CHANCE MECHANISMS UNDERLYING MULTIVARIATE DISTRIBUTIONS

by
C. TAILLIE, J. K. ORD, J. E. MOSIMANN, and G. P. PATIL

Reprinted from
STATISTICAL DISTRIBUTIONS IN ECOLOGICAL WORK
International Co-operative Publishing House
Fairland, Maryland

pp. 157-191

A publication from the
Satellite Program in Statistical Ecology
International Statistical Ecology Program

Statistical Ecology Series
Volume 4

International Co-operative Publishing House
P. O. Box 245
Burtonsville, Maryland 20730
Statistical Ecology Series

General Editor: G. P. Patil

Volume 1* SPATIAL PATTERNS AND STATISTICAL DISTRIBUTIONS

Volume 2* SAMPLING AND MODELING BIOLOGICAL POPULATIONS AND POPULATION DYNAMICS

Volume 3* MANY SPECIES POPULATIONS, ECOSYSTEMS, AND SYSTEMS ANALYSIS

Volume 4 STATISTICAL DISTRIBUTIONS IN ECOLOGICAL WORK

Volume 5 SAMPLING BIOLOGICAL POPULATIONS

Volume 6 ECOLOGICAL DIVERSITY IN THEORY AND PRACTICE

Volume 7 MULTIVARIATE METHODS IN ECOLOGICAL WORK

Volume 8 SPATIAL AND TEMPORAL ANALYSIS IN ECOLOGY
edited by R. M. Cormack and J. K. Ord (1979)

Volume 9 SYSTEMS ANALYSIS OF ECOSYSTEMS
edited by G. S. Innis and R. V. O’Neill (1979)

Volume 10 COMPARTMENTAL ANALYSIS OF ECOSYSTEM MODELS
edited by J. H. Matis, B. C. Patten, and G. C. White (1979)

Volume 11 ENVIRONMENTAL BIOMONITORING, ASSESSMENT, PREDICTION, AND MANAGEMENT — CERTAIN CASE STUDIES AND RELATED QUANTITATIVE ISSUES

Volume 12 CONTEMPORARY QUANTITATIVE ECOLGY AND RELATED ECOMETRICS
edited by G. P. Patil and M. L. Rosenzweig (1979)

Volume 13 QUANTITATIVE POPULATION DYNAMICS
edited by D. G. Chapman and V. Gallucci (1979)

*For these first three volumes, contact:
The Pennsylvania State University Press
University Park, PA 16802 USA

For all of the remaining volumes, contact:
International Co-operative Publishing House
P.O. Box 245
Burtonsville, MD 20730 USA
CHANCE MECHANISMS UNDERLYING MULTIVARIATE DISTRIBUTIONS

C. TAILLIE

Department of Statistics
The Pennsylvania State University
University Park, Pennsylvania 16802 USA

J. K. ORD
Department of Statistics
University of Warwick
Coventry CV4 7AL, ENGLAND

J. E. MOSIMANN
Division of Computer Research and Technology
National Institutes of Health
Bethesda, Maryland 20014 USA

G. P. PATIL
Department of Statistics
The Pennsylvania State University
University Park, Pennsylvania 16802 USA

SUMMARY. Many ecological studies deal with variables which vary simultaneously, and which are not readily, nor even desirably, subject to the control of the investigator. The statistical analysis of such data, whether counts or measurements, requires multivariate models. The problem of finding good families of multivariate distributions is both a challenging and vexing one. An ideal multivariate family should have enough flexibility to describe widely different patterns of covariation. In addition, it should have parameters which are easily interpreted in relation to that covariation, and it should permit many and relative simple inferential procedures for these parameters.

This paper is divided into two parts. The first treats multivariate discrete distributions, while the second treats continuous distributions.
A number of distributions for counts are discussed. Illustrative data include pollen and diatom counts. We give particular attention to chance mechanisms and to covariance properties.

None of the classical discrete distribution families encompasses a wide variety of patterns for covariation. However, two families of continuous distributions do. The first is the multivariate normal. The second is the multivariate lognormal. The latter is for positive variable only, but permits a wide range of inferential procedures. Some applications are considered.

KEY WORDS. multivariate distributions, inverse sampling, covariation, mixtures, discrimination, chance mechanisms.

TABLE OF CONTENTS

CHAPTER 1: DISCRETE MULTIVARIATE DISTRIBUTIONS

1.1 Introduction ...................................... 159
1.2 Multinomial and Multinomial–Dirichlet Distributions 160
  1.2.1 The Multinomial as a Conditional Model .... 161
  1.2.2 An Example .................................... 162
  1.2.3 The Multinomial–Dirichlet ...................... 164
1.3 Negative Multinomial and Negative Multinomial–Dirichlet Distributions .................................. 166
  1.3.1 The Negative Multinomial as a Mixture .... 167
  1.3.2 The Negative Multinomial–Dirichlet ........ 168
  1.3.3 An Example .................................... 170
1.4 Bivariate Logarithmic Series Distribution .......... 170
  1.4.1 An Example .................................... 172
1.5 Multivariate Hypergeometric and Negative Hypergeometric Distributions ................................. 175
  1.5.1 Inverse Sampling ............................... 176
1.6 Further Reading .................................. 177

CHAPTER 2: CONTINUOUS MULTIVARIATE DISTRIBUTIONS

2.1 Introduction ...................................... 179
2.2 Multivariate Normal Distribution .................. 179
2.3 Multivariate Lognormal Distribution .......... 182
  2.3.1 Choosing the Normal or Lognormal in Practice 186
2.4 Dirichlet Distribution ............................ 188

REFERENCES ........................................ 189
CHAPTER 1

DISCRETE MULTIVARIATE DISTRIBUTIONS

1.1 INTRODUCTION

Many of the distributions considered in Boswell, Ord, and Patil (1979) were derived on the basis of sampling considerations. Despite their origins as sampling models, these distributions, together with their limiting and compound forms, are quite flexible and adequately describe a wide variety of ecological data.

In this chapter we turn our attention to discrete multivariate data and we would like to have families of distributions which are capable of describing wide patterns of covariation among the variables. We will consider simple sampling schemes and derive multivariate analogues of the familiar univariate distributions. However, as we shall see, the parameters of these distributions are completely determined by their univariate marginals. This implies that the joint distribution is also determined by the marginals. In other words, these distributions have no flexibility whatsoever with regard to covariation.

The ideal multivariate family should have three properties: (1) the ability to describe wide patterns of covariation; (2) parameters that are easily interpreted in relation to that covariation; and (3) a parametric structure permitting many and relatively simple inferential procedures*. It seems safe to state that, at present, no known family of multivariate discrete distributions enjoys all three properties.

Nevertheless, the distributions described in this chapter are useful, in appropriate circumstances, as pure sampling models and as null models of the kind considered in Sections 1.2 and 1.3.

*All applied statisticians will immediately recognize which family is being described.
1.2 MULTINOMIAL AND MULTINOMIAL-DIRICHLET DISTRIBUTIONS

In this section we consider multivariate analogues of the binomial and the binomial-beta distributions.

Recall the sampling theory derivation of the binomial. A population is partitioned into two categories, nominally called 'success' and 'failure,' and a sample of size \( n \) is drawn with replacement. The number of successes in the sample then has a binomial distribution with parameters \( n \) and \( p \) where \( p \) is the population proportion of successes.

Now, instead of two categories, suppose the population is partitioned into \( k \) categories with respective proportions \( p_1, p_2, \ldots, p_k \) where \( \sum p_i = 1 \). Once again a sample of size \( n \) is drawn with replacement and we let \( X_i \) be the number of times that the \( i \)th category is represented in the sample. The vector \( \mathbf{X} = (X_1, X_2, \ldots, X_k) \) is then said to have a multinomial distribution with parameters \( n \) and \( \mathbf{p} = (p_1, p_2, \ldots, p_k) \). It may be noted that, due to the constraint \( X_1 + X_2 + \cdots + X_k = n \), one of the components of \( \mathbf{X} \) could be suppressed as redundant.

Many of the properties of the multinomial are immediate from the above description even without writing down the probability function. Thus, if some of the categories are combined, the resulting distribution is still multinomial. For example, \( (X_1+X_2, X_3+X_4+X_5, X_6, X_7, \ldots, X_k) \) is multinomial with parameters \( n \) and \( (p_1+p_2, p_3+p_4+p_5, p_6, p_7, \ldots, p_k) \). Similarly, each marginal \( X_i \) is binomial with parameters \( n \) and \( p_i \). In particular,

\[ E[X_i] = np_i \quad \text{and} \quad \text{Var}(X_i) = np_i(1-p_i). \]

Also, covariances and correlations can be calculated using the formula \( 2\text{Cov}(X_i, X_j) = \text{Var}(X_i + X_j) - \text{Var}(X_i) - \text{Var}(X_j) \). Thus, when \( i \neq j \),

\[ \text{Cov}(X_i, X_j) = -np_i p_j \quad \text{and} \quad \text{Corr}(X_i, X_j) = -\frac{p_i p_j}{\sqrt{(1-p_i)(1-p_j)}}. \]

Notice that the correlation depends only on the marginal expectations. More generally, the marginals completely determine the parameters \( n, p_1, p_2, \ldots, p_k \) and hence the joint distribution. The multinomial is therefore quite inflexible in its ability to describe patterns of covariation.
In the next section, we derive the multinomial probability function by conditioning upon the total in a multiple Poisson distribution. The derivation has the minor advantage of avoiding combinatorics, but, more importantly, shows that the dependence structure in the multinomial is due entirely to the constraint $X_1+X_2+\cdots+X_k = n$.

1.2.1 The Multinomial as a Conditional Model. Light traps are frequently used to capture certain flying insects such as moths. Typically the light trap is operated for a fixed time period and the analysis carried out conditional upon the total number of moths caught. The procedure may be formulated as follows: Suppose that there are $k$ species in the vicinity of the trap, with abundances $\lambda_1, \lambda_2, \ldots, \lambda_k$, and that the number of specimens $X_j$ of the $j$th species caught in time $t$ follows a Poisson distribution with mean $\lambda_jt$; that is

$$
\Pr(X_j = x_j) = e^{-\lambda_j t} \frac{x_j^{x_j}}{x_j!} \quad (x_j = 0, 1, 2, \ldots).
$$

Let $n = x_1 + x_2 + \cdots + x_k$ denote the total number of specimens captured. If the different species are trapped independently, then the joint conditional distribution is

$$
\Pr(X_1 = x_1, \ldots, X_k = x_k \mid \Sigma X_j = n) = \frac{e^{-\lambda t} \prod_{j=1}^k \frac{x_j^{x_j}}{x_j!}}{\prod_{j=1}^k \frac{x_j^{x_j}}{x_j!}} \cdot \frac{\lambda_1^{x_1} \lambda_2^{x_2} \cdots \lambda_k^{x_k}}{x_1! x_2! \cdots x_k!},
$$

where $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_k$ (note that $\Sigma X_j$ has a Poisson distribution with mean $\lambda t$). The right hand side of this equation simplifies to

$$
\left( \frac{n!}{x_1! x_2! \cdots x_k!} \right) \cdot \frac{\lambda_1^{x_1} \lambda_2^{x_2} \cdots \lambda_k^{x_k}}{p_1 \cdot p_2 \cdot \cdots \cdot p_k},
$$

where $p_j = \lambda_j / \lambda$ is the proportional abundance of the $j$th species.

This is the probability function of the multinomial distribution with parameters $n$ and $(p_1, p_2, \ldots, p_k)$. The expression involving factorials is called a multinomial coefficient and is often written as
\[
\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1! x_2! \cdots x_k!},
\]

where \( n = x_1 + x_2 + \cdots + x_k \).

1.2.2 An Example. Martin and Mosimann (1965) record pollen counts from a sediment core from the dry Willcox Playa of Lake Cochise in southeastern Arizona. Twelve categories of pollen were counted until 200 pollen grains were recorded for each of 78 levels in the core. Here we consider only the bottom 20 levels of core slug 96 and the 12 levels of core slug 97. These data are presented in Table 1.

It is reasonable to regard the counts for any given level as following a multinomial distribution with parameters \( n = 200 \) and \( \mathbf{p} = (p_1, p_2, \ldots, p_{12}) \). Primary interest will focus on how the vector of proportions \( \mathbf{p} \) varies with the depth, since it is from this variation that one attempts a reconstruction of past changes in vegetation. First it is useful to consider the biologically null situation of no variation, that is we will test the null hypothesis \( H_0 : \mathbf{p} \) is constant across depth. This is known as a hypothesis of homogeneity for which formal significance tests are available (c.f. Potthoff and Whittinghill, 1966). For simplicity, let us examine only the pine (\textit{pinus}) counts. If \( H_0 \) were true, then the counts for pine, recorded in the second column of Table 1, would be a random sample of size 32 from a binomial distribution with parameters \( n = 200 \) and \( p_1 \). Now \( p_1 \) may be estimated as \( \hat{p}_1 = 0.596 \) (i.e., the sum of the counts (3812) divided by 32 \( \times \) 200) and we may test \( H_0 \) by comparing the sample variance with the variance predicted by the binomial distribution. These are

\[
\text{sample variance} = s^2 = 288.0
\]
\[
\text{predicted variance} = np_1(1-p_1) = 48.2.
\]

Under \( H_0 \), the ratio of the sample sum of squares to the predicted variance is distributed approximately as \( \chi^2(31) \) so that the observed ratio of about 185 is highly significant, leading us to reject the hypothesis of homogeneity.

Mosimann (1970) has described a second biologically null model for the depth to depth variation in the pollen profile. In
<table>
<thead>
<tr>
<th>Depth, 20 meters</th>
<th>P</th>
<th>Q</th>
<th>J</th>
<th>PAP</th>
<th>G</th>
<th>CA</th>
<th>S</th>
<th>A</th>
<th>HSC</th>
<th>LSC</th>
<th>OA</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>.015</td>
<td>103</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>32</td>
<td>9</td>
<td>0</td>
<td>35</td>
<td>2</td>
<td>13</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>.022</td>
<td>112</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>35</td>
<td>11</td>
<td>0</td>
<td>15</td>
<td>3</td>
<td>22</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>.028</td>
<td>108</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>29</td>
<td>15</td>
<td>0</td>
<td>24</td>
<td>2</td>
<td>17</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.034</td>
<td>118</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>30</td>
<td>8</td>
<td>0</td>
<td>17</td>
<td>6</td>
<td>12</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>.041</td>
<td>131</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>30</td>
<td>7</td>
<td>0</td>
<td>15</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>.047</td>
<td>129</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>13</td>
<td>0</td>
<td>20</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.053</td>
<td>132</td>
<td>3</td>
<td>1</td>
<td>29</td>
<td>2</td>
<td>0</td>
<td>20</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>.060</td>
<td>134</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>28</td>
<td>6</td>
<td>0</td>
<td>20</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>.066</td>
<td>120</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>32</td>
<td>2</td>
<td>0</td>
<td>15</td>
<td>3</td>
<td>21</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>.072</td>
<td>79</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>43</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>3</td>
<td>23</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>.079</td>
<td>111</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>37</td>
<td>6</td>
<td>0</td>
<td>23</td>
<td>2</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>.085</td>
<td>134</td>
<td>1</td>
<td>0</td>
<td>29</td>
<td>5</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>.091</td>
<td>133</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>31</td>
<td>7</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.098</td>
<td>95</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>41</td>
<td>10</td>
<td>0</td>
<td>24</td>
<td>3</td>
<td>17</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>.104</td>
<td>95</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>36</td>
<td>16</td>
<td>1</td>
<td>23</td>
<td>3</td>
<td>10</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>.111</td>
<td>106</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>8</td>
<td>0</td>
<td>22</td>
<td>3</td>
<td>20</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.117</td>
<td>102</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>12</td>
<td>1</td>
<td>30</td>
<td>7</td>
<td>23</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>.123</td>
<td>129</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>23</td>
<td>7</td>
<td>0</td>
<td>20</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>.130</td>
<td>88</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>38</td>
<td>13</td>
<td>0</td>
<td>20</td>
<td>1</td>
<td>25</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>.136</td>
<td>136</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>13</td>
<td>9</td>
<td>0</td>
<td>15</td>
<td>4</td>
<td>15</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>.142</td>
<td>101</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>39</td>
<td>8</td>
<td>0</td>
<td>9</td>
<td>6</td>
<td>24</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>.149</td>
<td>131</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>24</td>
<td>8</td>
<td>0</td>
<td>7</td>
<td>2</td>
<td>14</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>.155</td>
<td>127</td>
<td>14</td>
<td>3</td>
<td>2</td>
<td>20</td>
<td>5</td>
<td>0</td>
<td>10</td>
<td>2</td>
<td>16</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>.161</td>
<td>126</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>5</td>
<td>0</td>
<td>15</td>
<td>1</td>
<td>15</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>.168</td>
<td>150</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>13</td>
<td>7</td>
<td>0</td>
<td>9</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.174</td>
<td>143</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>28</td>
<td>10</td>
<td>0</td>
<td>14</td>
<td>5</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.180</td>
<td>141</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>18</td>
<td>3</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td>16</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>.187</td>
<td>124</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>6</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>25</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>.193</td>
<td>110</td>
<td>13</td>
<td>7</td>
<td>1</td>
<td>29</td>
<td>11</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td>14</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>.199</td>
<td>129</td>
<td>14</td>
<td>3</td>
<td>1</td>
<td>16</td>
<td>8</td>
<td>0</td>
<td>10</td>
<td>3</td>
<td>12</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>.206</td>
<td>119</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>33</td>
<td>6</td>
<td>0</td>
<td>16</td>
<td>2</td>
<td>11</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>.212</td>
<td>116</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>30</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>6</td>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
this model, the vector of proportions $p$ is considered to be random, varying independently from level to level. The pollen counts then follow a compound, or mixed, multinomial distribution. A particular multinomial mixture, the multinomial-Dirichlet, is considered in the next section, after which we return to this example. In passing, we remark that large sample variances, such as encountered above, are typical when heterogeneity is present.

1.2.3 The Multinomial-Dirichlet. We may introduce heterogeneity into the multinomial distribution by allowing for random variation in the vector $p = (p_1, p_2, \ldots, p_k)$. The $p_j$ are constrained to be nonnegative and must sum to unity; otherwise any pattern of co-variation is possible. Here we concentrate upon a very simple model in which $p$ is assumed to follow a Dirichlet distribution with parameters $\alpha_1, \alpha_2, \ldots, \alpha_k > 0$ (see Section 2.4). The resulting mixture is known as the multinomial-Dirichlet distribution with parameters $n$ and $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ and was introduced by Mosimann (1962).

The probability function for this distribution is obtained by multiplying the Dirichlet density function into the multinomial probability function and then integrating with respect to $p$.

Although the resulting integral is not particularly difficult to evaluate, a simpler approach is possible. Consider once again the derivation of the multinomial given in Section 1.2.1. There the parameters $\lambda_1, \lambda_2, \ldots, \lambda_k$ were assumed constant. Instead, let them have independent gamma distributions with common scale parameter and respective shape parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$. From Boswell, Ord, and Patil (1979, Section 6.3), we see that the $X_j$ then have independent negative binomial distributions with

$$
Pr(X_j = x_j) = (-1)^x_j \binom{-\alpha_j}{x_j} \theta^x_j (1 - \theta)^{\alpha_j - 1} = \binom{\alpha_j + x_j - 1}{x_j} \theta^x_j (1 - \theta)^{\alpha_j - 1}.
$$

Further, $X = X_1 + \cdots + X_k$ has a negative binomial distribution with parameters $\alpha = \alpha_1 + \cdots + \alpha_k$ and $\theta$. Here $\theta$ depends on the scale parameter of the gamma distributions, but is the same for all $j$. 
MULTIVARIATE DISTRIBUTIONS

If we now evaluate the conditional distribution of $X_1, \cdots, X_k$, given that $X = n$, we obtain after some simplification the probability function for the multinomial-Dirichlet:

$$Pr(X_1 = x_1, \cdots, X_k = x_k \mid X = n) = \frac{\binom{-\alpha}{x_1} \binom{-\alpha_2}{x_2} \cdots \binom{-\alpha_k}{x_k}}{\binom{-\alpha}{n}},$$

where $n = x_1 + x_2 + \cdots + x_k$ and $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_k$.

Setting $p_j = \alpha_j / \alpha$ ($j=1, \cdots, k$), the moments of the multinomial-Dirichlet distribution are as follows:

$$E[X_j] = np_j, \quad Var(X_j) = Cnp_j(1-p_j),$$

$$Cov(X_i, X_j) = -Cnp_i p_j \quad (i \neq j),$$

$$Corr(X_i, X_j) = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}}, \quad (1)$$

where $C = (n+\alpha)/(1+\alpha) > 1$. It is instructive to compare these moments with the corresponding multinomial moments. Since $C > 1$, the multinomial-Dirichlet has larger variances for given marginal expectations $E[X_j]$ than does the multinomial - as expected. But for both distributions, the correlations are the same function of the proportions $p_j$. Thus the multinomial-Dirichlet has slightly more flexible marginals, but is just as inflexible as the multinomial in its ability to describe patterns of covariation.

We now return to the pollen example of Section 1.2.2 and ask if the depth-to-depth pollen profile is adequately described by the multinomial-Dirichlet distribution. In lieu of a goodness-of-fit test, we compare sample correlations with predicted correlations. The latter are obtained by inserting sample proportions into (1).

For pine (*pinus*) versus grass (*gramineae*) the sample correlation is -.773 as opposed to -.491 for the predicted correlation. Transformation to $Z$, while not strictly valid, indicates that the difference is highly significant. Correlations for the other pairs are summarized in the correlation web of Figure 1; a solid line indicates a correlation significantly greater than would be
predicted by the multinomial-Dirichlet, a broken line indicates a 
sample correlation significantly less than the predicted. Readers 
are referred to Martin and Mosimann (1965) for further discussion 
and possible biological interpretation of the pattern in the 
correlation web.

1.3 NEGATIVE MULTINOMIAL AND NEGATIVE 
MULTINOMIAL-DIRICHLET DISTRIBUTIONS

These distributions are multivariate analogues of the negative 
binomial and the negative binomial-beta. The discussion parallels 
that of Section 1.2 so we will generally be brief.

First we recall how the negative binomial arises as an inverse 
sampling distribution. A population is partitioned into two catego-
ries one of which, called the 'marker' category, is considered 
special. Let $p_0$ and $p = 1 - p_0$ denote the population propor-
tions of the marker category and the other category, respectively. 
The population is then sampled with replacement until the marker 
category has entered the sample a specified number of times, say 
r. The number of times the other category enters the sample then 
has a negative binomial distribution with parameters $r$ and $p$.

For the multi-category extension, the population is partitioned 
into the marker category and $k$ additional categories with
respective proportions $p_0, p_1, p_2, \ldots, p_k$ where $p_0 + p_1 + \cdots + p_k = 1$. The population is sampled with replacement until the marker category has appeared $r$ times, and we write $X_i$ for the number of times that the $i$th category appears ($i = 1, 2, \ldots, k$). The vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ is then said to have a negative multinomial distribution with parameters $r$ and $\mathbf{p} = (p_1, p_2, \ldots, p_k)$. Note that $\mathbf{p}$ is not a probability vector here; the sum of its components is $1 - p_0$ which is less than unity.

As was the case for the multinomial, many of the properties of the negative multinomial are obvious. Combining some of the non-marker categories results in a negative multinomial, while each marginal $X_i$ has a negative binomial distribution with parameters $r$ and $p_i/(p_0 + p_i)$. The moments are

$$E[X_i] = rp_i/p_0, \quad \text{Var}(X_i) = rp_i(p_0 + p_i)/p_0^2,$$

$$\text{Cov}(X_i, X_j) = rp_i p_j / p_0^2 \quad (i \neq j)$$

$$\text{Corr}(X_i, X_j) = \frac{p_i p_j}{(p_0 + p_i)(p_0 + p_j)}.$$

Here again the correlation depends only on the expected values. The reader is invited to convince himself, by intuitive arguments, that the correlations should indeed be positive.

We next derive the probability function of the negative multinomial by arguments similar those of Section 1.2.1.

1.3.1 The Negative Multinomial as a Mixture. We assume that the population contains $k + 1$ species, labeled 0, 1, 2, \ldots, $k$, and that the individual specimens of the different species arrive at the light trap according to independent Poisson processes with parameters $\lambda_0, \lambda_1, \ldots, \lambda_k$. The light trap is operated until exactly $r$ specimens of species 0 have been captured (a peculiar light trap!). The time $t$ that the light trap is in operation then follows a gamma distribution with shape parameter $r$:

$$f(t) = \lambda_0^r t^{r-1} e^{-\lambda_0 t} / \Gamma(r).$$
Given \( t \), the numbers captured for the other species follow independent Poisson distributions with means \( \lambda_j^t \) \((j = 1, 2, \ldots, k)\). Hence we arrive at the joint probability distribution

\[
\Pr(X_1 = x_1, \ldots, X_k = x_k) = \int_0^\infty \prod_{j=1}^k \left( \lambda_j^t \right)^{x_j} e^{-\lambda_j^t} /x_j! \right) f(t) \, dt
\]

\[
= \frac{\Gamma(r + \sum x_j)}{\Gamma(r) \prod_{j=1}^k x_j!} \prod_{j=1}^k \left( p_j \cdot \frac{x_j}{p_j} \right)^r
\]

where \( p_j = \lambda_j / (\lambda_0 + \lambda_1 + \cdots + \lambda_k) \) \((j = 0, 1, 2, \ldots, k)\). This is the probability function of the negative multinomial distribution. The inverse sampling interpretation required \( r \) to be a positive integer. However, the negative multinomial is a valid probability distribution for all positive real values of \( r \).

1.3.2 The Negative Multinomial-Dirichlet. This distribution was considered by Mosimann (1963) and is obtained from the negative multinomial by letting the probability vector \((p_0, p_1, p_2, \ldots, p_k)\) follow a Dirichlet distribution with parameters \( \alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_k \). The probability function is

\[
\Pr(X_1 = x_1, \ldots, X_k = x_k)
\]

\[
= \frac{\Gamma(\alpha) \Gamma(\alpha_0 + r)}{\Gamma(\alpha_0) \Gamma(\alpha + r)} \left( \frac{1}{z} \right)^{r + z - 1} \left( \frac{x_1}{\alpha_1} \right) \left( \frac{x_2}{\alpha_2} \right) \cdots \left( \frac{x_k}{\alpha_k} \right)
\]

where \( \alpha = \alpha_0 + \alpha_1 + \cdots + \alpha_k \) and \( z = x_1 + x_2 + \cdots + x_k \). The moments are as follows:

\[
\mathbb{E}[X_j] = r \alpha_j / (\alpha_0 - 1), \quad \text{Var}(X_j) = r \alpha_j (\alpha_j + \alpha_0 - 1) / (\alpha_0 - 1)^2,
\]

\[
\text{Cov}(X_i, X_j) = r \alpha_i \alpha_j / (\alpha_0 - 1)^2 \quad (i \neq j),
\]

\[
\text{Corr}(X_i, X_j) = \sqrt{\frac{\alpha_i \alpha_j}{(\alpha_i + \alpha_0 - 1)(\alpha_j + \alpha_0 - 1)}} \quad (3)
\]
TABLE 2: Non-forest pollen counts recorded while counting 100 grains of forest pollen from each of 73 levels of the Bellas Artes core. The taxa are: Grasses (G); Composites (C); Amaranthes (A); and Cypress (CY). (From Clisby and Sears, 1955.)

<table>
<thead>
<tr>
<th>G</th>
<th>C</th>
<th>A</th>
<th>CY</th>
<th>G</th>
<th>C</th>
<th>A</th>
<th>CY</th>
<th>G</th>
<th>C</th>
<th>A</th>
<th>CY</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>9</td>
<td>25</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>234</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>19</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>15</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>136</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>283</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>17</td>
<td>49</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>3</td>
<td>11</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>249</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>9</td>
<td>13</td>
<td>167</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>27</td>
<td>45</td>
<td>36</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>13</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>45</td>
<td>31</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>18</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>16</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>19</td>
<td>31</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>10</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>14</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>- .06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 3: Correlation coefficients from the Bellas Artes core. Expected correlations estimated assuming a negative multinomial-Dirichlet distribution. See Table 2 for codes.

<table>
<thead>
<tr>
<th>Expected</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>CY</td>
<td>1</td>
</tr>
</tbody>
</table>
where \( g = (\alpha_0 + r - 1)/(\alpha_0 - 2) \) and moments of order \( m \) exist only when \( \alpha_0 > m \).

1.3.3 An Example. Clisby and Sears (1955) record pollen counts for the Bellas Artes core from the valley of Mexico. Part of their data is reproduced in Table 2. These are the number of grains of grasses (Gramineae), Composites (Compositae), Armaranth (Ararmathaceae), and Cypress (Taxodium) noted while counting 100 grains of forest pollen.

We have 73 sets, each set representing a depth in the core, and each set with \( r = 100 \) as the number of 'marker' forest grains counted. Sample variances of counts of grasses, composites, amaranths, and cypress are 64.78, 56.78, 462.1, and 2827.2, respectively. Variances assuming constant \( p \)'s were estimated using the formula \( \bar{x}(x + r)/r \) (see equation (2)). These are respectively 5.43, 3.98, 7.24, and 15.7. Since the observed variances are substantially larger than corresponding negative multinomial variances, we conclude that the \( p \)'s vary with the depth. Sample correlation coefficients and those expected under the negative multinomial-Dirichlet hypothesis are shown in Table 3. The latter were estimated using equation (3). The expected correlations are all close to zero. It is obvious upon inspection that the observed correlations are not compatible with the hypothesis.

1.4 BIVARIATE LOGARITHMIC SERIES DISTRIBUTION

As a species frequency model, the univariate logarithmic series distribution (LSD) has enjoyed remarkable success*, particularly in view of the extreme simplicity of this distribution. In this section we describe the bivariate LSD and attempt to fit some bivariate species frequency data. For reasons of notational simplicity as well as a lack of suitable data, only the bivariate LSD is considered here; the reader is referred to Patil and Bildi- kar (1967) for a full discussion of the multivariate LSD.

Suppose two trapping devices (perhaps light traps) are in operation simultaneously and, for a given species with abundance \( \lambda \), let \( X_1, X_2 \) be the number of specimens captured by the two traps. We assume that:

* Small but consistent departures have been noted however. See Williams (1964) and Engen (1978).
i) conditional on $\lambda$, $X_1$ and $X_2$ have independent Poisson distributions with means $\lambda A_1$ and $\lambda A_2$. The constants $A_1$ and $A_2$ reflect the relative efficiencies of the traps.

ii) the abundances $\lambda$ vary from species to species in accordance with a gamma distribution having shape parameter $r > 0$ and scale parameter $A_0$.

With these assumptions, the joint distribution of $(X_1, X_2)$ is bivariate negative multinomial,

$$\Pr(X_1 = x_1, X_2 = x_2) = \frac{\Gamma(r + x_1 + x_2)}{\Gamma(r) x_1! x_2!} \theta_1^{x_1} \theta_2^{x_2} (1 - \theta_1 - \theta_2)^r,$$

where $\theta_i = A_i/(A_0 + A_1 + A_2)$ ($i = 1, 2$). However, species cannot be observed unless captured by at least one of the traps. Therefore we truncate this multinomial at the origin $(x_1, x_2) = (0, 0)$ and, as in the univariate case, take the limit as $r \to 0$. This yields the probability function of the bivariate LSD:

$$-\log \left(1 - \theta_1 - \theta_2\right) \frac{\Gamma(x_1 + x_2)}{x_1! x_2!} \theta_1^{x_1} \theta_2^{x_2}$$

$(x_1, x_2 = 0, 1, 2, \ldots; (x_1, x_2) \neq (0, 0))$.

It will be convenient to introduce the notation $\psi = \theta_1 + \theta_2$, $\gamma = 1 - \psi = 1 - \theta_1 - \theta_2$, and $L = -\log \gamma$. Then the moments of the bivariate LSD are

$$E[X_1] = \mu_1 = \theta_1 / (\gamma L), \quad \text{Var}(X_1) = \mu_1 + \mu_1^2 (L - 1),$$

$$\text{Cov}(X_1, X_2) = \mu_1 \mu_2 (L - 1),$$

$$\text{Corr}(X_1, X_2) = \frac{L - 1}{\sqrt{[L - 1 + \mu_1^{-1}] [L - 1 + \mu_2^{-1}]} \cdot}$$

Patil and Bildiker (1966) have discussed maximum likelihood estimation of $\theta_1$ and $\theta_2$. Let $(x_{1j}, x_{2j})$ ($j = 1, 2, \ldots, n$) be a random sample drawn from the bivariate LSD and let $\bar{x}_1 = (1/n) \sum_j x_{1j}$.
be the sample means. Also write $\bar{x} = \bar{x}_1 + \bar{x}_2$. The maximum like-
lihood estimates of $\theta_1$ and $\theta_2$ are then

$$\hat{\theta}_1 = \hat{\psi} \bar{x}_1 / \bar{x}, \quad \hat{\theta}_2 = \hat{\psi} \bar{x}_2 / \bar{x},$$

where $\hat{\psi}$ is the solution of the equation

$$\bar{x} = \frac{\hat{\psi}}{(1 - \hat{\psi})[-\log(1 - \hat{\psi})]}.$$  \hspace{1cm} (4)

The solution of (4) has been extensively tabulated in Patil and Wani (1965).

1.4.1 An Example. The Catherwood Diatometer is an instrument for
sampling the diatom population of rivers. In its original design
the diatoms collected on glass slides held in the diatometer, but
it was noticed that diatoms frequently collected in large numbers
on the styrofoam floats. In order to determine if styrofoam was
a better substrate than glass, Hohn and Hellerman (1963) installed
diatometers in several North American rivers. Their data for the
Potomac River is summarized in Table 4. Table 5 shows the result
of fitting the two zero-truncated marginals to univariate LSD's.

As measured by the $\chi^2$ criterion, the fit is excellent in both
cases, although the styrofoam marginal does have a large observed
frequency for the cell $x_2 = 2$.

In Table 6 we show the bivariate LSD fitted to this data. As
is immediately apparent, the fit is very poor. The observed fre-
quencies in the cells $x_2 = 0$ are considerably larger than
expected. The inadequate fit can also be seen by comparing
observed and expected variances. For glass these are 21,303 and
12,266, respectively, while for styrofoam the figures are 64,809
and 21,898. Similarly the observed correlation is .872 as compared
with a predicted correlation of .997.

Insofar as the authors are aware, no one has previously
attempted to fit bivariate species frequency data (to any distri-
bution!). It might be interesting and worthwhile to investigate
alternative models in the light of this example and any other
bivariate data that may be available.
### TABLE 4: Potomac River diatom counts. \( x_1 \) = \# specimens collected on glass slides, \( x_2 \) = \# collected on styrofoam. Each pair \((x_1, x_2)\) corresponds to a different species. Numbers of species found: 137 (total), 115 (on glass), 108 (on styrofoam). (From Hohn and Hellerman, 1963.)

<table>
<thead>
<tr>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>44</td>
<td>47</td>
<td>27</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>88</td>
<td>59</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>104</td>
<td>44</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>80</td>
<td>27</td>
<td>208</td>
<td>33</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>39</td>
<td>44</td>
<td>1</td>
<td>0</td>
<td>54</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>101</td>
<td>126</td>
</tr>
<tr>
<td>79</td>
<td>60</td>
<td>1</td>
<td>0</td>
<td>29</td>
<td>24</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>14</td>
<td>18</td>
<td>5</td>
<td>12</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>816</td>
<td>2412</td>
<td>80</td>
<td>45</td>
<td>0</td>
<td>2</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>28</td>
<td>153</td>
<td>778</td>
<td>662</td>
<td>55</td>
<td>65</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>6</td>
<td>25</td>
<td>29</td>
<td>136</td>
<td>189</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>34</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>20</td>
<td>31</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>240</td>
<td>234</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>19</td>
<td>21</td>
<td>12</td>
<td>4</td>
<td>36</td>
<td>99</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>88</td>
<td>99</td>
<td>2</td>
<td>0</td>
<td>1152</td>
<td>1413</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>56</td>
<td>126</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>58</td>
<td>99</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>9</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>288</td>
<td>720</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>336</td>
<td>423</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>12</td>
<td>14</td>
<td>2</td>
<td>3</td>
<td>41</td>
<td>153</td>
</tr>
<tr>
<td>104</td>
<td>40</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>22</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>368</td>
<td>405</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>61</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>17</td>
<td>23</td>
<td>10</td>
<td>14</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>208</td>
<td>261</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>13</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>24</td>
<td>23</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>11</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 5: Potomac River diatom counts. Zero-truncated marginals fitted to univariate LSD's.

<table>
<thead>
<tr>
<th>Glass slides</th>
<th>Expected</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>19.84</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>9.89</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>6.57</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>4.91</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>3.92</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>3.26</td>
<td>3</td>
</tr>
<tr>
<td>7-8</td>
<td>5.21</td>
<td>7</td>
</tr>
<tr>
<td>9-10</td>
<td>4.08</td>
<td>3</td>
</tr>
<tr>
<td>11-12</td>
<td>3.35</td>
<td>3</td>
</tr>
<tr>
<td>13-15</td>
<td>4.10</td>
<td>4</td>
</tr>
<tr>
<td>16-19</td>
<td>4.33</td>
<td>6</td>
</tr>
<tr>
<td>20-24</td>
<td>4.25</td>
<td>3</td>
</tr>
<tr>
<td>25-31</td>
<td>4.59</td>
<td>4</td>
</tr>
<tr>
<td>32-39</td>
<td>4.04</td>
<td>4</td>
</tr>
<tr>
<td>40-50</td>
<td>4.26</td>
<td>3</td>
</tr>
<tr>
<td>51-64</td>
<td>4.08</td>
<td>4</td>
</tr>
<tr>
<td>65-83</td>
<td>4.09</td>
<td>4</td>
</tr>
<tr>
<td>84-110</td>
<td>4.14</td>
<td>5</td>
</tr>
<tr>
<td>111-149</td>
<td>4.04</td>
<td>1</td>
</tr>
<tr>
<td>150-213</td>
<td>4.07</td>
<td>2</td>
</tr>
<tr>
<td>214-339</td>
<td>4.00</td>
<td>3</td>
</tr>
<tr>
<td>340+</td>
<td>4.00</td>
<td>4</td>
</tr>
</tbody>
</table>

$\hat{\theta} = 0.99691 \quad x^2_{20} = 9.46$

<table>
<thead>
<tr>
<th>Styrofoam</th>
<th>Expected</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17.38</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>8.67</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>5.77</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>4.32</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>3.45</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>2.87</td>
<td>4</td>
</tr>
<tr>
<td>7-8</td>
<td>4.60</td>
<td>4</td>
</tr>
<tr>
<td>9-10</td>
<td>3.61</td>
<td>5</td>
</tr>
<tr>
<td>11-13</td>
<td>4.27</td>
<td>4</td>
</tr>
<tr>
<td>14-17</td>
<td>4.38</td>
<td>6</td>
</tr>
<tr>
<td>18-22</td>
<td>4.20</td>
<td>4</td>
</tr>
<tr>
<td>23-29</td>
<td>4.48</td>
<td>7</td>
</tr>
<tr>
<td>30-38</td>
<td>4.33</td>
<td>3</td>
</tr>
<tr>
<td>39-49</td>
<td>4.01</td>
<td>5</td>
</tr>
<tr>
<td>50-64</td>
<td>4.11</td>
<td>3</td>
</tr>
<tr>
<td>65-84</td>
<td>4.05</td>
<td>1</td>
</tr>
<tr>
<td>85-112</td>
<td>4.09</td>
<td>3</td>
</tr>
<tr>
<td>113-152</td>
<td>4.06</td>
<td>1</td>
</tr>
<tr>
<td>153-212</td>
<td>4.01</td>
<td>3</td>
</tr>
<tr>
<td>213-313</td>
<td>4.00</td>
<td>2</td>
</tr>
<tr>
<td>314-531</td>
<td>4.01</td>
<td>2</td>
</tr>
<tr>
<td>532+</td>
<td>3.35</td>
<td>4</td>
</tr>
</tbody>
</table>

$\hat{\theta} = 0.99797 \quad x^2_{20} = 18.36$

TABLE 6: Potomac River diatom counts fitted to bivariate LSD. Integral numbers in the cells represent observed frequencies whereas decimal ones represent expected frequencies.

<table>
<thead>
<tr>
<th>$x_2$ STYROFOAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3+</td>
</tr>
<tr>
<td>0 19 (8.9)</td>
</tr>
<tr>
<td>1 13 (11.9)</td>
</tr>
<tr>
<td>2 5 (12.0)</td>
</tr>
<tr>
<td>3+ 8 (3.8)</td>
</tr>
<tr>
<td>3+ 10 (2.7)</td>
</tr>
<tr>
<td>2 5 (8.3)</td>
</tr>
<tr>
<td>3+ 68 (83.6)</td>
</tr>
</tbody>
</table>

$X^2_5 = 46.1$
1.5 MULTIVARIATE HYPERGEOMETRIC AND NEGATIVE HYPERGEOMETRIC DISTRIBUTIONS

While it is rare that the ecologist will literally sample with immediate replacement of each individual, the with replacement assumption frequently offers a convenient approximation and simplification. The aim in capture-recapture studies however is to estimate the size of a wildlife population and so sampling without replacement is of the essence. In the multivariate case we consider \( m \) species which are to be observed or trapped by the same method. Let \( N \) be the total population size and \( N_1, N_2, \ldots, N_m \) the size of the various species. After setting the traps on the first occasion the numbers \( M_1, \ldots, M_m \) of the \( m \) species are captured and tagged, leaving \( M_0 = (N - \sum_j M_j) \) untrapped.

When different trapping rates are assumed for each species then the problem must be formulated as \( m \) separate experiments. However, since the same traps are used for all individuals (that is, they 'compete' for the same trap space) a common multivariate formulation seems appropriate whenever species are 'equally catchable.' Then, on the second occasion, a total of \( n \) individuals are captured comprising \( x_0 \) untagged and \( x_1, \ldots, x_m \) tagged members of the \( m \) species. Just as in the univariate case, we may derive the probability function as

\[
\Pr(X_1 = x_1, \ldots, X_m = x_m) = \frac{\binom{M_0}{x_0} \binom{M_1}{x_1} \cdots \binom{M_m}{x_m}}{\binom{N}{n}},
\]

where \( 0 \leq x_j \leq n \) for all \( j \) and \( \sum_{j=0}^{m} x_j = n \). This is the multivariate hypergeometric distribution. Putting \( p_j = M_j/N \), the moments of this distribution are

\[
E[X_j] = np_j, \quad \text{Var}(X_j) = C n p_j (1-p_j),
\]

\[
\text{Cov}(X_i, X_j) = -C n p_i p_j,
\]

where \( C = (N-n)/(N-1) \) is the finite population correction. In particular we see that the counts are negatively correlated.
For the capture-recapture problem, \( N \) may be estimated by

\[
\tilde{N} = nM/z
\]

where \( M = \sum_{j=1}^{m} M_j \) and \( z = \sum_{j=1}^{m} x_j \), although the standard modified version

\[
\tilde{N} = (n + 1)(M + 1)/(z + 1) - 1
\]

may be preferred. If the equal catchability of species is assumed, so that \( E[M_j] = \alpha N_j \), then the \( N_j \) may be estimated as

\[
\tilde{N}_j = \frac{M_j}{\tilde{\alpha}},
\]

where \( \tilde{\alpha} = M/\tilde{N} \) is used to estimate \( \alpha \). Particularly when the numbers for individual species are small, the smoothed estimators \( \tilde{\alpha} \) may provide more satisfactory estimators than those given by \( \alpha \) separate studies.

1.5.1 Inverse Sampling. If a particular species is important to the study, then the investigator may resolve to continue sampling until \( k \) tagged members of that species have been recaptured. We let this 'marker' species be labelled with the subscript one. Then the probability that \( x_2, \ldots, x_m \) tagged members of the other species will be recaptured before the \( k \)th member of the marker species is given by the multivariate inverse hypergeometric distribution:

\[
\Pr(X_0 = x_0, X_2 = x_2, \ldots, X_m = x_m | X_1 = k) = \frac{\binom{M_0}{x_0} \binom{M_2}{x_2} \cdots \binom{M_m}{x_m}}{\binom{N}{k + z - 1}} \cdot \frac{\binom{M_1 - k + 1}{N - k - z + 1}}{N - k - z + 1}
\]

where \( z = x_0 + \cdots + x_m \). The last term arises since the final observation must be the \( k \)th individual from the marker species. We note that our notation is slightly different from the usual format (Patil and Joshi, 1968, p. 76) since we are interested in the particular problem of estimating population size. Otherwise, a reordering of subscripts makes the notation simpler.
The moments are as follows (for \( j = 0, 2, \cdots, m \)):

\[
E[X_j] = \frac{kM_j}{(M_1 + 1)},
\]

\[
\text{Var}(X_j) = \frac{kM_j (M_j + M_1 + 1)(M_1 - k + 1)}{(M_1 + 1)^2 (M_1 + 2)},
\]

\[
\text{Cov}(X_j, X_1) = \frac{kM_j M_1 (M_1 - k + 1)}{(M_1 + 1)^2 (M_1 + 2)}.
\]

Again, the correlations are positive for the inverse sampling scheme. Since \( E[X_0] = \frac{kM_0}{(M_1 + 1)} \), we have the minimum variance unbiased estimator for \( M_0 \) as

\[
\tilde{M}_0 = (M_1 + 1)x_0/k;
\]

hence

\[
\tilde{N} = \tilde{M}_0 + \sum_{j=1}^{m} M_j = \tilde{M}_0 + M.
\]

Again, using the assumption of equal catchability for each species so that \( E[M_j] = \alpha N_j \), we obtain \( \tilde{N}_j = M_j/\tilde{\alpha} \), where now

\[
\tilde{\alpha} = M/\tilde{N} = M/(M + \tilde{M}_0).
\]

The inverse scheme gives estimators with better sampling properties because of the guarantee that at least \( k \) tagged individuals will be recaptured. This emphasizes that when direct sampling schemes yield low numbers of recaptures, further sampling should be attempted whenever possible.

1.6 FURTHER READING

Our survey has been quite brief and only touches the surface of multivariate distribution theory. Readers wishing information about some of the many other discrete multivariate distributions should consult Johnson and Kotz (1969, chap. 11), Patil and Joshi (1968, p. 60-86) and Ord (1972, chap. 7).
We should also mention that multivariate extensions to the discussion on birth and birth-death processes (see Boswell, Ord, and Patil, 1979, chap. 3) are given by Patil and Boswell (1972). In particular, selected birth and equilibrium birth-death processes lead to the multiple Poisson, multinomial, and negative multinomial distributions, thus providing an interesting set of alternative chance mechanisms for these distributions. The underlying principles involved follow the univariate case although the mathematical arguments become more involved, so the details are omitted from our general survey. The results of Patil and Boswell (1972) extend beyond the distributions just mentioned and the interested reader should consult this paper and Boswell and Patil (1972) for further details.

Both univariate and multivariate distributions are considered in the context of urn models by Johnson and Kotz (1977). All the distributions considered in this chapter may be generated by the Polya-Eggenberger urn scheme.
CHAPTER 2

CONTINUOUS MULTIVARIATE DISTRIBUTIONS

2.1 INTRODUCTION

In this chapter we shall examine three families of continuous multivariate distributions: (1) the normal or Gaussian family; (2) the lognormal family; and (3) the Dirichlet family. The first family, the normal, is for variables which can take on negative as well as positive values. The second, the lognormal, is for positive variables only. The third, the Dirichlet, is for positive variables whose values are constrained to add to one.

As with discrete multivariate families of distributions, a major question we shall pose for each family is: Do its members represent widely different patterns of covariation among the variables? To anticipate, the answer to this question is yes for the normal and lognormal families. For the Dirichlet family it is a decided no.

2.2 MULTIVARIATE NORMAL DISTRIBUTION

The multivariate normal distribution has received a great deal of attention in the literature, and it is not our purpose to survey this vast literature here. Much of the literature concerns the use of the multivariate normal in statistical inference, and again it is not our purpose to discuss such uses. Instead we shall look briefly at certain properties of the multivariate normal model.

Consider first a univariate normal random variable $X_1$ with probability density function

$$f(x_1) = \exp\left[-\frac{1}{2} \sigma_{11}^{-1} (x_1 - \mu_1)^2\right]/\sqrt{(2\pi\sigma_{11})},$$

where $\sigma_{11} > 0$, $-\infty < \mu_1 < \infty$, and $-\infty < x_1 < \infty$. It is readily
shown that $E[X_i] = \mu_i$ and $\text{Var}(X_i) = \sigma_{i1}$, so that both the mean and variance of a normal variable are finite.

With the understanding that a univariate normal variable may be degenerate ($\sigma_{i1} = 0$), then Rao (1973, p. 518) gives the following definition of a multivariate normal distribution.

An $m$-dimensional random vector $\sim{X} = (X_1, \cdots, X_m)$ has an $m$-dimensional multivariate normal distribution if and only if every linear function of $\sim{X}$ has a univariate normal distribution. [The prime indicates transpose.]

In other words, if $a' = (a_1, \cdots, a_m)$ is any vector of real coordinates, then $\sim{X}$ is multivariate normal if and only if every scalar (univariate) random variable

$$a' \sim{X} = \sum_{i=1}^{m} a_i \sim{X}_i$$

is a univariate normal variable.

An immediate consequence of this definition is that each coordinate variable, $\sim{X}_i$, has a univariate normal distribution.

Further since a univariate normal variable has a finite mean and variance, it follows that the multivariate normal distribution has finite second moments and we can write the mean vector $\sim{\mu}' = (\mu_1, \cdots, \mu_m)$ and covariance matrix

$$
\Sigma = \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\
\sigma_{12} & \sigma_{22} & \cdots & \sigma_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{1m} & \sigma_{2m} & \cdots & \sigma_{mm}
\end{bmatrix} .
$$

It can be shown that a member of the multivariate normal family is completely determined by specifying $\sim{\mu}$ and $\sim{\Sigma}$, and we shall henceforth write that $\sim{X}$ is $\text{MVN}_m(\sim{\mu}, \sim{\Sigma})$ to denote that particular $m$-dimensional multivariate normal distribution specified by such a choice.

If the covariance matrix is nonsingular, so that $|\Sigma| > 0$, then an $m$-dimensional density for $\sim{X}$ exists and can be shown to be
\[ f(x_1, \ldots, x_m) = (2\pi)^{-\frac{m}{2}} \left| \Sigma \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (x-\mu)' \Sigma^{-1} (x-\mu) \right\}, \quad (5) \]

where: \(-\infty < x_i < \infty \) (i=1, \ldots, m), \( \mu \) is any real vector with finite coordinates, and \( \Sigma \), being a non-singular covariance matrix, is positive definite. The integral of this density over any subset of \( m \)-dimensional space gives the probability that \( \mathbf{X} \) takes its value in that subset. In particular, if \( A \) is some neighborhood of the \( m \)-dimensional point \( \mu \), then \( \Pr(A) > 0 \). In other words, in the multivariate model with \( |\Sigma| > 0 \), \( \mathbf{X} \) may occur theoretically anywhere in \( m \)-dimensional space, although for many regions the probability will be extremely small.

We now consider \( \mathbf{X} \), \( \text{MVN}(\mu, \Sigma) \) but where \( \Sigma \) is singular so that \( |\Sigma| = 0 \). Then, the multivariate density in (5) is not defined and in fact no density exists for \( \mathbf{X} \). To illustrate a theoretically important member of the normal family for which \( |\Sigma| = 0 \), we let \( m=2 \) so that \( \mathbf{X}' = (X_1, X_2) \) is bivariate normal.

A measure of the association of \( X_1 \) with \( X_2 \) is the 'population correlation coefficient'

\[ \rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11} \sigma_{22}}}, \quad (6) \]

and perfect linear association results if \( \rho_{12} = \pm 1 \). Now consider \((X_1, X_2) \) bivariate normal with \( \sigma_{11}, \sigma_{22} > 0 \) and \( \rho_{12} = 1 \). (For convenience let \( \mu = 0 \).) There is now a perfect linear association of \( X_1 \) with \( X_2 \) and, in fact,

\[ X_2 = \sqrt{\sigma_{22}/\sigma_{11}} X_1 \]

with probability one. Also using (6) the choice \( \rho_{12} = 1 \) implies that \( |\Sigma| = \sigma_{11} \sigma_{22} - \sigma_{12}^2 = 0 \). Consequently, the covariance matrix is singular and there is no two-dimensional density. Here any subset of two-dimensional space which does not intersect the line \( x_2 = \sqrt{\sigma_{22}/\sigma_{11}} x_1 \) has probability zero. While there is no two-dimensional density, there is, however, a one-dimensional normal density over the one-dimensional subspace defined by the line \( x_2 = (\sigma_{11}/\sigma_{22}) x_1 \). This is simply to illustrate that there are certainly distributions with singular covariance matrix which one
wishes to include in the multivariate normal family. Rao's definition of the multivariate normal distribution accomplishes this.

One cannot overemphasize the role of the multivariate normal family in statistics. As any statistical practitioner knows, most procedures for parametric inference can be traced to the normal or to distributions closely associated with it through sampling theory. The generality of this family results, in part, from the fact that every linear combination of a multivariate normal vector is univariate normal. Another source of its generality lies in the fact that, just as in the univariate case, there is a central limit theorem which shows that the limit distribution of many random vectors is multivariate normal.

The multivariate normal family of distributions is described in many texts, for example Anderson (1958), Dempster (1969, part 4), Johnson and Kotz, (1972), chap. 35), Morrison (1976), and Rao (1952, 1973).

2.3 THE MULTIVARIATE LOGNORMAL FAMILY OF DISTRIBUTIONS

The problem of obtaining a flexible multivariate distribution for a 'positive' random vector \( \mathbf{X} \) (whose coordinates are each a positive random variable) has been a vexing one. Two univariate distributions for a positive random variable are well known. One is the lognormal family whose density is

\[
f(x) = \frac{\exp[-(\log x - \mu)^2 / 2\sigma^2]}{x\sqrt{2\pi\sigma^2}} \quad (x > 0),
\]

where \(-\infty < \mu < +\infty, \sigma > 0\). The other is the gamma family with density

\[
f(x) = (\lambda^k / \Gamma(k)) x^{k-1} \exp(-\lambda x) \quad (x > 0),
\]

where \( \lambda, k > 0 \). With \( k=1 \) this gives the familiar exponential distribution, while the choice \( k=\nu/2 \) and \( \lambda=1/2 \) gives a chi-squared distribution with \( \nu \) degrees of freedom. Unfortunately, attempts to obtain general covariance structure in a multivariate family whose marginal distributions are all univariate gamma have not been very successful. In fact using characteristic properties of gamma distributions along with the properties of 'size and shape' variables, one can see that a truly general covariance structure cannot be obtained under the limitation of gamma marginals (Mosimann, 1975, p. 234).
At present the only truly flexible multivariate model for positive random variables is that of the multivariate lognormal family. The univariate lognormal distribution is discussed in some detail by Aitchison and Brown (1957).

The multivariate lognormal family is best defined through the multivariate normal family. Suppose $Y = (Y_1, \ldots, Y_m)$ is a multivariate normal vector with mean $\mu^\sim$ and covariance matrix $\Sigma^\sim$, and let $X_i = \exp(Y_i)$ or $Y_i = \log X_i$ ($i=1, \ldots, m$). Then $X = (X_1, \ldots, X_m)$ is a multivariate lognormal vector with parameters $\mu^\sim$ and $\Sigma^\sim$.

From Rao's definition of an $m$-dimensional multivariate normal distribution, we obtain the following equivalent definition for a multivariate lognormal distribution.

An $m$-dimensional positive random vector $X = (X_1, \ldots, X_m)$ has an $m$-dimensional multivariate lognormal distribution if and only if every product function, $\prod_{i=1}^{m} X_i^{a_i}$, has a univariate lognormal distribution.

In other words, if $(a_1, \ldots, a_m)$ is any real vector, then $X$ is multivariate lognormal if and only if every scalar (univariate) random variable $\prod X_i^{a_i}$ is a univariate lognormal variable.

In Figures 2 and 3 we contrast the bivariate normal and lognormal distributions. In the normal case the probability contours are concentric ellipses. The slope of the major axis of these ellipses has the same sign as the population correlation coefficient. When $(X_1, X_2)$ are lognormal, their positive probability tends to concentrate along a curve with equation of the form $x_2 = ax_1^b$, where $a > 0$ and where $b$ has the same sign as $\text{Corr}(X_1, X_2)$.

If $X$ is lognormal $(\mu^\sim, \Sigma^\sim)$, then $\mu^\sim$ and $\Sigma^\sim$ are not the mean vector and covariance matrix of $X$. Let $\Sigma = (\sigma_{ij})$. The moments of the $X_i$'s are

$$E[X_i] = e^{\mu_{ii} + (\sigma_{ii}/2)}$$
$$\text{Var}(X_i) = (E[X_i])^2(\sigma_{ii} - 1),$$
FIG. 2: In each of the six graphs $X_1, X_2$ have respective means of 70, 50 and respective standard deviations of 10, 5. The correlation coefficient is indicated. Normal confidence boundaries are ellipses and also are contours of the density. Lognormal confidence boundaries were obtained by fitting normal ellipses to $\log X_1, \log X_2$ and transforming these curves to $X_1, X_2$. They are not, however, contours of the lognormal density.
**FIG. 3:** Explanation as in Figure 2. Here all correlations are negative.
\[ \text{Cov}(X_i, X_j) = E[X_i]E[X_j](\sigma_{ij} - 1). \]

Interestingly, as in the multivariate normal, \( X_i \) and \( X_j \) are independent if and only if they are uncorrelated \( (\rho_{ij} = 0) \).

Two facts point to the generality of the multivariate lognormal family for positive variables: (1) it successfully accommodates negative correlations among its variables and (2) for every member of the normal family determined by a choice of \( \mu, \Sigma \), there is a corresponding member of the lognormal family and conversely. Finally in terms of characterizations based on size and shape variables the lognormal can be shown to have general properties (Mosimann, 1975, p. 234).

2.3.1 Choosing the Normal or Lognormal in Practice. While there are various tests, such as of skewness, which might be used on univariate distributions to distinguish the lognormal from the normal distribution, we would here like to illustrate a graphic approach used by Jolicoeur and Heusner (1971).

Figure 4a is a scatter diagram relating basal oxygen consumption \( (X_2) \) with body weight \( (X_1) \) in rats. The data are fit with a 95% concentration ellipse (solid line) under a bivariate normal assumption of \( X_1, X_2 \). While the ellipse covers the points well, there is an excess of points outside the ellipse at large values and a possible paucity of points inside the ellipse at low values. In Figure 4b the log data \( (Y_1, Y_2) \) are fit with a 95% ellipse under a bivariate normal assumption. This ellipse covers the points in a more symmetric fashion. The particularly telling graph however is that of the ellipse fit to \( Y_1, Y_2 \) but transformed to the \( X_1, X_2 \) space (Figure 4a, dashed line curve). Comparison of the two contours in Figure 4a clearly shows that the bivariate lognormal assumption is preferable to that of the normal.

In cases of size and shape studies (Mosimann and James, 1979), the lognormal assumption for \( X \) allows exact tests for allometry. Figure 5 shows that there is little difference between the assumption of lognormal or normal in the case of the data (morphometric data on blackbirds). Since the lognormal affords better theory in this case, it is chosen for pragmatic reasons.
FIG. 4: Body weight $x_1$ and basal oxygen consumption $x_2$ of 252 rats. (a) Solid curve is 95% equal likelihood ellipse calculated from $x_1, x_2$; broken curve is antilog of 95% equal likelihood ellipse calculated from $y_1 = \log_{10} x_1$ and $y_2 = \log_{10} x_2$. Latter ellipse shown in (b). (From Jolicoeur and Heusner, 1971.)

FIG. 5: Tail and wing length for male blackbirds; solid and broken curves as in Figure 4a (from Mosimann and James, 1979).
INTERNATIONAL STATISTICAL ECOLOGY PROGRAM

The International Statistical Ecology Program (ISEP) consists of the activities of the Statistical Ecology Section of the International Association for Ecology and of the Liaison Committee on Statistical Ecology of the International Statistical Institute, the Biometric Society, and the International Association for Ecology. The ISEP is a non-profit program formulated to serve the needs of interdisciplinary research and training in the newly emerging fields of Statistical Ecology and Ecological Statistics.

SATELLITE PROGRAM IN STATISTICAL ECOLOGY

The Second International Congress of Ecology was held in Jerusalem during September 1978. In this connection, ISEP organized a Satellite Program in Statistical Ecology during 1977 and 1978. The emphasis was on research, review, and exposition concerned with the interface between quantitative ecology and relevant quantitative methods. Both theory and application of ecology and econometrics received attention. The Satellite Program consisted of instructional coursework, seminar series, thematic research conferences, and collaborative research workshops.

Research papers and research-review-expositions were specially prepared for the program by concerned experts and expositors. These materials have been refereed and revised, and are now available in a series of ten edited volumes listed on page ii of this volume.

The Satellite Program takes as its theme the better melding of fundamental ecological concepts with rigorous empirical quantification. The overall result should be progress toward a stronger body of general ecologic and econometric theory and practice.

FUTURE DIRECTIONS

The satellite-like-programs help create and sustain enthusiasm, inward strength, and working efficiency of those who desire to meet a contemporary social need in the form of some interdisciplinary work. It should be only proper and rewarding for everyone involved that such programs are planned from time to time.

Plans are being made for a satellite program in conjunction with the next Biennial Conference of the International Statistical Institute and the next International Congress of Ecology. Care should be exercised that the next program not become a mere replica of the present one, however successful it has been. Instead, the next program should be organized so that it helps further the evolution of statistical ecology as a productive field.

The next program is being discussed in terms of subject area groups. Each subject group is to have a coordinator assisted by small committees, such as a program committee, a research committee, an annual review committee, a journal committee, and an education committee. This approach is expected to respond to the need for a journal on statistical ecology, and also to the need of bringing out well planned annual review volumes. The education committee would formulate plans for timely modules and monographs. Interested readers may feel free to communicate their ideas and interests to those involved in planning the next program. The mailing address is: International Statistical Ecology Program, P. O. Box 218, State College, PA 16801, USA.
International Statistical Ecology Program
Satellite Program in Statistical Ecology

CONTENTS OF EDITED VOLUMES

BACKGROUND

Close to 350 lectures and discussions were held during 50 days in the middle of the summer seasons of
1977 and 1978. The program was endowed with stimulating local setting as well as with atmosphere
conducive for scholarship. Participants represented broad background and expertise. Advice was sought
from a number of special advisors. The advice received was immensely helpful in guiding selection of the
best experts in the field to achieve as representative and balanced a coverage as possible.

The editors together with the referees took a rather critical and constructive attitude from initial to final
stages of preparation of papers by offering specific suggestions concerning the suitability, and also the
structure, content, and size. These efforts of coordination and revision were intensified through editorial
sessions at the program itself as a necessary step for the benefit of both the readers and the participants.
Everyone went by scientific interests often at the expense of personal preferences. The program atmosphere
became truly creative and friendly, and this remarkable development contributed to the maximal cohesion
of the program and its proceedings within the limited time period available.

The edited research papers and research-review-expositions prepared for the program are organized in
ten volumes. Altogether, they consist of an estimated 4,000 pages of research, review, and exposition in
addition to a common foreword in each, followed by individual volume introductions. Subject and author
indexes are also prepared for each volume. May this ten volume set serve as a useful stimulus to further
research and training in statistical ecology for the nineteen eighties.

CONTENTS OF THE VOLUME

corresponding to this reprint

STATISTICAL DISTRIBUTIONS IN ECOLOGICAL WORK

J. K. Ord, G. P. Patil, and C. Taillie (editors)

400 pp. approx.

Patil, Chance Mechanisms Underlying Univariate Distributions. C. Taillie, J. K. Ord,

Ecological Advances Involving Statistical Distributions: M. T. Boswell and B. Dennis, Ecological
Contributions Involving Statistical Distributions. W. G. Warren, Some Recent Developments Relating
to Statistical Distributions in Forestry and Forest Products Research.

R. Hengeveld, On the Use of Abundance and Species Abundance Curves. S. Engen and C.
Taillie, Abundance Models and Community Description. S. Engen, Abundance Models and Sampling
Results. R. Hengeveld, S. Koohman, and C. Taillie, A Spatial Model Explaining the
Species-Abundance Curve. S. Kobayashi, Species-Area Curves.

Research Papers: T. Caraco, Ecological Response of Animal Group Size Frequencies. J. Derr and
J. K. Ord, Feeding Preferences of a Tropical Seedfeeding Insect: An Analysis of the Distribution of
Insects and Seeds in the Field. R. A. Kempton, The Statistical Analysis of Frequency Data Obtained
from Sampling an Insect Population Grouped by Stages. M. Pennington, Fitting a Growth Curve to
Field Data. M. Yang, Some Results on Size and Utilization Distribution of Home Range.

To place your order, contact:

International Co-operative Publishing House
P. O. Box 245
Burtonsville, Maryland 20730