JOURNAL OF MULTIVARIATE ANALYSIS 32, 256–268 (1990)

Some Representations of the Multivariate Bernoulli and Binomial Distributions

JOZEF L. TEUGELS

Katholieke Universiteit Leuven, B-3030 Heverlee, Belgium

Communicated by the Editors

Multivariate but vectorized versions for Bernoulli and binomial distributions are established using the concept of Kronecker product from matrix calculus. The multivariate Bernoulli distribution entails a parameterized model, that provides an alternative to the traditional log-linear model for binary variables.

1. INTRODUCTION

Assume $\{X_i,\ i=1, 2, ..., n\}$ is a sequence of Bernoulli random variables, i.e., for $i=1, 2, ..., n$,

$$P\{X_i=0\}=q_i, \qquad P\{X_i=1\}=p_i,$$

where $0<p_i=1-q_i<1$. Note that $EX_i=1-q_i<1$.

We look for an algebraically convenient representation for the *multivariate Bernoulli distribution*

$$p_{k_1,k_2,...,k_n}:=P\{X_1=k_1, X_2=k_2, ..., X_n=k_n\}, \tag{1.1}$$

where we assume that $k_i \in \{0, 1\}$, $i=1, 2, ..., n$. The representation should be amply parameterized by the n mean values $\{p_i, i=1, 2, ..., n\}$ and by other parameters expressing the possible dependencies among the $\{X_i, i=1, 2, ..., n\}$.

EXAMPLE 1.1. Let $n=2$. Here $p_{00}=P\{X_1=0, X_2=0\}$, $p_{10}=P\{X_1=1, X_2=0\}$, $p_{01}=P\{X_1=0, X_2=1\}$, and $p_{11}=P\{X_1=1, X_2=1\}$. Furthermore $p_{00}+p_{10}+p_{01}+p_{11}=1$ so that the distribution can be

Received February 13, 1989; revised August 16, 1989.

AMS 1980 subject classifications: 62E20, 62H17.

Key words and phrases: multivariate Bernoulli distribution, multivariate binomial distribution, log-linear models, categorical data.

0047-259X/90 $3.00

characterized by three parameters. However there are three "natural" parameters for $n=2$, i.e.,

$$p_1 = EX_1 = P\{X_1 = 1\}$$
$$p_2 = EX_2 = P\{X_2 = 1\}$$
$$\sigma_{12} := E(X_1 - p_1)(X_2 - p_2).$$

Alternatively we can use $\mu_{12} := EX_1X_2 = \sigma_{12} + p_1 \cdot p_2 = p_{11}$. Solving for p_{00}, p_{10}, p_{01}, and p_{11} we obtain the following two representations

$$\begin{aligned}
p_{00} &= q_1q_2 + \sigma_{12} = 1 - p_1 - p_2 + \mu_{12} \\
p_{10} &= p_1q_2 - \sigma_{12} = p_1 - \mu_{12} \\
p_{01} &= q_1p_2 - \sigma_{12} = p_2 - \mu_{12} \\
p_{11} &= p_1p_2 + \sigma_{12} = \mu_{12}.
\end{aligned}$$

Notice that X_1 and X_2 will be independent iff $\sigma_{12} = 0$; the first representation nicely separates the independence part from the dependency quantity σ_{12}. We can cast the representation in matrix form:

$$\begin{bmatrix} p_{00} \\ p_{10} \\ p_{01} \\ p_{11} \end{bmatrix} = \begin{bmatrix} q_1q_2 & -q_2 & -q_1 & 1 \\ p_1q_2 & q_2 & -p_1 & -1 \\ q_1p_2 & -p_2 & q_1 & -1 \\ p_1p_2 & p_2 & p_1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \sigma_{12} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & -1 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ p_1 \\ p_2 \\ \mu_{12} \end{bmatrix}.$$

The latter can in turn be rewritten in terms of *Kronecker products*. Put $\mathbf{p}^{(2)}$ for the vector on the left; then we have

$$\mathbf{p}^{(2)} = \begin{bmatrix} q_2 & -1 \\ p_2 & 1 \end{bmatrix} \otimes \begin{bmatrix} q_1 & -1 \\ p_1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \sigma_{12} \end{bmatrix}$$

and

$$\mathbf{p}^{(2)} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ p_1 \\ p_2 \\ \mu_{12} \end{bmatrix}.$$

EXAMPLE 1.2. Let $n=3$. Put

$$\mathbf{p}^{(3)}=(p_{000}, p_{100}, p_{010}, p_{110}, p_{001}, p_{101}, p_{011}, p_{111})^{\mathrm{T}}$$

where $(\ )^{\mathrm{T}}$ stands for the transpose. We need seven parameters to characterize this distribution. On the other hand, seven natural parameters are

$$\begin{aligned}
&p_i = EX_i, \qquad i=1,2,3\\
&\sigma_{ij}=E(X_i-p_i)(X_j-p_j), \qquad (i,j)\in\{(1,2),(1,3),(2,3)\},\\
&\theta := E(X_1-p_1)(X_2-p_2)(X_3-p_3).
\end{aligned}$$

In the same way as above one can find the following representation

$$\begin{aligned}
p_{000}&=q_1q_2q_3+q_3\sigma_{12}+q_2\sigma_{13}+q_1\sigma_{23}-\theta\\
p_{100}&=p_1q_2q_3-q_3\sigma_{12}-q_2\sigma_{13}+p_1\sigma_{23}+\theta\\
p_{010}&=q_1p_2q_3-q_3\sigma_{12}+p_2\sigma_{13}-q_1\sigma_{23}+\theta\\
p_{110}&=p_1p_2q_3+q_3\sigma_{12}-p_2\sigma_{13}-p_1\sigma_{23}-\theta\\
p_{001}&=q_1q_2p_3+p_3\sigma_{12}-q_2\sigma_{13}-q_1\sigma_{23}+\theta\\
p_{101}&=p_1q_2p_3-p_3\sigma_{12}+q_2\sigma_{13}-p_1\sigma_{23}-\theta\\
p_{011}&=q_1p_2p_3-p_3\sigma_{12}-p_2\sigma_{13}+q_1\sigma_{23}-\theta\\
p_{111}&=p_1p_2p_3+p_3\sigma_{12}+p_2\sigma_{13}+p_1\sigma_{23}+\theta.
\end{aligned}$$

Again the Kronecker form can be given

$$\mathbf{p}^{(3)}=\begin{bmatrix} q_3 & -1\\ p_3 & 1\end{bmatrix}\otimes\begin{bmatrix} q_2 & -1\\ p_2 & 1\end{bmatrix}\otimes\begin{bmatrix} q_1 & -1\\ p_1 & 1\end{bmatrix}\cdot(1,0,0,\sigma_{12},0,\sigma_{13},\sigma_{23},\theta)^{\mathrm{T}}.$$

The latter expression seems easily generalizable.

2. THE MULTIVARIATE BERNOULLI DISTRIBUTION

2.1. *Derivation*

We start by introducing a general notation for $\mathbf{p}^{(n)}$, a vector containing 2^n components. For $1\leqslant k\leqslant 2^n$ write k in a *binary expansion*, i.e.,

$$k=1+\sum_{i=1}^{n} k_i 2^{i-1}, \tag{2.1}$$

where $k_i\in\{0,1\}$. This expansion induces a 1–1 correspondence

$$k\leftrightarrow(k_1, k_2, ..., k_n)$$

so that with (1.1)

$$p_k^{(n)} = p_{k_1,k_2,\ldots,k_n}, \qquad 1 \leqslant k \leqslant 2^n.$$

If we introduce

$$\bar{X}_i = 1 - X_i, \qquad i = 1, 2, \ldots, n,$$

then we can write

$$p_k^{(n)} = P\left\{\bigcap_{i=1}^{n} [X_i = k_i]\right\} = E\left\{\prod_{i=1}^{n} X_i^{k_i} \bar{X}_i^{1-k_i}\right\}.$$

The expression on the right can be considered as an element from a Kronecker product. In general we have the following formula for the Kronecker product of 2×1 vectors:

$$\left[\begin{bmatrix} a_n \\ b_n \end{bmatrix} \otimes \begin{bmatrix} a_{n-1} \\ b_{n-1} \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}\right]_k = \prod_{i=1}^{n} a_i^{1-k_i} b_i^{k_i}, \qquad 1 \leqslant k \leqslant 2^n, \tag{2.2}$$

where k is given by (2.1). Putting $a_i = \bar{X}_i$ and $b_i = X_i$, we obtain the starting formula for $\mathbf{p}^{(n)}$,

$$\mathbf{p}^{(n)} = E\left[\begin{bmatrix} \bar{X}_n \\ X_n \end{bmatrix} \otimes \begin{bmatrix} \bar{X}_{n-1} \\ X_{n-1} \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} \bar{X}_1 \\ X_1 \end{bmatrix}\right]. \tag{2.3}$$

We want to express $\mathbf{p}^{(n)}$ in terms of the vector of *ordinary moments*,

$$\boldsymbol{\mu}^{(n)} = (\mu_1^{(n)}, \mu_2^{(n)}, \ldots, \mu_{2^n}^{(n)})^{\mathrm{T}},$$

where

$$\mu_k^{(n)} = E\left(\prod_{i=1}^{n} X_i^{k_i}\right) = E\left[\begin{bmatrix} 1 \\ X_n \end{bmatrix} \otimes \begin{bmatrix} 1 \\ X_{n-1} \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ X_1 \end{bmatrix}\right]_k \tag{2.4}$$

and k is given by (2.1). Relation (2.4) follows from (2.2) by the choice $a_i = 1$, $b_i = X_i$.

Another representation makes use of the *central moments* expressed in the *dependency vector*

$$\boldsymbol{\sigma}^{(n)} = (\sigma_1^{(n)}, \sigma_2^{(n)}, \ldots, \sigma_{2^n}^{(n)})^{\mathrm{T}},$$

where

$$\sigma_k^{(n)} = E\left(\prod_{i=1}^{n} (X_i - p_i)^{k_i}\right) = E\left[\begin{bmatrix} 1 \\ Y_n \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{n-1} \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ Y_1 \end{bmatrix}\right]_k$$

and $Y_i = X_i - p_i$, $i = 1, 2, \ldots, n$.

Both representations are combined in our first result.

THEOREM 1. *With* $\mathbf{p}^{(n)}$, $\boldsymbol{\mu}^{(n)}$, *and* $\boldsymbol{\sigma}^{(n)}$ *defined above, we have*:

(i)

$$\mathbf{p}^{(n)}=\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}^{\otimes n}\boldsymbol{\mu}^{(n)} \tag{2.5}$$

and

$$\mu^{(n)}=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}^{\otimes n}\mathbf{p}^{(n)}; \tag{2.6}$$

(ii)

$$\mathbf{p}^{(n)}=\begin{bmatrix}q_n & -1\\ p_n & 1\end{bmatrix}\otimes\begin{bmatrix}q_{n-1} & -1\\ p_{n-1} & 1\end{bmatrix}\otimes\cdots\otimes\begin{bmatrix}q_1 & -1\\ p_1 & 1\end{bmatrix}\boldsymbol{\sigma}^{(n)} \tag{2.7}$$

and

$$\boldsymbol{\sigma}^{(n)}=\begin{bmatrix}1 & 1\\ -p_n & q_n\end{bmatrix}\otimes\begin{bmatrix}1 & 1\\ -p_{n-1} & q_{n-1}\end{bmatrix}\otimes\cdots\otimes\begin{bmatrix}1 & 1\\ -p_1 & q_1\end{bmatrix}\mathbf{p}^{(n)}. \tag{2.8}$$

Proof. (i) Note that

$$\begin{bmatrix}\bar{X}_i\\ X_i\end{bmatrix}=\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}\begin{bmatrix}1\\ X_i\end{bmatrix}, \qquad i=1, 2, ..., n.$$

Filling this into (2.3), we have

$$\mathbf{p}^{(n)}=E\left[\left[\begin{matrix}1 & -1\\ 0 & 1\end{matrix}\right]\right]\otimes\begin{bmatrix}1\\ X_n\end{bmatrix}\cdot\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}$$
$$\otimes\begin{bmatrix}1\\ X_{n-1}\end{bmatrix}\cdots\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}\otimes\left[\begin{bmatrix}1\\ X_1\end{bmatrix}\right].$$

Apply the *mixed product rule* of matrix calculus [7],

$$(A\otimes B)\cdot(C\otimes D)=(A\cdot C)\otimes(B\cdot D), \tag{2.9}$$

$n-1$ times to obtain

$$\mathbf{p}^{(n)}=E\left[\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}^{\otimes n}\begin{bmatrix}1\\ X_n\end{bmatrix}\otimes\begin{bmatrix}1\\ X_{n-1}\end{bmatrix}\otimes\cdots\otimes\begin{bmatrix}1\\ X_1\end{bmatrix}\right].$$

By an application of the linearity of E we find (2.5) by (2.4). The other formula derives from the rule of matrix calculus [7],

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \tag{2.10}$$

and easy algebra.

(ii) Clearly

$$\begin{bmatrix} 1 \\ X_i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ p_i & 1 \end{bmatrix} \begin{bmatrix} 1 \\ Y_i \end{bmatrix}, \qquad i = 1, 2, ..., n,$$

and hence,

$$\begin{bmatrix} \bar{X}_i \\ X_i \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ X_i \end{bmatrix} = \begin{bmatrix} q_i & -1 \\ p_i & 1 \end{bmatrix} \begin{bmatrix} 1 \\ Y_i \end{bmatrix}.$$

The remaining calculations are as in case (i). ▮

A slightly different proof of (ii) could be based on the following formulas, linking $\boldsymbol{\mu}^{(n)}$ and $\boldsymbol{\sigma}^{(n)}$:

$$\boldsymbol{\mu}^{(n)} = \begin{bmatrix} 1 & 0 \\ p_n & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ p_{n-1} & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ p_1 & 1 \end{bmatrix} \boldsymbol{\sigma}^{(n)}$$

and

$$\boldsymbol{\sigma}^{(n)} = \begin{bmatrix} 1 & 0 \\ -p_n & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ -p_{n-1} & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ -p_1 & 1 \end{bmatrix} \boldsymbol{\mu}^{(n)}.$$

For future reference we mention another property of the Kronecker product [7],

$$(A \otimes B)^{\mathrm{T}} = A^{\mathrm{T}} \otimes B^{\mathrm{T}}. \tag{2.11}$$

The above formulas easily lead to expressions for the *marginal distributions.*

EXAMPLE 2.1. Let again $n = 3$. Then the three possible marginal distributions containing two of the three r.v.'s are given by (I_m is an $m \times m$ identity matrix):

(i) $(1, 1) \otimes I_2 \otimes I_2 \mathbf{p}^{(3)}$ for X_1 and X_2,

(ii) $I_2 \otimes (1, 1) \otimes I_2 \mathbf{p}^{(3)}$ for X_1 and X_3,

(iii) $I_2 \otimes I_2 \otimes (1, 1) \mathbf{p}^{(3)}$ for X_2 and X_3.

2.2. *Generating Function*

We now derive a number of equivalent expressions for the (*probability*) *generating function* of the multivariate Bernoulli distribution

$$\varphi(z_1, z_2, ..., z_n) = E(z_1^{X_1} z_2^{X_2} \cdots z_n^{X_n}),$$

where $z_1, z_2, ..., z_n$ are n (complex) numbers.

THEOREM 2. (i) $\varphi(z_1, z_2, ..., z_n) = \sum_{k=1}^{2^n} \omega_k(z_1, z_2, ..., z_n) \mu_k^{(n)}$, *where k is given by* (2.1) *and*

$$\omega_k(z_1, z_2, ..., z_n) = \prod_{i=1}^{n} (z_i - 1)^{k_i}, \qquad 1 \leqslant k \leqslant 2^n;$$

(ii) $\varphi(z_1, z_2, ..., z_n) = \sum_{k=1}^{2^n} \theta_k(z_1, z_2, ..., z_n) \sigma_k^{(n)}$, *where k is given by* (2.1) *and*

$$\theta_k(z_1, z_2, ..., z_n) = \prod_{i=1}^{n} (q_i + p_i z_i)^{1-k_i} (z_i - 1)^{k_i}, \qquad 1 \leqslant k \leqslant 2^n.$$

Proof. By the fact that $X_i \in \{0, 1\}$ we see that

$$z_i^{X_i} = \bar{X}_i + z_i X_i = (1, z_i) \begin{pmatrix} \bar{X}_i \\ X_i \end{pmatrix}.$$

Then

$$\varphi(z_1, z_2, ..., z_n) = E\left(\prod_{i=1}^{n} (1, z_i) \cdot \begin{pmatrix} \bar{X}_i \\ X_i \end{pmatrix} \right)$$

$$= E\left((1, z_n) \begin{pmatrix} \bar{X}_n \\ X_n \end{pmatrix} \otimes \cdots \otimes (1, z_1) \begin{pmatrix} \bar{X}_1 \\ X_1 \end{pmatrix} \right).$$

Hence by the mixed product rule and (2.3) we have

$$\varphi(z_1, z_2, ..., z_n) = (1, z_n) \otimes \cdots \otimes (1, z_1) \cdot \mathbf{p}^{(n)}.$$

(i) Use this relation together with (2.5) and the mixed product rule again to find

$$\varphi(z_1, z_2, ..., z_n) = (1, z_n) \cdot \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \otimes \cdots \otimes (1, z_1) \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \boldsymbol{\mu}^{(n)}.$$

But

$$(1, z_i) \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = (1, z_i - 1), \qquad i = 1, 2, ..., n$$

and so

$$\varphi(z_1, z_2, ..., z_n) = (1, z_n - 1) \otimes \cdots \otimes (1, z_1 - 1)\boldsymbol{\mu}^{(n)}.$$

To the last formula we apply the transposed version of (2.2) with $a_i = 1$, $b_i = z_i - 1$, to obtain the first formula.

(ii) The proof is entirely similar. ▮

EXAMPLE 2.2. For $n = 2$ we have the formulas,

$$\varphi(z_1, z_2) = \begin{cases} p_{00} + z_1 p_{10} + z_2 p_{01} + z_1 z_2 p_{11}, \\ 1 + (z_1 - 1) p_1 + (z_2 - 1) p_2 + (z_1 - 1)(z_2 - 1)\mu_{11}, \\ (q_1 + p_1 z_1)(q_2 + p_2 z_2) + (z_1 - 1)(z_2 - 1)\sigma_{11}. \end{cases}$$

The last expression best refers to the case of eventual independence, when $\varphi(z_1, z_2) = \varphi(z_1, 1)\,\varphi(1, z_2)$ iff $\sigma_{11} = 0$.

2.3. *Some Comments on the Multivariate Bernoulli Distribution*

(i) The representation (2.5) in Theorem 1 contains $2^n - 1$ parameters since $\mu_1^{(n)} = 1$. The number of parameters in (2.7) is also $2^n - 1$; $2^n - n - 1$ are obtained in $\boldsymbol{\sigma}^{(n)}$, since $\sigma_k^{(n)} = 0$ whenever $k_1 + k_2 + \cdots + k_n = 1$; the n remaining parameters are, of course, $\{p_i, 1 \leqslant i \leqslant n\}$. The representation (2.7) can be written as $\mathbf{p}^{(n)} = A_n \boldsymbol{\sigma}^{(n)}$; here $A_n \cdot \mathbf{e}_1^{(n)}$ is the representation of $\mathbf{p}^{(n)}$ under the assumption of *independence* of all r.v. $\{X_i, i = 1, 2, ..., n\}$. In general the non-null components of $\boldsymbol{\sigma}^{(n)}$ express in a transparent way the $2^n - n - 1$ possible dependencies that might exist between any subset (of at least 2) of the r.v. $\{X_i, i = 1, 2, ..., n\}$. The term dependence vector seems therefore appropriate.

(ii) In the analysis of cross-tabulated data one often uses *log-linear models*. See, for example, [3, 12]. For $n = 3$ the fully saturated model has the following log-linear representation in terms of Kronecker products; for $n = 2$, see [4];

$$\mathbf{v}^{(3)} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{\otimes 3} \boldsymbol{\lambda}^{(3)},$$

where $v_{ijk} = \log p_{ijk}$, $i, j, k \in \{0, 1\}$ and where

$$(\boldsymbol{\lambda}^{(3)})^{\mathrm{T}} = (\mu, \lambda_1, \lambda_2, \lambda_{12}, \lambda_3, \lambda_{13}, 0, \varphi).$$

Compare this representation with that of Example 1.2. The number of parameters is also seven, but the direct interpretation is different.

The representation $\mathbf{p}^{(n)} = A_n \boldsymbol{\sigma}^{(n)}$ allows $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0$ without $\theta = 0$; hence there is no need for hierarchical structures like with the log-linear representation. Also no particular problems arise when one or more of p_{ijk} happen to be zero, a situation hardly possible in the log-linear case.

Let us stress that the multivariate Bernoulli distribution only uses 0–1 variables while log-linear models have a much wider applicability.

(iii) To express the dependence between the different r.v.'s, a variety of *measures of association* can be defined. Let us restrict attention to the case $n=3$. A tedious calculation using the notation of Example 1.2 shows that

$$p_{111}p_{001}-p_{101}p_{011}=\sigma_{12}p_3^2-\sigma_{13}\sigma_{23}+p_3\theta$$
$$p_{110}p_{000}-p_{100}p_{010}=\sigma_{12}q_3^2-\sigma_{13}\sigma_{23}-q_3\theta.$$

The quantities [3, 12]

$$\frac{p_{111}p_{001}}{p_{101}p_{011}},\qquad \frac{p_{110}p_{000}}{p_{100}p_{010}}$$

are often used to describe the interaction among X_1, X_2, and X_3. From our model it seems natural to look at differences rather than at ratios. For example, X_1 and X_2 will be conditionally independent, given X_3, iff the two ratios are one, or iff

$$\sigma_{12}p_3^2-\sigma_{13}\sigma_{23}+p_3\theta=0$$
$$\sigma_{12}q_3^2-\sigma_{13}\sigma_{23}-q_3\theta=0.$$

This can be rewritten in the form

$$\sigma_{12}(p_3-q_3)+\theta=0$$
$$\rho_{13}\cdot\rho_{23}=\rho_{12},$$

where $\rho_{ij}=\sigma_{ij}/\sqrt{p_iq_ip_jq_j}$, $i, j\in\{1, 2, 3\}$, are the *correlation coefficients.*

3. The Multivariate Binomial Distribution

Our starting point is the multivariate Bernoulli distribution where $p_i=EX_i$ and $\boldsymbol{\sigma}^{(n)}$ is the dependency vector. We take a sample of size m from $(X_1, X_2, ..., X_n)$, say,

$$\{(X_{1i}, X_{2i}, ..., X_{ni}), i=1, 2, ..., m\};$$

we form the multivariate sum

$$(S_{1m}, S_{2m}, ..., S_{nm}),$$

where $S_{im} = X_{i1} + \cdots + X_{im}$, $i = 1, 2, ..., n$. We write for the joint distribution of $(S_{1m}, ..., S_{nm})$ where $0 \leqslant r_i \leqslant m$, $1 \leqslant i \leqslant n$,

$$\pi^{(m)}_{r_1, r_2, ..., r_n} = P\{S_{1m} = r_1, S_{2m} = r_2, ..., S_{nm} = r_n\}.$$

To vectorize this n-dimensional scheme we write

$$k = 1 + \sum_{i=1}^{n} r_i (m+1)^{i-1}, \qquad r_i \in \{0, 1, ..., m\}.$$

Then we associate with the n-tuple $(r_1, r_2, ..., r_n)$ the kth element of a vector

$${}_n\mathbf{q}^{(m)} := ({}_nq_1^{(m)}, {}_nq_2^{(m)}, ..., {}_nq_{(m+1)^2}^{(m)})^{\mathrm{T}}$$

by the 1–1 correspondence,

$${}_nq_k^{(m)} := \pi^{(m)}_{r_1, r_2, ..., r_n}, \qquad 1 \leqslant k \leqslant (m+1)^n.$$

The vector ${}_n\mathbf{q}^{(m)}$ will be called the (vectorized) *multivariate binomial distribution.*

The fundamental relationship between $\pi^{(m)}$ and ${}_n\mathbf{q}^{(m)}$ is most easily seen by looking at the generating function. The proof of the following lemma can be obtained by induction.

LEMMA 3. *For* $m \in \{1, 2, ...\}$,

$$\begin{aligned} \varphi^m(z_1, z_2, ..., z_n) &= \sum_{r_1=0}^{m} z_1^{r_1} \sum_{r_2=0}^{m} z_2^{z_2} \cdots \sum_{r_n=0}^{m} z_n^{r_n} \pi^{(m)}_{r_1, r_2, ..., r_n} \\ &= (\mathbf{v}_n^{(m)} \otimes \mathbf{v}_{n-1}^{(m)} \otimes \cdots \otimes \mathbf{v}_1^{(m)})^{\mathrm{T}} \, {}_n\mathbf{q}^{(m)}, \end{aligned} \tag{3.1}$$

where for $i \in \{1, 2, ..., n\}$ *and* $m \geqslant 1$, $\mathbf{v}_i^{(m)} := (1, z_i, ..., z_i^m)^{\mathrm{T}}$. *Note that for any* $i \in \{1, 2, ..., n\}$,

$$(q_i + p_i z_i)(\mathbf{v}_i^{(m)})^{\mathrm{T}} = (\mathbf{v}_i^{(m+1)})^{\mathrm{T}} \cdot Q_{i,m+1}, \tag{3.2}$$

where

$$Q_{i,m+1} = \left| \begin{array}{cccccc} q_i & 0 & 0 & \cdots & 0 & 0 \\ p_i & q_i & 0 & \cdots & 0 & 0 \\ 0 & p_i & q_i & \cdots & 0 & 0 \\ & \cdots & & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & p_i & q_i \\ 0 & 0 & 0 & \cdots & 0 & p_i \end{array} \right| \tag{3.3}$$

is a $(m+2)\times(m+1)$ *matrix, depending solely on* p_i. *In a similar manner,*

$$(z_i-1)(\mathbf{v}_i^{(m)})^{\mathrm{T}}=(\mathbf{v}_i^{(m+1)})^{\mathrm{T}}\,J_{m+1},\tag{3.4}$$

where

$$J_{m+1}=\begin{vmatrix} -1 & 0 & 0 & \cdots & 0 & 0\\ 1 & -1 & 0 & \cdots & 0 & 0\\ 0 & 1 & -1 & \cdots & 0 & 0\\ & \cdots & & & \cdots & \cdots\\ 0 & 0 & 0 & \cdots & 1 & -1\\ 0 & 0 & 0 & \cdots & 0 & 1 \end{vmatrix}.\tag{3.5}$$

We now derive a vector form for ${}_n\mathbf{q}^{(m)}$.

THEOREM 4. *For any* $n\geqslant 2$ *and* $m\geqslant 1$ *the multivariate binomial distribution is given by*

$${}_n\mathbf{q}^{(m)}=V_m^{(n)}\cdot V_{m-1}^{(n)}\cdot\cdots\cdot V_1^{(n)},$$

where

$$V_i^{(n)}=\sum_{k=1}^{2^n}\sigma_k^{(n)}R_{n,i}^{(k)}\otimes\cdots\otimes R_{1,i}^{(k)}$$

and

$$R_{m,i}^{(k)}=\begin{cases}Q_{i,m} & \text{if } k_i=0\\ J_m & \text{if } k_i=1;\end{cases}$$

the matrices $Q_{i,m}$ *and* J_m *are given by* (3.3) *and* (3.5), *respectively.*

Proof. By the independence of the elements of the sample,

$$\varphi^{m+1}(z_1,z_2,...,z_n)=\varphi(z_1,z_2,...,z_n)\,\varphi^m(z_1,z_2,...,z_n).\tag{3.6}$$

Now Theorem 2(ii) gives us a relationship allowing us to write $\varphi(z_1,...,z_n)$ in terms of the elements of the dependency vector $\boldsymbol{\sigma}^{(n)}$. Combine (3.6) and (ii) of Theorem 2 to obtain

$$(\mathbf{v}_n^{(m+1)}\otimes\cdots\otimes\mathbf{v}_1^{(m+1)})^{\mathrm{T}}\,{}_n\mathbf{q}^{(m+1)}=\sum_{k=1}^{2^n}\sigma_k^{(n)}\theta_k(z_1,...,z_n)\times(\mathbf{v}_n^{(m)}\otimes\cdots\otimes\mathbf{v}_1^{(m)})^{\mathrm{T}}\,{}_n\mathbf{q}^{(m)}.\tag{3.7}$$

Apply (2.11) and Theorem 2(ii) to write

$$\theta_k(z_1,z_2,...,z_n)(\mathbf{v}_n^{(m)})^{\mathrm{T}}\otimes\cdots\otimes(\mathbf{v}_1^{(m)})^{\mathrm{T}}=w_n\otimes w_{n-1}\otimes\cdots\otimes w_1,\tag{3.8}$$

where

$$w_i := (q_i + p_i z_i)^{1-k_i} (z_i - 1)^{k_i} (\mathbf{v}_i^{(m)})^{\mathrm{T}}.$$

If $k_i = 0$, then (3.2) applies; if $k_i = 1$, (3.4) applies. By the definition of $R_{m,i}^{(k)}$ we combine both statements into the single

$$w_i = (\mathbf{v}_i^{(m+1)})^{\mathrm{T}} R_{m+1,i}^{(k)}. \tag{3.9}$$

Combine (3.7)–(3.9) with the mixed product rule to find

$${}_n\mathbf{q}^{(m+1)} = \sum_{k=1}^{2^n} \sigma_k^{(n)} R_{m+1,n}^{(k)} \otimes \cdots \otimes R_{m+1,1}^{(k)} \, {}_n\mathbf{q}^{(m)} = V_{m+1}^{(n)} \, {}_n\mathbf{q}^{(m)}.$$

An easy induction completes the proof. ∎

For the special case of the *bivariate binomial distribution* one can replace Lemma 3 by the simpler formula

$${}_2\mathbf{q}^{(m)} = \operatorname{vec} \pi^{(m)},$$

where $\pi^{(m)}$ is the matrix with elements $\pi_{r_1,r_2}^{(m)}$ and *vec* refers to the vec-operator introduced by Neudecker in [11] and utilized in [7, 8]. There results the expression

$$\begin{aligned} {}_2\mathbf{q}^{(m)} = {} & (Q_{2,m} \otimes Q_{1,m} + \sigma J_1^{\otimes 2}) \cdots (Q_{2,2} \otimes Q_{1,2} + \sigma J_2^{\otimes 2}) \\ & \times (Q_{2,1} \otimes Q_{1,1} + \sigma J_1^{\otimes 2}), \end{aligned}$$

where $Q_{i,m}$ and J_m are defined by (3.3) and (3.5), respectively.

Note that

$$\begin{aligned} Q_{2,1} \otimes Q_{1,1} + \sigma J_1^{\otimes 2} &= \begin{pmatrix} q_2 \\ p_2 \end{pmatrix} \otimes \begin{pmatrix} q_1 \\ p_1 \end{pmatrix} + \sigma \begin{pmatrix} -1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} q_2 & -1 \\ p_2 & 1 \end{pmatrix} \otimes \begin{pmatrix} q_1 & -1 \\ p_1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ \sigma \end{pmatrix} \end{aligned}$$

by an easy calculation. For earlier attempts, see [1, 2, 5, 6, 9, 10].

4. Concluding Remarks

(i) The bivariate binomial has appeared a couple of times in the literature, mostly as a stepping stone towards a bivariate Poisson distribution. See [5, 9, 10].

(ii) The vectorization used in Section 3 is necessarily restricted to distributions with a finite number of non-zero probabilities. For the bivariate case one can still use infinite matrices like in (i) above. A bivariate version of the geometric and of the negative binomial distribution is also tractable using matrix theory. See [10]. Any further multivariate extension seems more difficult.

(iii) In [13] Wishart derives expressions for the *cumulants* of the multivariate multinomial distribution. They are obtained by expanding $\log \varphi(e^{t_1}, e^{t_2}, ..., e^{t_n})$ with respect to increasing powers of t_i, $i = 1, 2, ..., n$. We failed in finding a relatively easy tensor formulation of his results.

Acknowledgments

This work was done while the author was on sabbatical leave at the University of California, Santa Barbara. The author thanks J. Gani and the members of the Statistics and Applied Probability program for their kind hospitality. We take pleasure in thanking M. K. Sadaghiani and D. Nicholls for a number of helpful discussions. We also express our gratitude to M. Marcus for helping us find our way in the vast literature on multivariate algebra. Finally, we thank the referee whose comments greatly improved the presentation of the paper.

References

[1] Aitken, A. C. (1936). A further note on multivariate selection. *Proc. Edinburg Math. Soc.* **5** 37–40.

[2] Aitken, A. C., and Gonin, H. T. (1935). On fourfold sampling with and without replacement. *Proc. Royal Soc. Edinburgh* **55** 114–125.

[3] Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* M.I.T. Press, Cambridge, MA.

[4] Bonett, D. G., and Woodward, J. A. (1987). Application of the Kronecker product and Wald test in log-linear models. *Comput. Statist. Quart.* **4** 235–243.

[5] Campbell, J. T. (1934). The Poisson correlation function. *Proc. Edinburgh Math. Soc.* **4** 18–26.

[6] Capobianco, M. F. (1964). *On the Bivariate Binomial Distribution and Related Problems.* Ph.D. thesis, Polytechnic Inst. Brooklyn.

[7] Graham, A. (1981). *Kronecker Products and Matrix Calculus: With Applications.* Ellis Horwood, New York.

[8] Henderson, H. V., and Searle, S. R. (1979). Vec. and vech operators for matrices with some uses in Jacobian and multivariate statistics. *Canad. J. Statist.* **7** 65–81.

[9] Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika* **51** 241–245.

[10] Johnson, N. L., and Kotz, S. (1969). *Discrete Distributions.* Wiley, New York.

[11] Neudecker, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *J. Amer. Statist. Assoc.* **64** 953–963.

[12] Upton, G. J. G. (1978). *The Analysis of Cross-tabulated Data.* Wiley, Chichester.

[13] Wishart, J. (1949). Cumulants of multivariate multinomial distributions, *Biometrika* **36** 47–58.