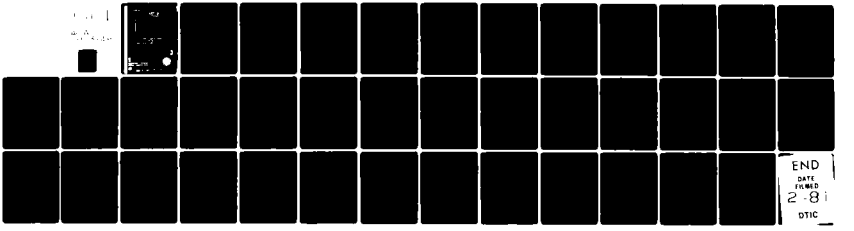
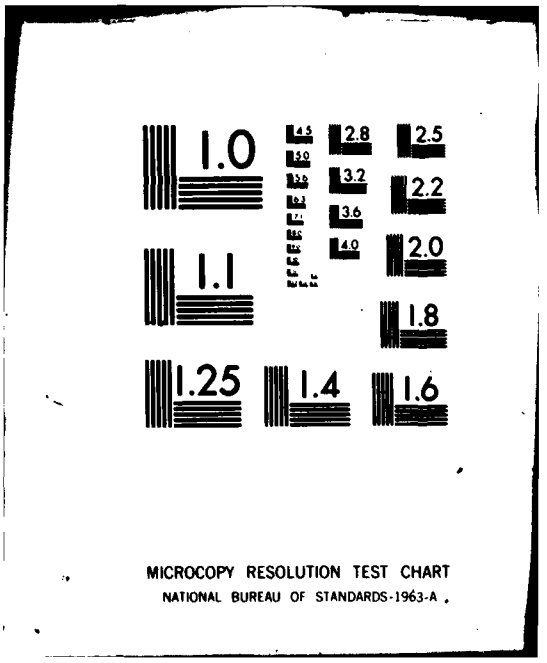
AD-A093 148 CALIFORNIA UNIV DAVIS INTERCOLLEGE DIV OF STATISTICS F/G 12/1
MODELING AND INFERENCE FOR POSITIVELY DEPENDENT VARIABLES IN DI--ETC(U)
AUG 80 R A BOYLES, F J SAMANIEGO AFOSR-77-3180
UNCLASSIFIED TR-19 AFOSR-TR-80-1160 NL



END
DATE FILMED
2-81
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-80-1160	2. GOVT ACCESSION NO.: AD-A093148
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): MODELING AND INFERENCE FOR POSITIVELY DEPENDENT VARIABLES IN DICHOTOMOUS EXPERIMENTS	5. TYPE OF REPORT & PERIOD COVERED: Interim rept.
	6. PERFORMING ORG. REPORT NUMBER: 14 TR-191
7. AUTHOR(s): Russell A. Boyles and Francisco J. Samaniego	8. CONTRACT OR GRANT NUMBER(s): AFOSR-77-3180
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of California, Dvision of Statistics, Davis, California 95616	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientifc Research/NM, Bolling AFB, Washington, DC 20332	12. REPORT DATE: August 1980
	13. NUMBER OF PAGES: 37
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

multivariance, Bernoulli, maximum likelihood estimation, invariance, consistenct, efficiency

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Multivariatemodels with positively correlated components have found wide applicability in reliability and biostatistics. Perhaps the best known and most widely used model is the multivariate exponential distribution due to Marshall and Olkin (JASA, 1967). We study a discrete analogue of the latter model. Specifically, we consider a model for random vectors Y whose components are positively correlated and hve Bernoulli marginal distributions. The construction of the model reflects the fact that the k compcenet system under study may be subjected to independent shocks selectively fatal to any

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

411206

subset of components. A special representation of the probability function of Y is developed which proves useful in the inference questions pursued. While maximum likelihood estimation of the model parameters proves intractable, we obtain in closed form an alernative estimator which we show to be asymptotically equivalent to the MLE and, in fact, equals the MLE with limiting probability one. Similar results are obtained for a natural submodel whose parameter space is of substantially lower dimension. A Monte Carlo study sheds light on the sample size needed for the asymptotic results to take hold.

Modeling and Inference for Positively Dependent Variables in Dichotomous Experiments*

by Russell A. Boyles and Francisco J. Samaniego

Technical Report No. 19

August 1980

*This research is supported in part by the Air Force Office of Scientific Research under grant AFOSR 77-3180.

MODELING AND INFERENCE FOR POSITIVELY DEPENDENT VARIABLES IN DICHOTOMOUS EXPERIMENTS

ABSTRACT

Multivariate models with positively correlated components have found wide applicability in reliability and biostatistics. Perhaps the best known and most widely used such model is the multivariate exponential distribution due to Marshall and Olkin (JASA, 1967). We study a discrete analogue of the latter model. Specifically, we consider a model for random vectors $\underset{\sim}{Y}$ whose components are positively correlated and have Bernoulli marginal distributions. The construction of the model reflects the fact that the k component system under study may be subjected to independent shocks selectively fatal to any subset of components. A special representation of the probability function of $\underset{\sim}{Y}$ is developed which proves useful in the inference questions pursued. While maximum likelihood estimation of the model parameters proves intractable, we obtain in closed form an alternative estimator which we show to be asymptotically equivalent to the MLE and, in fact, equals the MLE with limiting probability one. Similar results are obtained for a natural submodel whose parameter space is of substantially lower dimension. A Monte Carlo study sheds light on the sample size needed for the asymptotic results to take hold.

I. INTRODUCTION

Consider the vectors in $\{0,1\}^k$ as being ordered from the smallest, $(0,0,0,\ldots,0)$ to the largest $(1,1,1,\ldots,1)$ according to the usual ordering, that is, lexicographically. Attach to the i^{th} such vector the probability $p_i \in [0,1]$, for $i=1,2,\ldots,2^k$, subject to the constraint $\sum_{i=1}^{2^k} p_i = 1$. The description above defines the most general multivariate Bernoulli distribution, and any k-variate distribution with Bernoulli marginals can be fully described by identifying the vector $\underset{\sim}{p}$ above. A useful reference on such distributions is the paper by Bahadur (1961). Inference questions for the general multivariate Bernoulli distribution are easily resolved; for example, the maximum likelihood estimate of p_i from a sample of size n is simply the relative frequency of occurence of the i^{th} vector in $\{0,1\}^k$. There are, however, a number of interesting subclasses of multivariate Bernoulli distributions which arise naturally in applications but lend themselves less readily to statistical analysis. This paper is dedicated to the study of one such class.

Multivariate models which postulate positive dependence among the components of a random vector have found substantial applicability in reliability theory and in biostatistics. For example, the lifetimes of items on test may well be positively correlated, particularly when these items are subject to shocks that threaten two or more items simultaneously, or when some items experience wear which serves to increase the load on other items on test. The multivariate exponential distribution of

Marshall and Olkin (1967) arises naturally in several reliability contexts, including a scenario involving a collection of shocks selectively fatal to one or more components in a coherent system. In constructing competing risk models in problems arising in the health sciences, one often encounters positive dependence due to the reduction of resistance to one disease that might be caused by the presence of a second disease. The model to be studied here may be viewed as a discrete analogue to the multivariate exponential distribution in that we will motivate it in much the same way. As we shall see, however, the multivariate Bernoulli model we study is quite general and may be viewed as a discretized version of any continuous multivariate model describing the joint lifetime of several components subjected to selectively fatal shocks. Alternative notions of positive dependence in reliability have been studied by several authors, including Esary, et al. (1972) and Shaked (1975). It is well known that these notions are all equivalent when dealing with Bernoulli variables

Consider a k-component system whose lifetime is under study. Suppose the status of the system is to be observed at some distinguished time T_0 which could, for example, be the so-called mission time of the system or of the individual components. The observation may be recorded as a vector $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_k)$ where, for each i, Y_i is one or zero depending upon whether the i^{th} component is working or has failed by time T_0. We will formulate a model for the distribution of $\underset{\sim}{Y}$ which reflects the fact that the system may be subjected to independent shocks

selectively fatal to any subset of the k components. To this end, define $\mathcal{A}_k = \{0,1\}^k - \{\underset{\sim}{0}\}$, that is, let $\mathcal{A}_k$ be the collection of all k-tuples of zeros and ones with the exception of the zero vector. Each element $\underset{\sim}{s} \in \mathcal{A}_k$ corresponds to a shock selectively fatal to those components i for which $s_i = 1$. For each $\underset{\sim}{s} \in \mathcal{A}_k$, define

$$Z_{\underset{\sim}{s}} = \begin{cases} 0 & \text{if the shock corresponding to } \underset{\sim}{s} \text{ occurs by time } T_0. \\ 1 & \text{otherwise.} \end{cases} \tag{1.1}$$

and define

$$p_{\underset{\sim}{s}} = P(Z_{\underset{\sim}{s}} = 1). \tag{1.2}$$

(Our notation is precisely that of Marshall and Olkin (1967).) Denoting the binomial distribution with parameters n and p by $\mathcal{B}(n,p)$, we have that $Z_{\underset{\sim}{s}} \sim \mathcal{B}(1,p_{\underset{\sim}{s}})$ for each $\underset{\sim}{s} \in \mathcal{A}_k$. We may now define the aforementioned vector $\underset{\sim}{Y}$ componentwise as

$$Y_i = \prod_{\{\underset{\sim}{s} \in \mathcal{A}_k : s_i = 1\}} Z_{\underset{\sim}{s}} = \min_{\{\underset{\sim}{s} \in \mathcal{A}_k : s_i = 1\}} Z_{\underset{\sim}{s}}. \tag{1.3}$$

We will use the notation MVB(2^k-1) to denote the distribution of $\underset{\sim}{Y}$ in equation (1.3), and, in general, the notation MVB(n), for various integers n, to denote particular submodels of MVB(2^k-1) with exactly n parameters. Equation (1.3) may be interpreted to mean that the i^{th} component will survive until time T_0 if and only if no shock that is fatal to the

i^{th} component occurs by time T_0. The shock that is fatal to the i^{th} component alone has the dual interpretation of being the embodiment of all factors that might prove fatal to the i^{th} component but have no effect whatever on any other component.

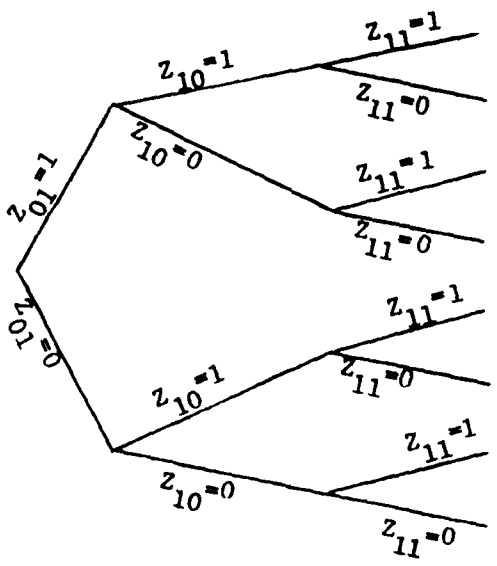
In section II, we study the properties of the shock model we have described in the preceding paragraph. In particular, a representation of the probability mass function of $\underset{\sim}{Y}$ (or any subvector of $\underset{\sim}{Y}$) is developed. This representation proves to be quite useful in the statistical inference we pursue in later sections. In section III, we consider maximum likelihood estimation of the parameters of our multivariate Bernoulli distribution. We investigate and comment on the difficulties involved in obtaining the MLE in closed form, and we then produce (in closed form) an alternative estimator which we show is asymptotically equivalent to the MLE. In fact, the proposed estimator is equal to the MLE with limiting probability one. In section IV, we investigate maximum likelihood estimation for a natural submodel of $MVB(2^k-1)$, and derive results which are comparable to those in section III. In section V, we summarize the results of a modest Monte Carlo study which sheds light on the sample size required to get reasonable results. The simulation study motivates our discussion of the primary domain of applicability of our results, namely, that of systems consisting of components subject to "rare shocks." We summarize our results, indicate some promising directions for future work, and make a number of concluding remarks in section VI.

II. CHARACTERISTICS OF THE MVB(2^k-1) MODEL

To motivate the distribution theory presented in the Lemma and Theorem below, we begin by examining the simplest model, that for a two-component system. When k=2, we are concerned with the representation of $P(\underset{\sim}{Y} = \underset{\sim}{y})$ for $\underset{\sim}{y} \in \{0,1\}^2$ in terms of the three model parameters p_{01}, p_{10} and p_{11}. Clearly, the following representation is one possibility:

$$\begin{aligned} P(\underset{\sim}{Y} = (1,1)) &= p_{10}p_{01}p_{11} \\ P(\underset{\sim}{Y} = (0,1)) &= (1-p_{10})p_{01}p_{11} \\ P(\underset{\sim}{Y} = (1,0)) &= (1-p_{01})p_{10}p_{11} \\ P(\underset{\sim}{Y} = (0,0)) &= (1-p_{11}) + p_{11}(1-p_{10})(1-p_{01}). \end{aligned} \tag{2.1}$$

The character of such representations for arbitrary k is already apparent from equation (2.1). It would seem that in general the probability $P(\underset{\sim}{Y} = \underset{\sim}{y})$ could be written as a sum of terms, each term being a product of <u>certain</u> parameters $p_{\underset{\sim}{s}}$ or their complements $(1-p_{\underset{\sim}{s}})$. A systematic approach to this sort of representation would proceed as follows: Consider the tree consisting of the eight branches which represent all possible combinations of the three dichotomies $\{Z_{\underset{\sim}{s}}=0 \text{ or } Z_s=1 | \underset{\sim}{s} \in \mathcal{S}_2\}$. This tree is pictured below:



	Probability	$\underset{\sim}{Y}$
	$p_{01}p_{10}p_{11}$	(1,1)
	$p_{01}p_{10}(1-p_{11})$	(0,0)
	$p_{01}(1-p_{10})p_{11}$	(1,0)
	$p_{01}(1-p_{10})(1-p_{11})$	(0,0)
	$(1-p_{01})p_{10}p_{11}$	(0,1)
	$(1-p_{01})p_{10}(1-p_{11})$	(0,0)
	$(1-p_{01})(1-p_{10})p_{11}$	(0,0)
	$(1-p_{01})(1-p_{10})(1-p_{11})$	(0,0)

(2.2)

The probabilities associated with each of the four possible values of $\underset{\sim}{Y}$ may be obtained from this tree simply by grouping branch probabilities appropriately. A general representation for the probability mass function of $\underset{\sim}{Y} \sim \text{MVB}(2^k-1)$ may be obtained in the same manner. However, since there are 2^k-1 shocks, the tree from which probabilities are to be identified will have 2^{2^k-1} branches. Thus, the representation along these lines, while conceptually simple, poses some practical difficulties. This approach can be made more appealing by applying combinatorial arguments which facilitate the counting of ways in which a specific value of $\underset{\sim}{Y}$ can arise. We do not pursue this further, however, because we have found this type of representation less useful in inference problems than the representation described in Theorem 2.1. As a closing remark on the representation discussed above, we note that the probabilities

associated with the 2^{2^k-1} branches of the aforementioned tree are simply the terms of the polynomial expansion of the right hand side of the equation below

$$1 \equiv \prod_{\{\underset{\sim}{s} \in \mathcal{S}_k\}} (p_{\underset{\sim}{s}} + (1-p_{\underset{\sim}{s}})) \tag{2.3}$$

Before establishing an alternative representation of $P(\underset{\sim}{Y} = \underset{\sim}{y})$ in the general model, we introduce the following notation, where all vectors are k-dimensional.

$$\underset{\sim}{0} \equiv (0,0,0,\ldots,0)$$

$$\mathcal{S}_k' \equiv \mathcal{S}_k \cup \{\underset{\sim}{0}\} \tag{2.4}$$

$$\underset{\sim}{1} \equiv (1,1,1,\ldots,1)$$

$$\underset{\sim}{y}^* \equiv \underset{\sim}{1} - \underset{\sim}{y} \tag{2.5}$$

$$|\underset{\sim}{y}| \equiv |(y_1,\ldots,y_k)| \equiv \sum_{i=1}^{k} |y_i| \tag{2.6}$$

$$\underset{\sim}{x}\,\underset{\sim}{y} = (x_1y_1, x_2y_2, \ldots, x_ky_k) \tag{2.7}$$

$$I_k(\underset{\sim}{y}) = \{s \in \mathcal{S}_k' : y_i=1 \Rightarrow s_i=1,\ i=1,\ldots,k\} \tag{2.8}$$

Now, let j be an integer less than or equal to k, and let $N_j=\{n_i,\ i=1,\ldots,j\}$ be a set of j integers such that $1 \le n_1 < n_2 < \ldots < n_j \le k$. We will have occasion to consider mappings $\pi_{N_j} : \mathcal{S}_k' \to \mathcal{S}_j'$ of the form

$$\pi_{N_j}\underset{\sim}{s} \equiv \pi_{N_j}(s_1,s_2,\ldots,s_k) = (s_{n_1},\ldots,s_{n_j}). \tag{2.9}$$

When confusion seems unlikely, the mapping π_{N_j} will be denoted by π_j or simply by π.

We need the following preliminary result.

Lemma 2.1. For any fixed but arbitrary set N_j of j ordered positive integers, with $j \leq k$ and $n_j \leq k$,

$$P(Y_{n_1} = Y_{n_2} = \ldots = Y_{n_j} = 1) = \prod_{\underset{\sim}{s} \in A} p_{\underset{\sim}{s}}, \tag{2.10}$$

where $A = \{\underset{\sim}{s} \in \mathcal{S}_k : s_{n_i} = 1 \text{ for some } i=1,\ldots,j\}$

Proof. $P(Y_{n_1} = \ldots = Y_{n_j} = 1)$

$$= P\left(\prod_{s_{n_1}=1} Z_{\underset{\sim}{s}} = \ldots = \prod_{s_{n_j}=1} Z_{\underset{\sim}{s}} = 1\right)$$

$$= P(Z_{\underset{\sim}{s}} = 1 \quad \forall \underset{\sim}{s} \ni s_{n_i} = 1 \text{ for some } i)$$

$$= \prod_{\underset{\sim}{s} \in A} P(Z_{\underset{\sim}{s}} = 1),$$

the last step being a consequence of the independence of the Z variables.

Theorem 2.1. Let N_j be a fixed but arbitrary set of j ordered positive integers, with $j \leq k$ and $n_j \leq k$. For every $\underset{\sim}{y} \in \mathcal{S}'_j$, we have

$$P(\pi \underset{\sim}{Y} = \underset{\sim}{y}) = \sum_{\underset{\sim}{w} \in I_j(\underset{\sim}{y})} (-1)^{|\underset{\sim}{w} - \underset{\sim}{y}|} \prod_{A_{\pi,\underset{\sim}{w}}} p_{\underset{\sim}{s}} \tag{2.11}$$

where $\pi \equiv \pi_{N_j}$ and $A_{\pi,\underset{\sim}{w}} = \{\underset{\sim}{s} \in \mathcal{S}_k : (\pi \underset{\sim}{s})\, \underset{\sim}{w} \neq \underset{\sim}{0}\}$.

Proof: The proof proceeds by induction on $|\underset{\sim}{y}^*|$, the number of zero components of the vector $\underset{\sim}{y}$, where the dimension j of $\underset{\sim}{y}$ may vary between 1 and k. We show that for each fixed k, the representation in (2.11) holds for each $j=1,\ldots,k$ and for all values of $|y^*| \leq k$. If $|y^*| = 0$, (2.11) follows from Lemma 2.1. Now suppose that $|\underset{\sim}{y}^*| = h$, where $1 \leq h \leq k$. Assume that for each $j=1,\ldots,k$, and for all possible mappings π_j of the form (2.9), the representation in (2.11) holds for $P(\pi_j \underset{\sim}{Y} = \underset{\sim}{s})$ whenever $\underset{\sim}{s} \in \mathcal{A}'_j$ is such that $|s^*| < h$. Since $\underset{\sim}{y}$ has at least one zero, we may assume without loss of generality that $y_1 = 0$. Define $\pi^* : \mathcal{A}'_j \to \mathcal{A}'_{j-1}$ by

$$\pi^* \underset{\sim}{s} = \pi^*(s_1,\ldots,s_j) = (s_2,s_3,\ldots,s_j),$$

and define $\underset{\sim}{y}^o \in \mathcal{A}'_j$ by

$$y_i^o = \begin{cases} y_i & \text{if } i > 1 \\ 1 & \text{if } i = 1. \end{cases}$$

We then have

$$P(\pi^* \pi_j \underset{\sim}{Y} = \pi^* \underset{\sim}{y}) = P(Y_{n_2} = y_2,\ldots,Y_{n_j} = y_j)$$

$$= P(\pi \underset{\sim}{Y} = \underset{\sim}{y}) + P(\pi \underset{\sim}{Y} = \underset{\sim}{y}^o)$$

or

$$P(\pi \underset{\sim}{Y} = \underset{\sim}{y}) = P(\pi^* \pi \underset{\sim}{Y} = \pi^* \underset{\sim}{y}) - P(\pi \underset{\sim}{Y} = \underset{\sim}{y}^o).$$

Since both $\pi^* \underset{\sim}{y}$ and $\underset{\sim}{y}^o$ have exactly $h-1$ zero components, we have by the induction hypothesis that

$$P(\pi \underset{\sim}{Y} = \underset{\sim}{y}) = \sum_{\underset{\sim}{v} \in I_{j-1}(\pi^* \underset{\sim}{y})} (-1)^{|\underset{\sim}{v} - \pi^* \underset{\sim}{y}|} \prod_{A_{\pi^* \pi, \underset{\sim}{v}}} p_{\underset{\sim}{s}}$$

$$- \sum_{\underset{\sim}{w} \in I_j(\underset{\sim}{y}^o)} (-1)^{|\underset{\sim}{w} - \underset{\sim}{y}^o|} \prod_{A_{\pi, \underset{\sim}{w}}} p_{\underset{\sim}{s}} \qquad (2.12)$$

The two sums in (2.12) can be combined upon noting the following:

(i) For any $\underset{\sim}{w} \in I_j(\underset{\sim}{y}^o)$, $|\underset{\sim}{w} - \underset{\sim}{y}^o| = |\underset{\sim}{w} - \underset{\sim}{y}| - 1$, and $I_j(\underset{\sim}{y}^o) = \{\underset{\sim}{w} \in I_j(\underset{\sim}{y}) : w_1 = 1\}$.

(ii) Each $\underset{\sim}{v} \in I_{j-1}(\pi^* \underset{\sim}{y})$ may be associated with the vector $\underset{\sim}{w}(\underset{\sim}{v}) = (0, v_1, \ldots, v_{j-1}) \in I_j(\underset{\sim}{y})$; moreover, $|\underset{\sim}{v} - \pi^* \underset{\sim}{y}| = |\underset{\sim}{w}(\underset{\sim}{v}) - \underset{\sim}{y}|$ and $\{\underset{\sim}{s} \in \mathcal{A}_k : (\pi^* \pi_j \underset{\sim}{s})\underset{\sim}{v} = \underset{\sim}{0}\} = \{\underset{\sim}{s} \in \mathcal{A}_k : [\pi_j \underset{\sim}{s}][\underset{\sim}{w}(\underset{\sim}{v})] = \underset{\sim}{0}\}$.

(iii) As $\underset{\sim}{v}$ ranges over $I_{j-1}(\pi^* \underset{\sim}{y})$, $\underset{\sim}{w}(\underset{\sim}{v})$ traces out exactly one copy of the set $\{w \in I_j(\underset{\sim}{y}) : w_1 = 0\}$.

We may thus write

$$P(\pi \underset{\sim}{Y} = \underset{\sim}{y}) = \sum_{\substack{\underset{\sim}{w} \in I_j(\underset{\sim}{y}) \\ \ni w_1 = 0}} (-1)^{|\underset{\sim}{w} - \underset{\sim}{y}|} \prod_{A_{\pi, \underset{\sim}{w}}} p_{\underset{\sim}{s}}$$

$$+ \sum_{\substack{\underset{\sim}{w} \in I_j(\underset{\sim}{y}) \\ \ni w_1 = 1}} (-1)^{|\underset{\sim}{w} - \underset{\sim}{y}|} \prod_{A_{\pi, \underset{\sim}{w}}} p_{\underset{\sim}{s}}$$

$$= \sum_{\underset{\sim}{w} \in I_j(\underset{\sim}{y})} (-1)^{|\underset{\sim}{w} - \underset{\sim}{y}|} \prod_{A_{\pi, \underset{\sim}{w}}} p_{\underset{\sim}{s}} ,$$

completing the proof.

The representation in (2.11) gives the distribution of the random vector Y and shows that the marginal distribution of any subvector has the same form. We state as a corollary the representation result to be used in our study of interence questions.

Corollary 2.1. Let $Y \sim MVB(2^k-1)$ for $k \geq 2$. Then for any $y \in \mathcal{J}_k'$,

$$P(\underset{\sim}{Y}=\underset{\sim}{y}) = \sum_{\underset{\sim}{w} \in I_k(\underset{\sim}{y})} (-1)^{|\underset{\sim}{w}-\underset{\sim}{y}|} \prod_{\substack{\{\underset{\sim}{s} \in \mathcal{J}_k: \\ \underset{\sim}{s}\underset{\sim}{w} \neq 0\}}} p_{\underset{\sim}{s}} \tag{2.13}$$

We have mentioned above that the model under discussion has the property that components of a vector $\underset{\sim}{Y}$ described by the model are positively correlated. This is easily demonstrated by noting that

$$Y_i Y_j \sim B\left(1, \prod_{\substack{\cup \\ r=i,j}} p_{\underset{\sim}{s}}\right), \text{ so that } \{\underset{\sim}{s} \in \mathcal{J}_k : s_r=1\}$$

$$\mathrm{Cov}(Y_i,Y_j) = \prod_{\substack{\cup \\ r=i,j}} p_{\underset{\sim}{s}} - \left(\prod_{\{\underset{\sim}{s} \in \mathcal{J}_k : s_i=1\}} p_{\underset{\sim}{s}}\right)\left(\prod_{\{\underset{\sim}{s} \in \mathcal{J}_k : s_j=1\}} p_{\underset{\sim}{s}}\right) \geq 0,$$

with equality holding if and only if $p_{\underset{\sim}{s}} = 0$ for some $\underset{\sim}{s} \in \bigcup_{r=i,j} \{\underset{\sim}{s} \in \mathcal{J}_k : s_r=1\}$ or $p_{\underset{\sim}{s}} = 1$ for all $\underset{\sim}{s} \in \mathcal{J}_k$ for which $s_i = s_j = 1$.

The simplest possible multivariate Bernoulli distribution is the model in which components Y_i are independent and identically distributed as B(1,p). Conditions on the parameters of the model MVB(2^k-1) which result in partial or complete simulation of the model with i.i.d. components are as follows: (i) Dependent components with identical marginal distributions result if and only if $p_{\underset{\sim}{s}} = p_{\underset{\sim}{s}'}$ for any $\underset{\sim}{s}, \underset{\sim}{s}' \in \mathcal{S}_k$ for which $|\underset{\sim}{s}| = |\underset{\sim}{s}'|$, (ii) Independent components with nonidentical, nondegenerate, marginal distributions result if and only if $p_{\underset{\sim}{s}} = 1$ for every $\underset{\sim}{s} \in \mathcal{S}_k$ for which $|\underset{\sim}{s}| > 1$ (that is, only shocks fatal to a single component have a positive probability of occurrence), (iii) The Y_i, i=1,...,k are nondegenerate, independent and identically distributed iff the conditions on $\underset{\sim}{s} \in \mathcal{S}_k$ in (i) and (ii) above are simultaneously satisfied.

III. ESTIMATION FOR MVB(2^k-1).

We begin this section with an examination of the problem of maximum likelihood estimation of the parameter $\underset{\sim}{p}$ for the model MVB(2^k-1). To fix ideas, we briefly discuss the case k=2, that is, the two component model. Let $\underset{\sim}{Y}_1,\ldots,\underset{\sim}{Y}_n$ be a sample of size n from MVB(2^2-1), and for each $\underset{\sim}{y} \in \{0,1\}^2$, let $N_{\underset{\sim}{y}}$ be the frequency of the observation $\underset{\sim}{Y} = \underset{\sim}{y}$. A complete discussion of the case k=2 involves consideration of the 15 data configurations where each $N_{\underset{\sim}{y}}$ is zero or nonzero, with at least one being nonzero. Suppose, for example, $N_{\underset{\sim}{y}} > 0$ for each $\underset{\sim}{y}$. Then the likelihood function may be written as

$$L = [(1 - p_{11}) + p_{11}(1 - p_{10})(1 - p_{01})]^{N_{00}}[(1 - p_0)p_{10}p_{11}]^{N_{10}}$$
$$\cdot [(1 - p_{10})p_{01}p_{11}]^{N_{01}}[p_{10}p_{01}p_{11}]^{N_{11}}. \qquad (3.1)$$

It is clear even in the simple case being considered that maximum likelihood estimation is fairly cumbersome. The system of equations to be solved, that is, $\frac{\partial}{\partial p_{\underset{\sim}{s}}} \ell n\, L = 0$ for $\underset{\sim}{s} \in \mathcal{S}_2$, are highly nonlinear and each involves all parameters in a complex way. It can be shown (by direct maximization) that the MLE when $N_{\underset{\sim}{y}} > 0 \quad \forall \underset{\sim}{y}$ is given by

$$\begin{aligned} \hat{p}_{10} &= \frac{N_{11}}{N_{11}+N_{01}} \\ \hat{p}_{01} &= \frac{N_{11}}{N_{11}+N_{10}} \\ \hat{p}_{11} &= \frac{N_{11}}{n\,\hat{p}_{10}\hat{p}_{01}} \end{aligned} \tag{3.2}$$

when $N_{11}N_{00} \geq N_{01}N_{10}$, and is

$$\begin{aligned} \hat{p}_{10} &= \frac{N_{11}+N_{10}}{n} \\ \hat{p}_{01} &= \frac{N_{11}+N_{01}}{n} \\ \hat{p}_{11} &= 1 \end{aligned} \tag{3.3}$$

when $N_{11}N_{00} < N_{01}N_{10}$. We are, in fact, able to find the MLE in closed form for the case k=2, but will not pursue direct maximum likelihood estimation further for several reasons. It is clear that direct maximum likelihood estimation is unpromising, involving, in the case of arbitrary k, the examination of $2^{2^k}-1$ data configurations, each associated with a highly complex system of (2^k-1) equations. Moreover, there is no guarantee

that any one of these systems has a closed form solution. Indeed, we have found that there is no closed form solution of the likelihood equations in some of the special cases we have studied. We will thus turn our attention to the construction of an alternative estimator. We show in the sequel that this estimator is asymptotically optimal and, in fact, is equal to the MLE with limiting probability one.

Let $\underset{\sim}{Y}_1,\ldots,\underset{\sim}{Y}_n$ be a random sample from MVB(2^k-1), and define $N_{\underset{\sim}{y}}$ for each $\underset{\sim}{y} \in \mathcal{I}_k'$ as

$$N_{\underset{\sim}{y}} = \text{frequency of occurrence of } \underset{\sim}{Y} = \underset{\sim}{y}.$$

If $Q_{\underset{\sim}{y}} \equiv P(\underset{\sim}{Y} = \underset{\sim}{y})$, then the vector $\underset{\sim}{N} = (N_{\underset{\sim}{0}},\ldots,N_{\underset{\sim}{1}})$, with components lexicographically ordered, has a multinomial distribution with parameters n and $\underset{\sim}{Q}$. The MLE of $\underset{\sim}{Q}$ is given by

$$\hat{Q}_{\underset{\sim}{y}} = \frac{N_{\underset{\sim}{y}}}{n} \qquad \text{for } \underset{\sim}{y} \in \mathcal{I}_k'. \tag{3.4}$$

The estimator we develop is derived from an attempt to invert the functional relationship

$$Q_{\underset{\sim}{y}} = \sum_{\underset{\sim}{w} \in I_k(\underset{\sim}{y})} (-1)^{|\underset{\sim}{w} - \underset{\sim}{y}|} \prod_{\{\underset{\sim}{s} \in \mathcal{I}_k : \underset{\sim}{s}\underset{\sim}{w} \neq \underset{\sim}{0}\}} p_{\underset{\sim}{s}} \tag{3.5}$$

for each $\underset{\sim}{y} \in \mathcal{I}_k'$ to find $\underset{\sim}{p}$ as a function of $\underset{\sim}{Q}$. Under well-known conditions, the estimate of $\underset{\sim}{p}$ obtained from (3.5) and (3.4) will be the MLE. These conditions do not obtain in the present problem; we are nevertheless able to use the invariance property of MLE's to advantage. We proceed with the inversion of the relationship in (3.5).

Let $\underset{\sim}{y} \in \mathcal{A}_k$ be such that $|\underset{\sim}{y}| = 1$. Then $|\underset{\sim}{y}^*| = k-1$, and we obtain from (3.5)

$$Q_{\underset{\sim}{y}^*} = \prod_{\underset{\sim}{s} \neq \underset{\sim}{y}} p_{\underset{\sim}{s}} - \prod_{\underset{\sim}{s} \in \mathcal{A}_k} p_{\underset{\sim}{s}} \tag{3.6}$$

Since $Q_{\underset{\sim}{1}} = \prod_{\underset{\sim}{s} \in \mathcal{A}_k} p_{\underset{\sim}{s}}$, we obtain from (3.6) that, provided $Q_{\underset{\sim}{1}} \neq 0$,

$$\begin{aligned} p_{\underset{\sim}{y}} &= \left[\frac{Q_{\underset{\sim}{y}^*}}{Q_{\underset{\sim}{1}}} + 1 \right]^{-1} \\ &= \frac{Q_{\underset{\sim}{1}}}{Q_{\underset{\sim}{1}} + Q_{\underset{\sim}{y}^*}} \end{aligned} \tag{3.7}$$

Now let $\underset{\sim}{y} \in \mathcal{A}_k$ be such that $|\underset{\sim}{y}| = j < k$, and suppose $p_{\underset{\sim}{s}}$ is expressed as a function of $\underset{\sim}{Q}$ for all $\underset{\sim}{s}$ for which $|\underset{\sim}{s}| < j$. From (3.5) we have

$$\frac{Q_{\underset{\sim}{y}^*}}{Q_{\underset{\sim}{1}}} = \sum_{\underset{\sim}{w} \in I_k(\underset{\sim}{y}^*)} (-1)^{|\underset{\sim}{w} - \underset{\sim}{y}^*|} \prod_{\{\underset{\sim}{s} \in \mathcal{A}_k \,:\, \underset{\sim}{s}\,\underset{\sim}{w}=0\}} p_{\underset{\sim}{s}}^{-1}$$

$$= \sum_{i=0}^{|\underset{\sim}{y}|} \sum_{\{\underset{\sim}{w} \in I_k(\underset{\sim}{y}^*) \,:\, |\underset{\sim}{w} - \underset{\sim}{y}^*| = i\}} (-1)^i \prod_{\{\underset{\sim}{s} \in \mathcal{A}_k \,:\, \underset{\sim}{s}\,\underset{\sim}{w}=0\}} p_{\underset{\sim}{s}}^{-1} =$$

$$= \prod_{\{\underset{\sim}{s} \in \mathcal{A}_k : \underset{\sim}{s}\,\underset{\sim}{\sim}\,\underset{\sim}{y}^* = 0\}} p_{\underset{\sim}{s}}^{-1}$$

$$+ \sum_{i=1}^{|\underset{\sim}{y}|-1} \sum_{\{\underset{\sim}{w} \in I_k(\underset{\sim}{y}^*) : |\underset{\sim}{w}-\underset{\sim}{y}^*| = i\}} (-1)^i \prod_{\{\underset{\sim}{s} \in \mathcal{A}_k : \underset{\sim}{s}\,\underset{\sim}{\sim}\,\underset{\sim}{w} = 0\}} p_{\underset{\sim}{s}}^{-1} + (-1)^{|\underset{\sim}{y}|}$$

$$= p_{\underset{\sim}{y}}^{-1} \prod_{\{\underset{\sim}{s} \in \mathcal{A}_k : \underset{\sim}{s}\,\underset{\sim}{\sim}\,\underset{\sim}{y}^* = 0,\ |\underset{\sim}{s}| < |\underset{\sim}{y}|\}} p_{\underset{\sim}{s}}^{-1}$$

$$+ \sum_{i=1}^{|\underset{\sim}{y}|-1} \sum_{\{\underset{\sim}{w} \in I_k(\underset{\sim}{y}^*) : |\underset{\sim}{w}-\underset{\sim}{y}^*| = i\}} (-1)^i \prod_{r=1}^{|\underset{\sim}{y}|-i} \prod_{\{\underset{\sim}{s} \in \mathcal{A}_k : \underset{\sim}{s}\,\underset{\sim}{\sim}\,\underset{\sim}{w} = 0,\ |\underset{\sim}{s}| = r\}} p_{\underset{\sim}{s}}^{-1} + (-1)^{(\underset{\sim}{y})} \tag{3.8}$$

Verification of the last equation above is facilitated by noting that $|\underset{\sim}{y}|-i$ is the number of zero components of $\underset{\sim}{w}$ when $|\underset{\sim}{w} - \underset{\sim}{y}^*| = i$, and that $r \geq 1$ since $\mathcal{A}_k$ does not contain $\underset{\sim}{0}$. Equating the first and last terms in the preceding chain of equalities, and solving for $p_{\underset{\sim}{y}}$, we obtain

$$p_{\underset{\sim}{y}} = \left\{ \left[\frac{Q_{\underset{\sim}{y}^*}}{Q_{\underset{\sim}{1}}} - (-1)^{|\underset{\sim}{y}|} \right] \prod_{\{\underset{\sim}{s} \in \mathcal{A}_k : \underset{\sim}{s}\,\underset{\sim}{y}^* = \underset{\sim}{0}, |\underset{\sim}{s}| < |\underset{\sim}{y}|\}} p_{\underset{\sim}{s}} \right.$$

$$\left. - \sum_{i=1}^{|\underset{\sim}{y}|-1} \sum_{\{\underset{\sim}{w} \in I_k(\underset{\sim}{y}^*) : |\underset{\sim}{w}-\underset{\sim}{y}^*| = i\}} (-1)^i \prod_{\substack{\{\underset{\sim}{s} \in \mathcal{A}_k : |\underset{\sim}{s}| \le |\underset{\sim}{y}|-i, \\ \underset{\sim}{s}\,\underset{\sim}{y}^* = \underset{\sim}{0} \neq \underset{\sim}{s}\,\underset{\sim}{w}\}}} p_{\underset{\sim}{s}} \prod_{\substack{\{\underset{\sim}{s} \in \mathcal{A}_k : |\underset{\sim}{y}|-i \\ \le |\underset{\sim}{s}| < |\underset{\sim}{y}|, \\ \underset{\sim}{s}\,\underset{\sim}{y}^* = \underset{\sim}{0}\}}} p_{\underset{\sim}{s}} \right\}^{-1} . \qquad (3.9)$$

Equation (3.9) may be derived from (3.8) by noting that for $\underset{\sim}{w} \in I_k(\underset{\sim}{y}^*)$, $\underset{\sim}{s}\,\underset{\sim}{w} = \underset{\sim}{0} \Rightarrow \underset{\sim}{s}\,\underset{\sim}{y}^* = \underset{\sim}{0}$. In equation (3.9), $p_{\underset{\sim}{y}}$ is expressed recursively as a function of $\underset{\sim}{Q}$ and $\{p_{\underset{\sim}{s}} : |\underset{\sim}{s}| < |\underset{\sim}{y}|\}$, where $1 \le |\underset{\sim}{y}| < k$.

We may determine $p_{\underset{\sim}{1}}$ from the equation

$$Q_{\underset{\sim}{1}} = \prod_{\underset{\sim}{s} \in \mathcal{A}_k} p_{\underset{\sim}{s}} ,$$

that is, as

$$p_{\underset{\sim}{1}} = Q_{\underset{\sim}{1}} \prod_{\{\underset{\sim}{s} \in \mathcal{A}_k : |\underset{\sim}{s}| < k\}} p_{\underset{\sim}{s}}^{-1} \qquad (3.10)$$

The estimate $\hat{\underset{\sim}{p}}^I$ is obtained by replacing $Q_{\underset{\sim}{y}}$ and $p_{\underset{\sim}{y}}$ with $\hat{Q}_{\underset{\sim}{y}}$ and $\hat{p}_{\underset{\sim}{y}}^I$ in equations (3.7), (3.9) and (3.10).

Several remarks are in order. First, it is clear that $\hat{\underset{\sim}{p}}^I$ is undefined when $n\hat{\underset{\sim}{Q}}_1 = \underset{\sim}{N}_1 = 0$. In fact, the process by which $\hat{\underset{\sim}{p}}^I$ was obtained requires that $p_{\underset{\sim}{s}} \neq 0 \quad \forall\, \underset{\sim}{s} \in \mathcal{S}_k$ (indeed, this implies $Q_{\underset{\sim}{1}} \neq 0$ by (3.10)). Finally, even when $\hat{\underset{\sim}{p}}^I$ is well defined by the process described above, it may well happen that $\hat{\underset{\sim}{p}}^I \notin (0,1]^{2^k-1}$, and in such cases, the estimate can clearly be improved. On the other hand, if $\hat{\underset{\sim}{p}}^I \in (0,1]^{2^k-1}$, then it is the MLE of the vector $\underset{\sim}{p}$, and if $\hat{\underset{\sim}{p}}^I \in (0,1]^{2^k-1}$ with sufficiently high probability, then the estimate $\hat{\underset{\sim}{p}}^I$ may well have good statistical properties. In order to have a well-defined estimator, let

$$\hat{\underset{\sim}{p}} = \begin{cases} \hat{\underset{\sim}{p}}^I & \text{if } \hat{\underset{\sim}{p}}^I \in (0,1]^{2^k-1} \\ \underset{\sim}{0} & \text{otherwise.} \end{cases} \tag{3.11}$$

We show below that the estimator $\hat{\underset{\sim}{p}}$ is in fact asymptotically optimal. (Were $\hat{\underset{\sim}{p}}$ to be more carefully defined when $\hat{\underset{\sim}{p}}^I$ exists but lies outside $(0,1]^{2^k-1}$, one could undoubtedly obtain an asymptotically equivalent estimator with better small sample properties.)

Let $\hat{p}_L$ denote the MLE of $\underset{\sim}{p}$, and let $\mathcal{I}(\underset{\sim}{p})$ denote the information matrix of the MVB(2^k-1) model, that is,

$$\left[\mathcal{I}(\underset{\sim}{p})\right]_{\underset{\sim}{u}\,\underset{\sim}{v}} = -E\left[\frac{\partial^2 \ln Q_{\underset{\sim}{Y}}}{\partial p_{\underset{\sim}{u}}\, \partial p_{\underset{\sim}{v}}}\right] \tag{3.12}$$

for $u, v \in \mathcal{S}_k$. The computation of $\mathcal{I}(\underset{\sim}{p})$ is straightforward and is omitted.

Theorem 3.1. Let $\underset{\sim}{Y}_1, \ldots, \underset{\sim}{Y}_n, \ldots$ be a sequence of iid random vectors distributed according to MVB(2^k-1). Let $\underset{\sim}{p}$ be the parameter vector of the model, and assume that $p_{\underset{\sim}{y}} \in (0,1) \quad \forall\, \underset{\sim}{y} \in \mathcal{S}_k$. Then

(i) $\hat{\underset{\sim}{p}} \xrightarrow{a.s.} \underset{\sim}{p}$, and

(ii) $\sqrt{n}\ (\hat{\underset{\sim}{p}} - \underset{\sim}{p}) \xrightarrow{\mathcal{L}} X \sim N(\underset{\sim}{0}, \mathcal{I}^{-1}(\underset{\sim}{p}))$,

that is, $\hat{\underset{\sim}{p}}$ is strongly consistent and asymptotically equivalent to the MLE.

Proof. The regularity of the MVB(2^k-1) model is easy to verify. We first prove the consistency of $\hat{\underset{\sim}{p}}$. The relationship between $\underset{\sim}{p}$ and $\underset{\sim}{Q}$ given in (3.7), (3.9) and (3.10) makes clear that there is a one-to-one relationship between $\underset{\sim}{p}$ and $\underset{\sim}{Q}$ when each $p_{\underset{\sim}{y}} > 0$ and $Q_{\underset{\sim}{l}} > 0$, inequalities which hold by hypothesis. Moreover, the relationship is a continuous one in a neighborhood of the true $\underset{\sim}{p}$ and $\underset{\sim}{Q}$. Thus $\underset{\sim}{p} = f(\underset{\sim}{Q})$, where f is continuous. Now since $\hat{\underset{\sim}{Q}} \xrightarrow{a.s.} \underset{\sim}{Q}$, it follows that $f(\hat{\underset{\sim}{Q}}) \xrightarrow{a.s.} f(\underset{\sim}{Q})$. Thus, outside of a null set, the sequences $\{f(\hat{\underset{\sim}{Q}})\}$ converges to $f(\underset{\sim}{Q})$ as $n \to \infty$. For n sufficiently large, $\hat{\underset{\sim}{p}} = f(\hat{\underset{\sim}{Q}})$ and the sequence $\{\hat{\underset{\sim}{p}}\}$ will thus converge to $\underset{\sim}{p}$. Therefore, outside of the same null set, $\hat{\underset{\sim}{p}} \to \underset{\sim}{p}$ and we may write $\hat{\underset{\sim}{p}} \xrightarrow{a.s.} \underset{\sim}{p}$. Now let $\hat{\underset{\sim}{p}}_L$ be the maximum likelihood estimate of $\underset{\sim}{p}$. Using $\|\cdot\|$ for Euclidean distance, we have

$$P(\sqrt{n}\ \|\hat{\underset{\sim}{p}} - \hat{\underset{\sim}{p}}_L\| \geq \epsilon) \leq 1 - P(\hat{\underset{\sim}{p}} \in (0,1)^{2^k-1}) \to 0$$

since $\hat{\underset{\sim}{p}} \xrightarrow{P} \underset{\sim}{p} \in (0,1)^{2^k-1}$. Thus

$$\sqrt{n}\ (\hat{\underset{\sim}{p}} - \hat{\underset{\sim}{p}}_L) \xrightarrow{P} 0.$$

Since

$$\sqrt{n}\ (\hat{\underset{\sim}{p}} - \underset{\sim}{p}) = \sqrt{n}\ (\hat{\underset{\sim}{p}}_L - \underset{\sim}{p}) + \sqrt{n}\ (\hat{\underset{\sim}{p}} - \hat{\underset{\sim}{p}}_L),$$

we have that $\hat{\underset{\sim}{p}}$ and $\hat{\underset{\sim}{p}}_L$ are asymptotically equivalent, or,

$$\sqrt{n}\ (\hat{\underset{\sim}{p}} - \underset{\sim}{p}) \xrightarrow{\mathcal{L}} X \sim N(\underset{\sim}{0}, \mathcal{I}^{-1}(\underset{\sim}{p})).$$

IV. INFERENCE FOR SUBMODELS

Our treatment of the general MVB model is relatively complete. We are able to give a simple, finite recursive scheme which specifies a strongly consistent and asymptotically efficient estimator of the parameter vector. Nonetheless, the general model is unsatisfactory in one key respect. For systems with more than a handful of components, the general model has so many parameters that the sample size required for our asymptotic results to take hold could well be prohibitive. This difficulty can be circumvented in situations where one can justify the use of a submodel of MVB(2^k-1) whose parameter space is of substantially lower dimension. In this section, we treat the estimation of the parameter vector of a reasonable submodel with a k-dimensional parameter space.

Let $\underset{\sim}{Y}$ be distributed according to the MVB(2^k-1) distribution, subject to the following parametric restrictions:

$$p_{\underset{\sim}{s}} = p_{\underset{\sim}{t}} \qquad \forall\, \underset{\sim}{s},\underset{\sim}{t} \in \mathcal{A}_k \ni |\underset{\sim}{s}| = |\underset{\sim}{t}|. \tag{4.1}$$

If $|\underset{\sim}{s}| = |\underset{\sim}{t}| = j$, the common parameter value in (4.1) will be denoted by p_j. We may rewrite the mass function of $\underset{\sim}{Y}$ in terms of the parameters p_j, $j=1,2,\ldots,k$, as follows.

$$P(\underset{\sim}{Y}=\underset{\sim}{y}) = \sum_{\ell=|\underset{\sim}{y}|} \sum_{\substack{\{\underset{\sim}{w}\in I_k(\underset{\sim}{y}):\\ |\underset{\sim}{w}|=\ell\}}} (-1)^{\ell-|\underset{\sim}{y}|} \prod_{j=1}^{k} \prod_{\substack{\{\underset{\sim}{s}\in\mathcal{A}_k:\\ \underset{\sim}{s}\underset{\sim}{w}\neq\underset{\sim}{0}\\ |\underset{\sim}{s}|=j\}}} p_{\underset{\sim}{s}} .$$

Using the convention $\binom{r}{t} = 0$ if $t > r$, the cardinality of the set $\{\underset{\sim}{s} \in \mathcal{A}_k : |\underset{\sim}{s}| = j, \underset{\sim}{s}\underset{\sim}{w} \neq \underset{\sim}{0}\}$ may be seen to be

$$\binom{k}{j} - \binom{k-\ell}{j}$$

for any $\underset{\sim}{w} \in \mathcal{A}_k$ for which $|\underset{\sim}{w}| = \ell$. Moreover, the cardinality of the set $\{\underset{\sim}{w} \in I_k(\underset{\sim}{y}) : |\underset{\sim}{w}| = \ell\}$ is

$$\binom{k-|\underset{\sim}{y}|}{\ell-|\underset{\sim}{y}|}.$$

It follows that

$$P(\underset{\sim}{Y} = \underset{\sim}{y}) = \sum_{\ell=|\underset{\sim}{y}|} \binom{k-|\underset{\sim}{y}|}{\ell-|\underset{\sim}{y}|} (-1)^{\ell-|\underset{\sim}{y}|} \prod_{j=1}^{k} p_j^{\binom{k}{j} - \binom{k-\ell}{j}}$$

which we rewrite as

$$P(\underset{\sim}{Y} = \underset{\sim}{y}) = \sum_{t=0}^{k-|\underset{\sim}{y}|} \binom{k-|\underset{\sim}{y}|}{t} (-1)^t \prod_{j=1}^{t} p_j^{\binom{k}{j} - \binom{k-|\underset{\sim}{y}|-t}{j}}. \tag{4.2}$$

The model whose probability function is displayed in (4.2) is a k-parameter multivariate Bernoulli distribution which we henceforth refer to as MVB(k). This model retains some of the essential features of the general model we have studied. In particular, if $\underset{\sim}{Y} \sim$ MVB(k), then the components of $\underset{\sim}{Y}$ are positively correlated. Moreover, the submodel preserves the notion that shocks of varying gravity may occur while making the simplifying assumption that shocks of the same gravity (i.e., simultaneously fatal to a fixed number of components) are equiprobable.

One can easily deduce from this assumption that the components of $\underset{\sim}{Y}$ have identical marginal distributions. Indeed, all subvectors of $\underset{\sim}{Y}$ of a fixed dimension will be identically distributed. The model should be applicable in reliability experiments in which the components of a system are inherently similar when viewed one at a time (e.g., the cells of an automobile battery), yet tend to be positively dependent due to the increased load on working cells as each failure occurs.

We proceed with our development of an efficient estimator of the parameter vector of MVB(k). Note that the probability function in (4.2) depends on an observed vector $\underset{\sim}{y}$ only through $|\underset{\sim}{y}|$, the number of ones in the vector $\underset{\sim}{y}$. Thus, the likelihood function L is proportional to

$$\prod_{x=0}^{k} [P(|\underset{\sim}{Y}| = x)]^{N_x}$$

where, for $x = 0,1,\ldots,k$,

$$Q_x \equiv P(|\underset{\sim}{Y}| = x) = \sum_{\{\underset{\sim}{y} : |\underset{\sim}{y}| = x\}} P(\underset{\sim}{Y} = \underset{\sim}{y})$$

$$= \binom{k}{x} \sum_{t=0}^{k-x} \binom{k-x}{t} (-1)^t \prod_{j=1}^{k} p_j^{\binom{k}{j} - \binom{k-x-t}{j}}, \qquad (4.3)$$

and N_x is the number of occurrences of $|\underset{\sim}{Y}| = x$ in the sample. If $\underset{\sim}{Q}$ was completely unconstrained, the MLE of the vector $(Q_0,Q_1,\ldots,Q_k)$ would be $(\frac{1}{n}N_0, \frac{1}{n}N_1,\ldots,\frac{1}{n}N_k)$. As in the previous section, we will obtain an estimator of the vector $\underset{\sim}{p}$ by inverting the functional relationship (4.3). Since

$$Q_k = \prod_{j=1}^{k} p_j^{\binom{k}{j}},$$

we have (assuming $Q_k > 0$)

$$\frac{Q_{k-1}}{Q_k} = k\left[\frac{1}{p_1} - 1\right]$$

which yields

$$p_1 = \frac{Q_k}{Q_k + (1/k)Q_{k-1}} . \tag{4.4}$$

Now, suppose $p_1,\ldots,p_{x-1}$ have been expressed as functions of the Q's for a fixed x between 2 and k-1. We then solve the equation

$$\frac{Q_{k-x}}{Q_k} = \binom{k}{k-x}\left[\sum_{t=0}^{x-1}\binom{x}{t}(-1)^t\prod_{j=1}^{x-t}p_j^{-\binom{x-t}{j}} + (-1)^x\right] \tag{4.5}$$

for p_x. We note that (4.5) may be rewritten as

$$\frac{Q_{k-x}}{Q_k} = \binom{k}{x}\left[\prod_{j=1}^{x}p_j^{-\binom{x}{j}} + \sum_{t=1}^{x-1}\binom{x}{t}(-1)^t\prod_{j=1}^{x-t}p_j^{-\binom{x-t}{j}} + (-1)^x\right],$$

so that

$$p_x = \left\{\left[\frac{Q_{k-x}}{\binom{k}{x}Q_k} - (-1)^x\right]\prod_{j=1}^{x-1}p_j^{\binom{x}{j}} - \sum_{t=1}^{x-1}\binom{x}{t}(-1)^t\prod_{j=1}^{x-1}p_j^{\binom{x}{j}-\binom{x-t}{j}}\right\}^{-1} \tag{4.6}$$

Finally, we may solve the equation

$$Q_k = \prod_{j=1}^{k} p_j^{\binom{k}{j}}$$

to obtain

$$p_k = Q_k \prod_{j=1}^{k-1} p_j^{\binom{k}{j}} \tag{4.7}$$

An estimate for $\underset{\sim}{p}$ is obtained by replacing $\underset{\sim}{Q}$ and $\underset{\sim}{p}$ by $\hat{\underset{\sim}{Q}}$ and $\hat{\underset{\sim}{p}}$ in equations (4.4) - (4.7). Such an estimator is, of course, undefined when $N_k = 0$. However, if we define $\hat{\underset{\sim}{p}}$ to be the estimator obtained from equations (4.4) - (4.7) when the estimator exists and lies in $[0,1]^k$ and define $\hat{\underset{\sim}{p}} \equiv 0$ otherwise, we can establish the following

Theorem 4.1. Let $\hat{\underset{\sim}{p}}_L$ be the maximum likelihood estimator of $\underset{\sim}{p}$, and assume that $\underset{\sim}{p} \in (0,1)^k$. Then $\hat{\underset{\sim}{p}}$ is strongly consistent and is asymptotically equivalent to $\hat{\underset{\sim}{p}}_L$. In particular,

$$\sqrt{n}\,(\hat{\underset{\sim}{p}} - \underset{\sim}{p}) \xrightarrow{\mathcal{L}} X \sim N(\underset{\sim}{0}, \mathcal{J}^{-1}(\underset{\sim}{p})),$$

where $\mathcal{J}(\underset{\sim}{p})$ is the information matrix for the model MVB(k).

The proof of Theorem 4.1 is similar to that of Theorem 3.1, and is omitted.

In their paper on maximum likelihood estimation for the multivariate exponential distribution, Proschan and Sullo (1976) study in detail the submodel in which the only shocks which occur with positive probability are shocks fatal to single components only or the universal shock which

is fatal to all components simultaneously. The analogue of this submodel in our context is the MVB distribution subject to the restriction that

$$p_s = 1 \qquad \forall\, \underset{\sim}{s} \ni 1 < |\underset{\sim}{s}| < k.$$

This is a $k+1$ parameter submodel for which we have obtained the following results: Direct maximum likelihood succeeds to the extent that the MLE may be identified explicitly as a function of the smallest root of a certain k^{th} degree polynomial. Moreover, we are able to give a simple iterative procedure which converges to the desired root. The approach taken in our study of this submodel differs completely from that taken in this paper. Full details appear in Boyles and Samaniego (1980).

The two submodels of MVB(2^k-1) discussed above serve to illustrate the fact that one may model the behavior of positively dependent components in reliability experiments in such a way that (1) the parameter space has a manageable dimension and (2) efficient estimation of parameters is tractable and computationally feasible. Moreover, the submodels considered here correspond to realistic constraints, that is, represent situations in which (a) components seem to have the same probabilistic behavior when viewed one at a time, or (b) among shocks fatal to more than one component, only the catastrophic or universal shock is deemed important or at all likely.

V. A MONTE CARLO STUDY

In the preceding sections we have derived the asymptotic properties of the estimator $\hat{\underset{\sim}{p}}$ for the parameter $\underset{\sim}{p}$ of the MVB(2^k-1) and MVB(k) models. In particular, $P\{\hat{\underset{\sim}{p}} = MLE\} \to 1$ as sample size $\to \infty$. We now employ Monte Carlo simulations to investigate the rate of this convergence as a function of sample size and the true parameter $\underset{\sim}{p}$.

Table 1 below summarizes the bivariate case $k = 2$ for the fixed sample size $n = 50$. For each parameter vector, $\hat{P}\{\hat{\underset{\sim}{p}} = MLE\}$ is the observed frequency of the event $\hat{\underset{\sim}{p}} = MLE$, based on 100 simulations.

Table 1 shows that $\hat{\underset{\sim}{p}}$ "works well," i.e., equals the MLE with probability near 1, when shocks are "rare," i.e., when the components of $\underset{\sim}{p}$ are "close" to 1. In other situations, it is evident that $\hat{\underset{\sim}{p}}$ may perform quite poorly. As we should expect from symmetries in the MVB parametrization, the performance of $\hat{\underset{\sim}{p}}$ appears to be invariant with respect to permutations of p_{10} and p_{01}. One might also suspect that $\hat{\underset{\sim}{p}}$ should work better in the submodel than the full model, since the former has fewer parameters. However, Table 1 does not support this conjecture. Table 1 does show that the performance of $\hat{\underset{\sim}{p}}$ is sensitive to certain order relations among the components of $\underset{\sim}{p}$. In particular, $\hat{\underset{\sim}{p}}$ performs best when at least one of p_{10} and p_{01} is greater than p_{11}.

Table 1. $100 \times \hat{P}\{\hat{\underset{\sim}{p}} = MLE\}$ (k=2,n=50).

p_1	p_2	p_{10}	p_{01}	p_{11}	Full Model	Submodel
0.1	0.1	0.1	0.1	0.1	7	6
0.1	0.5	0.1	0.1	0.5	17	24
0.1	0.9	0.1	0.1	0.9	37	37
		0.1	0.5	0.1	24	
		0.1	0.5	0.5	65	
		0.1	0.5	0.9	55	
		0.1	0.9	0.1	34	
		0.1	0.9	0.5	89	
		0.1	0.9	0.9	72	
		0.5	0.1	0.1	23	
		0.5	0.1	0.5	73	
		0.5	0.1	0.9	57	
0.5	0.1	0.5	0.5	0.1	66	65
0.5	0.5	0.5	0.5	0.5	98	97
0.5	0.9	0.5	0.5	0.9	69	75
		0.5	0.9	0.1	87	
		0.5	0.9	0.5	100	
		0.5	0.9	0.9	95	
		0.9	0.1	0.1	34	
		0.9	0.1	0.5	92	
		0.9	0.1	0.9	83	
		0.9	0.5	0.1	88	
		0.9	0.5	0.5	100	
		0.9	0.5	0.9	93	
0.9	0.1	0.9	0.9	0.1	98	98
0.9	0.5	0.9	0.9	0.5	100	100
0.9	0.9	0.9	0.9	0.9	100	99

A similar study was carried out for the trivariate case $k = 3$ and the fixed sample size $n = 50$, with results similar to those of the preceding case. To save space, we present only a small portion of this study in Table 2 below.

Table 2. $100 \times \hat{P}\{\hat{\underset{\sim}{p}} = MLE\}$ $(k=3, n=50)$.

$p_1 = p_{100} = p_{010} = p_{001}$	$p_2 = p_{110} = p_{101} = p_{011}$	$p_3 = p_{111}$	Full Model	Sub-Model
0.70	0.70	0.70	82	78
0.70	0.70	0.90	88	68
0.70	0.90	0.70	56	80
0.70	0.90	0.90	58	73
0.90	0.70	0.70	98	93
0.90	0.70	0.90	97	77
0.90	0.90	0.70	88	100
0.90	0.90	0.90	95	93

Table 2 shows again that $\hat{\underset{\sim}{p}}$ works well in the "rare shocks" region of the parameter space, although a comparison of Tables 1 and 2 shows that the boundaries of this region depend on the dimension of the model. Once again the submodel does not show a clear advantage over the full model. The submodel performs best when the ordering $p_1 \geq p_2 \geq p_3$ obtains, but the full model does not appear to be sensitive to order restrictions in this case.

Tables 3 and 4 below illustrate the large-sample behavior of $\hat{\underset{\sim}{p}}$.

Table 3. $100 \times \hat{P}\{\hat{\underset{\sim}{p}} = \text{MLE}\}$ (Full model, k=3, large samples).

$P_{111}P_{010}P_{001}$	$P_{110}P_{101}P_{011}$	P_{111}	Sample Size 100	200	300	400	500	600	700	800	900	1000
0.30	0.30	0.30	4	4	8	14	13	13	17	24	18	22
0.30	0.50	0.70	22	37	49	55	61	67	70	69	71	73
0.50	0.30	0.70	23	34	45	57	68	74	78	85	88	91
0.50	0.50	0.50	44	82	82	92	91	94	92	95	97	99
0.50	0.70	0.30	47	69	80	79	80	90	88	92	93	93
0.70	0.50	0.30	65	90	95	99	100	100	99	100	100	100
0.70	0.70	0.70	94	100	100	100	100	100	100	100	100	100
0.90	0.90	0.90	100	100	100	100	100	100	100	100	100	100

Table 4. $100 \times \hat{P}\{\hat{\underset{\sim}{p}} = \text{MLE}\}$ (Submodel, k=3, large samples).

P_1	P_2	P_3	Sample Size 100	200	300	400	500	600	700	800	900	1000
0.30	0.30	0.30	1	3	0	4	12	12	16	18	18	19
0.30	0.50	0.70	18	21	31	42	39	53	66	62	67	70
0.50	0.30	0.70	21	39	38	49	50	56	64	69	68	67
0.50	0.50	0.50	47	71	83	89	94	93	96	93	97	98
0.50	0.70	0.30	69	86	90	90	91	94	92	98	100	97
0.70	0.50	0.30	67	89	96	100	100	100	100	100	100	100
0.70	0.70	0.70	96	98	100	100	100	100	100	100	100	100
0.90	0.90	0.90	100	100	100	100	100	100	100	100	100	100

The rate of convergence of $\hat{P}\{\hat{\underset{\sim}{p}} = MLE\}$ to 1 increases dramatically in both models as we move toward rare shock parameter values. The parameter vectors are arranged in increasing lexicographical order, and in both models this rate of convergence is evidently increasing with respect to this ordering.

Our final study involves the performance of $\hat{\underset{\sim}{p}}$ for small and moderate sample sizes. Our simulations have shown that here $\hat{\underset{\sim}{p}}$ will perform very poorly except in the rare shock case. In the case of very rare shocks (i.e., all components of $\underset{\sim}{p}$ greater than 0.90) we found that the event $\hat{\underset{\sim}{p}} = MLE$ obtained in more than 84% of 170,100 simulations performed with very small samples ($n = 1,2,\ldots,7$), and in more than 67% of 72,900 simulations performed with moderate sample sizes ($n = 10,20,30$). It is notable that our estimation procedure works fairly well even when the sample size is less than the dimension of the parameter vectors. The explanation for this unexpected result is related to the fact that $\hat{\underset{\sim}{p}}$ appears to work better in very small samples than for moderate sample sizes. This may be attributed to the fact that, for very rare shocks and very small samples, only a few data configurations possess a non-negligible probability of occurring, and that in fact those turn out to be configurations for which $\hat{\underset{\sim}{p}}$ works. With moderate sample sizes this restriction is dropped, and the asymptotic properties have not yet taken hold, hence we observe a decline in performance.

Our simulations have shown that $\hat{\underset{\sim}{p}}$ performs very well in large samples whenever the assumption of moderately rare shocks (e.g., all components of $\underset{\sim}{p} \geq 0.70$) is appropriate. Under stronger assumptions (e.g., all components of $\underset{\sim}{p} \geq 0.90$) our estimation procedure works fairly well at small and moderate sample sizes. The simulations also show that the estimation procedure is more sensitive to the parameter values corresponding to low-gravity shocks than to the parameter values corresponding to high-gravity shocks. This should be taken into account when assessing the usefulness of our procedure in a given situation.

VI. DISCUSSION AND CONCLUSIONS

We have described in this paper a multivariate Bernoulli distribution which may be viewed as a shock model in reliability. It is a discrete analogue of the multivariate exponential distribution of Marshall and Olkin (1967) but, in fact, serves as a discrete version of any continuous model for which component lifetimes are modeled as minima of waiting times for selectively fatal shocks. Asymptotically optimal estimators have been given in closed form, both for the general model MVB(2^k-1) and for a natural submodel MVB(k) with a parameter space of substantially lower dimension. The estimators produced are only of value when they are equal to the MLE, since we have defined them quite arbitrarily elsewhere. We have investigated the rate at which $P(\hat{\underset{\sim}{p}} = \text{MLE})$ tends to one through a Monte Carlo study.

The Monte Carlo study makes the domain of applicability of our results rather clear. We have found that $P(\hat{\underset{\sim}{p}} = \text{MLE})$ is high for moderate sample sizes only when the shocks in the model (or its submodel) are rare. This has been found to be particularly true for shocks affecting only one or possibly a small number of components. The method of estimation is somewhat more robust with respect to the rareness of shocks of greater gravity. As Beuhler (1957) has pointed out, reliability experiments often deal with highly reliable systems in which the probability of failure of any given component is very low. It naturally follows that the shocks affecting such components occur only rarely. Thus, the method of estimation advanced in this paper tends to work well in a class of problems that occur frequently in practice.

The Monte Carlo study has also revealed that the rate at which $P(\hat{\underset{\sim}{p}} = \text{MLE}) \to 1$ is about the same for the full model and for its submodel. In a sense, this is a disappointing result, since one would hope to gain some advantage when one dramatically reduces the dimension of the parameter space. We should emphasize, however, that the rate of convergence studied in Section V does not discredit the submodel. It is a comment only on the inversion process involved in obtaining our estimator. There is a second convergence rate of interest in our problem, namely, the rate at which the asymptotic theory of maximum likelihood estimation takes hold. It is well known that this rate is affected by the size of the model, and it is in this domain that we expect the advantage of MVB(k) over MVB(2^k-1) to show up. When MVB(k) is appropriate as a model, we expect that the use of MVB(2^k-1) would be very costly in terms of the efficiency of the MLE.

There are a number of issues that we postpone for future investigations. Of particular interest is the development of modeling and inference for larger classes of MVB distributions with positive dependence. In this regard, inference for the class of all MVB distributions with positively correlated components may be feasible using the general representation theorem for MVB distributions developed by Bahadur (1961) and Lazarsfeld.

REFERENCES

Bahadur, R.R. (1961). A representation of the joint distribution of responses to n dichotomous items. In Studies In Item Analysis and Prediction, H. Solomon, Editor. Stanford University Press, Stanford, CA.

Boyles, R. and Samaniego F. (1980). Maximum likelihood estimation for a discrete multivariate shock model. Technical Report No. 21, Division of Statistics, University of California, Davis.

Buehler, R.J. (1957). Confidence intervals for the product of two binomial probabilities. Journal of the American Statistical Association, 52, 482-93.

Esary, J.D. and Proschan, F. (1972). Relationships among some concepts of bivariate dependence. Annals of Mathematical Statistics, 43, 651-5.

Marshall, A. and Olkin, I. (1967). A multivariate exponential distribution. Journal of the American Statistical Association, 62, 30-44.

Proschan, F. and Sullo, P. (1976). Maximum likelihood estimation for a multivariate exponential distribution. Journal of the American Statistical Association, 71, 465-72.

Shaked, M. (1975). On concepts of dependence for multivariate distributions. Ph.D. thesis, Department of Statistics, University of Rochester, Rochester, N.Y.