Derivation of a Class of Frequency Distributions via Bayes's Theorem

By CARL-ERIK SÄRNDAL

University of Michigan

[Received November 1964. Revised March 1965]

SUMMARY

Being in some respects expository in character, this paper uses Bayes's theorem to derive a k-variate discrete probability distribution which is fundamental inasmuch as it proliferates, under various sets of asymptotic conditions, to yield as its limiting forms an array of other k-variate distributions including the multinomial, independent Poisson, independent negative binomial, independent gamma, Dirichlet and multivariate normal. A systematic arrangement of these distributions thereby suggests itself, and a number of properties pertaining to the distributions are discussed within the framework created by this arrangement.

1. INTRODUCTION

CONSIDER a model population the elements of which are known to classify into $k+1$ mutually exclusive classes, where $k \geqslant 1$ is a fixed integer. In this paper we find it convenient to think in terms of an urn containing balls of $k+1$ different colours, each colour representing a class. Denote the (unknown) number of balls of the ith colour by s_i $(i = 1, ..., k+1)$; the (known) total number of balls in the urn is $s = \Sigma_{i=1}^{k+1} s_i$. A first sample of balls (of fixed size m) has been drawn at random and without replacement, and the observed frequencies of the various colours are m_i $(i = 1, ..., k+1)$, and $m = \Sigma_1^{k+1} m_i$. From the remainder population a second random sample of balls (of fixed size n) is drawn without replacement, and the observed frequencies are n_i $(i = 1, ..., k+1)$, and $n = \Sigma_1^{k+1} n_i$. Denote the frequency vectors by $\mathbf{s} = (s_1, ..., s_k), \mathbf{m} = (m_1, ..., m_k), \mathbf{n} = (n_1, ..., n_k)$. Corresponding to these observed vectors there are vector random variables

$$\mathbf{S} = (S_1, ..., S_k), \quad \mathbf{M} = (M_1, ..., M_k), \quad \mathbf{N} = (N_1, ..., N_k),$$

respectively. (The sum of the $k+1$ frequencies being a fixed integer, it is sufficient to take only k of them into account when defining the vector. Here we choose to consider the frequency of the $(k+1)$th class as redundant.) For brevity we shall throughout the paper use a notation such that, for instance, $P(\mathbf{n} \mid \mathbf{m}, \mathbf{s})$ represents the probability that $\mathbf{N} = \mathbf{n}$ given $\mathbf{M} = \mathbf{m}$ and $\mathbf{S} = \mathbf{s}$. Also, throughout the paper, capital letters will denote random variables, and their small counterparts will denote observed values.

We shall investigate the distribution of the class frequency vector $\mathbf{N}$ of the second sample utilizing the information available to us by the fact that a first sample has previously been drawn, and that thereby we have observed the vector $\mathbf{M} = \mathbf{m}$. In such a situation the first sample represents, literally speaking, our prior knowledge of the population. The distribution of $\mathbf{N}$ will, however, also be influenced by the subjective choice of a prior distribution for **S**. Two prior distributions which are later referred to as the Bose–Einstein prior and the Maxwell–Bolzmann prior are to be considered.

Attention will be focused on the case where the Bose–Einstein prior is used. This prior distribution is a uniform one, and it produces a distribution of **N** (given $\mathbf{M} = \mathbf{m}$) which has as its limiting forms, under various assumptions on the quantities m, m_i, n and n_i, a spectrum of well-known distributions including the multinomial, Poisson, negative binomial, gamma, Dirichlet and multivariate normal distributions. The various assumptions made on m, m_i, n and n_i, suggest a systematic arrangement of these distributions. Sections 3, 4 and 5 are devoted to developing this system and to discussing a number of properties relating to the distributions, and the way in which these properties fit into the system.

2. The Joint Distribution of S, M and N

Without specifying its form for the moment, let $P(\mathbf{s})$ denote the prior distribution of **S**. By elementary probability laws,

$$P(\mathbf{m}\,|\,\mathbf{s}) = \prod_{i=1}^{k+1} \binom{s_i}{m_i} \Big/ \binom{s}{m},$$

where

$$0 \leqslant m_i \leqslant m \quad (i = 1, \ldots, k), \qquad \sum_{i=1}^{k} m_i \leqslant m, \tag{2.1}$$

and

$$P(\mathbf{n}\,|\,\mathbf{m},\mathbf{s}) = \prod_{i=1}^{k+1} \binom{s_i - m_i}{n_i} \Big/ \binom{s-m}{n},$$

where

$$0 \leqslant n_i \leqslant n \quad (i = 1, \ldots, k), \qquad \sum_{i=1}^{k} n_i \leqslant n. \tag{2.2}$$

(Note that some outcomes **m** for which (2.1) holds may actually be impossibilities, i.e. $P(\mathbf{m}\,|\,\mathbf{s}) = 0$. Likewise, for some outcomes **n** for which (2.2) holds we may have $P(\mathbf{n}\,|\,\mathbf{m},\mathbf{s}) = 0$. In such cases the probability expressions would contain one or more binomial factors $\binom{a}{b}$ with $b > a$, which are to be interpreted as zero.)

We find

$$\begin{aligned} P(\mathbf{s},\mathbf{m},\mathbf{n}) &= P(\mathbf{s})\,P(\mathbf{m}\,|\,\mathbf{s})\,P(\mathbf{n}\,|\,\mathbf{m},\mathbf{s}) \\ &= P(\mathbf{s}) \prod_{i=1}^{k+1} \binom{s_i}{m_i+n_i} \Big/ \binom{s}{m+n} \prod_{i=1}^{k+1} \binom{m_i+n_i}{n_i} \Big/ \binom{m+n}{n}, \end{aligned} \tag{2.3}$$

where, in addition to (2.1) and (2.2),

$$m_i + n_i \leqslant s_i \leqslant s \quad (i = 1, \ldots, k), \qquad \sum_{i=1}^{k} s_i \leqslant s. \tag{2.4}$$

We recall that s, m and n are fixed integers, and they remain so throughout the paper.

At this point we shall make some assumption as to the particular form of the prior distribution, $P(\mathbf{s})$, and then proceed to find the conditional distribution of **N**, given $\mathbf{M} = \mathbf{m}$.

3. Bose–Einstein Prior Distribution

Consider s indistinguishable balls, and assume that $k+1$ different colours are available. Let each of the s balls be assigned one of the colours at random. Letting s_i denote the frequency of balls of the ith colour resulting from this procedure, the number of different frequency configurations $\mathbf{s} = (s_1, \ldots, s_k)$ that can arise is $\binom{s+k}{k}$. Under Bose–Einstein statistics (see, for example, Feller, 1957, p. 38) all the configurations are equally likely. Thus, the probability of any configuration is given by

$$P(\mathbf{s}) = 1 \bigg/ \binom{s+k}{k} \tag{3.1}$$

for

$$0 \leqslant s_i \leqslant s \quad (i = 1, \ldots, k), \qquad \sum_{i=1}^{k} s_i \leqslant s. \tag{3.2}$$

The distribution (3.1) will be referred to as a Bose–Einstein prior for **S**.

Inserting (3.1) into (2.3) we obtain

$$P(\mathbf{s}, \mathbf{m}, \mathbf{n}) = \prod_{i=1}^{k+1} \binom{s_i}{m_i+n_i} \bigg/ \binom{s+k}{m+n+k} \prod_{i=1}^{k+1} \binom{m_i+n_i}{n_i} \bigg/ \frac{(m+n+k)!}{m!\,n!\,k!}.$$

We eliminate $\mathbf{s}$ by summing over (2.4), whereupon the first product expression becomes unity, yielding

$$P(\mathbf{m}, \mathbf{n}) = \prod_{i=1}^{k+1} \binom{m_i+n_i}{n_i} \bigg/ \frac{(m+k+n)!}{m!\,k!\,n!}.$$

The marginal distributions are

$$P(\mathbf{m}) = 1 \bigg/ \binom{m+k}{k}, \quad P(\mathbf{n}) = 1 \bigg/ \binom{n+k}{k},$$

defined over (2.1) and (2.2), respectively. Thus, a Bose–Einstein prior for **S** produces marginal distributions of the Bose–Einstein type for **M** and **N**. However, **M** and **N** are not independent, and the conditional distribution of **N** is found to be

$$P(\mathbf{n} \mid \mathbf{m}) = \prod_{i=1}^{k+1} \binom{m_i+n_i}{n_i} \bigg/ \binom{m+k+n}{n}, \tag{3.3}$$

defined over (2.2). This distribution was also discussed in Särndal (1964), although in a somewhat different context. A univariate version of (3.3) is given, with different notation in Fisher (1956, p. 111).

The distribution (3.3) has a number of limiting distributions which are arrived at under varying assumptions on m, m_i, n and n_i $(i = 1, \ldots, k+1)$. We introduce the following notation for the distribution (3.3) and its limiting forms. Let

$$A = \prod_{i=1}^{k+1} \binom{m_i+n_i}{n_i} \bigg/ \binom{m+k+n}{n}$$

be the k-variate distribution (3.3). As a matter of convenience we shall refer to the integers m_i+1 $(i = 1, \ldots, k+1)$, rather than to the m_i themselves, as the parameters of this distribution. We may also consider

$$m+k+1 = \sum_{i=1}^{k+1} (m_i+1)$$

as a parameter. The random variables N_i take on values n_i which are restricted by (2.2). The first two moments are

$$E(N_i|\mathbf{m}) = n\pi_i,$$

$$\operatorname{var}(N_i|\mathbf{m}) = n^2\left(1+\frac{m+k+1}{n}\right)\frac{\pi_i(1-\pi_i)}{m+k+2},$$

$$\operatorname{cov}(N_i, N_j|\mathbf{m}) = -n^2\left(1+\frac{m+k+1}{n}\right)\frac{\pi_i\pi_j}{m+k+2},$$

where

$$\pi_i = (m_i+1)/(m+k+1). \tag{3.4}$$

Let

$$B' = \left(n!\Big/\prod_{i=1}^{k+1} n_i!\right)\prod_{i=1}^{k+1} p_i^{n_i}$$

be the k-variate multinomial distribution where n and p_i $(i = 1, \ldots, k+1)$ are parameters such that

$$0<p_i<1; \quad \sum_{i=1}^{k+1} p_i = 1.$$

The distribution is defined for non-negative integers n_i fulfilling (2.2).

Let

$$B_1 = \left\{(m+k)!\Big/\prod_{i=1}^{k+1} m_i!\right\}\prod_{i=1}^{k+1} y_i^{m_i}$$

denote the k-variate Dirichlet distribution. The integers m_i+1 $(i = 1, \ldots, k+1)$ are called the parameters of the distribution which is defined over the k-dimensional simplex

$$0\leqslant y_i\leqslant 1 \quad (i = 1, \ldots, k), \qquad \sum_{i=1}^{k} y_i\leqslant 1.$$

Let

$$A_0 = \prod_{i=1}^{k}\binom{m_i+n_i}{n_i}\lambda^{m_i+1}(1-\lambda)^{n_i}$$

be the joint probability of k independent discrete random variables N_i $(i = 1, \ldots, k)$, each having a negative binomial distribution with the same parameter λ, where $0<\lambda<1$, but possibly differing m_i. Again we shall refer to m_i+1 (rather than to m_i) as a parameter. The variables N_i take on values n_i restricted by

$$0\leqslant n_i<\infty \quad (i = 1, \ldots, k). \tag{3.5}$$

Let

$$C' = \prod_{i=1}^{k}\frac{\theta_i^{n_i}}{n_i!}e^{-\theta_i}$$

denote the joint probability of k independent discrete random variables

$$N_i \quad (i = 1, \ldots, k),$$

each having a Poisson distribution with parameter $\theta_i > 0$ $(i = 1, \ldots, k)$. The n_i are subject to (3.5).

Let

$$C_1 = \prod_{i=1}^{k} \frac{x_i^{m_i}}{m_i!} e^{-x_i}$$

be the joint density of k independent continuous random variables X_i $(i = 1, \ldots, k)$, each having a gamma distribution with parameter $m_i + 1$. The range of X_i is $0 < x_i < \infty$ $(i = 1, \ldots, k)$.

Let

$$D = (2\pi)^{-\frac{1}{2}k} |\mathbf{R}|^{\frac{1}{2}} \exp\{-\tfrac{1}{2}(\mathbf{z}-\boldsymbol{\mu})\mathbf{R}(\mathbf{z}-\boldsymbol{\mu})'\}$$

be the k-variate normal density for the random vector $\mathbf{Z} = (Z_1, \ldots, Z_k)$, where $\boldsymbol{\mu} = E(\mathbf{Z}) = (\mu_1, \ldots, \mu_k)$ is the mean vector and $\mathbf{R}^{-1} = \mathbf{V}$ is the $k \times k$ variance-covariance matrix of $Z_1, \ldots, Z_k$ with elements $v_{ii} = \text{var}(Z_i)$, $v_{ij} = \text{cov}(Z_i, Z_j)$ $(i \neq j)$. In particular, if $v_{ij} = 0$ $(i \neq j)$, let

$$D_0 = \prod_{i=1}^{k} (2\pi v_{ii})^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(z_i - \mu_i)^2 / v_{ii}\},$$

be the joint density of k independently distributed continuous random variables Z_i, each being normal (μ_i, v_{ii}).

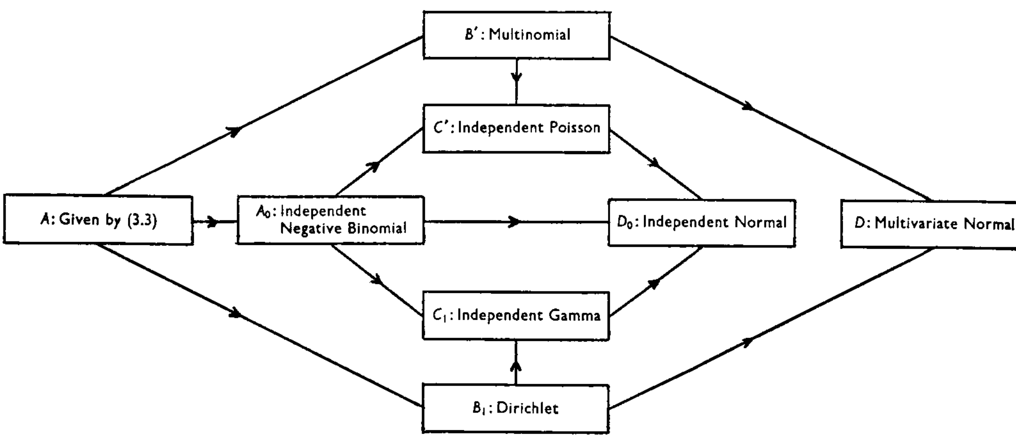

FIG. 1. Arrow scheme showing limiting relations between various k-variate distributions.

The first two moments of B', B_1, A_0, C' and C_1 are given in Table 2. The distributions A, B', B_1, A_0, C', C_1, D_0 and D are interconnected (in a certain limiting sense) in a way which is described by the arrow scheme shown in Fig. 1.

Under certain limiting conditions, the distribution of one box approaches (after possible multiplication by a constant) that of another box connected with the first one by a line and an arrow pointing toward the limit distribution. The proofs of all these relations are elementary, and need not be further discussed here, and, indeed, some of them provide standard material for many textbooks in elementary theoretical statistics. However, for completeness the exact nature of the limiting conditions represented by each arrow in Fig. 1 are listed in Tables 1 and 2. Table 1 also clarifies how the newly introduced parameters and random elements (i.e. p_i, λ, y_i, θ_i and x_i) depend in the limit upon the quantities m, m_i, n and n_i.

TABLE 1

Conditions for some limiting relations displayed in Fig. 1

No.	*Limiting relation*	*Conditions*	*Quantities remaining finite*
1	$A \to B'$	$m \to \infty$; $m_i/m \to p_i > 0$ $(i = 1, \ldots, k+1)$	n, n_i $(i = 1, \ldots, k+1)$
2	$A \to A_0$	$m, m_{k+1}, n, n_{k+1} \to \infty$; $m/(m+n) \to \lambda > 0$	m_i, n_i $(i = 1, \ldots, k)$
3	$n^k A \to B_1$	$n \to \infty$; $n_i/n \to y_i > 0$ $(i = 1, \ldots, k+1)$	m, m_i $(i = 1, \ldots, k+1)$
4	$B' \to C'$	$n, n_{k+1} \to \infty$; $p_i \to 0$; $np_i \to \theta_i > 0$ $(i = 1, \ldots, k)$	n_i $(i = 1, \ldots, k)$
5	$A_0 \to C'$	$m_i \to \infty$ $(i = 1, \ldots, k)$; $\lambda \to 1$; $m_i(1-\lambda)/\lambda \to \theta_i > 0$ $(i = 1, \ldots, k)$	n_i $(i = 1, \ldots, k)$
6	$\lambda^{-k} A_0 \to C_1$	$n_i \to \infty$ $(i = 1, \ldots, k)$; $\lambda \to 0$; $n_i \lambda/(1-\lambda) \to x_i > 0$ $(i = 1, \ldots, k)$	m_i $(i = 1, \ldots, k)$
7	$m^{-k} B_1 \to C_1$	$m, m_{k+1} \to \infty$; $y_i \to 0$; $my_i \to x_i > 0$ $(i = 1, \ldots, k)$	m_i $(i = 1, \ldots, k)$

TABLE 2

Conditions under which distributions B', C', A_0, C_1 and B_1 approach a multivariate normal distribution

Limiting relation	*Conditions*	*Elements of* $\mathbf{z}$, $\boldsymbol{\mu}$ *and* $\mathbf{V}$ $(i, j = 1, \ldots, k)$			
		z_i	μ_i	v_{ii}	v_{ij} $(i \neq j)$
$B' \to D$	$n, n_i \to \infty$ $(i = 1, \ldots, k+1)$; $n \sum_{i=1}^{k+1} \frac{(n_i/n - p_i)^3}{p_i^2} \to 0$	n_i	np_i	$np_i(1-p_i)$	$-np_i p_j$
$C' \to D_0$	$n_i \to \infty$, $\frac{(n_i - \theta_i)^3}{\theta_i^2} \to 0$ $(i = 1, \ldots, k)$	n_i	θ_i	θ_i	0
$A_0 \to D_0$	$m_i, n_i \to \infty$, $\frac{\{n_i - (m_i+1)(1-\lambda)/\lambda\}^3}{(m_i+1)^2} \to 0$ $(i = 1, \ldots, k)$	n_i	$\frac{(m_i+1)(1-\lambda)}{\lambda}$	$\frac{(m_i+1)(1-\lambda)}{\lambda^2}$	0
$C_1 \to D_0$	$m_i \to \infty$, $\frac{\{x_i - (m_i+1)\}^3}{(m_i+1)^2} \to 0$ $(i = 1, \ldots, k)$	x_i	m_i+1	m_i+1	0
$B_1 \to D$	$m, m_i \to \infty$ $(i = 1, \ldots, k+1)$; $(m+k+2) \sum_{i=1}^{k+1} \frac{(y_i - \pi_i)^3}{\pi_i^2} \to 0$	y_i	π_i	$\frac{\pi_i(1-\pi_i)}{m+k+2}$	$-\frac{\pi_i \pi_j}{m+k+2}$

In this table, π_i is given by (3.4).

4. A DISCUSSION OF THE DISTRIBUTIONS IN FIG. 1

A number of observations are readily made by studying the arrow scheme of Fig. 1 and the distributions associated with it.

(1) The set of distributions can be divided into four subsets each containing two distributions, the upper set (B', C'), the lower set (C_1, B_1), the left-hand set (A, A_0), and the right-hand set (D_0, D), such that the distributions of each subset are similar with respect to the behaviour of the quantities m_i and n_i $(i = 1, \ldots, k)$. For the left-hand set, the m_i and the n_i are finite, while for the right-hand set the m_i as well as the n_i are infinite. The upper set refers to the situation where the m_i are infinite but the n_i are finite, whereas for the lower set the n_i are infinite and the m_i finite.

(2) The distributions can be divided into an "outer set" consisting of A, B', B_1 and D and an "inner set" consisting of A_0, C', C_1 and D_0. The common characteristic of the outer set is the fact that its members are joint distributions of dependent random variables, whereas the inner set is made up of joint distributions of independent random variables.

Under (3) to (7) below, as well as in Section 5, a number of properties relating to the distributions of Fig. 1 are discussed. The majority of these properties are well known and often cited, and our main concern is to observe how they fit into the pattern set by Fig. 1.

Property (3) deals with the pair B', B_1, of the outer set, while (4) brings forth a corresponding property pertaining to the pair C', C_1 of the inner set.

(3) B' and B_1 are "mirror images" of each other in the following sense. If the N_i (given $P_i = p_i$) have the multinomial distribution B', and the parameters p_i are considered random variables P_i possessing a uniform prior distribution, then the posterior distribution of the P_i is a k-variate Dirichlet distribution of the type B_1. Conversely, if the variables Y_i (given $M_i = m_i$) have the distribution B_1 and the M_i are assumed to have a uniform prior, then their posterior distribution is of the type B'.

(4) Similarly, C' and C_1 are mirror images of each other. If N_i has a Poisson distribution with parameter θ_i, which is considered to be random having a uniform prior distribution, then the posterior distribution of this parameter is a gamma density. Or, conversely, if X_i (given the parameter $M_i+1 = m_i+1$) has a gamma distribution, and M_i is assumed to be random with a uniform prior, then the posterior distribution of M_i is Poisson.

Property (5) relates to the distributions of the outer set, while its counterpart for the inner set is stated as property (6).

(5) Combining B' and B_1 in a certain way leads to A. Let the distribution of the N_i (given $P_i = p_i$) be given by B'. Assume that the parameters P_i are of random nature having (prior) density B_1 with parameters m_i+1 $(i = 1, \ldots, k+1)$. Then the (marginal) joint density of the N_i is given by A. Furthermore, the posterior density of the P_i (after having observed the n_i) is a density B_1 with parameters

$$m_i+n_i+1 \; (i = 1, \ldots, k+1).$$

(In a Bayesian's language, the kernel of the prior density combines with the sample kernel in the same way that two sample kernels would combine; see Raiffa and Schlaifer (1961).)

(6) Combining C' and C_1 in a similar way leads to A_0. If N_i has a Poisson distribution with parameter $T_i = \theta_i$ and this parameter is considered to be random such that T_i/γ (where $\gamma > 0$ is an arbitrary scale factor) has a (prior) gamma density with parameter m_i+1, then the (marginal) distribution of N_i is negative binomial with

parameters m_i+1 and $\lambda = 1/(1+\gamma)$. Furthermore, the posterior density of T_i/γ is a gamma density with parameter m_i+1+n_i, a convolution corresponding to the one mentioned in (5).

Property (7) illustrates a relationship by which members of the inner set of k-variate distributions are turned into $(k-1)$-variate distributions belonging to the outer set.

(7) Given that the sum of the k variables in A_0, C' and C_1 is fixed, the joint density of the resulting $k-1$ non-redundant variables is given by $(k-1)$-variate correspondences of A, B' and B_1, respectively. We can express this in more specific terms as (7a)–(7c), where

$$m' = \sum_{i=1}^{k} m_i, \quad \theta' = \sum_{i=1}^{k} \theta_i.$$

(7a) If the N_i $(i = 1, \ldots, k)$ have the distribution A_0, and given that $\Sigma_1^k N_i = N'$ is fixed at, say, $N' = n'$, then $N_1, \ldots, N_{k-1}$ are distributed according to a $(k-1)$-variate version of A with parameters m_i+1 $(i = 1, \ldots, k)$, i.e.

$$\prod_{i=1}^{k} \binom{m_i+n_i}{n_i} \Big/ \binom{m'+k-1+n'}{n'}.$$

(7b) If the variables N_i $(i = 1, \ldots, k)$ are jointly distributed according to C', and given that $\Sigma_1^k N_i = N'$ equals n', then the resulting joint distribution of $N_1, \ldots, N_{k-1}$ is a $(k-1)$-variate version of B', namely the $(k-1)$-variate multinomial distribution with parameters n', θ_i/θ' $(i = 1, \ldots, k)$.

(7c) If the variables X_i $(i = 1, \ldots, k)$ are jointly distributed according to C_1, and $\Sigma_1^k X_i = X'$ is held fixed, then the joint density of the variables X_i/X' $(i = 1, \ldots, k-1)$ is a $(k-1)$-variate distribution of the type B_1, namely, a $(k-1)$-variate Dirichlet distribution with parameters m_i+1 $(i = 1, \ldots, k)$; $m'+k = \Sigma_1^k (m_i+1)$.

5. Additive Properties

A number of well-known additive (or convolution) properties relating to the multinomial, Poisson, negative binomial, gamma and Dirichlet distributions are considered in this Section. They can all be viewed in the light of an additivity property that pertains to the distribution (3.3).

Consider the set of integers $U = \{1, 2, \ldots, k+1\}$, and let q be an integer such that $1 \leqslant q \leqslant k$. Furthermore, consider an arbitrary division of U into $q+1$ mutually exclusive non-empty subsets $T_1, \ldots, T_{q+1}$. Let

$$\sum_{i \in T_j} m_i = m_{T_j}, \qquad \sum_{i \in T_j} n_i = n_T$$

be the class frequencies for a new system of $q+1$ classes which is obtained by pooling those classes i for which $i \in T_j$ $(j = 1, \ldots, q+1)$. Each of the new classes is then comprehending one or more of the old ones. Denote the number of elements in T_j by t_j. We have

$$t_j \geqslant 1 \quad (j = 1, \ldots, q+1); \qquad \sum_{j=1}^{q+1} t_j = k+1.$$

Let

$$\mathbf{m}_T = (m_{T_1}, \ldots, m_{T_{q+1}}), \quad \mathbf{n}_T = (n_{T_1}, \ldots, n_{T_{q+1}}),$$

be the observed vectors of class frequencies for the new system of classes, and let $\mathbf{M}_T$ and $\mathbf{N}_T$ be the corresponding vector random variables. The distribution of $\mathbf{N}_T$ given $\mathbf{M}_T = \mathbf{m}_T$ is (under the assumption of a Bose–Einstein prior) obtained from (3.3) as a q-variate version of A, namely

$$P(\mathbf{n}_T \,|\, \mathbf{m}_T) = \prod_{j=1}^{q+1} \binom{m_{T_j} + t_j - 1 + n_{T_j}}{n_{T_j}} \bigg/ \binom{m+k+n}{n}, \tag{5.1}$$

the parameters being $m_T + t_j$ $(j = 1, \ldots, q+1)$.

We shall consider this additive property in more detail for the case $q = 1$, and in this connection we introduce some notation that will be used in the remainder of this Section. Let T be an arbitrary subset of U containing t elements. Let i_0 be an arbitrary element in T, and let i_1 be an arbitrary element in the complement set $\bar{T} = U - T$. Also, let

$$T' = T - i_0, \quad \bar{T}' = \bar{T} - i_1,$$

$$n_T = \sum_{i \in T} n_i, \quad m_T = \sum_{i \in T} m_i.$$

The distribution of

$$N_T = \sum_{i \in T} N_i$$

(given $M_T = m_T$) is then obtained as

$$\Sigma P(\mathbf{n} \,|\, \mathbf{m}) = P(n_T \,|\, m_T), \tag{5.2}$$

where the summation is a multiple one extending over all non-negative n_i $(i \in T' + \bar{T}')$ such that

$$\sum_{i \in T'} n_i \leqslant n_T, \quad \sum_{i \in \bar{T}'} n_i \leqslant n - n_T.$$

The distribution of N_T (given $M_T = m_T$) is denoted by E, say, and is given by

$$E = P(n_T \,|\, m_T) = \binom{m_T + t - 1 + n_T}{n_T} \binom{m - m_T + k - t + n - n_T}{n - n_T} \bigg/ \binom{m+k+n}{n} \tag{5.3}$$

for $0 \leqslant n_T \leqslant n$. It is, of course, a special case of (5.1).

Consider again equation (5.2) and let us apply to both members those limiting conditions (as stated in Table 1) that turned the distribution $P(\mathbf{n} \,|\, \mathbf{m}) = A$ appearing in the left member into a k-variate multinomial, independent Poisson, independent negative binomial, independent gamma or Dirichlet distribution. (In Table 1 these conditions are found under (1), (4), (2), (7) and (3), respectively.) Letting these limiting conditions operate one by one on the right member of (5.2), i.e. on the distribution of N_T, we obtain in the limit a binomial, Poisson, negative binomial, gamma or beta distribution, respectively. Using the definitions of p_i, λ, y_i, θ_i and x_i introduced in Table 1, the limiting forms of E given by (5.3) turn out to be as follows (the figure in parentheses refers to the relation number in Table 1 under which the limiting conditions are stated).

$$(1)\qquad \binom{n}{n_T} p_T^{n_T}(1-p_T)^{n-n_T};\quad p_T = \lim \frac{m_T}{m} = \sum_{i\in T} p_i,$$

$$(4)\qquad \frac{\theta_T^{n_T}}{n_T!} e^{-\theta_T};\quad \theta_T = \lim np_T = \sum_{i\in T} \theta_i,$$

$$(2)\qquad \binom{m_T+t-1+n_T}{n_T} \lambda^{m_T+t}(1-\lambda)^{n_T};\quad m_T+t = \sum_{i\in T} (m_i+1),$$

$$(7)\qquad \frac{x_T^{m_T+t-1}}{(m_T+t-1)!} e^{-x_T};\quad x_T = \lim my_T = \sum_{i\in T} x_i,$$

$$(3)\qquad \frac{(m+k)!}{(m_T+t-1)!\,(m-m_T+k-t)!} y_T^{m_T+t-1}(1-y_T)^{m-m_T+k-t};\quad y_T = \lim \frac{n_T}{n} = \sum_{i\in T} y_i.$$

(Note that nE, rather than E itself, approaches the beta density of the last line. If this beta density is denoted by E_1, then $m^{-1} E_1$ approaches the gamma density associated with condition (7). In these two cases of continuous random variables, the summation in (5.2) becomes an integration over two simplexes of dimensions $t-1$ and $k-t$, respectively, after the limiting processes have been performed.)

A number of additive properties are thereby illustrated. For a sum of independent random variables N_i or X_i ($i \in T$) these are as follows.

(A1) If each N_i is Poisson distributed with parameter θ_i, then ΣN_i is Poisson distributed with parameter $\Sigma\theta_i$.

(A2) If each N_i is distributed according to a negative binomial distribution with parameters λ and m_i+1, then ΣN_i has a negative binomial distribution with parameters λ and $\Sigma(m_i+1)$.

(A3) If each X_i is gamma distributed with parameter m_i+1, then ΣX_i is gamma distributed with parameter $\Sigma(m_i+1)$.

For a sum of dependent random variables N_i or Y_i ($i \in T$), we have the following additive properties:

(B1) If the joint distribution of the N_i is multinomial such that the marginal distribution of each N_i is binomial with parameters n, p_i, then ΣN_i is binomial with parameters n, Σp_i.

(B2) If the joint distribution of the Y_i is a Dirichlet distribution such that the marginal distribution of each Y_i is a beta density with parameters m_i+1, $m+k+1-(m_i+1)$, i.e. the density is

$$f(y_i) = \frac{(m+k)!}{m_i!\,(m+k-m_i-1)!} y_i^{m_i}(1-y_i)^{m+k-(m_i+1)}\quad (0 \leqslant y_i \leqslant 1),$$

then ΣY_i has a beta distribution with parameters $\Sigma(m_i+1)$, $m+k+1-\Sigma(m_i+1)$.

Finally, by the connections with the normal distributions brought forth by Fig. 1, it is clear that similar additive properties can be formulated for a sum of normally distributed random variables.

6. Maxwell–Bolzmann Prior Distribution

This final Section gives one brief illustration of a situation in which a non-uniform prior is chosen for S.

Consider s balls that are distinguishable so that, for instance, each ball is labelled with an integer from 1 to s. Each of the balls is assigned at random one out of $k+1$ different colours. The number of different assignments is $(k+1)^s$, and, under Maxwell–Bolzmann statistics (see Feller, 1957, p. 38) all of them are equally likely. If s_i is the number of balls having the ith colour, the probability of the frequency configuration $\mathbf{s} = (s_1, \ldots, s_k)$ then becomes

$$P(\mathbf{s}) = \left(s! \Big/ \prod_{i=1}^{k+1} s_i!\right)(k+1)^{-s}$$

defined over (3.2). This we shall call a Maxwell–Bolzmann prior; clearly it is a multinomial with parameters s and equal $p_i = 1/(k+1)$ $(i = 1, \ldots, k+1)$. Using it as a prior distribution for $\mathbf{S}$ in (2.3), we find

$$P(\mathbf{s}, \mathbf{m}, \mathbf{n}) = \frac{(s-m-n)!}{\prod_{i=1}^{k+1}(s_i - m_i - n_i)!} \frac{m!}{\prod_{i=1}^{k+1} m_i!} \frac{n!}{\prod_{i=1}^{k+1} n_i!}(k+1)^{-s}.$$

Eliminating s by summing over (2.4), we obtain

$$P(\mathbf{m}, \mathbf{n}) = \frac{m!}{\prod_{i=1}^{k+1} m_i!} \frac{n!}{\prod_{i=1}^{k+1} n_i!}(k+1)^{-(m+n)}.$$

Consequently **M** and **N** are independently distributed random vectors, each having a Maxwell–Bolzmann type marginal distribution. In particular,

$$P(\mathbf{n}) = P(\mathbf{n} \mid \mathbf{m}) = \left(n! \Big/ \prod_{i=1}^{k+1} n_i\right)(k+1)^{-n},$$

i.e. a multinomial distribution with parameters n and $p_i = 1/(k+1)$ $(i = 1, \ldots, k+1)$.

Choosing a Maxwell–Bolzmann prior distribution for **S** thus results in a situation in which the first sample does not provide any information towards the outcome of the second sample.

ACKNOWLEDGEMENT

This research was supported by the United States Army Research Office, Durham, North Carolina.

REFERENCES

FELLER, W. (1957), *An Introduction to Probability Theory and Its Applications, Vol. I*, 2nd ed. New York: Wiley.

FISHER, R. A. (1956), *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.

RAIFFA, H. and SCHLAIFER, R. (1961), *Applied Statistical Decision Theory*. Boston: Harvard University Graduate School of Business Administration.

SÄRNDAL, C-E. (1964). "A unified derivation of some nonparametric distributions", *J. Amer. statist. Ass.*, **59**, 1042–53.