknowns. As $p_{ij} \geq 0$, it follows that

$$\left.\begin{array}{c} p_{13} = p_1 - p_{11} - p_{12} \geq 0 \\ p_{21} = q_1 - p_{11} \geq 0 \\ p_{22} = q_2 - p_{12} \geq 0 \\ p_{23} = q_3 - p_1 + p_{11} + p_{12} \geq 0 \end{array}\right\}. \tag{3.23}$$

The above may be expressed as four inequalities for p_{11} and p_{12}. These are

$$\left.\begin{array}{c} p_{11} + p_{12} \leq p_1 \\ p_{11} \leq q_1 \\ p_{12} \leq q_2 \\ p_{11} + p_{12} \geq p_1 - q_3 \end{array}\right\}. \tag{3.24}$$

In addition, we have two obvious inequalities which are

$$p_{11} \geq 0 \qquad \text{and} \qquad p_{12} \geq 0.$$

These six inequalities may be illustrated by the diagram above. The feasible region of bivariate distributions is a hexagon. However, if either q_1 or q_2 exceeds p_1, the region is then reduced to a pentagon. If both q_1 and q_2 exceed p_1, then the region is a quadrilateral. If both q_1 and q_2 are smaller or equal to $p_1 - q_3$, then the region is a triangle. If one of q_1 and q_2 is less than $p_1 - q_3$ whereas the other one exceeds $p_1 - q_3$, then the resulting region is a quadrilateral.

Note that the intersections of the boundary lines are the extreme points. In this example, there are three to six extremal points.

Well-established mathematical fact: Let $A_i (i = 1, 2, \ldots, n)$ be the extreme points of a compact convex set M. Then any element B of M can be written as $B = \sum_{i=1}^{n} \alpha_i A_i$ where $\sum_{i=1}^{n} \alpha_i = 1$.

It follows that we can generate a discrete bivariate distribution after the extreme points are identified.

3.10.1 Finding extreme points

From the above discussion, it is clear that it is easy to generate a bivariate distribution with specified marginals if we can identify the extremal points of M. For example, suppose we have $p_1 = \frac{1}{3}, p_2 = \frac{2}{3}; q_1 = \frac{1}{4}, q_2 = \frac{1}{2}, q_3 = \frac{1}{4}$. As $q_2 = \frac{1}{2} > p_1 = \frac{1}{3}$, the region is a pentagon. It follows from the above diagram that one of the intersections is $p_{11} = p_1 - q_3 = \frac{1}{12}, p_{12} = 0$. It follows from (3.22) and (3.23) that one of the extreme points of M is

$$\begin{bmatrix} \frac{1}{12} & 0 & \frac{1}{4} \\ \frac{1}{6} & \frac{1}{2} & 0 \end{bmatrix}.$$

The other four extreme points can be found similarly.

Let m be the number of $p_i > 0$ and n be the number of $q_i > 0$. This contingency table has $(m-1)(n-1)$ degrees of freedom. In general, we have $(m+n)$ equations and $(m+n-1)$ unknowns (i.e., one equation is always redundant). These $(m+n-1)$ equations are expressed in terms of p_i, q_i and the $(m-1)(n-1)$ free parameters (with dependent parameters on the left of the equations, and p_i, q_i and free parameters on the right). Also as the free parameters $p_{ij} \geq 0$, we therefore have $(m+n-1)+(m-1)(n-1) = mn$ inequalities. Hence they form a polygon with a maximum of mn sides.

Oluyede (1994) obtained a family of bivariate binomial distributions generated by extreme bivariate Bernoulli distributions.

3.11 Generalized Distributions

The adjective "generalized" has often been used for discrete distributions, however, its meaning is not uniquely defined. In the literature, there is no clear-cut discrimination between the terms "compound" and "generalized." Moreover, the word "generalized" in this discussion is also used with other meanings such as extension. For example, we used the term "generalized inverse Gaussian" in Section 3.3 to denote a distribution which includes the inverse Gaussian as its special case.

We now define "generalized" in a restricted sense.

Suppose the pgf (probability-generating function) of a distribution F_1 is $g_1(s)$. If the argument s is replaced by the pgf $g_2(s)$ of another distribution F_2, then the resulting generating function $g_1(g_2(s))$ is also a probability-generating function. This distribution is called a generalized F_1 distribution. More precisely, it is called an F_1 distribution generalized by the generalizer (or generalizing distribution) F_2. It may be written in the symbolic form

$$F_1 \vee F_2; \tag{3.25}$$

see Johnson and Kotz (1969, p. 202).

In the univariate case, the generalized distribution is simply a compound distribution.

3.11.1 Generalized bivariate distributions

The above idea may be extended to the bivariate case. In a general setting, there are at least two ways of "generalizing."

(i) Let $G(s)$ be the pgf of the original distribution F_1 and $\pi(s,t)$ be the joint pgf of the bivariate distribution of F_2. Then a generalized bivariate distribution can be obtained by replacing s of G by $\pi(s,t)$ to give

$$g(s,t) = G(\pi(s,t)). \tag{3.26}$$

(ii) Let $G(s,t)$ be the original pgf of a bivariate distribution F_1. Replace the arguments s and t of G by the univariate pgf's $\pi_1(s)$ and $\pi_2(t)$, respectively, so that the resulting generalized distribution has pgf

$$g(s,t) = G(\pi_1(s), \pi_2(t)). \tag{3.27}$$

(iii) The third way may be obtained by combining the trivariate reduction technique together with the "generalized" method. Let π_i be the pgf of the generalizer X_i and G_i be the pgf of the distribution that generalizes $X_i, i = 1, 2, 3$. Let $(X, Y) = (X_1+X_3, X_2+X_3)$. Then the resulting generalized bivariate distribution of (X, Y) has pgf given by

$$g(s,t) = G_1(\pi_1(s))G_2(\pi_2(t))G_3(\pi_3(st)). \tag{3.28}$$

3.11.2 Generalized bivariate Poisson distributions

(i) Bivariate Neyman type A distributions

Holgate (1966) constructed three types of bivariate Neyman A distributions.

Type I: This corresponds to (3.26) with G being the pgf of a Poisson and $\pi(s,t)$, the pgf of the bivariate Poisson given by

$$\pi(s,t) = \exp\{\lambda_1(s-1) + \lambda_2(t-1) + \lambda_3(st-1)\}. \tag{3.29}$$

Type II: This corresponds to (3.27) where G is the pgf of the bivariate Poisson given by (3.29) and $\pi_1(s) = \exp\{\phi_1(s-1)\}$ and $\pi_2(t) = \exp\{\phi_2(t-1)\}$.

Type III: This is obtained via the trivariate reduction method such that $X = X_1+X_3$ and $Y = X_2+X_3$ where X_i $(i = 1, 2, 3)$ are independent Neyman A distributions. Alternatively, let $G_i(s) = \exp\{\lambda_i(s-1)\}$ and $\pi_i(s) = \exp\{\phi_i(s-1)\}$. By applying (3.28), we obtain this distribution.

(ii) Bivariate Poisson binomial distributions

Charalambides and Papageorgiou (1981a) also derived three types of bivariate Poisson binomial distributions based on the "generalized" method.

3.11.3 Generalized bivariate general binomial distributions

Three types of bivariate generalized general binomials were derived by Charalambides and Papageorgiou (1981b).

For other examples, see Papageorgiou and Kemp (1983).

3.12 Canonical Correlation Coefficients and Semigroups

3.12.1 Diagonal expansion

The diagonal expansion of a bivariate distribution involves representing it as

$$dH(x,y) = dF(x)dG(y)\sum_{i=1}^{\infty}\rho_i\xi_i(x)\eta_j(y), \tag{3.30}$$

ξ_i and η_i being known as the canonical variables and the ρ_i as the canonical correlations.

When X and Y have finite moments of all orders, sets of orthonormal polynomials $\{P_n\}$ and $\{Q_n\}$ can be constructed with respect to F and G; for example, the Krawtchouk polynomials for binomial marginals, the Meixner polynomials for negative binomial marginals, and the Poisson–Charlier polynomials for Poisson marginals.

If

$$\left.\begin{array}{rl} E[X^n|Y=y] = & \text{a polynomial of degree } n \\ E[Y^n|X=x] = & \text{a polynomial of degree } n \end{array}\right\}, \tag{3.31}$$

then H has a diagonal expression in terms of F and G and their respective orthonormal polynomials.

3.12.2 Canonical correlation coefficients and positive definite sequence

Suppose now X and Y are two exchangeable variables so that the two sets of orthonormal polynomials $\{P_n\}$ and $\{Q_n\}$ are identical. A sequence $\{t_n\}$ is said to be positive definite with respect to $\{Q_n\}$ if for all M (integer), all $x = 0, 1, 2, \ldots$ and all sequences $\{a_n\}$ of real numbers, $\sum_{n=0}^{M} a_n Q_n(x)$ implies that $\sum_{n=0}^{M} a_n t_n Q_n(x)$. (We assume here $t_0 = 1$.)

For finite discrete bivariate distributions, Eagleson (1969) showed that every canonical sequence $\{\rho_n : \sum_{i=0}^{\infty} \rho_i^2 < \infty\}$ is a positive definite sequence. Griffiths (1970) generalized the result to the case when the support of X is unbounded.

3.12.3 Moment sequence and canonical correlation coefficient

A sequence $\{b_n\}$ is said to be a moment sequence if it can be expressed as $b_n = \int t^n dG(t)$ for some distribution function G. Assume again that the support of X is unbounded and X and Y are exchangeable. Tyan and Thomas (1975) showed that every sequence of canonical correlation coefficients is a moment sequence on $[0, 1]$ or $[-1, 1]$. If X is non-negative, then the moment sequence is defined on $[0, 1]$. Conversely, if $\{\rho_n = \rho^n\}$ is a sequence of canonical correlation coefficients, it is easy to show that every moment sequence is a sequence of canonical correlation coefficients. For the binomial and Poisson, the sequence $\{\rho^n\}$ is indeed a sequence of canonical correlation coefficients; see, for example, Lancaster (1983).

3.12.4 Constructions of bivariate distributions via canonical sequences

Let C denote the set of all sequences of canonical correlation coefficients.

- It is easy to see that C is convex. Hence, if $\{a_n\}$ and $\{b_n\}$ are two sequences of canonical correlation coefficients, then $\{\rho_n = \lambda a_n + (1-\lambda) b_n\}$ is also a sequence of canonical correlation coefficients for a new bivariate distribution having the same set of marginals.

- As positive definite sequences are closed under termwise multiplication, C forms a semigroup with respect to termwise multiplication. For finite discrete distribution, this result was proved by Vere-Jones (1971). Vere-Jones's result can be easily generalized to the case with unbounded support. In other words, $\{\rho_n = a_n b_n\}$ is a sequence of canonical correlation coefficients if $\{a_n\}$ and $\{b_n\}$ are. In this way, numerous bivariate distributions can be constructed.

3.13 Bivariate Distributions from Accident Models

In Section 20.3 and Section 21, Hutchinson and Lai (1990) considered the joint distribution of the severities of injury to two people in the same road accident. It was found that a bivariate normal distribution, generated by the method of variables in common, may be used to model such injury. Here, we are concerned with the number of injury accidents rather than the amount of injury.

Let X denote the number of injury accidents on a given stretch of highway and Z_i denote the number of fatalities in the i^{th} accident, $i = 1, 2, \ldots, X$. Also, let Y denote the total number of fatalities recorded among the X accidents. In other words, we may represent them in the following manner:

$$Y = Z_1 + Z_2 + \cdots + Z_X \tag{3.32}$$

The question of interest is to find the joint distribution of X and Y. Unlike the bivariate distributions we have discussed so far, the two marginals are, in general, of different types of univariate distributions.

Following the pioneering work of Edwards and Gurland (1961) in using a discrete bivariate distribution (i.e., a bivariate negative binomial) to model accident data, Leiter and Hamdan (1973), Cacoullos and Papageorgiou (1980, 1982) and others developed several models to represent the joint distribution of (X, Y) as specified in (3.32).

3.13.1 The Poisson-Poisson, Poisson-binomial, and Poisson-Bernoulli methods

Suppose X has a Poisson distribution. By letting Z_i (assuming they are i.i.d), we obtain

- Poisson-Bernoulli model when Z_i has a Bernoulli distribution [Leiter and Hamdan (1973)].
- Poisson-Binomial model when Z_i has a binomial distribution [Cacoullos and Papageorgiou (1980)].
- Poisson-Poisson model when Z_i has a Poisson distribution [Leiter and Hamdan (1973)].
- Poisson-geometric model when Z_i has a geometric distribution [Papageorgiou (1985b)].

3.13.2 Negative binomial-Poisson and negative binomial-Bernoulli models

It has been pointed out by many authors [see Kemp (1970)] that the number of accidents is more adequately described by a negative binomial (i.e., the Poisson distribution whose parameter λ has a gamma distribution). For this reason, Cacoullos and Papageorgiou (1982) constructed the following bivariate distribution assuming X to have a negative binomial distribution.

- Negative binomial-Poisson model where Z_i has a Poisson distribution.
- Negative binomial-Bernoulli models where Z has a Bernoulli distribution. The joint distribution of (X, Y) is a special case of the bivariate negative binomial of Edwards and Gurland (1961).

3.14 Bivariate Distributions Generated from Weight Functions

Let $f(x, y)$ be the probability function of (X, Y). Kocherlakota (1995), and Gupta and Tripathi (1996) defined the probability function of the weighted distribution with the weight function $W(x, y)$ as

$$h_W(x, y) = \frac{f(x, y)W(x, y)}{E[W(X, Y)]}.$$

In particular, they considered the multiplicative weight function of the form

$$W(x, y) = x^{(\alpha)}y^{(\beta)},$$

where $x^{(\alpha)} = x(x-1)\cdots(x-\alpha+1)$. The weighted bivariate Poisson, weighted bivariate binomial, weighted bivariate negative binomial, and weighted bivariate logarithmic series distributions were obtained by this method; see also Section 43.5 of Johnson *et al.* (1997) for other details.

3.15 Marginal Transformations Method

The marginal transformation method to generate a continuous bivariate distribution from another continuous bivariate distribution can be implemented easily. Suppose (X, Y) has a joint cumulative distribution function $H(x, y)$

with marginal $F(x)$ and $G(y)$. If we transform $X \to X^*$ and $Y \to Y^*$, then the joint distribution function of (X^*, Y^*) is given by

$$H^*(x^*, y^*) = H(F^{-1}[F^*(x^*)], G^{-1}[G^*(y^*)]), \tag{3.33}$$

where F^* and G^* are the distribution functions of X^* and Y^*, respectively. The key to this method lies on the fact that $U = F(X), V = G(Y)$ as well as $U' = F^*(X^*), V' = G^*(Y^*)$ are all uniformly distributed for continuous marginals. Thus, the method cannot be readily applied to construct discrete bivariate distributions as discrete random variables cannot be transformed into uniform random variables.

It appears that the method can be transportable if $H(x, y)$ is continuous, whereas X^* and Y^* are two discrete random variables with finite or countable values. Then, the $H^*(x^*, y^*)$ can be expressed as

$$H^*(x^*, y^*) = \int_{-\infty}^{F^{-1}(x^*)} \int_{-\infty}^{G^{-1}(y^*)} h(x, y)\, dxdy, \tag{3.34}$$

where $h(x, y)$ is the joint density function of (X, Y).

Van Ophem (1999) has constructed a discrete bivariate distribution in this manner assuming $h(x, y)$ to be the standard bivariate normal density function with correlation coefficient ρ. Lee (2001) derived the range of correlation coefficients of a discrete bivariate distribution and showed that the discrete bivariate distribution of Van Ophem (1999) has a flexible correlation coefficient.

3.16 Truncation Methods

Similar to its continuous counterpart, discrete bivariate distributions may be obtained through truncations. Truncations may be necessary where certain values are missing or may not be recorded in the data sets. Piperigou and Papageorgiou (2003) gave a unified treatment of three types of zero class truncation:

- The zero cell $(0, 0)$ is not recorded.
- The zero class for the variable X, $\{(0, y),\ y = 0, 1, \ldots\}$, is not recorded.
- The zero class for both X and Y, $\{(0, y),\ y = 0, 1, \ldots; (x, 0),\ x = 0, 1, \ldots\}$, is not recorded.

Using the probability-generating function approach, various properties of the truncated discrete bivariate distributions are then examined.

3.17 Construction of Positively Dependent Discrete Bivariate Distributions

There are various concepts of positive dependence for a bivariate distribution. We consider only two of these here.

A pair of random variables, X and Y, are said to be positively quadrant dependent (PQD) if the following inequality holds, that is, if

$$\Pr(X \le x, Y \le Y) \ge \Pr(X \le x)\Pr(Y \le y). \tag{3.35}$$

The variable Y is said to be positive regression dependent (PRD) on X if $\Pr(Y > x|X = x)$ is increasing in x for every y.

For other concepts of stochastic dependence, one may see, for example, Chapter 12 of Hutchinson and Lai (1990).

3.17.1 Positive quadrant dependent distributions

We shall begin with construction of a pair of PQD binary variables. A binary random variable may be used to indicate the state of a component (or a system) which is either functioning or not functioning. More specifically, we let the binary variable X_i denote the state of the ith component such that

$$X_i = \begin{cases} 1 & \text{if it is functioning} \\ 0 & \text{otherwise.} \end{cases} \tag{3.36}$$

Then, $\Pr(X_i = 1)$ is the static reliability of the component at a given time instant.

Suppose X and Y are two identically distributed binary random variables having the joint probability function given as follows:

$$\Pr(X = 0) = a + b, \quad \Pr(X = 1) = 1 - a - b$$

and

$$\Pr(Y = 0) = a + b, \quad \Pr(Y = 1) = 1 - a - b.$$

Table 3.2: Joint probabilities

$\Pr(X = 0, Y = 0) = a$	$\Pr(X = 0, Y = 1) = b$
$\Pr(X = 1, Y = 0) = b$	$\Pr(X = 1, Y = 1) = 1 - a - 2b$

We now proceed to construct a pair of PQD binary variables as follows:

Clearly, for $(x, y) = (0,1), (1,0)$, or $(1,1)$, inequality (3.35) readily holds without requiring any condition. Thus, the binary pair X and Y are positively quadrant dependent if and only if

$$\Pr(X = 0, Y = 0) \geq \Pr(X = 0)\Pr(Y = 0) \tag{3.37}$$

which is equivalent to the condition

$$(a + b)^2 \leq a. \tag{3.38}$$

It is clear that for a given b, $0 < b < 1$, we can solve for a so that (3.38) holds. It is easy to show that

$$0 \leq \frac{(1 - 2b) - \sqrt{1 - 4b}}{2} < a < \frac{(1 - 2b) + \sqrt{1 - 4b}}{2}. \tag{3.39}$$

Now, let X and Y be two discrete non-negative integer valued random variables with $\Pr(X = i, Y = j) = p_{ij}$, $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$.
Holzsager (1996) has proved that if

$$p_{i+1\,j+1}\Pr(X \leq i, Y \leq j) \geq \Pr(X \leq i, Y = j+1)\Pr(X = i+1, Y \leq j), \tag{3.40}$$

then X and Y are PQD. Thus, (3.40) provides a mechanism to construct a pair of discrete PQD random variables.

Rao and Subramanyam (1990) provided a mechanism to identify the extreme points of the set of all discrete PQD bivariate distributions when the marginal distributions have finite support. It is easy to see that we can utilize this idea to generate PQD discrete distributions with finite marginals.

3.17.2 Positive regression dependent distributions

Subramanyam and Rao and (1996) also provided an algorithm to identify the extreme points of the set of all discrete PRD bivariate distributions when the marginal distributions have finite support. After identifying these points, positive regression dependent discrete bivariate distributions can be constructed.

References

1. Aki, S. (1985). Discrete distributions of order k on a binary sequence, *Annals of the Institute of Statistical Mathematics*, **37**, 205–224.

2. Aki, S., and Hirano, K. (1994). Distributions of number of failures and successes until the first k consecutive successes, *Annals of the Institute of Statistical Mathematics*, **46**, 193–202.

3. Aki, S., and Hirano, K. (1995). Joint distributions of numbers of success-runs and failures until the first k consecutive successes, *Annals of the Institute of Statistical Mathematics*, **47**, 225–235.

4. Antzoulakos, D. L., and Philippou, A. N. (1991). A note on multivariate negative binomial distribution of order k, *Communications in Statistics—Theory and Methods*, **20**, 1389–1399.

5. Arnold, B. C., Castillo, E., and Sarabia, J. M. (1999). *Conditional Specification of Statistical Methods*, Springer-Verlag, New York.

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

7. Balakrishnan, N. (2005). Discrete multivariate distributions, In *Encyclopedia of Statistical Sciences*, 2nd ed., (Eds., N. Balakrishnan, C. Read, and B. Vidakovic), John Wiley & Sons, Hoboken, NJ (to appear).

8. Balakrishnan, N., and Koutras, M. V. (2002). *Runs and Scans with Applications*, John Wiley & Sons, New York.

9. Bates, G. E., and Neyman, J. (1952). Contribution to the theory of accident proneness I, *University of California Publications in Statistics*, **1**, 215–254.

10. Blischke, W. R. (1978). Mixtures of distributions, *International Encyclopedia of Statistics*, Vol 1, 174-179.

11. Cacoullos, T., and Papageorgiou, H. (1980). On some bivariate probability models applicable to traffic accidents and fatalities, *International Statistical Review*, **48**, 345–356.

12. Cacoullos, T., and Papageorgiou, H. (1982). Bivariate negative binomial-Poisson and negative binomial-Bernoulli models with an application to accident data. In *Statistics and Probability: Essays in Honor of C. R. Rao* (Eds., G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh), pp. 155–168, North-Holland, Amsterdam.

13. Cacoullos, T., and Papagergiou, H. (1983). Characterisations of discrete distributions by a conditional distribution and a regression function, *Annals of the Institute of Statistical Mathematics*, **35**, 95–104.

14. Charalambides, Ch. A., and Papageroiou, H. (1981a). Bivariate Poisson binomial distributions, *Biometrical Journal*, **23**, 437–450.

15. Charalambides, Ch. A., and Papageorgiou, H. (1981b). On bivariate generalised binomial and negative binomial distributions, *Metrika*, **28**, 83–92.

16. Consael, R. (1952). Sur les processus composes de Poisson a deux variables aleatories, *Academie Royale de Belgigue, Classe des Sciences, Memoires*, **27**, 4–43.

17. Dahiya, R. C., and Korwar, R. M. (1977). On characterising some bivariate discrete distributions by linear regression, *Sankhyā, Series A*, **39**, 124–129.

18. David, K. M., and Papageorgiou, H. (1994). On compounded bivariate Poisson distributions, *Naval Research Logistics*, **41**, 203–214.

19. Eagleson, G. K. (1969). A characterization theorem for positive definite sequences on the Krawtchouk polynomials, *Australian Journal of Statistics*, **11**, 29–38.

20. Edwards, C. B., and Gurland, J. (1961). A class of distributions applicable to accidents, *Journal of the American Statistical Association*, **56**, 503–517.

21. Fréchet, M. (1951). Sur le tableaux de correlation dont les marges sont donness, *Annales de l'Universite de Lyon, Serie 3*, **14**, 53–77.

22. Gelman, A., and Speed, T. P. (1993). Characterizing a joint probability distributions by conditionals, *Journal of the Royal Statistical Society, Series B*, **55**, 185–188.

23. Goodman, L. A., and Kruskal, W. H. (1959). Measures of association for cross classifications: II, Further discussion and references, *Journal of the American Statistical Association*, **54**, 123–163.

24. Griffiths, R. C. (1970). Positive definite sequences and canonical correlation coefficients, *Australian Journal of Statistics*, **12**, 162–165.

25. Griffiths, R. C., Milne, R. K., and Wood, R. (1979). Aspects of correlation in bivariate Poisson distributions and processes, *Australian Journal of Statistics*, **21**, 238–255.

26. Gupta, R. C., and Tripathi, R. C. (1996). Weighted bivariate logarithmic series distributions, *Communications in Statistics—Theory and Methods*, **25**, 2517–2539.

27. Hoeffding, W. (1940). Masstabinvariate Korrelations-theorie, *Schriffen des Mathematischen Instituts und des Instituts fur Angewandte Mathematik der Universitat Berlin*, **5**, 179–233.

28. Holgate, P. (1966). Bivariate generalisations of Neyman's Type A distribution, *Biometrika*, **53**, 241–244.

29. Holzsager, R. (1996). Positive quadrant dependent random variables, *American Mathematical Monthly*, **103**, 350–351.

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

31. Janardan, K. G. (1972). A unified approach for a class of multivariate hypergeometric models, *Sankhyā, Series A*, **35**, 363–376.

32. Janardan, K. G. (1973). Chance mechanisms for multivariate hypergeometric models, *Sankhyā, Series A*, **35**, 465–478.

33. Janardan, K. G. (1975). Certain inference problems for multivariate hypergeometric models, *Communications in Statistics*, **4**, 375–388.

34. Janardan, K. G. (1976). Certain estimation problems for multivariate hypergeometric models, *Annals of the Institute of Statistical Mathematics*, **28**, 429–444.

35. Janardan, K. G., and Patil, G. P. (1970). On the multivariate Polya distribution: a model of contagion for data with multiple counts, In *Random Counts in Scientific Work* (Ed., G. P. Patil), Vol. 3, pp. 143–162, The Pennsylvania State University Press, University Park, PA.

36. Janardan, K. G., and Patil, G. P. (1971). The multivariate inverse Polya distribution: a model of contagion for data with multiple counts on inverse sampling, *Studi di Probabilita Statistica e Ricera Operative in Onore de G. Pompilj, Toreno*, 327–341.

37. Janardan, K. G., and Patil, G. P. (1972). A unified approach for a class of multivariate hypergeometric models, *Sankhyā, Series A*, **34**, 363–376.

38. Johnson, N. L. and Kotz, S. (1969). *Distributions in Statistics: Discrete Distributions*, Houghton Mifflin, Boston.

39. Johnson, N. L., and Kotz, S. (1977). *Urn Models and Their Applications*, John Wiley & Sons, New York.

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

41. Kemp, C. D. (1970). Accident proneness and discrete distribution theory, In *Random Counts in Scientific Work* (Ed., G. P. Patil), pp. 41–64, The Pennsylvania University Press, Philadelphia, PA.

42. Kemp, C. D., and Papageorgiou, H. (1982). Bivariate Hermite distributions, *Sankhyā, Series A*, **44**, 269–280.

43. Kocherlakota, S. (1988). On the compounded bivariate Poisson distribution: a unified approach, *Annals of the Institute of Statistical Mathematics*, **40**, 61–76.

44. Kocherlakota, S. (1995). Discrete bivariate weighted distributions under multiplicative weight function, *Communications in Statistics—Theory and Methods*, **25**, 533–551.

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

46. Kocherlakota, S., and Kocherlakota, K (1998). *Bivariate Discrete Distributions.* In *Encyclopedia of Statistical Sciences* (Eds., S. Kotz, C. B. Read and D. L. Banks), Update Vol. 2, pp. 68–83, John Wiley & Sons, New York.

47. Korwar, R. M. (1975). On characterizing some discrete distributions by linear regression, *Communication in Statistics*, **4**, 1133-1147.

48. Korwar, R. M. (1988). On the observed number of classes from multivariate distributions, *Sankhyā, Series B*, **50**, 39–59.

49. Kyriakoussis, A. (1988). Characterizations of bivariate discrete distributions, *Sankhyā, Series A*, **50**, 286–287.

50. Kyriakoussis, A., and Papageorgiou, H. (1989). On characterization of power series distribution by a marginal distribution and a regression function, *Annals of the Institute of Statistical Mathematics*, **41**, 671–676.

51. Lai, C. D. (1995). Construction of bivariate distributions by a generalised trivariate reduction technique, *Statistics & Probability Letters*, **25**, 265–270.

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

53. Lancaster, H. O. (1983). Special joint distributions of Meixner variables, *Australian Journal of Statistics*, **25**, 298–309.

54. Lee, L. F. (2001). On the range of correlation coefficients of bivariate ordered discrete random variables, *Econometric Theory*, **17**, 247–256.

55. Leiter, R. E., and Hamdan, M. A. (1973). Some bivariate probability models applicable to traffic accidents and fatalities, *International Statistical Review*, **41**, 87-100.

56. Ling, K. D., and Tai, T. H. (1990). On bivariate binomial distributions of order k, *Soochow Journal of Mathematics*, **16**, 211–220.

57. Marida, K. V. (1970). *Families of Bivariate Distributions*, Charles Griffin, London.

58. Marshall, A. W., and Olkin, I. (1985). A family of bivariate distributions generated by the bivariate Bernoulli distribution, *Journal of the American Statistical Association*, **80**, 332–338.

59. Marshall, A. W., and Olkin, I. (1990). Bivariate distributions generated from Polya-Eggenberger urn models, *Journal of Multivariate Analysis*, **35**, 48–65.

60. Nelsen, R. B. (1987). Discrete bivariate distributions with given marginals and correlation. *Communications in Statistics—Simulation and Computation*, **16**, 199–208.

61. Oluyede, B. O. (1994). A family of bivariate binomial distributions generated by extreme Bernoulli distributions, *Communications in Statistics—Theory and Methods*, **23**, 1531–1547.

62. Panaretos, J., and Xekalaki, E. (1986). On generalised binomial and multinomial distributions and their applications to generalised Poisson distributions, *Annals of the Institute of Statistical Mathematics*, **38**, 223–231.

63. Papageorgiou, H. (1983). On characterizing some bivariate discrete distributions, *Australian Journal of Statistics*, **25**, 136–144.

64. Papageorgiou, H. (1984). Characterizations of multinomials and negative multinomial mixtures by regression, *Australian Journal of Statistics*, **26**, 25-29.

65. Papageorgiou, H. (1985a). On characterizing some discrete distributions by a conditional distribution and a regression function, *Biometrical Journal*, **27**, 473–479.

66. Papageorgiou, H. (1985b). On a bivariate Poisson-geometric distribution. *Zastowania Mathematyki*, **18**, 541–547.

67. Papageorgiou, H. (1997). Multivariate discrete distributions, In *Encyclopedia of Statistical Sciences* (Eds., S. Kotz, C. B. Read and D. L. Banks), Update Vol. 1, pp. 408–419, John Wiley & Sons, New York.

68. Papageorgiou, H., and David, K. M. (1994). On the countable mixture of bivariate binomial distributions, *Biometrical Journal*, **36**, 581–601.

69. Papageorgiou, H., and Kemp, C. D. (1983). Conditionality in bivariate generalized distributions, *Biometrical Journal*, **25**, 757–763.

70. Patil, G. P. (1965). On a characterization of multivariate distribution by a set of its conditional distributions, In *Handbook of 35th International Statistical Institute Conference in Belgrade*, International Statistical Institute.

71. Patil, G. P. (1986). Polya distribution, multivariate, In *Encyclopedia of Statistical Sciences* (Eds., S. Kotz and N. L. Johnson), Vol. 7, pp. 59–63, John Wiley & Sons, New York.

72. Patil, G. P., Rao, C. R., and Ratnaparkhi, M. V. (1986). On discrete weighted distributions and their use in model choice for observed data, *Communications in Statistics—Theory and Methods*, **15**, 907–918.

73. Philippou, A. N., and Antzoulakos, D. L. (1990). Multivariate distributions of order k on a generalised sequence, *Statistics & Probability Letters*, **9**, 453–463.

74. Philippou, A. N., Antzoulakos, D. L., and Tripsiannis, G. A. (1989). Multivariate distributions of order k, *Statistics & Probability Letters*, **7**, 207–216.

75. Philippou, A. N., Antzoulakos, D. L., and Tripsiannis, G. A. (1990). Multivariate distributions of order k, Part II, *Statistics & Probability Letters*, **10**, 29–35.

76. Philippou, A. N., and Tripsiannis, G. K. (1991). Multivariate Polya and inverse Polya distributions of order k, *Biometrical Journal,* **33**, 225–236.

77. Piperigou, V. E., and Papageorgiou, H. (2003). On truncated bivariate discrete distributions: A unified treatment, *Metrika*, **58**, 221–233.

78. Rao, M. B., and Subramanyam, K. (1990). The structure of some classes of bivariate distributions and some applications, *Computational Statistics & Data Analysis,* **10**, 175–187.

79. Stein, G. Z., and Juritz, J. M. (1987). Bivariate compound distributions, *Communications in Statistics—Theory and Methods*, **16**, 3591–3607.

80. Strandskov, H. H., and Edelen, E. W. (1946). Monozygotic and dizygotic twin birth frequencies in the total of the "white" and the "coloured" US population, *Genetics*, **31**, 438–446.

81. Subramanyam, K., and Rao, M. B. (1996). Analysis of regression dependence in $2 \times n$ bivariate distributions and some applications in contingency tables, *Mathematics and Applications*, Vol. 359, pp. 385–400, Kluwer Academic Publishers, Dordrecht.

82. Tyan, S., and Thomas, J. B. (1975). Characterization of a class of bivariate distribution functions, *Journal of Multivariate Analysis*, **5**, 227–235.

83. Van Ophem, H. (1999). A general method to estimate correlated discrete random variables, *Econometric Theory*, **15**, 228–237.

84. Vere-Jones, D. (1971). Finite bivariate distributions and semigroups of non-negative matrices, *The Quarterly Journal of Mathematics*, **22**, 247–270.

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

86. Zheng, Q., and Matis, J. H. (1993). Approximating discrete multivariate distributions from known moments, *Communications in Statistics—Theory and Methods*, **22**, 3553–3567.